Contributions à la dynamique de frappe au clavier : multibiométrie, biométrie douce et mise à jour de la référence

Romain Giot

UNIVERSITÉ de CAEN/BASSE-NORMANDIE
U.F.R. de Sciences
École doctorale S.I.M.E.M

THÈSE

présentée par

Romain Giot

et soutenue

le 23 octobre 2012

En vue de l'obtention du

Doctorat de l'Université de Caen Basse-Normandie
Spécialité : informatique et applications

(Arrêté du 7 août 2006)

Contributions à la dynamique de frappe au clavier : multibiométrie, biométrie douce et mise à jour de la référence

Membres du Jury

Hubert Cardot	Professeur des universités à l'université de Tours	(Rapporteur)
Andrzej Drygajlo	Associé de recherche à l'Ecole Polytechnique Fédérale de Lausanne, Suisse	(Rapporteur)
Jean-luc Dugelay	Professeur à EURECOM	(Examinateur)
Alain Rakotomamonjy	Professeur des universités à l'université de Rouen	(Examinateur)
Bernadette Dorizzi	Professeur à l'Institut Mines Telecom/Telecom sudParis	(Co-directeur)
Christophe Rosenberger	Professeur des universités à l'ENSICAEN	(Directeur)

À mes enfants

Rose
et Léo

et à mon épouse

Christelle

Résumé

La dynamique de frappe au clavier est une modalité biométrique comportementale qui permet d'authentifier des individus selon leur façon de taper au clavier. Un tel système est peu coûteux, car il ne nécessite pas de matériel d'acquisition autre que le clavier de l'ordinateur, et est facilement accepté par l'utilisateur. Nous nous sommes principalement intéressé aux systèmes statiques où le texte saisit par l'utilisateur est connu à l'avance par la machine. Malheureusement, les performances de cette modalité sont plutôt médiocres en raison de la forte variabilité de la donnée biométrique. Cette variabilité est due à l'état émotionnel de la personne, l'apprentissage de la façon de taper, ...

Nous proposons dans cette thèse différentes contributions permettant d'améliorer les performances de reconnaissance de systèmes de dynamique de frappe au clavier (DDF). Nous effectuons également une analyse des bases publiques permettant d'évaluer la performance de nouveaux systèmes de reconnaissance. Une contribution est la mise au point d'un système de DDF par mot de passe partagé. Nous étudions ensuite la fusion multibiométrique avec la dynamique de frappe au clavier et la reconnaissance faciale afin d'augmenter les performances des deux systèmes. Nous montrons, sur deux jeux de données différents, qu'il est possible de reconnaitre le genre d'un individu suivant sa façon de taper au clavier. Enfin, nous présentons une nouvelle méthode de mise à jour de la référence biométrique qui permet de prendre en compte le vieillissement de la donnée biométrique, afin de ne pas avoir une diminution des performances de reconnaissance au cours du temps.

Abstract

Keystroke dynamics is a behavioural biometry which allows to authenticate individuals through there way of typing on a keyboard. Such systems are cheap, as they do not need specific devices different from the keyboard of the computer. They are also well accepted by the user. We are mainly interested in static systems where the text typed by the user is known in advance by the machine. Sadly, the performance of this modality are rather mediocre because of the high variability of the biometric data which comes from emotional state of the individual, the learning of they way to type, ...

In this thesis, we propose various contributions which allow to improve the recognition performance of keystroke dynamics systems. We also do an analysis of the public datasets allowing to evaluate the performance of new recognition systems. One contribution is the creation of a system which allows the authentication of users with a shared password. Then, we study the biometric fusion with face recognition and keystroke dynamics in order to increase the performance of the two systems. We show, on two different datasets, that it is possible to guess the gender of an individual through its way of typing to a keyboard. Finally, we present a new template update method which allows to take into account the ageing of the biometric data in order to not observe a decrease of performance overtime.

Remerciements

Premièrement, je veux remercier Christophe Rosenberger qui a su me faire confiance en 2008, en m'employant comme ingénieur de recherche dans son équipe, pour me permettre d'avoir une expérience dans la recherche et compenser mon absence de master recherche. Je tiens également à le remercier d'avoir fait tout ce qui lui était possible pour me financer ces deux années de thèse, après deux refus d'attribution de bourse de ministérielle par le comité d'attribution du GREYC.

Ensuite, je veux remercier Bernadette Dorizzi d'avoir accepté d'être co-directrice de ma thèse et d'avoir ajouté un niveau de rigueur supplémentaire à mon travail.

Je remercie Hubert Cardot et Andrzej Drygajlo qui ont accepté de rapporter ce manuscrit de thèse et m'ont permis de l'améliorer grâce à leurs remarques, ainsi que Jean-luc Dugelay et Alain Rakotomamonjy qui ont accepté d'examiner ce travail.

Je tiens à remercier Christophe Charrier et Maxime Descoteaux de m'avoir permis de travailler avec eux sur une thématique de recherche totalement différente de mon sujet de thèse : la classification de voxels d'IRM de diffusion.

J'aimerai remercier les lecteurs anonymes (ou non) des différents publications soumises tout au long de ces années, ainsi que du manuscrit de cette thèse, pour leurs remarques et échanges constructifs qui m'ont toujours permis d'améliorer la qualité de mon travail. Je tiens également à remercier les membres du feu DRI (pas de chances pour vous, vous êtes trop nombreux pour que je vous cite) et mes différents collègues de bureau : Alexandre, Baptiste, Mohamad et Syed. J'aimerai aussi remercier Valentine et Nicole pour avoir pris l'initiative de commencer à préparer le pot de thèse pendant la soutenance, les différentes personnes m'ayant aider à ranger et Aude pour avoir fait un super gateau. J'aimerai remercier mes parents qui sont venu assister à ma soutenance.

J'aimerai également remercier la scène démomaking sur Amstrad CPC, pour m'avoir appris, il y a plus de 12 ans, l'assembleur z80, la notion de compétition qui est utile pour la recherche, et des notions d'améliorations et d'optimisations perpétuelles qui semblent oubliées et inutiles dans l'informatique moderne.

Enfin merci à toutes les personnes que je n'ai pas citées ici et qui se reconnaîtront dans ces quelques lignes, ainsi qu'à toi lecteur.

Pour finir, j'aimerais remercier mon épouse et mes deux enfants d'avoir accepté mon manque fréquent de disponibilité durant les quatre dernières années.

Table des matières

Table des figures

Liste des tableaux

Liste des abréviations

ACP	Analyse en Composantes Principales.
AG	Algorithmes Génétiques.
AUC	aire sous la courbe ROC (*Area Under the ROC Curve*).
DDF	Dynamique De Frappe au clavier.
EER	taux d'égales erreurs (*Equal Error Rate*).
FAR	taux de fausses acceptations (*false acceptance rate*).
FMR	taux de fausses correspondances (*false match rate*).
FNMR	taux de fausses non-correspondances (*false non-match rate*).
FRR	taux de faux rejets (*false rejection rate*).
FTAR	taux d'échec à l'acquisition (*failure-to-acquire rate*).
GUMR	taux d'oublis de clients pour la mise à jour (*Genuine Update Miss Rate*).
IUSR	taux de sélection d'imposteurs pour la mise à jour (*Impostor Update Selection Rate*).
KPPV	K Plus Proches Voisins.
LDA	Linear Discriminent Algorithm.
PG	programmation génétique.
PP	intervalle de temps entre la pression de touches successives (Press-Press).
PR	intervalle de temps entre la pression et le relâchement d'une touche, autrement dit : durée de pression de la touche (Press-Release).
RF	Reconnaissance Faciale.
ROC	caractéristique de performance (*Receiver Operating Characteristic*).
RP	intervalle de temps entre le relâchement d'une touche et la pression de la suivante (Release-Press).
RR	intervalle de temps entre le relâchement de touches successives (Release-Release).
SVM	Séparateur à Vaste Marge.

1. Introduction

Contexte

Le CRÉDOC (Centre de Recherche pour l'Étude et l'Observation Des Conditions de vie) a réalisé une étude en 2011 sur l'utilisation de la téléphonie mobile et de l'informatique par les Français [Bigot et Croutte, 2011]. Cette étude montre que 78% des foyers ont au moins un ordinateur à domicile et que 75% des Français ont internet chez eux. De plus, 85% de la population est équipé d'un téléphone mobile et 24% s'en sert pour naviguer sur internet. Ces différents outils (téléphone, ordinateur, tablette, ...), ainsi que leurs applications (connexion à la machine, utilisation d'un logiciel, achats en ligne, ...) sont souvent sécurisés à l'aide d'un mot de passe. Cependant, la sécurité des systèmes de protection par mot de passe est relativement faible à cause de leur mauvaise utilisation. Pour cette raison, il est nécessaire de trouver des moyens alternatifs à ce mécanisme d'authentification basique, sans pour autant perturber l'expérience utilisateur ; certaines biométries peuvent être une alternative intéressante.

C'est pourquoi nous nous sommes intéressés, dans cette thèse, à l'utilisation de systèmes biométriques à *bas coût*, *non intrusifs*, *permanents* et *faiblement contraints pour l'utilisateur* comme mécanisme d'authentification pour les applications de *contrôle d'accès logique* sur ordinateur :

- un *système à bas coût* permet d'augmenter le nombre d'acquéreurs du système ;
- un *système non intrusif* permet d'augmenter son acceptation par ses utilisateurs en leur permettant de garder leurs habitudes ;
- un *système permanent* fonctionne convenablement tout au long de l'utilisation du système ;
- un *système faiblement contraint* ne nécessite pas une initialisation compliquée et fastidieuse pour l'utilisateur et l'administrateur.

Dans un contexte biométrique, un système à bas coût n'utilise pas de capteurs ou ne nécessite pas de capteurs supplémentaires à ceux déjà présents dans un ordinateur standard (autrement dit le clavier ou la souris, ainsi que la webcam pour les ordinateurs les plus récents). Les biométries non intrusives fonctionnent sans demander une intervention spécifique ou inhabituelle à l'utilisateur. La Dynamique De Frappe au clavier (DDF) statique correspond à ce critère ; il s'agit d'une modalité biométrique comportementale qui permet d'authentifier un individu selon sa façon de saisir un texte sur un clavier standard. C'est donc la modalité que nous étudions dans cette thèse, car elle correspond aux différents critères :

- comme aucun capteur supplémentaire n'est nécessaire (seul le clavier de l'ordinateur est nécessaire), elle est relativement peu coûteuse ;
- comme pour une authentification classique par mot de passe, l'utilisateur se contente de saisir son mot de passe ce qui en fait une modalité peu intrusive ;
- comme nous nous limitons aux méthodes nécessitant peu de données d'enrôlement, son utilisation est peu contraignante ;
- pour finir, comme nous allons étudier des mécanismes de mise à jour de la référence biométrique, le système fonctionne de façon pérenne.

Cependant, utiliser peu de données pour l'enrôlement diminue les performances de reconnaissance car ces données ne sont pas suffisamment nombreuses et représentatives pour permettre de modéliser convenablement le comportement de l'utilisateur. Ce problème est d'autant plus

pénalisant pour la DDF étant donné qu'elle a des performances peu élevées du fait qu'il s'agisse d'une modalité comportementale. Par conséquent, il est nécessaire d'étudier des mécanismes additionnels afin d'améliorer les performances de reconnaissance. Ceux que nous étudions dans cette thèse concernent :

- l'utilisation de systèmes multibiométriques qui effectuent la vérification de l'identité de l'utilisateur à l'aide de plusieurs facteurs biométriques différents ;
- l'utilisation de données de type « biométrie douce » afin de disposer de données supplémentaires qui permettent de différencier plusieurs catégories d'utilisateurs ou d'augmenter les performances de reconnaissance de systèmes d'authentification par DDF existants ;
- et la mise à jour semi-supervisée du modèle biométrique tout au long de la vie du système de reconnaissance par DDF, afin de prendre en compte la variabilité de la façon de saisir son mot de passe au cours du temps.

Organisation et contributions de cette thèse

Ce manuscrit de thèse est organisé selon six chapitres principaux pour lesquels la figure 1.1 représente graphiquement les dépendances.

Chapitre 1 : Introduction Ce chapitre présente l'intérêt et la nécessité des travaux présentés dans cette thèse. Il présente également le contenu du manuscrit.

Chapitre 2 : Positionnement du problème Ce chapitre présente les différentes généralités sur la biométrie. Il décrit le fonctionnement de tels systèmes, les compare les uns aux autres et explique comment les évaluer. Les informations ne sont donc pas spécifiques à la DDF, mais vont permettre de comprendre le fonctionnement général d'un système biométrique et de comprendre l'intérêt des systèmes de DDF.

Chapitre 3 : Dynamique de frappe au clavier Ce chapitre présente la DDF et décrit les travaux originaux de cette thèse la concernant. On y découvre le fonctionnement de la DDF ainsi qu'un historique des études les plus importantes sur le domaine. Nous y présentons également la liste des bases de données publiques disponibles pour travailler sur ce thème. La première contribution consiste à comparer objectivement et subjectivement les différentes bases de données de DDF afin d'aider à la sélection ou création de bases utiles pour des études précises. La deuxième contribution présentée est une méthode d'authentification par mot de passe partagé, qui est plus performante que les méthodes de l'état de l'art : soit sur le taux d'erreurs, soit sur le temps de calcul. Une étude comparative des systèmes de l'état de l'art y est également présentée.

Chapitre 4 : Multimodalité et biométrie douce Le chapitre précédent a permis de découvrir la DDF ainsi que ses contraintes. Nous avons vu que les performances ne sont pas très élevées et qu'il est nécessaire de les améliorer. Comme la multimodalité peut être une des solutions, ce chapitre permet de présenter le fonctionnement des systèmes multimodaux. Par la suite, nous nous intéressons principalement aux systèmes de fusion de scores. Nous abordons également le concept de biométrie douce qui peut être utilisée dans des systèmes multimodaux. Ensuite nous présentons nos contributions en multibiométrie et l'intérêt d'utiliser des systèmes multibiométriques en DDF. Notre première contribution est annexe à la multibiométrie, mais elle a un intérêt majeur pour les contributions suivantes. Il s'agit de la création d'une méthode d'approximation rapide du taux d'égales erreurs (*Equal Error Rate*) (EER), ce qui nous permet d'accélérer l'optimisation de la configuration des paramètres de nos contributions suivantes. La seconde

contribution montre l'intérêt et l'efficacité de fusionner des systèmes de Reconnaissance Faciale (RF) avec des systèmes de reconnaissance de DDF afin d'obtenir des systèmes avec une efficacité acceptable et à très bas coût. La troisième contribution présente l'intérêt d'utiliser des informations de type biométrie douce lors de la fusion. Nous montrons qu'il est possible de déterminer le genre d'un individu qui saisit une phrase fixe, et qu'utiliser cette information peut augmenter les performances de reconnaissance. Nous montrons également, qu'à partir de texte libre, nous pouvons déterminer le genre et la catégorie d'âge de l'individu l'ayant saisi.

Chapitre 5 : Mise à jour de la référence biométrique Ce chapitre décrit les travaux originaux de cette thèse concernant la mise à jour de la référence biométrique. La première section présente le fonctionnement des systèmes adaptatifs. Bien que la majorité des travaux de la littérature concernent d'autres modalités que la DDF, les principes restent les mêmes. Nous définissons ce qu'est un modèle biométrique ainsi que le fonctionnement des systèmes de mise à jour de la référence. Nous essayons ensuite de définir les modalités d'évaluation des systèmes de mise à jour. Les sections suivantes présentent nos contributions à la mise à jour du modèle. La première contribution analyse le comportement d'un système de mise à jour semi-supervisée pour la DDF. C'est la première fois qu'une telle étude est réalisée en DDF, car les études précédentes utilisaient des systèmes de mise à jour supervisée. La seconde contribution est la proposition d'une architecture hybride de système de mise à jour de la référence biométrique monomodal. Ce système donne de meilleures performances que les techniques d'auto-apprentissage. Bien qu'il ait été évalué dans un contexte de DDF, le système a été développé pour être également utilisé avec d'autres modalités biométriques.

Chapitre 6 : Conclusion Nous récapitulons les travaux effectués tout au long de cette thèse et donnons leurs perspectives.

Annexe A - Enquête sur l'usage d'un système biométrique Cette annexe présente les résultats d'une étude évaluant l'usage d'un système de RF et d'un autre de DDF. Ils montrent que malgré la différence de performance des deux systèmes (à l'avantage de la RF), le système de DDF reste mieux accepté par les utilisateurs.

Annexe B - Rappels statistiques Cette annexe effectue différents rappels statistiques sur les outils utilisés lors de cette thèse.

Annexe C - Flôt max/Coupe min Cette annexe présente l'algorithme du flôt max et de la coupe min qui est utile pour comprendre le fonctionnement des méthodes de mise à jour de la référence biométrique qui utilisent un mécanisme de propagation d'étiquettes dans un graphe.

Annexe D - Les méthodes d'édition Cette annexe présente les méthodes d'édition qui sont des techniques utiles pour limiter le nombre d'individus à conserver dans une galerie dans des mécanismes utilisant les K Plus Proches Voisins (KPPV). Elles peuvent donc être utilisées dans des mécanismes de mise à jour de galerie.

Annexe E - Machines à vecteur support Cette annexe présente le fonctionnement des machines à vecteur support ou Séparateur à Vaste Marge (SVM). Il s'agit d'un outil de classification supervisée souvent utilisé dans les travaux de cette thèse ainsi que dans la littérature.

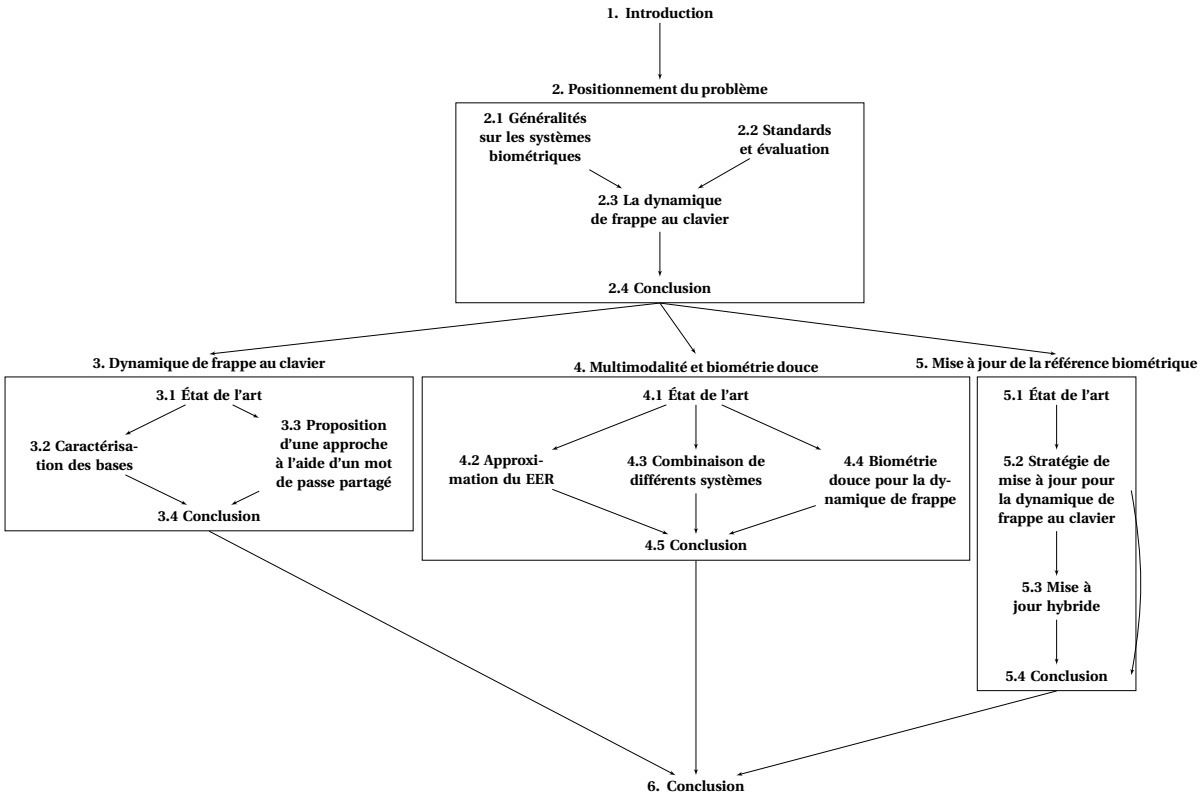


FIG. 1.1: Dépendances entre les chapitres de la thèse

Annexe F - Publications de l'auteur Cette annexe énumère les publications issues de nos contributions dans cette thèse. Nous avons également mis à disposition de la communauté scientifique : une application de création de base de données de DDF, ainsi que deux des bases de DDF les plus conséquentes de l'état de l'art.

2. Positionnement du problème

Sommaire

Présentation

Qu'est-ce que la biométrie ? Quelle est la position de la dynamique de frappe dans la biométrie ? Ce chapitre répond à ces questions en présentant le fonctionnement général des systèmes biométriques, la façon de les évaluer, et les différentes modalités possibles. La Dynamique De Frappe au clavier (DDF) est ensuite succinctement présentée et comparée aux autres modalités biométriques.

Mots clés

Biométrie, évaluation, dynamique de frappe au clavier.

Organisation du chapitre

Ce chapitre est organisé de la façon suivante. La section 2.1 présente d'abord les généralités sur les systèmes biométriques. Puis, la section 2.2 aborde les standards et la façon d'évaluer les systèmes biométriques, tandis que la section 2.3 présente quelques généralités sur la DDF.

2.1. Généralités sur les systèmes biométriques

Dans cette section, nous présentons brièvement les notions générales utilisées en biométrie.

2.1.1. Les modalités biométriques

La biométrie est initialement la science de « mesure du vivant », mais dans notre cas il s'agit de techniques « d'authentification et d'identification d'individus ». Les différentes modalités biométriques peuvent être hiérarchisées selon trois familles principales : les modalités (i) biologiques, (ii) comportementales, et (iii) morphologiques.

Les modalités biologiques sont basées sur l'analyse de *données biologiques* liées à l'individu. Nous pouvons citer l'analyse de l'ADN [Hashiyada, 2004], de l'odeur [Korotkaya, 2003], du sang ou de différents signaux physiologiques [Palaniappan, 2006; Phua *et al.*, 2008; Riera *et al.*, 2008].

Les modalités comportementales sont basées sur l'analyse des *comportements* de l'individu tels que la DDF [Gaines *et al.*, 1980; Giot *et al.*, 2011c], l'analyse de la signature [Fierrez et Ortega-Garcia, 2008], de la façon d'utiliser la souris [Weiss *et al.*, 2007], de la voix [Petrovska-Delacretaz *et al.*, 2007], de la façon de marcher [Han et Bhanu, 2006] ou de conduire [Benli *et al.*, 2008].

Les modalités morphologiques sont basées sur l'identification de *traits physiques* particuliers, qui, pour toutes personnes, sont permanents et uniques tels que le visage [Turk et Pentland, 1991], les empreintes digitales [Maltoni *et al.*, 2009], la forme de la main [Kumar et Zhang, 2006], le fond de l'œil [Xu *et al.*, 2005], ...

Il faut noter que certaines modalités peuvent être classées dans plusieurs catégories ; la voix est une modalité morphologique si on prend en compte le mécanisme des cordes vocales, ou une modalité comportementale si on prend en compte l'impact lié au stress de l'individu. La DDF, qui est la modalité qui nous intéresse dans cette thèse, est une modalité comportementale : il s'agit d'analyser le comportement de l'utilisateur pour saisir son texte au clavier de son ordinateur. Les modalités morphologiques sont à la fois les modalités les plus souvent utilisées, et les modalités ayant les meilleures performances (minimisation des erreurs de reconnaissance).

2.1.2. Cadre d'utilisation

Nous pouvons observer deux types de systèmes biométriques différents [Jain *et al.*, 2004b] qui permettent, (i) l'*identification*, ou (ii) la *vérification*.

L'identification permet au système biométrique de reconnaître un individu parmi l'ensemble des utilisateurs du système. Il s'agit d'une comparaison de type « 1:n », où l'utilisateur n'a pas besoin de clamer son identité. Le système retourne l'identité de l'individu identifié (dans un système d'identification ouverte), ou le rejette s'il ne correspond à aucun utilisateur du système (dans un système d'identification fermée).

La vérification permet au système biométrique de vérifier si l'identité de l'individu est bien celle qu'il clame être. Cette fois ci, il s'agit d'une comparaison « 1:1 ». Le système retourne la décision d'accepter ou rejeter l'utilisateur.

La figure 2.1 présente le fonctionnement de ces deux types de systèmes. Dans nos travaux, nous nous intéressons essentiellement à la vérification (ou authentification) biométrique.

2.1.3. Propriétés d'un système biométrique

Une modalité biométrique est considérée comme intéressante et exploitable, si la donnée biométrique utilisée satisfait différentes propriétés qui permettent d'obtenir des performances non négligeables [Jain *et al.*, 2004b] : (i) l'universalité, (ii) l'unicité, (iii) la permanence, (iv) la collectabilité, et (v) l'acceptabilité.

L'universalité implique que tous les individus qui doivent être identifiés avec le système possèdent la donnée biométrique utilisée par le système.

L'unicité implique que la donnée biométrique doit être relativement différente d'un individu à l'autre, afin de pouvoir les différencier.

La permanence implique que la donnée peut être collectée tout au long de la vie de l'individu, ou de l'utilisation du système, de l'individu.

La possibilité de collecte implique que la donnée biométrique doit pouvoir être collectée et mesurée afin de pouvoir être utilisée comme moyen de comparaison.

L'acceptabilité implique que les utilisateurs doivent accepter le système et être d'accord de l'utiliser[1].

En réalité, les modalités ne respectent pas nécessairement toutes les propriétés. Le tableau 2.1 présente les propriétés des principaux systèmes biométriques, et le tableau 2.2 présente leur performance. Lorsque l'unicité ne peut pas être vérifiée, la modalité reste malgré tout intéressante : elle peut être utilisée pour faire du *soft biometrics* (ou *biométrie douce*) [Jain *et al.*, 2004a]. Dans ce cas, la modalité ne permet pas de vérifier les individus, et encore moins de les identifier, mais elle permet de les catégoriser et elle peut être combinée avec un système plus général afin d'en améliorer les performances.

TAB. 2.1: Propriétés des modalités biométriques (le nombre d'étoiles dans la colonne performance est relié à l'efficacité de la reconnaissance)

Modalité	*Universalité*	*Unicité*	*Permanence*	*Collectabilité*	*Acceptabilité*	*Performance*
Modalités biologiques						
ADN	Oui	Oui	Oui	Faible	Faible	⋆⋆⋆⋆⋆
Groupe sanguin	Oui	Non	Oui	Faible	Non	⋆
Signal du cerveau (EEG)	Oui	Oui	Oui	Faible	Non	⋆⋆⋆⋆
Modalités comportementales						
Démarche	Oui	Non	Faible	Oui	Oui	⋆⋆⋆
Signature dynamique	Oui	Oui	Faible	Oui	Oui	⋆⋆⋆⋆
Dynamique de frappe	Oui	Oui	Faible	Oui	Oui	⋆⋆⋆
Voix	Oui	Oui	Faible	Oui	Oui	⋆⋆⋆⋆
Modalités morphologiques						
Iris	Oui	Oui	Oui	Oui	Un peu	⋆⋆⋆⋆⋆
Rétine	Oui	Oui	Oui	Oui	Un peu	⋆⋆⋆⋆⋆
Visage	Oui	Non	Faible	Oui	Oui	⋆⋆⋆⋆
Géométrie de la main	Oui	Non	Oui	Oui	Oui	⋆⋆⋆⋆
Veines de la main	Oui	Oui	Oui	Oui	Oui	⋆⋆⋆⋆⋆
Oreille	Oui	Oui	Oui	Oui	Oui	⋆⋆⋆⋆⋆
Empreinte digitale	Oui	Oui	Oui	Oui	Oui	⋆⋆⋆⋆

2.1.4. Fonctionnement d'un système biométrique

Les systèmes biométriques basés sur la vérification de l'identité d'un individu sont composés de deux modules principaux : (i) l'enregistrement et (ii) la vérification. *Le module d'enregistrement* consiste à effectuer les tâches suivantes :

1. la *capture* d'un ou plusieurs échantillons ;
2. le *pré-traitement* et l'*extraction* des données utiles sur chacun des échantillons ;
3. l'*apprentissage* et la génération d'une *référence biométrique* (également appelé *modèle*) de l'individu (voir 5.1.2);
4. le *stockage* du modèle, de préférence dans un conteneur de confiance.

1. Ainsi, un système très performant mais non accepté par les utilisateurs pour différentes raisons, n'est pas considéré comme étant bon.

TAB. 2.2: Ordre d'idée des taux aux de performance des différentes technologies biométriques selon l'état de l'art (les valeurs peuvent changer en fonction des bases de données)

Modalité	EER
Signal du cerveau (EER)	16 - 28%
Démarche	19 - 37%
Voix	5%
Dynamique de frappe	5%
Signature dynamique	5%
Empreinte digitale	2%
Veine de la main	1.15%
Iris	$\approx 0\%$
Visage	5 - 10%

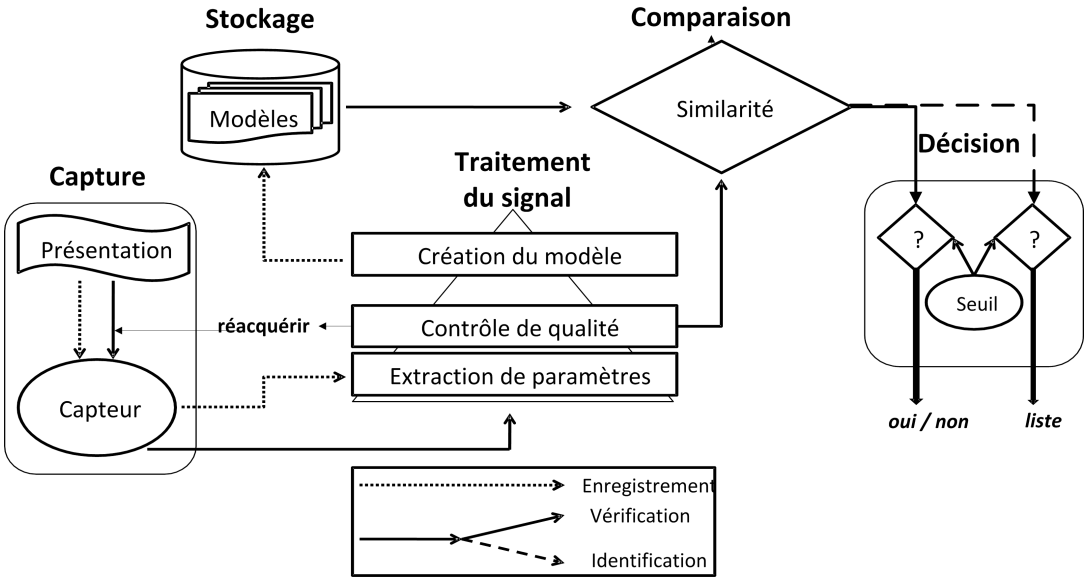


FIG. 2.1: Différence de fonctionnement entre l'identification et l'authentification biométrique

Le module de vérification consiste à effectuer les points suivants :

1. la *capture* d'un échantillon. Cet échantillon est appelé la *requête* ;
2. l'*extraction* des informations utiles ;
3. la *comparaison* de la capture avec la référence biométrique ;
4. la prise de la *décision* d'acceptation ou de rejet de l'individu ;
5. éventuellement, la mise à jour du modèle à l'aide de la nouvelle capture, afin de prendre en compte l'évolution temporelle de la donnée biométrique.

Comme nous ne traitons pas d'identification dans cette thèse, nous ne présentons pas le module d'identification.

2.2. Standards et évaluation

Afin de comparer les différents systèmes biométriques entre eux, il est nécessaire de pouvoir les évaluer. Il existe entre autre trois types d'évaluations : (i) les *évaluations objectives* qui consistent à quantifier le système biométrique ; (ii) les *évaluations subjectives* qui permettent d'analyser

le ressenti de l'utilisateur ; (iii) les *évaluations sur la sécurité du système biométrique* qui permettent de connaître les failles du système. Cette section n'est qu'une présentation succincte de l'évaluation ; pour plus d'informations, se rapporter à [El-Abed, 2011; ISO, 2006].

2.2.1. Évaluation objective

Plusieurs métriques peuvent être utilisées, mais comme elles sont trop nombreuses pour être toutes listées, nous présentons celles qui nous semblent les plus importantes :

- Les mesures d'erreurs :
  - le *Taux d'Échec à l'Enrôlement* (*Failure To Enroll Rate*, *FTER*) qui est la proportion des individus pour lesquels le système ne peut pas créer de modèle biométrique (ceci mesure le degré de non-universalité du système) ;
  - le *Taux d'Échec à l'Acquisition* (*Failure to Acquire Rate*, *FTA*) qui est la proportion des tentatives de capture de données biométriques qui échouent ;
  - le *Faux Rejet* (*False Rejection, FR*) lorsque le système rejette le demandeur, alors qu'il s'agit du bon utilisateur ;
  - le *Taux de Faux Rejet* (*False Rejection Rate*, *FRR*, spécifique à la vérification) qui est la proportion des transactions des demandeurs légitimes rejetées par erreur. Si le système comporte N_u transactions légitimes d'utilisateurs, le taux est estimé de la façon suivante :
  $$\widehat{FRR} = \frac{FR}{N_u} \tag{2.1}$$
  - la *Fausse Acceptation* (*False Acceptance, FA*) lorsque le système déclare l'individu comme étant l'utilisateur attendu, alors qu'il s'agit d'un imposteur ;
  - le *Taux de Fausse Acceptation* (*False Acceptance Rate*, *FAR*, spécifique à la vérification) qui est la proportion des transactions des imposteurs acceptées par erreur. Si le système comporte N_i transactions illégitimes d'imposteurs, le taux est estimé de la façon suivante :
  $$\widehat{FAR} = \frac{FA}{N_i} \tag{2.2}$$
  - le *Taux d'Erreur Égale* (*Error Equal Rate*, *EER*) qui est le taux d'erreur lorsque le système est configuré de telle façon à avoir le *FAR* égal au *FRR*.
- le temps de traitement pour les phases :
  - d'acquisition : temps nécessaire pour capturer la donnée biométrique ;
  - d'enregistrement : temps nécessaire pour générer le modèle de l'individu ;
  - de vérification : temps nécessaire pour l'acquisition de la donnée biométrique et sa comparaison avec le modèle de référence ;
  - d'identification : temps nécessaire pour l'acquisition de la donnée biométrique et sa comparaison avec les modèles de la base.
- La taille mémoire requise pour
  - les algorithmes ;
  - et les modèles.
- Le comportement global d'un système
  - La *Courbe caractéristique de performance (Receiver Operating Characteristic) (ROC)* permet d'évaluer la performance globale d'un système biométrique. Ce type de courbe est obtenue en faisant varier le seuil de décision de l'algorithme et en affichant la valeur du *FRR* en fonction du *FAR* ;

- La *Courbe Cumulative de Correspondance* (*Cumulative Match Characteristic Curve*, *CMC*) présente les valeurs du rang d'identification et les probabilités d'une identification correcte inférieure ou égale à ces valeurs. Elle permet de comparer les performances des systèmes d'identification.
- *Courbe de coût* (*Cost Curve*) permet de donner un coût général du système biométrique en fonction d'un seuil, en donnant des poids différents aux taux de faux rejets et de fausses acceptations.

2.2.2. Évaluation subjective

L'évaluation subjective permet de récolter l'avis des utilisateurs dans l'étude d'un système biométrique et de le prendre en compte pour améliorer le système tout en quantifiant son utilisabilité. Plusieurs propriétés peuvent être vérifiées :

- *L'acceptabilité* : il s'agit de savoir si l'utilisateur est prêt à utiliser le système biométrique.
- *La confiance* : il s'agit de déterminer le degré de confiance que l'utilisateur accorde au système (ce degré de confiance est bien souvent lié aux a priori de l'utilisateur sur le système, plutôt qu'aux performances réelles du système, voir l'annexe A) ;
- *La facilité d'utilisation* : il s'agit de savoir si l'utilisateur trouve le système simple à utiliser.

Ces différents points sont importants, car ce sont eux qui permettent de savoir si un système sera globalement accepté et utilisé par les utilisateurs. L'évaluation subjective peut donner des résultats différents de ceux issus de métriques de performance [El-Abed *et al.*, 2012].

2.2.3. Évaluation de la sécurité

Il est également important de prendre en compte la sécurité du système biométrique lors de son évaluation ou de son acquisition : en effet, il parait aberrant d'utiliser un système biométrique très performant, avec une forte acceptabilité, mais présentant de grosses failles de sécurité (par exemple, authentification faciale détournée avec l'utilisation d'une photo du visage du véritable utilisateur). Ratha *et al.* [2001a] présentent différentes attaques possibles. Les 8 classes recensées sont les suivantes :

1. *Données biométriques falsifiées* : une reproduction de la donnée biométrique utilisée est présentée au capteur biométrique (*cf.*, une copie d'une signature).
2. *Transmission de données biométriques interceptées* : une ancienne donnée biométrique enregistrée est rejouée dans le système sans passer par le capteur biométrique (*cf.*, présentation d'une ancienne copie de l'image de l'empreinte).
3. *Attaque sur le module d'extraction de paramètres* : ce module est remplacé par un cheval de Troie lui permettant de produire des informations choisies par l'attaquant.
4. *Altération de paramètres extraits* : après leur extraction par le module d'extraction de paramètres, les données sont altérées voire remplacées par d'autres données choisies par l'attaquant.
5. *Le module de comparaison est remplacé par un autre malveillant* : ce module peut être remplacé par un cheval de Troie afin de produire artificiellement des scores hauts ou bas.
6. *Altération de la base de données* : la base de données de modèles est disponible localement, à distance ou distribuée sur plusieurs serveurs. Dans ce type d'attaques, l'attaquant modifie un ou plusieurs modèles afin d'autoriser un imposteur ou d'empêcher un utilisateur légitime d'accéder à la base.

7. *Attaque sur le canal entre la base de données et le module de comparaison* : dans ce type d'attaque, les modèles sont altérés sur le canal reliant la base de données et le module de comparaison.
8. *Altération des décisions (accepté/rejeté)* : ce type d'attaques altère la décision (Oui/Non) prise par le module de comparaison. Cette attaque est considérée comme très dangereuse. En effet, même si le système est robuste en termes de performance, il est rendu inutile par ce type d'attaque.

2.3. La dynamique de frappe au clavier

2.3.1. Comparaison aux autres modalités

Les tableaux 2.1 et 2.2 présentent respectivement les propriétés et performances des modalités majeures, afin de les comparer à celles de la DDF.

Globalement, les performances des systèmes utilisant la DDF sont inférieures à celles des systèmes morphologiques majeurs. Cependant, le coût d'utilisation d'une telle modalité est nettement inférieur en raison de l'absence d'un capteur spécifique. Ces plus faibles performances peuvent s'expliquer par plusieurs raisons. En tant que modalité comportementale, elle est *de facto* plus sujette à une variabilité au cours du temps par rapport aux modalités biologiques ou morphologiques. De plus, la façon de taper au clavier est dépendante de l'état émotionnel de l'individu [Epp, 2010]. Une autre raison est que peu de recherches ont été effectuées sur cette modalité, et, par conséquent il y a un plus faible nombre d'études pouvant apporter des améliorations à cette modalité. Le figure 2.2 présente, à titre informatif, le nombre de papiers référencés par « Google scholar » pour différentes biométries.

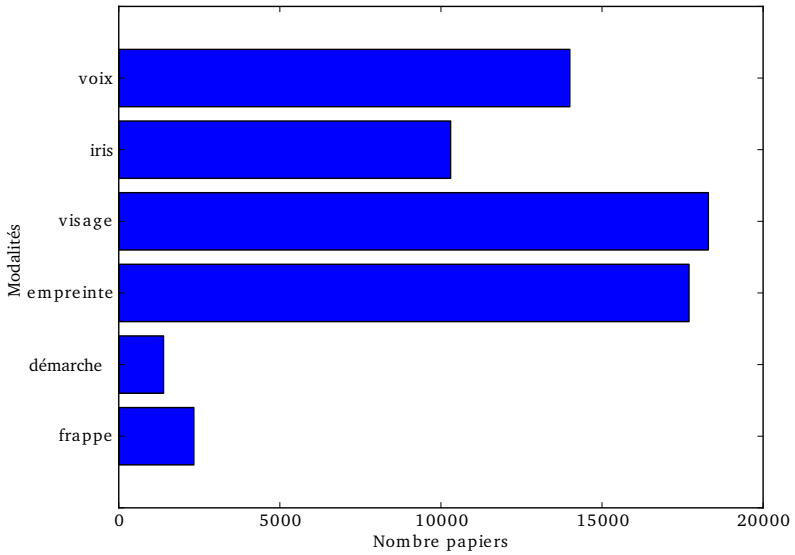


FIG. 2.2: Nombre de documents référencés par © Google Scholar pour chaque modalité biométrique

2.3.2. Applications de la dynamique de frappe au clavier

Dans notre thèse, nous nous intéressons à la DDF pour vérifier l'identité d'individus qui ont saisi leur propre mot de passe. Cependant, les informations liées à la DDF peuvent être utilisées dans d'autres types d'applications. Cette sous-section présente un ensemble de cas d'utilisations possibles abordés dans la littérature.

2.3.2.1. Authentification forte

En plus d'un mot de passe, un système de DDF peut être utilisé comme moyen d'authentification forte. En effet, l'utilisateur est authentifié car les deux éléments suivants sont présents :

- *ce que l'on connaît* : l'utilisateur connaît le mot de passe ;
- *ce que l'on sait faire* : l'utilisateur a été reconnu à sa façon de taper, il ne s'agit pas d'un imposteur.

Le système devient donc moins sensible aux vols de mots de passe, tant que la façon de faire de l'utilisateur reste unique et non reproductible. Dans le cas où *ce que l'on connait* est finalement connu par un tiers, celui-ci ne pourra pas se connecter car il ne saura pas reproduire *ce que l'on sait faire*. Dans un environnement nécessitant un faible niveau de sécurité, *ce que l'on connaît* pourrait même être connu de tout le monde et ne pas faire partie du système de sécurité.

2.3.2.2. Faciliter la vie de l'utilisateur

Les systèmes dont la sécurité est basée sur une authentification par mots de passe statiques (saisie d'un mot de passe connu à l'avance) sont très pratiques (coté service). Cependant, l'utilisation de tels systèmes peut être compliquée et peu agréable à utiliser pour les utilisateurs en raison des différentes contraintes qu'il est nécessaire de respecter pour en améliorer la sécurité. Pour qu'un système classique à mot de passe soit performant, plusieurs points doivent être respectés [Conklin *et al.*, 2004] :

1. l'utilisateur doit changer régulièrement son mot de passe ;
2. le mot de passe doit être suffisamment long ;
3. le mot de passe doit être compliqué et utiliser des caractères non alphanumériques ;
4. le mot de passe doit rester secret.

Ces différents points sont généralement perçus comme une contrainte pour l'utilisateur qui ne les respecte donc pas forcément, et diminue la sécurité du système en les contournant. De plus, toujours en raison de ces contraintes, la probabilité que l'utilisateur oublie son mot de passe est relativement importante, notamment si celui-ci ne l'utilise pas régulièrement. Dans le cas de l'authentification sur un site internet, il est généralement possible de demander au service de fournir en clair le mot de passe, ou d'en saisir un nouveau. Cependant, cette procédure induit deux problèmes supplémentaires :

- souvent, le mot de passe est envoyé par courrier électronique, qui est un canal non sécurisé la plupart du temps ;
- l'utilisateur n'a aucune maitrise de la vie du mot de passe (Transit en clair sur le réseau ? Stockage en clair dans la base de données ?);
- afin de vérifier si l'utilisateur qui fait la demande du mot de passe est le bon, le système pose des questions, définies à la création du compte. Malgré tout, l'utilisateur peut très bien en avoir également oublié la réponse ou l'attaquant peut trouver cette information sur un réseau social.

En utilisant un système d'authentification par DDF sur texte statique, il n'est plus nécessaire de respecter ces contraintes, d'autant plus que nous avons montré [Giot *et al.*, 2012b] (dans des travaux non présentés dans ce manuscrit) que le système de reconnaissance n'est pas plus performant en utilisant un mot hors du dictionnaire par rapport à un mot du dictionnaire (d'autant plus que le taux d'échec à l'acquisition (*failure-to-acquire rate*) (FTAR) est plus élevé si on utilise un mot hors du dictionnaire).

2.3.2.3. Surveillance et authentification continue

Un autre exemple d'application est la surveillance de la machine [Rao, 2005; Song *et al.*, 1997]. Il s'agit de vérifier en temps réel si l'utilisateur utilisant le clavier de la machine est bien celui qui s'y est connecté. Ainsi, il est possible au cours de l'utilisation de la machine de se rendre compte si un utilisateur oublie de se déconnecter et qu'un autre a pris sa place ce qui permet de le déconnecter automatiquement.

Un autre type de surveillance pouvant être mis en place est la surveillance de saisies invalides pouvant indiquer que l'utilisateur est assoupi, stressé, perturbé, distrait ou a adopté un comportement inhabituel [Monrose et Rubin, 2000]. Ce type de surveillance peut avoir un intérêt dans les systèmes où l'utilisateur doit rester concentré (*cf.*, pour la surveillance du trafic aérien).

La surveillance peut poser de nombreux problèmes en raison de la difficulté de configurer de tels systèmes et de l'existence de faux négatifs. La DDF peut donc également être utilisée dans un cadre d'audit. Les données ne sont utilisées qu'a posteriori en cas de problème afin de vérifier si, à l'instant t, l'utilisateur ayant utilisé le clavier était bien la bonne personne.

2.3.2.4. Surveillance ssh afin de deviner le mot de passe

Song *et al.* [2001] présentent une application écoutant le réseau et capable de deviner, ou diminuer le domaine de recherche pour deviner, le mot de passe de l'utilisateur à travers une session SSH. L'attaque se base sur le fait que chaque événement pression d'une touche du clavier est envoyé dans un paquet IP. La surveillance de l'envoi de ces paquets est susceptible de donner des informations. Il faut cependant noter que les techniques de gestion de DDF ne sont pas utilisées pour faire de l'authentification, mais, pour deviner quelles touches ont été saisies.

2.3.2.5. Attaque de *Pretty Good Privacy*

Pretty Good Privacy (PGP) est un ensemble de logiciels permettant la sécurisation des messages électroniques. Sa sécurité est basée sur l'utilisation de signatures électroniques, de chiffrement et de hachage. PGP utilise les temps de latence au clavier pour générer les clés (lors de la génération de la clé, il est demandé à l'utilisateur d'utiliser son clavier). Monrose et Rubin [1997] pensent qu'il est possible d'utiliser la DDF pour générer une clé identique en faisant une saisie identique.

2.3.2.6. Détection automatique de programmes malicieux

La DDF peut également être utilisée pour différencier un utilisateur humain d'une machine afin de permettre la détection des robots ayant pris le contrôle d'une machine afin d'empêcher d'effectuer certaines actions si l'utilisateur n'a pas utilisé le clavier. Il est possible de monitorer les évènements du clavier de la machine et de différencier ceux générés par un robot, et ceux générés par un humain [Stefan et Yao, 2008].

2.3.2.7. Gestion de licences logicielles

Une autre utilisation possible de la DDF, mais n'ayant aucune illustration pour le moment, est la gestion de licence logicielle. L'utilisateur saisit plusieurs fois un mot de passe lors de l'acquisition du logiciel sur un site de vente en ligne, puis, lors de l'installation et l'utilisation du logiciel, ce même mot de passe lui est demandé. Si le mot de passe n'est pas saisi de la même façon que lors de l'enregistrement de la licence, l'utilisateur n'est pas considéré comme légitime et l'accès au logiciel lui est interdit.

2.3.2.8. Personnalisation d'interface

Nous verrons en section 4.4 qu'il est possible de catégoriser les utilisateurs en fonction de leur catégorie d'âge ou leur genre. Il est peut-être possible de les catégoriser suivant d'autres critères. Ces informations peuvent être utiles pour permettre à certaines applications de proposer des fonctionnalités spécifiques à certaines catégories. Ainsi l'expérience utilisateur peut être améliorée. Il n'existe pas encore, à notre connaissance, de produits utilisant cette technique.

2.4. Conclusion du positionnement du problème

Nous avons vu que la DDF est une modalité comportementale intéressante grâce à son faible coût et sa forte acceptabilité. En contrepartie, ses performances de reconnaissance sont plus faibles a priori que celles des modalités majoritairement étudiées dans la littérature. Il reste donc à améliorer les performances de systèmes d'authentification basés sur la DDF, ce qui n'est pas forcément possible à cause du caractère intrinsèquement variable de la DDF. Dans le cas de systèmes à mot de passe partagé, nous proposons d'utiliser les données d'imposteur pour améliorer les performances. Le chapitre 3 présente cette contribution ainsi qu'une analyse des bases de DDF disponibles et futures.

Une autre solution est l'utilisation de systèmes multimodaux associant la DDF avec une ou plusieurs autres modalités. Ainsi la décision d'acceptation d'un individu est basée à la fois sur le système de reconnaissance par DDF, et sur un autre système utilisant une autre modalité supposée indépendante. Cependant, le coût d'un tel système est nécessairement plus élevé. Une autre alternative est l'utilisation de données de types biométrie douce acquises grâce à la donnée biométrique de DDF. De cette façon, de meilleures performances peuvent être obtenues sans augmenter le coût du système. Le chapitre 4 défini les contributions sur ce sujet.

Un des problèmes majeurs de la DDF est sa sensibilité au vieillissement de la donnée biométrique. Plus le temps passe, moins la référence biométrique correspond à la donnée biométrique réelle de l'utilisateur. En effet, la référence biométrique représente l'utilisateur au moment où celui-ci a effectué l'étape d'acquisition, mais entre-temps, sa façon de saisir le mot de passe a évolué. Il est donc nécessaire d'appliquer des mécanismes de mise à jour du modèle. Le chapitre 5 propose des solutions à ce problème.

3. Dynamique de frappe au clavier

Sommaire

Présentation

Comment fonctionne la DDF ? Quels sont les travaux existants dans le domaine ? Quelles sont les données publiques disponibles pour travailler sur le sujet ? Voici les trois questions qui sont traitées au début de ce chapitre et permettront d'illustrer l'intérêt des travaux réalisés par la suite. Comment comparer ou sélectionner les bases de DDF pour répondre à une problématique particulière ? Quelle est la performance des algorithmes de l'état de l'art, capables de fonctionner en environnement faiblement contraint, évaluée sur un jeu de données commun ? Comment obtenir de meilleures performances de reconnaissance dans les systèmes utilisant un mot de passe commun à tous les utilisateurs ? Voici les trois questions qui concernent notre contribution à la DDF et sont traitées dans la suite de ce chapitre.

Nous insistons sur la mauvaise qualité des bases de DDF publiques, ainsi que sur la difficulté à comparer les études entre elles en raison de la grande différence de protocoles d'acquisition et d'expérimentation. Nous faisons une comparaison des études (les plus proches de notre contribution et de nos contraintes) de l'état de l'art sur une base publique que nous avons créées dans ce but. Cette base est une des deux bases publiques les plus complètes et pertinentes pour évaluer objectivement les recherches du domaine. Nous développons également dans ce chapitre notre méthode de reconnaissance par texte partagé.

Mots clés

Dynamique de frappe au clavier, état de l'art, comparaison de bases, mot de passe partagé.

Contributions de ce chapitre

- Une méthodologie de comparaison de bases de données d'authentification par DDF, afin d'aider à sélectionner, ou créer, des bases pertinentes pour des études précises.
- Une nouvelle base de données de DDF de meilleure qualité que la majorité des bases publiques.
- La comparaison de différentes méthodes de l'état de l'art, respectueuses de nos contraintes sur la base que nous fournissons.
- La proposition d'une méthode de reconnaissance par mot de passe partagé, plus performante que les méthodes de l'état de l'art : en termes de taux d'erreur et de temps de calcul.

Organisation du chapitre

Les trois premières sections présentent l'état de l'art de la DDF. La section 3.1.1 présente l'historique des travaux sur la DDF, des années 80 à aujourd'hui et la section 3.1.2 présente les principes généraux de fonctionnement de la DDF.

Les sections suivantes présentent notre contribution à la DDF. La section 3.2 présente les bases publiques pour l'authentification statique par DDF, ainsi qu'une méthode pour les caractériser objectivement afin de faciliter : la sélection, la comparaison, ou la création de bases pour de nouvelles études. La section 3.3 présente l'approche que nous avons développée dans le cas d'une authentification par mot de passe partagé, ainsi que sa comparaison aux méthodes de l'état de l'art.

3.1. État de l'art de la dynamique de frappe au clavier

3.1.1. Rappels historiques

Un historique de très bonne qualité sur les différentes études de DDF a été fait par Zhou [2008] ; nous nous en sommes grandement inspiré pour écrire cette partie. Nous présentons uniquement les études qui ont eu un impact dans l'histoire de l'authentification par DDF.

3.1.1.1. 1980-1989 : premières études et brevets

Les premières recherches dans l'étude de la DDF datent de 1980 avec le rapport de la Rand Corporation [Gaines *et al.*, 1980]. Le gouvernement américain a financé cette étude afin de vérifier si, comme dans le cas des télégraphes, il est possible de différencier les individus en fonction de leur façon de taper au clavier. Les premières techniques sont donc apparues dans cette étude préliminaire. C'est ainsi que les *digraphes* ont été présentés : il s'agit du temps écoulé entre le moment où une touche est pressée et où la suivante est relâchée. Ainsi, dans cette étude, les digraphes des différents participants ont été mesurés, et les données suivantes calculées : la *moyenne*, la *variance*, le *kurtosis*. La classification des utilisateurs est ensuite faite à l'aide d'un *t-test*. Les chercheurs ont pu démontrer, à l'aide d'analyses statistiques, que la DDF est bel et bien une donnée biométrique. Les temps des digraphes supérieurs à 500ms n'ont pas été pris en compte.

Suite à cette étude, d'autres chercheurs ont tenté d'améliorer le procédé. En 1985, Umphress et Williams ont fait une expérience donnant plus de crédibilité à la DDF [1] [Umphress et Williams, 1985]. Une nouvelle étude a été faite en 1988 par Williams et Leggett puis en 1989 par Umphress,

Williams et Leggett. Ces études utilisent des outils statistiques pour comparer la saisie de l'utilisateur avec le modèle de l'utilisateur attendu. Elles ont une nouvelle fois permis de démontrer que les digraphes sont une bonne mesure pour la DDF, et qu'il est préférable de ne pas prendre en compte les temps supérieurs à 500ms.

En 1986, un brevet américain déposé par Garcia décrit un schéma dans lequel les utilisateurs saisissent leur nom afin de s'authentifier [Garcia, 1986]. Il suppose que le mot de passe est forcément facile à retenir et que la saisie sera plus constante du fait de saisir quelque chose d'habituel. Une idée proposée dans le brevet est d'utiliser le vecteur moyen des temps de digraphes comme modèle, une distance de Mahalanobis étant ensuite effectuée avec la requête. Si cette distance est supérieure à 100, l'utilisateur est rejeté, tandis que si elle est inférieure à 50, l'utilisateur est accepté. Dans le cas, où elle est entre 50 et 100, il est nécessaire qu'il fasse une nouvelle saisie. Le brevet décrit un autre système où les utilisateurs saisissent 10 fois 1000 des mots les plus courants afin de leur générer un profil, tandis que l'authentification se fait en saisissant une phrase générée aléatoirement.

Un autre brevet est déposé en 1989 [Young et Hammon, 1989]. Celui-ci mentionne l'utilisation des temps de latence et de la pression exercée sur le clavier. La méthode d'authentification utilise également comme modèle un vecteur de digraphes. Cependant, la mesure utilisée est une distance euclidienne entre la requête et le modèle.

La plupart des recherches utilisent des méthodes équivalentes à celles présentées dans les brevets. Le principe général étant la mesure de digraphes, leur stockage dans des vecteurs et le calcul d'une distance entre le modèle de l'utilisateur attendu et la requête. L'utilisateur est accepté si la distance ne dépasse pas une certaine valeur (dans le cas d'une identification, il s'agit de l'utilisateur ayant la plus petite distance).

3.1.1.2. 1990-1999 : amélioration des performances et de l'utilisabilité

Les années 90 ont permis aux chercheurs d'améliorer les performances des méthodes de DDF tout en améliorant leur utilisabilité : donc de diminuer les taux de FAR et FRR ainsi que le nombre d'exemples utilisés pour créer le modèle. En 1990, Joyce et Gupta déclarent que le FRR en dessous de 1% et le FAR en dessous de 5% est acceptable. Ils ont utilisé dans leur expérience le nom de la personne, un mot de passe et deux phrases pour créer le modèle d'un utilisateur. À l'aide d'une distance euclidienne, ils ont obtenu un FAR de 6,67% et un FRR inférieur à 1%. Ils proposent d'utiliser leur principe comme moyen de sécurité pour détecter une intoxication ou une fatigue.

Différentes études ont été effectuées à l'aide de méthodes utilisant un calcul de *distance minimum* ou un *classifieur bayésien* [Bleha *et al.*, 1990]. Bleha *et al.* affirment que plus le mot de passe est long, plus l'erreur d'identification est faible ; plus le nombre d'exemples utilisés pour calculer le modèle est important, plus l'erreur est faible ; utiliser une réduction de dimension du vecteur de caractéristiques permet d'améliorer les résultats (ici, analyse discriminante de *Fisher*).

Rogers et Brown proposent en 1996 d'utiliser des réseaux de neurones comme méthode de classification dans un brevet qu'ils déposent [Rogers et Brown, 1996]. Ils concluent que la taille idéale d'un mot de passe est entre 11 et 25 frappes, tandis que le nombre d'exemples d'apprentissage est de 20.

En 1997, Monrose et Rubin ont travaillé sur l'analyse des textes libres [Monrose et Rubin, 1997]. Ils recommandent de séparer les utilisateurs en différents groupes (en fonction de leur vitesse de frappe) afin d'accélérer le temps d'identification. Leurs techniques sont équivalentes aux précédentes. Bien que l'idée soit intéressante, elle n'a été reprise que dans une seule autre

1. Les résultats de [Gaines *et al.*, 1980] étaient considérés avec scepticisme du fait de la faible quantité de personnes ayant pris part au protocole.

étude plus récente [Hocquet *et al.*, 2006], alors qu'il pourrait s'agir d'une piste de recherche utile. Pendant ce temps, Obaidat et Sadoun ont également procédé à différentes études sur des méthodes statistiques et des réseaux de neurones [Obaidat et Sadoun, 1997]. Les meilleures performances ont été obtenues en utilisant à la fois les temps de latence et la durée de pression. Ils ont obtenu un taux d'erreur de 0% à l'aide d'un des réseaux de neurones et un FAR de 1,9% et FRR de 0,7% en utilisant une mesure de distance.

Song *et al.* [1997] présentent un travail utilisant la DDF pour surveiller les individus et fermer la session de l'utilisateur s'il est détecté comme étant un imposteur. Les données sont stockées sous forme de bigrames (temps entre deux touches), trigrames (temps entre trois touches) et motgrame (temps pour un mot). Un modèle pour chacune de ces données est généré. La prédiction d'appartenance au modèle utilisateur d'une touche est basée à la fois sur la prédiction précédente et la comparaison de l'évènement clavier avec le modèle.

Robinson, Liang, Chambers, et MacKenzie ont confirmé l'idée que les temps de pression sont supérieurs aux temps de latences [Robinson *et al.*, 1998]. Pour leur étude, ils ont utilisé des données issues d'authentifications réelles d'étudiants (cependant, ils n'ont sélectionné que 10 étudiants sur 140 pour leur étude). Les meilleurs résultats obtenus présentent un FAR de 10% et un FRR de 9%.

3.1.1.3. Depuis 2000 : nouvelles idées et remises en question

Les recherches effectuées ces dix dernières années en DDF ont fait émerger de nouvelles idées et différents produits commerciaux basés sur de tels systèmes ont été développés.

En 2000, les chercheurs ont continué à améliorer l'utilisation des réseaux de neurones afin de les rendre utilisables en dehors d'un laboratoire. Cho, Han, et Kim ont essayé de faire en sorte que les réseaux de neurones classent les utilisateurs sans avoir besoin de grande quantité de données d'imposteurs[2] [Cho *et al.*, 2000]. Des mots de passe courts (7 caractères) ont été utilisés et les résultats du réseau de neurones ont été comparés à ceux d'un classifieur statistique (k plus proches voisins, avec $k = 1$) ; le réseau de neurones a été plus performant avec un FAR de 4% et un FRR de 0%.

Un nouvel usage de la DDF a été présenté par Monrose, Reiter, et Wetzel : ils ont proposé l'idée de l'utiliser afin de durcir les mots de passe. Dans ce cas, la DDF n'est pas utilisée directement en tant que moyen d'authentification, mais fait partie d'un processus plus global.

Une autre étude présentée en 2002 a permis d'améliorer les performances de la DDF par texte libre. Bergadano, Gunetti, et Picardi ont proposé l'idée d'utiliser une mesure relative à la différence d'une mesure absolue [Bergadano *et al.*, 2002]. De plus, il s'agit de l'étude ayant le plus grand nombre de participants (même aujourd'hui) y ayant pris part : 154 volontaires ! Cette nouvelle technique a permis d'obtenir un FAR de 4% et un FRR de 0,01%. L'idée sous-jacente est d'utiliser comme vecteur de référence les mesures de trigraphes et de les ordonner du plus petit au plus grand. Le vecteur de l'utilisateur s'authentifiant est trié de la même façon et une mesure de désordre est calculée entre les deux. Les trigraphes ont été reconnus comme étant plus performants que les digraphes (du moins dans cette méthode).

Améliorer la performance des systèmes est de plus en plus difficile, cependant il est toujours possible d'améliorer la consistance de la frappe de l'utilisateur afin de diminuer les erreurs. Hwang, Lee, et Cho soutiennent que la qualité des mesures utilisées pour créer le vecteur de référence est plus importante que leur quantité [Hwang *et al.*, 2006]. Ainsi, ils augmentent la consistance de la donnée biométrique en utilisant des pauses synchronisées à l'aide de signaux et ainsi améliorent les performances d'authentification sans modifier les algorithmes.

2. Dans un cas réel d'authentification par mot de passe, ces données ne sont pas disponibles.

Hocquet, Ramel, et Cardot présentent en 2005 trois méthodes différentes (basées sur des adaptations et améliorations de méthodes existantes) utilisées en fusion afin d'améliorer les performances. Les méthodes sont de type statistique, basées sur le rythme de frappe, ou la mesure du désordre. Le taux d'égales erreurs (*Equal Error Rate*) (EER) obtenu avoisine les 5%, mais il a été calculé avec une base relativement petite. Sang, Shen, et Fan quant à eux, ont testé la DDF en utilisant un Séparateur à Vaste Marge (SVM). Ils ont testé des SVM à 1 et 2 classes. Dans le cas du SVM à 2 classes, les données des imposteurs ont été générées à partir des données de l'utilisateur. Seulement 10 utilisateurs et 5 imposteurs ont pris part à l'étude. Le SVM à une classe a de meilleurs performances et un temps de calcul plus faible que le système à base de réseaux de neurones.

En 2006, Hocquet, Ramel, et Cardot proposent une méthode permettant de configurer automatiquement les paramètres des algorithmes pour chaque utilisateur (seuils des algorithmes, paramètres divers, poids pour les fusions) [Hocquet *et al.*, 2006]. En effet, il est connu qu'utiliser des paramètres individuels donne de meilleures performances qu'utiliser des paramètres globaux. L'idée étant de disposer de classes d'utilisateurs et de trouver les paramètres optimums pour chacune d'entre elles. Ensuite, chaque utilisateur est automatiquement assigné à une classe et utilise ses paramètres.

Des modèles à base de mélanges de gaussiennes [Hosseinzadeh et Krishnan, 2008] ont également été utilisés. Cette méthode reste relativement efficace, mais il est nécessaire de disposer de 30 exemples d'enregistrement pour créer le modèle. Des études ont également représenté les données extraites de DDF sous forme de chaines d'acides aminés [Revett, 2009] de façon telle à pouvoir utiliser des méthodes de comparaison de chaines de caractères. Bien que les deux méthodes présentées précédemment donnent des résultats intéressants, il faut noter qu'elles permettent plusieurs essais à l'utilisateur en cas d'échec.

Cette décennie a également permis à des chercheurs de commencer à s'intéresser au lien entre les performances et le matériel utilisé. Ainsi, plusieurs études ont vérifié le lien entre les performances d'un système biométrique et la résolution de l'horloge [Killourhy et Maxion, 2008; Narainsamy *et al.*, 2010]. Les résultats montrent que plus la fréquence d'horloge est élevée, plus la performance des systèmes est importante, mais que cette fréquence dépend fortement du matériel et du système d'exploitation utilisé. Killourhy et Maxion [2011] montrent également que la majorité des études sur le sujet manquent de rigueur scientifique. Cela implique une faible confiance à accorder aux résultats (peu d'études présentent une validation statistique) et une impossibilité de comparaison (les comparaisons sont inexistantes ou faites sur des bases privées).

3.1.1.4. Brevets et produits commerciaux

De nombreux brevets ont également été déposés durant cette décennie [Bender et Postley, 2007; Cho et Han, 2000; Cho, 2006; Cho et Hwang, 2009; Davis *et al.*, 2009; Gilfix *et al.*, 2008; Paula *et al.*, 2005; Phoha *et al.*, 2005; Pohoa *et al.*, 2009; Primeaux *et al.*, 2001; Revett, 2007; Ross, 2004; Schreiber et Knox, 2007; Serpa, 2005; Zilberman, 2002]. Les informations présentées ne sont pas fondamentalement différentes de ce que l'on peut trouver dans la littérature scientifique. Il existe également quelques produits commerciaux, mais nous ne disposons pas de réelles évaluations de ceux-ci.

Authenware Corp. (`http://www.authenware.com/whatis.php`) est une société proposant des authentifications à base de DDF.

bioChec^{TM} (`http://www.bioChec.com`) propose une application brevetée. Il semble également que le système utilise des mécanismes d'adaptation de modèle (il n'y a pas d'enregistrement, les données sont capturées au fil des authentifications).

Biopassword doit être l'un des plus anciens produits. `http://www.admitonesecurity.com/`

La société Delfigo Security (`http://www.delfigosecurity.com/products`) propose aussi des authentifications à base de DDF, mais ne donne pas d'information à ce sujet.

La société Deepnet Security propose l'application Typesense (`http://www.deepnetsecurity.com/products2/TypeSense.asp`) qui utilise un mécanisme d'apprentissage incrémental.

iMagic Software (`http://www.imagicsoftware.com`) propose des solutions basées sur la DDF (cependant, ils nomment la technique "trustable passwords") fonctionnant en environnement web. L'entreprise détient au moins un brevet [Bender et Postley, 2007].

KeystrokeID (`http://www.idcontrol.net`) est également une application commerciale utilisant la DDF.

Probayes (`http://www.probayes.com/index.php/en/products/applications/keystroke-dynamics`) semble être la seule entreprise française proposant également de la DDF. Cette application est également liée à un brevet et repose sur la théorie bayésienne.

La société Psylock (`http://www.psylock.com`) propose des applications utilisant la DDF. Cependant, il semble que contrairement aux concurrents, l'authentification utilise un secret partagé plutôt qu'un mot de passe (la création du modèle bénéficie des exemples des autres imposteurs).

Nous pouvons donc voir que plusieurs entreprises proposent des applications de DDF. Un certains nombre d'entre elles peuvent fonctionner en environnement web (un point qui n'a pas été particulièrement abordé dans la littérature). Cependant, les taux d'erreurs qu'elles présentent sur leur site web sont généralement inférieurs à ce que l'on peut trouver dans la littérature. Naturellement, chacune des sociétés présente son produit comme étant le meilleur. Il est possible que des brevets soient associés à ces logiciels, mais cette information n'est pas toujours mise en avant par les éditeurs.

3.1.2. Principe de fonctionnement de la dynamique de frappe au clavier

Cette sous-section présente le fonctionnement de l'authentification par DDF, et la figure 3.1 présente la taxonomie des différents systèmes de dynamique de frappe au clavier. Deux familles principales cohabitent : (a) l'authentification par texte libre où l'utilisateur ne saisit pas toujours la même chose, et (b) l'authentification par texte statique où l'utilisateur tape toujours le même texte. Concernant à l'utilisation par texte libre, il peut exister l'authentification continue qui consiste à vérifier en permanence l'identité de l'utilisateur et l'authentification par défi qui consiste à demander à l'utilisateur de saisir un texte qu'il ne connait pas à l'avance. Nous nous intéressons principalement au cas de textes statiques avec mot de passe (ou passphrase). Nous désignons par passphrase un mot de passe commun à tous les utilisateurs (une sorte de secret partagé).

3.1.2.1. Capture et extraction des données

3.1.2.1.1. Capture des données brutes La capture consiste à enregistrer les données biométriques *brutes* de l'utilisateur. Le moyen de capture peut être différent en fonction du contexte de cette capture (enregistrement, identification ou vérification). Elle peut être effectuée sur n'importe quelle machine disposant d'un clavier. Cependant, même s'il existe des études sur appareils mobiles [Clarke et Furnell, 2006; Hwang *et al.*, 2008; Karatzouni et Clarke, 2007; Myung, 2004], il n'existe pas, à notre connaissance, d'études conséquentes utilisant un écran tactile faisant office de clavier afin de pouvoir prendre en compte des paramètres supplémentaires (par exemple, la pression du doigt, la taille du doigt, la localisation du doigt[3]) ; il s'agit donc d'une piste qu'il serait intéressant d'explorer. Néanmoins, plusieurs études ont déjà été réalisées sur des claviers

3. cf. android API `http://developer.android.com/reference/android/view/MotionEvent.html`

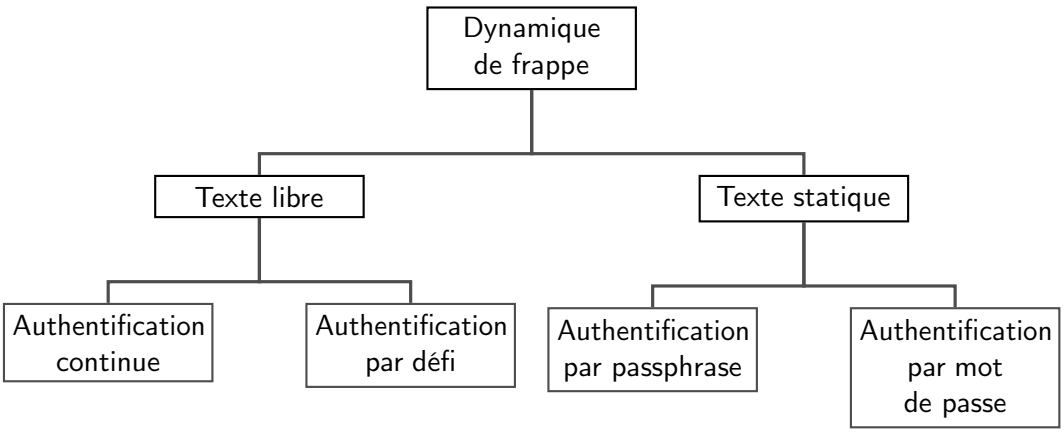


FIG. 3.1: Taxonomie des différents types de systèmes de dynamique de frappe

intégrant des capteurs de pression, ou des pavés tactiles simulant le comportement d'un écran tactile [Allen, 2010; Eltahir *et al.*, 2008; Grabham et White, 2008]. Il est également possible d'utiliser le son des touches du clavier [Dozono *et al.*, 2007] ou une vidéo de la frappe [**?**]. Comme nous nous sommes intéressés uniquement à des systèmes à bas coût, nos travaux utilisent un clavier d'ordinateur standard. Dans le cas des systèmes d'authentification par texte statique, la capture des données est effectuée grâce à la saisie du mot de passe de l'utilisateur. La plupart du temps, cette capture consiste à enregistrer les données brutes suivantes :

- le *type d'évènement* clavier (pression ou relâchement de la touche générant l'évènement) ;
- le *code* de la touche pressée. Malheureusement, celui-ci peut dépendre du type de clavier ou de la bibliothèque graphique utilisés, et pourrait poser des problèmes d'interopérabilité sur des systèmes hétérogènes (le code d'une même touche peut être différent) ;
- l'*instant* de l'évènement, qui est souvent une date exprimée en milli-secondes, mais dépend, à la fois, du système d'exploitation de la machine et de la résolution de son horloge [Killourhy et Maxion, 2008; Narainsamy *et al.*, 2010].

3.1.2.1.2. Extraction des données utiles Les données brutes ne sont pas directement utilisées pour la reconnaissance, il est donc nécessaire de passer par une étape *d'extraction des données*. Celle-ci consiste à construire les données utilisées par les algorithmes de reconnaissance. D'une façon générale, les données suivantes sont extraites :

- le temps de pression des touches (noté PR). Il se calcule en soustrayant l'instant du relâchement de la touche et l'instant de sa pression ;
- le temps de latence entre l'appui des touches (noté PP). Il se calcule en soustrayant l'instant de la pression d'une touche à l'instant de la pression de la touche précédente.

Rogers et Brown [1996] présentent également d'autres temps de latence qui peuvent être discriminants :

- le temps entre le relâchement d'une touche et la pression d'une autre (noté RP);
- le temps entre deux relâchements de touches (noté RR);
- ainsi que la pression sur le clavier[4] est une donnée brute capturée.

4. qui nécessite un clavier spécial, et donc sort de notre cadre d'utilisation (d'autant que peu d'études utilisent ce mécanisme)

Ainsi, RR et RP sont deux autres temps de latence qui sont utilisés dans certains articles. Souvent, le mode de calcul du temps de latence utilisé dans le papier n'est pas explicité et peut correspondre à n'importe lequel des trois. De façon plus épisodique, les données suivantes ont également été utilisées [Ilonen, 2003] :

- la vitesse de frappe de l'intégralité de la chaîne de caractères ;
- la fréquence des fautes de frappe[5] (utilisation de la touche « retour chariot », par exemple). Cependant, à notre connaissance, cette technique n'a été utilisée que dans les authentifications à base de texte libre ;
- l'ordre de frappe des touches lors de la présence de majuscules.

Ces valeurs permettent, la plupart du temps, la création d'un vecteur de temps qui est utilisé pour les traitements suivants (plusieurs types de temps sont souvent concaténés pour former un plus grand vecteur). À titre d'exemple, la figure 3.2 présente les données de DDF extraites par le même utilisateur 90 fois. Un pic correspond la plupart du temps à une hésitation de l'utilisateur (engendrant un temps plus grand du paramètre mesuré).

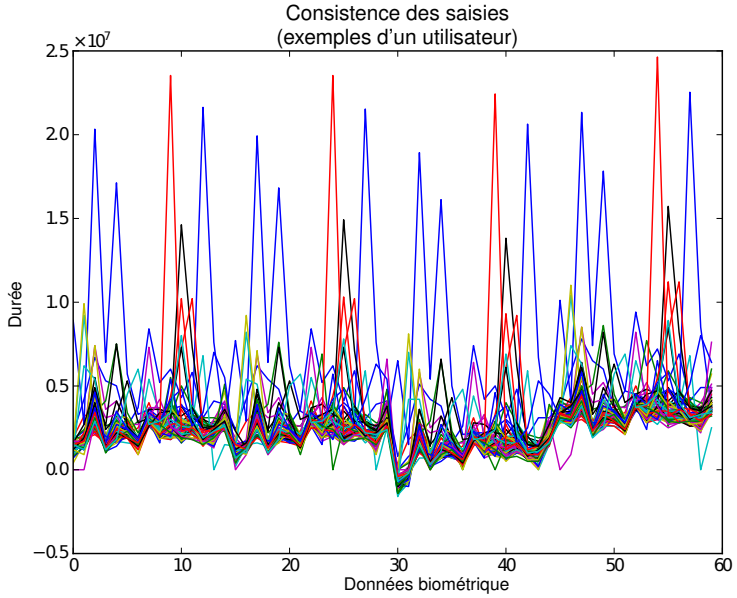


FIG. 3.2: Ensemble de captures du même mot de passe par le même utilisateur 90 fois. On peut voir, que globalement les temps de saisie sont proches. Cependant, il reste plusieurs écarts à la moyenne engendrant des erreurs de reconnaissance

Suivant les études, d'autres informations peuvent être extraites à partir des mêmes données brutes. Par exemple Gaines *et al.* [1980]; Umphress et Williams [1985] utilisent des digraphes. Il s'agit du temps nécessaire à la saisie de deux lettres (le délai entre la pression de la première touche et le relâchement de la suivante). Il faut noter que d'autres études nomment également digraphe le temps de latence. Dans d'autres travaux [Bergadano *et al.*, 2002], il s'agit de trigraphes, voire plus. Dans l'étude [Umphress et Williams, 1985], les données capturées sont les six premiers temps de chaque mot, le reste est ignoré. Une moyenne des temps est faite lorsque la

5. volontaires (ou pas) pour ajouter de l'entropie

touche majuscule est utilisée. Plusieurs données différentes peuvent être acquises et utilisées ; le choix des données influe naturellement sur la performance des algorithmes.

3.1.2.2. Enrôlement d'un utilisateur

L'enregistrement est l'étape d'enregistrement de la personne sur le système qui calcule la référence biométrique de l'utilisateur à partir d'une ou plusieurs captures. Dans la littérature, ce nombre de captures est souvent supérieur ou égal à 20. La quantité de captures nécessaires à la création de la référence peut donc devenir pénible pour les utilisateurs (d'autant plus s'ils sont sujets à faire de nombreuses fautes de frappe). En fonction des études, soit le mot de passe de l'utilisateur est demandé, soit un ou plusieurs textes identiques à tous les utilisateurs sont demandés [Bergadano *et al.*, 2002; Gaines *et al.*, 1980]. Généralement, le premier cas sert à mettre en place des mécanismes d'authentification statique (au lancement de la session de l'utilisateur), tandis que dans le second, il sert plus à une authentification continue (la machine vérifie perpétuellement si l'utilisateur actuel est celui qui s'est connecté)

3.1.2.2.1. Pré-traitement des données Afin d'obtenir une référence biométrique la plus représentative possible de l'utilisateur, diverses méthodes utilisent une étape de nettoyage des données qui peut être manuelle ou automatisée. Ce pré-traitement va donc principalement consister à :

- ne pas conserver les captures considérées comme erronées, dans le cas d'une authentification statique (et donc proposer une nouvelle saisie à l'utilisateur) ;
- ne pas conserver les digraphes considérés comme erronés, dans le cas d'une authentification continue (ce qui est totalement transparent pour l'utilisateur) ;
- normaliser les données [Hocquet, 2007] ;
- discrétiser les données Revett [2007] ;
- diminuer la dimension de l'espace des données capturées.

Dans [Gaines *et al.*, 1980], le filtrage se fait en supprimant les temps de saisie supérieurs à 500ms, tandis que dans [Umphress et Williams, 1985], ce sont les temps supérieurs à 750ms. Dans [Rogers et Brown, 1996], le nettoyage est fait: en utilisant un réseau de neurones (Kohonen) avec les données des imposteurs et avec une méthode statistique.

Cho et Hwang [2006] ne cherchent pas à filtrer les données, mais à faire en sorte que l'utilisateur saisisse son mot de passe de la façon la plus unique possible. Ils proposent différentes métriques afin de juger la qualité d'un ensemble de captures en fonction de :

son unicité définie par rapport à l'éloignement des motifs des imposteurs aux motifs d'enregistrement (dépend de la façon de taper). Soient $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$, respectivement les vecteurs d'enregistrement de l'utilisateur, de test de l'utilisateur et les vecteurs de tous les imposteurs. Avec N_x le nombre de données d'apprentissage, N_y le nombre de données de validation de l'utilisateur, N_z le nombre de données d'imposteurs, $\mathbf{m} = \sum_i^{N_x} \mathbf{x_i}/N_x$ le vecteur moyen des données d'apprentissage de l'utilisateur, l'unicité est définie de la façon suivante :

$$U = \sum_{k=1}^{N_z} \frac{|\mathbf{z_k} - \mathbf{m}|}{N_z} - \sum_{i=1}^{N_x} \frac{|\mathbf{x_i} - \mathbf{m}|}{N_x} \tag{3.1}$$

sa consistance définie par rapport à la ressemblance du motif de test avec les motifs d'enregistrement pour un utilisateur donné (dépend de l'habileté à taper et de la concentration).

L'inconsistence est donnée dans la formule suivante :

$$I = \sum_{j=1}^{N_y} \frac{|\mathbf{y_j} - \mathbf{m}|}{N_y} - \sum_{i=1}^{N_x} \frac{|\mathbf{x_i} - \mathbf{m}|}{N_x} \tag{3.2}$$

sa discriminalité consiste en la différence entre la distance minimale aux imposteurs et maximale à l'utilisateur :

$$D = \min_k |\mathbf{z_k} - \mathbf{m}| - \max_j |\mathbf{y_j} - \mathbf{m}| \tag{3.3}$$

Différentes techniques de réduction de dimension de l'espace des données peuvent être appliquées, mais cette étape n'est effectuée que dans une minorité des travaux. Un mécanisme de réduction de dimension présenté dans [Bleha et Obaidat, 1991] consiste à garder $n-1$ dimensions avec n utilisateurs dans la base. Comme il n'y a que 9 utilisateurs, on voit clairement que cette approche n'est pas viable en dehors de ce papier. Yu et Cho [2004] utilisent un algorithme basé sur les SVM et des algorithmes génétiques pour réduire la taille des vecteurs et ne garder que les valeurs clés pour chaque utilisateur. D'autres techniques similaires sont présentes dans la littérature [Akila *et al.*, 2012; Chen et Lin, 2005; Shanmugapriya et Padmavathi, 2011]. Une idée originale est de transformer les données en les discrétisant sous la forme de chaînes d'*Acides Aminés* et en utilisant des algorithmes de comparaison de chaînes de caractères utilisés dans le monde de la bioinformatique [Revett *et al.*, 2007a]. Chang [2006a] transforment les données dans le domaine fréquentiel à l'aide d'ondelettes.

3.1.2.2.2. Apprentissage de la référence biométrique Il existe différents types de méthodes pour vérifier si une requête correspond à un modèle particulier. Certaines d'entre elles sont basées sur des méthodes *statistiques*, d'autres sur des méthodes de *fouille de données*. En général, les méthodes de fouille de données nécessitent un nombre relativement conséquent de captures pour pouvoir créer le modèle (de l'ordre de quelques centaines dans le cas des réseaux de neurones). Les principales techniques utilisées sont :

- le calcul des vecteurs moyens et écart type [Umphress et Williams, 1985] ;
- le stockage pur et simple des données d'enregistrement, afin de les utiliser dans des algorithmes de *k plus proches voisins* [Rao, 2005] (les variations étant sur les méthodes de calcul de distance [Kang et Cho, 2009]) ;
- les classifieurs bayésiens [Janakiraman et Sim, 2007; Rao, 2005] ;
- l'utilisation de l'algorithme des *K-moyennes* [Hwang *et al.*, 2006; Obaidat et Sadoun, 1997] ;
- l'apprentissage de *Modèle de Markov Caché* (MMC) [6] [Rodrigues *et al.*, 2006] ;
- l'apprentissage de réseaux de neurones [Clarke et Furnell, 2006; Obaidat et Sadoun, 1997; Rogers et Brown, 1996] ;
- l'utilisation de *SVM* [Rao, 2005; Sang *et al.*, 2004; Yu et Cho, 2004].

Il existe donc plusieurs façons de calculer la référence biométrique d'un utilisateur. La plupart d'entre elles utilisent uniquement les données de l'utilisateur pour créer sa référence, il s'agit donc de problèmes à 1 classe ; cependant, quelques autres utilisent à la fois les données de l'utilisateur et des données d'imposteur. Dans le second cas, soit les données d'imposteurs sont générées à la volée à partir des données de l'utilisateur [Sang *et al.*, 2004], soit il s'agit des données d'enregistrement [Clarke et Furnell, 2006; Obaidat et Sadoun, 1997] d'autres utilisateurs. Ce procédé n'est pas forcément applicable en production en raison de la faible quantité de

6. Hidden Markov Model (HMM), en anglais

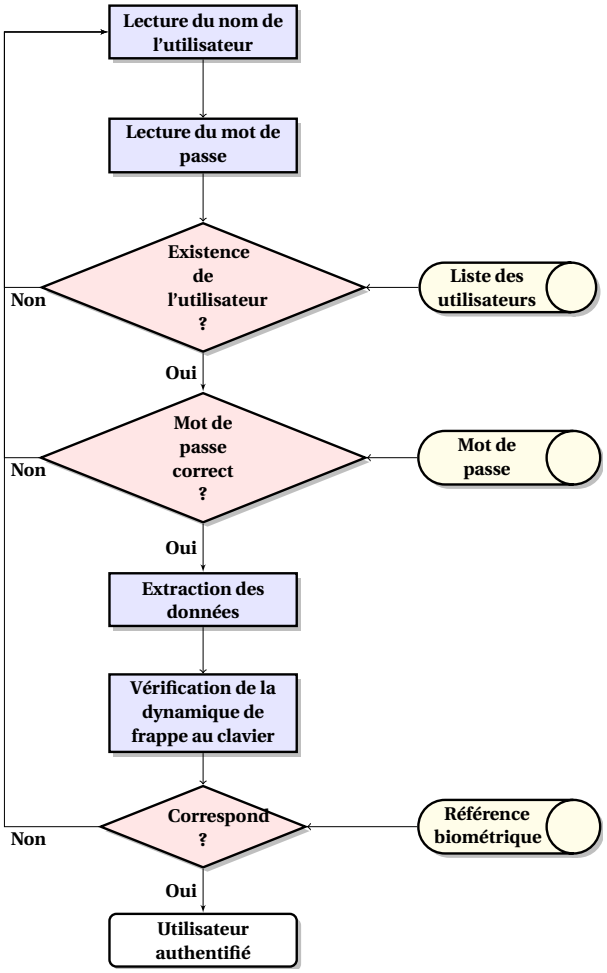


FIG. 3.3: Mécanisme de l'authentification en utilisant la dynamique de frappe au clavier

données disponibles en général pour l'enregistrement. Un mécanisme de génération de données artificielles a même été mis au point afin d'augmenter la quantité de données disponibles pour l'enregistrement, ainsi que les performances du système [Chang, 2006b].

3.1.2.3. Vérification de l'utilisateur

La vérification consiste à demander à l'utilisateur de fournir son identifiant et d'effectuer une capture du mot de passe. Une fois les données de la capture extraites, le mécanisme consiste à vérifier si cette signature correspond bien au modèle enregistré dans le système. Si les deux concordent, l'utilisateur est authentifié, sinon il est rejeté. Le plus souvent, ce critère de décision est basé sur un seuil qui est paramétré dans le système. La figure 3.3 présente les différentes étapes de l'authentification.

Afin d'augmenter les performances des algorithmes, une technique est présentée dans [Bleha et Obaidat, 1991] : elle consiste à créer une capture à partir de deux en les fusionnant ; ce mécanisme permet de filtrer les légères hésitations (mais nécessite à l'utilisateur d'effectuer deux saisies correctes au lieu d'une seule). Plusieurs méthodes de calcul de score existent, elles sont toutes

dépendantes de la façon dont se déroule l'enregistrement, et donc similaires à la liste précédente. Les principales familles de calcul sont [Guven et Sogukpinar, 2003]:

- Le calcul de la distance minimale. Dans [Monrose et Rubin, 1997], le principe consiste à calculer une distance euclidienne entre le vecteur de test et chacun des vecteurs d'enregistrement, puis à conserver la distance moyenne pour calculer le score. Ainsi, pour un vecteur test $\boldsymbol{v}$ de taille n, le calcul du score est le suivant :

$$score = \min_{\forall \boldsymbol{u} \in enregistrement} \left(\sqrt{\sum_{i=1}^{n} (u_i - v_i)^2} \right) \tag{3.4}$$

- Les méthodes statistiques. Une des méthodes les plus anciennes est basée sur les probabilités bayésiennes [Bleha *et al.*, 1990]. Pour un vecteur colonne $\boldsymbol{v}$, et un vecteur moyen $\boldsymbol{\mu}$, la formule est :

$$score = \frac{(\boldsymbol{v} - \boldsymbol{\mu})^t (\boldsymbol{v} - \boldsymbol{\mu})}{||\boldsymbol{v}||.||\boldsymbol{\mu}||} \tag{3.5}$$

  Les auteurs proposent également une version normalisée de cette formule. La méthode statistique présentée dans [Hocquet *et al.*, 2007] permet de calculer un score dépendant du vecteur moyen et de l'écart type. Le score global est la moyenne des scores de chacun des éléments du vecteur. Pour un vecteur de taille n, avec v_i la i^e donnée et μ_i et σ_i respectivement la valeur moyenne et son écart type :

$$score = 1 - \frac{1}{n} \sum_{i=1}^{n} e^{-\frac{|v_i - \mu_i|}{\sigma_i}} \tag{3.6}$$

  Une autre méthode testée est présentée dans [Filho et Freire, 2006]. Il s'agit également d'une méthode basée sur le calcul d'un score. Le calcul est le suivant : on note $\boldsymbol{\mu}$ le vecteur moyen des vecteurs d'enregistrement :

$$score = ||\boldsymbol{v} - \boldsymbol{\mu}||^2 \tag{3.7}$$

- La vérification d'appartenance à une classe (*cf.*, méthodes citées plus haut) ;
- Des méthodes basées sur le degré de désordre des vecteurs [Bergadano *et al.*, 2002] ;
- Des méthodes basées sur la discrétisation des temps [Hocquet *et al.*, 2007] ;
- Des méthodes de bioinformatique basées sur la recherche de motifs dans un texte [Revett *et al.*, 2007a].

Le figure 3.4 présente la taxonomie des méthodes de comparaison. Les méthodes statistiques simples ont été beaucoup plus étudiées que les autres méthodes. Ceci s'explique sûrement par le fait que peu de données sont disponibles à l'enregistrement, et que des méthodes plus complexes nécessiteraient plus de captures. Plusieurs méthodes peuvent être fusionnées pour augmenter la performance du système :

- dans [Bleha *et al.*, 1990], un classifieur bayésien est associé à un calcul de distance minimale entre le vecteur de test et le vecteur d'enregistrement (constitué des valeurs minimales de l'ensemble des 30 vecteurs d'enregistrement [7]) ;
- dans [Hocquet *et al.*, 2007], une fusion de trois méthodes est utilisée, ce qui améliore grandement les performances.
- une somme pondérée est utilisée dans [Teh *et al.*, 2007].

7. les expériences ont montré que la valeur minimale est plus efficace que la moyenne

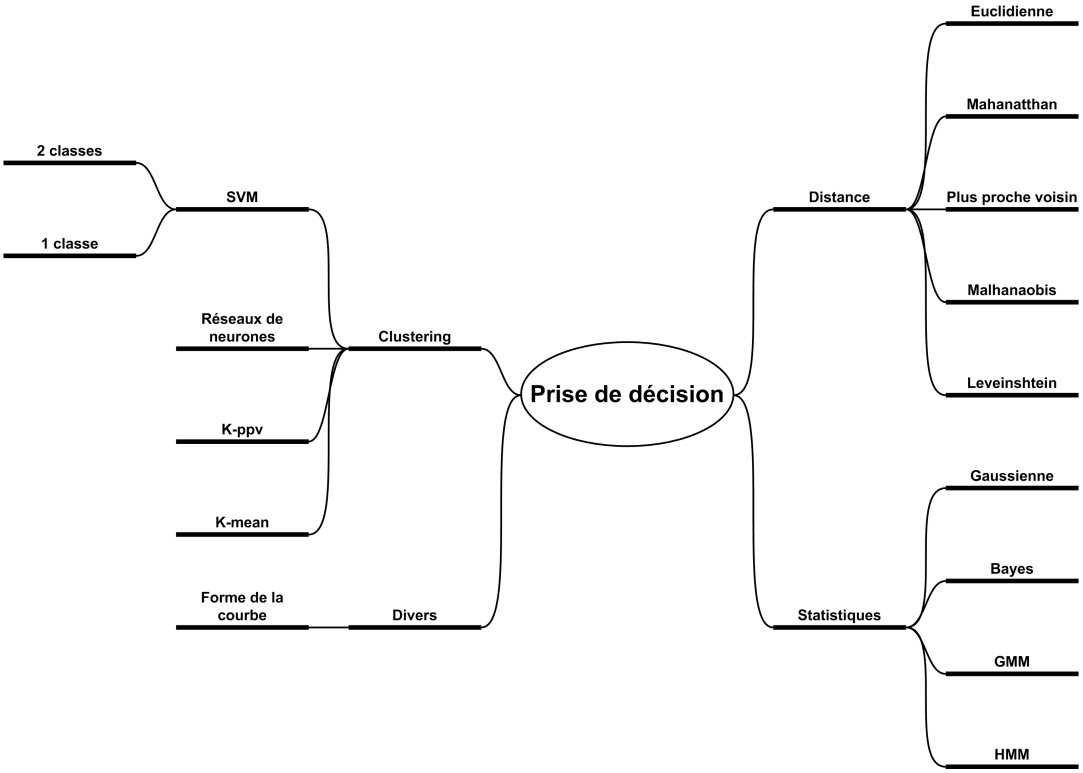


FIG. 3.4: Taxonomie des méthodes d'authentification

Dans le cas d'une vérification continue [Rao, 2005], cette étape est faite en permanence pendant que l'utilisateur utilise sa machine.

3.1.2.4. Identification de l'utilisateur

L'identification consiste, à l'aide d'une capture, à déterminer l'identité de la personne sans qu'elle ne la fournisse. Les algorithmes d'identification peuvent être spécifiques pour l'identification ou consister à comparer la capture à la totalité des modèles du système et retenir celui ayant obtenu la distance la plus faible, ou rejeter l'utilisateur si cette distance est supérieure à un seuil. Dans Bleha *et al.* [1990], un classifieur bayésien est utilisé pour identifier l'utilisateur. L'identification n'a pas été beaucoup étudiée dans la littérature. Ceci s'explique probablement par les faibles performances en vérification de la modalité.

3.1.3. Discussion

La DDF est une modalité moins stable que celles basées sur la morphologie de la personne. Cela peut s'expliquer par le fait que la façon de taper au clavier évolue avec le temps ou le contexte. Cependant, ce fait n'est pas inhérent à la DDF, mais, à l'ensemble des modalités comportementales. C'est pour cette raison qu'il est intéressant d'utiliser des systèmes de mise à jour du modèle qui seront présentés dans le chapitre 5. Les articles de la littérature montrent que la performance du système dépend de nombreux facteurs. Pour que le système soit le plus performant possible, il est préférable que le nombre d'exemples d'enregistrement, pour le modèle, soit le plus important possible et que le nombre de données dans un enregistrement (qui est

proportionnel à la taille du texte à saisir) soit également le plus important possible, sans dépasser une certaine taille [Hosseinzadeh et Krishnan, 2008].

Actuellement, un système performant nécessite donc un grand nombre d'enregistrements à partir d'un texte relativement grand. On voit clairement qu'un tel système est difficilement utilisable en pratique. Personne n'a le temps, ni le courage, de saisir autant de données. Ces systèmes sont donc impossibles à utiliser dans un environnement opérationnel. De plus, le temps nécessaire pour certains des algorithmes peut être relativement prohibitif. Par exemple, dans le cas de systèmes utilisant les réseaux de neurones, il est nécessaire à chaque ajout ou suppression d'utilisateurs dans le système, de refaire une étape d'apprentissage. Cette étape peut être très longue et même prohibitive (plusieurs centaines de secondes)[8]. Il est donc nécessaire d'axer les recherches futures en DDF, sur l'amélioration des performances avec un nombre minimal de données ; amélioration des performances, à la fois en termes de taux d'erreurs et de temps de traitement.

De plus, la majorité des techniques nécessitent que les fautes de frappes de l'utilisateur soient filtrées. Plus tard (voir en subsection 3.3.2.2.2, page 49), nous verrons que le nombre d'erreurs de saisies est relativement important pour une majorité d'utilisateurs. Il est donc nécessaire d'avoir un système qui soit le moins sensible possible aux corrections orthographiques [Bergadano *et al.*, 2002]. Nous pouvons trouver différents conseils dans la littérature pour augmenter les performances de systèmes de reconnaissance :

- plus le mot de passe est long, plus l'erreur d'identification est faible [Bleha *et al.*, 1990; Giot *et al.*, 2012c] ;
- plus le nombre d'exemples d'apprentissage est important, plus l'erreur est faible [Bleha *et al.*, 1990; Hocquet *et al.*, 2007] ;
- les résultats sont meilleurs sur les noms des personnes que des phrases (car les utilisateurs sont plus habituées à les saisir) [Bleha *et al.*, 1990];
- il est important d'utiliser une réduction des données [Bleha *et al.*, 1990; Yu et Cho, 2004] ;
- il est utile d'adapter le seuil à chacun des utilisateurs [Kacholia et Pandit, 2003], ou d'adapter les paramètres des algorithmes à certaines classes d'utilisateurs, voir [Hocquet *et al.*, 2007] ;
- une égalisation des temps de saisie augmente les performances [Filho et Freire, 2006] ;
- l'utilisation de pauses associées à des signaux (faisant office de métronome) [Hwang *et al.*, 2006] augmente également les performances du système, même pour les personnes n'ayant pas l'habitude d'utiliser un clavier ;
- la fusion des résultats de plusieurs méthodes de classification à partir des mêmes données augmente également les performances [Hocquet *et al.*, 2007].

Obaidat et Sadoun [1997] montrent que le taux d'erreurs en utilisant certains types de réseau de neurones est quasiment nul, mais le nombre de données doit être relativement important (*cf.*, remarque plus haut). Rogers et Brown [1996] indiquent que la taille idéale d'un mot de passe est entre 11 à 25 frappes, tandis que le nombre de saisies minimum pour l'apprentissage doit être de 20. Dans le cas de méthodes utilisant des *n-graphes*, il est plus performant de se baser sur des tri-graphes [Bergadano *et al.*, 2002]. Les performances des systèmes dépendent également de la résolution de l'horloge utilisée pour capturer les temps [Killourhy et Maxion, 2008]. Il est donc nécessaire d'avoir une horloge suffisamment précise pour améliorer les performances du système.

Cependant, comme nous pouvons le remarquer dans les tableaux 3.1 et 3.3, les paramètres des différentes études sur la DDF sont totalement différents. Pour cette raison, il est quasiment

8. Une de nos études a nécessité plusieurs semaines de calculs avec les réseaux de neurones au lieu de quelques heures pour l'ensemble de toutes les autres méthodes

impossible de comparer les études entre elles et de dire quelle méthode est la plus performante. Il est nécessaire d'effectuer nous-mêmes une étude comparative... De plus, Killourhy et Maxion [2011] montrent [9] que seulement 53.75% des papiers sur la DDF utilisent une étude comparative et 7.5% tirent des conclusions statistiques. Nous ne pouvons donc pas faire confiance à une grande partie des résultats de la littérature. Le tableau 3.1 présente, pour certaines études [10], le nombre de personnes y ayant pris part, la plateforme utilisée pour l'étude et sa durée. Il montre que la majorité des études ont été faites pour des ordinateurs de bureau.

On remarque également que, la plupart du temps, le nombre de personnes ayant pris part aux tests est largement insuffisant pour tirer des conclusions pouvant être généralisées à une mise en production dans le monde industriel.

Le tableau 3.3 présente quelques informations sur les résultats de méthodes de différentes études de l'état de l'art. Lorsque dans une même étude, plusieurs méthodes ont été testées, seule la plus performante est indiquée. Il présente :

- la quantité d'informations utilisée à l'enregistrement ;
- la quantité d'informations utilisée à la vérification ;
- le taux d'erreur du dispositif.

On voit clairement sur le tableau 3.3 que les mesures d'erreur des différentes études ne sont pas forcément les mêmes, ce qui rend difficile leur comparaison. Le nombre de données nécessaires pour l'apprentissage varie énormément d'une séquence à l'autre.

3.2. Caractérisation des bases de données disponibles et futures

La plupart des études effectuées en DDF sont validées sur une base privée : seuls les créateurs de l'étude y ont accès et les algorithmes développés sont optimisés pour la base en question. Cependant, il existe plusieurs bases publiques qui sont présentées en section 3.2.1. Malheureusement, celles-ci ont rarement été utilisées par d'autres scientifiques que leurs créateurs. La section 3.2.2 propose un ensemble d'indices permettant de caractériser une base de données de DDF. Cet ensemble d'indices va nous permettre de comparer les bases existantes et nous aider à comprendre pourquoi elles ne sont pas souvent utilisées (section 3.2.3). Ces indices peuvent également aider à créer de nouvelles bases de données de qualité.

3.2.1. Bases publiques existantes et proposées

Pour qu'une modalité biométrique soit convenablement utilisée, il est nécessaire de disposer de différentes bases de validation afin de faciliter la comparaison des algorithmes, et d'éviter d'optimiser un algorithme pour une base particulière. Nous listons les bases publiques de la littérature, ainsi que celles que nous proposons. Les bases publiques sont celles pour lesquelles les auteurs ont fourni dans leur article un lien sur la page web donnant les informations pour télécharger la base.

3.2.1.1. Le projet BioChaves

Les participants du projet BioChaves ont utilisé plusieurs fois la même base dans différents papiers [Filho et Freire, 2006]. Ils rendent public le contenu de la base de données sur le site web du projet: `http://itabi.infonet.com.br/biochaves/br/download.htm`. Quatre sous-bases

9. `http://www.cs.cmu.edu/~keystroke/cset-2011/`

10. nous avons choisi celles qui semblent les plus importantes d'un point vue historique et de performance

TAB. 3.1: Ce tableau présente, pour chaque étude analysée, le nombre d'utilisateurs ayant participés à cette étude, la plateforme et la durée d'acquisition des données

Étude	**Participants**	**Plateforme**	**Durée**	**Méthode**
Gaines *et al.* [1980]	7	PC	4 mois	distance
Umphress et Williams [1985]	17	PC	plusieurs jours	statistique
Bleha *et al.* [1990]	9, 10 et 26	PC	9, 5 et 8 semaines	distance minimum, classeur bayésien
Rogers et Brown [1996]		PC		réseau de neurones
Obaidat et Sadoun [1997]	15	MS-DOS	8 semaines	K-moyen, réseau de neurones, Cosine Measure Algorithm
Song *et al.* [1997]	10	X-Window		statistique
Robinson *et al.* [1998]	10	MS-DOS		statistique
Coltell *et al.* [1999]	10	PC		Méthode statistique
Monrose et Rubin [2000]	63	X-Window	11 mois	K-NN
Bergadano *et al.* [2002]	154	X-Window	1 mois	mesure du désordre
Monrose *et al.* [2002]	20	Applet Java	6 mois	
Kacholia et Pandit [2003]	20	PC		distribution de cauchy
Guven et Sogukpinar [2003]	12	PC		Mesure de similarité par des cosinus
Myung [2004]	5	téléphone mobile		loi de fitt
Yu et Cho [2004]	36	PC		SVM
Revett *et al.* [2005]	100	PC	7 jours	"Rough Set"
Rodrigues *et al.* [2006]	20	PC	4 sessions	H-MM
Rao [2005]	20	PC		fusion de KPPV, SVM, Bayes
Clarke et Furnell [2006]	30	téléphone mobile		réseau de neurones
Filho et Freire [2006]	18 et 15	PC	2 semaines	statistiques eet HMM
Hwang *et al.* [2006]	25	PC	3 mois	SVM, KPPV,K-moyen fenêtres de gauss et parzen
Zhao [2006]				Instanced Based Learning, K-Star, Classifieur bayésien, One-R
Obaidat *et al.* [2006]	6	PC	6 semaines	réseau de neurones
Clarke et Furnell [2007]	30	téléphone mobile		réseau de neurones
Hocquet *et al.* [2007]	38	PC		fusion de méthodes statistiques
Lee et Cho [2007]	21 et 25	PC	2 ans	réseau de neurones, SVM, Gauss et Parzen
Janakiraman et Sim [2007]	22	PC	2 semaines	classeur bayésien
Revett *et al.* [2007a]	30	PC	14 jours	bioinformatique
Revett *et al.* [2007b]	50	PC	14 jours	réseau de neurones probabiliste
Killourhy et Maxion [2008]	51	PC	8 sessions	distances et réseau de neurones
Rybnik *et al.* [2008]	37	PC		distance de Manhattan
Hosseinzadeh et Krishnan [2008]	41	PC		Mélange de Gaussiennes
Revett [2009]	20	PC	7 jours	Distance de levenstein

TAB. 3.3: Ce tableau présente pour chaque étude analysée leur performance. La première colonne présente l'étude, la suivante le nombre de captures nécessaire pour l'enregistrement, la seconde la taille des motifs et la dernière présente la performance du système ainsi que la méthode mise en œuvre

Étude	**Apprentissage**	**Test**	**Performance**
Gaines *et al.* [1980]	3 textes		
Umphress et Williams [1985]	1400 caractères	300 caractères	FRR : 12% FAR : 6%
Bleha *et al.* [1990]	30 séquences	20 caractères	TFA : 2,8% TFR : 8,1%
Obaidat et Sadoun [1997]	20 séquences	15 caractères	TBC =97,5% TCB=95,8% TCB=89,17%
Robinson *et al.* [1998]	10	login	EER=10%
Coltell *et al.* [1999]	20 motifs	20 caractères	TFA : 5% TFR : 30%
Monrose et Rubin [2000]	Texte libre	Texte libre	TBC=92,14 %
Bergadano *et al.* [2002]	4 fois un texte de 683 caractères	1 séquence de reconnaissance	TFA : 0,04 % TFR : 4%
Kacholia et Pandit [2003]	12 motifs	11 caractères	TFA : 1 % TFR : 4,8%
Guven et Sogukpinar [2003]	1 motif	8 caractères	TFA : 1 % TFR : 11,7%
Yu et Cho [2004]	50 motifs	mot entre 6 et 10 caractères	TFA : 0 % TFR : 6,28%
Revett *et al.* [2005]			Reconnaissance=97%
Rodrigues *et al.* [2006]	30 motifs	8 chiffres	EER = 3.6
Clarke et Furnell [2006]	4 séquences	11 chiffres, 4 chiffres, texte	EER=5%, EER=9%, EER=15%
Filho et Freire [2006]	15 séquences	4 mots	EER=6%
Hwang *et al.* [2006]	30 séquences	mot de passe	EER=0.25%
Zhao [2006]			Reconnaissance=91%
Hocquet *et al.* [2007]		mot de passe entre 8 et 30 caractères	EER=5%
Lee et Cho [2007]	de 76 à 388 motifs	mot de passe	erreur intégré 0.40%
Janakiraman et Sim [2007]	de 30000 à 2000000 de caractères	à la volée	
Revett *et al.* [2007a]	10 couples	logins et mots de passe 8 caractères	FAR=0.15 FRR=0.2
Rybnik *et al.* [2008]	Texte de 110 événements	55 événements	Reconnaissance:73%
Hosseinzadeh et Krishnan [2008]			FRR : 4,8% FAR=4,3%
Uzun et Bicakci [2012]	200 exemples positifs, 10000 négatifs	1 exemple	EER=7.73%

ont été créées, la plupart d'entre elles ayant été construites sur deux sessions séparées d'une semaine ou d'un mois. Le nombre maximum d'utilisateurs dans une sous base est 15, et, chaque utilisateur propose 10 exemples. Les bases contiennent les données brutes et sont composées d'un couple de valeurs : le code ASCII de la touche concernée, et le temps écoulé depuis le dernier évènement pression. L'information de relâchement d'une touche n'est pas capturée.

3.2.1.2. DSL2009

Killourhy et Maxion [2009] proposent une base de 51 utilisateurs ayant fourni 400 saisies sur 8 sessions. Le délai d'attente entre chaque session est d'une journée au minimum, mais la valeur moyenne n'est pas précisée (on peut s'attendre à ce que cela soit différent en fonction des utilisateurs). Il s'agit du jeu public avec le plus grand nombre d'exemples par utilisateurs, mais beaucoup de saisies sont faites sur une période relativement courte (50 saisies par session). Chaque saisie est faite sur le mot de passe « tie5Roanl ». La base contient uniquement les données extraites : temps de pression, latence entre deux pressions, de latence entre le relâchement d'une touche et la pression de la suivante. Cette base est accessible à l'adresse suivante : `http://www.cs.cmu.edu/~keystroke/` et est stockée dans plusieurs formats différents (texte, csv, Excel).

3.2.1.3. GREYC2009

Dans [Giot *et al.*, 2009], nous proposons la base de données la plus importante de l'état de l'art en nombre d'utilisateurs. Elle contient 133 utilisateurs, et 100 d'entre eux ont participé à au moins 5 sessions. Chaque utilisateur a saisi la phrase « greyc laboratory » 12 fois sur 2 claviers durant chaque session (ce qui donne 60 exemples pour les utilisateurs ayant participé à 5 sessions). La base contient à la fois les données brutes et extraites. Elle est disponible à l'adresse suivante `http://www.ecole.ensicaen.fr/~rosenber/keystroke.html` et est stockée dans un fichier sqlite. Cette base a été distribuée au même moment que la base DSN2009 et créée pour les mêmes raisons : le manque de base publique conséquente. La section 3.3.2.2 présente plus en détails la création de cette base.

3.2.1.4. Pressure-Sensitive Keystroke Dynamics Dataset

Allen [2010] a créé une base de données en utilisant un clavier contenant des capteurs de pression. La base est accessible dans les formats csv et sqlite à l'adresse suivante : `http://jdadesign.net/2010/04/pressure-sensitive-keystroke-dynamics-dataset/`. Elle contient les données brutes suivantes : code de la touche, date de la pression, date du relâchement, et force de la pression. Bien que 104 utilisateurs soient présents dans la base, seuls 7 d'entre eux ont fourni une quantité conséquente d'exemples (entre 89 et 504), alors que les 97 autres ont fourni très peu d'exemples (entre 3 et 15). Trois mots de passes différents ont été saisis: « pr7qlz », « jeffrey allen » et « drizzle ».

3.2.1.5. Keystroke100

Loy *et al.* [2007] proposent une base contenant, en plus des informations temporelles, des informations de pression. Dans cette base, 100 utilisateurs ont fourni 10 captures chacun. La base est disponible à l'adresse suivante `http://www.eecs.qmul.ac.uk/~ccloy/downloads_keystroke100.html` dans des fichiers texte.

3.2.1.6. WEBGREYC{A,B}

Nous proposons également une seconde base [Giot *et al.*, 2012b] qui est la plus importante en nombre de sessions. 118 utilisateurs ont eu la possibilité d'effectuer l'acquisition de leur donnée biométrique une fois par semaine durant 1 an et demi. L'utilisateur ayant fourni le plus de sessions en a 47. Un autre avantage de ce jeu de données est qu'il contient deux types de données biométriques : (i) des exemples imposés et communs à tous les utilisateurs, comme pour 100% des autres bases publiques, (ii) des exemples choisis par l'utilisateur avec des exemples d'impostures fournis par les autres utilisateurs, à la différence des autres bases de données publiques. Il s'agit donc (i) de la première base de données publique où chaque utilisateur a son propre mot de passe, (ii) de la base la plus réaliste. Cette base, disponible à l'adresse suivante : `http://www.epaymentbiometrics.ensicaen.fr/index.php/app/resources/84`, est stockée dans un ensemble de fichiers texte contenant à la fois les données brutes et les données extraites.

3.2.1.7. Discussion sur les bases

Plusieurs bases de DDF sont disponibles et nous pouvons remarquer que la majorité d'entre elles sont plutôt récentes alors que la modalité est étudiée depuis plusieurs décennies. Celles-ci ne concernent que l'authentification sur un ordinateur personnel (il n'existe pas encore de base disponible pour téléphone mobile). Cependant, les bases n'ont été utilisées que par leurs auteurs, et la majorité des études ont été faites sur des bases privées utilisées nulle part ailleurs. Killourhy et Maxion [2011] nomment les expériences utilisant ce type de données, des « one-off evaluations » et montrent qu'elles donnent des résultats erronés. Il faut espérer que ces bases seront utilisées dans le futur afin de permettre des études comparatives qui seront non biaisées.

Dans ce chapitre, nous définissons un ensemble de critères utiles pour la caractérisation des bases de données en DDF. La caractérisation permet : de comparer les bases entre elles, indépendamment des performances de systèmes biométriques ; d'aider à choisir une base pour une étude spécifique ; et d'aider à créer une nouvelle base correspondant à des attentes précises.

3.2.2. Indices de caractérisation des bases

Avant de différencier les bases, il est nécessaire de fournir des indices de comparaison. Les points de divergence entre jeux de données sont présents à différentes localisations [Giot *et al.*, 2011b; Killourhy et Maxion, 2009] : la procédure d'acquisition ; la distribution de la population d'individus et de données biométriques ; la complexité du mot de passe utilisé ; et les performances du jeu de données.

3.2.2.1. Procédure d'acquisition

Plusieurs paramètres changent entre les procédures d'acquisition des jeux de données ; nous pensons que ces paramètres peuvent impacter la qualité du jeu de données et ses performances.

3.2.2.1.1. Durée d'acquisition de la base de données La DDF est une biométrie comportementale ayant une forte variabilité temporelle (section 5.2). Nous nous attendons à obtenir de meilleures performances sur des bases de données acquises sur de courtes périodes plutôt que sur de longues périodes. Les jeux de données acquis sur de longues périodes peuvent être utilisés dans des études sur la mise à jour du modèle. Notez que cette information est différente du nombre de sessions, car plusieurs sessions peuvent être capturées sur une durée relativement courte. Cette information peut être expliquée en nombre de jours entre la première et dernière acquisition.

3.2.2.1.2. Les erreurs de saisie sont-elles autorisées ? Si un individu fait une faute de frappe lors de la capture de sa donnée, la correction implique le changement de la façon de taper au clavier (*c.-à-d.*, plus de saisies en utilisant la touche retour clavier, ou les flèches). Ainsi, la donnée capturée ne correspond pas du tout au modèle. Nous n'avons pas trouvé de travaux résolvant ce problème, dans la littérature (bien que des HMM sembleraient appropriés). Pour contourner ce problème, la plupart des outils de collecte de bases de données forcent l'utilisateur à saisir correctement. Cette information peut être expliquée à l'aide d'un booléen.

3.2.2.1.3. Acquisition faite en environnement calme et stable Une fois de plus, nous espérons obtenir des données plus stables dans les environnements calmes où l'utilisateur n'est pas perturbé. Cette information peut être expliquée à l'aide de mots.

3.2.2.1.4. L'acquisition est-elle contrôlée par un opérateur ? Lorsque l'acquisition est contrôlée par un opérateur, nous affirmons que la base ne contient aucune (ou peu) donnée erronée et que les individus participant à la collecte ont respecté clairement le protocole d'acquisition. Lorsque l'acquisition n'est pas contrôlée, nous ne pouvons pas faire confiance à la qualité du jeu de données. Cette information est spécifiée en utilisant un booléen.

3.2.2.1.5. Chaque utilisateur a-t-il un mot de passe qui lui est propre ? Est-ce que chaque utilisateur a saisi son propre mot de passe ? Ou est-ce que tous les utilisateurs saisissent le même mot de passe ? Il est difficile d'acquérir un jeu de données avec suffisamment d'utilisateurs fournissant un mot de passe différent ; c'est pourquoi la plupart des bases publiques sont collectées avec un seul mot de passe. Lorsqu'un seul mot de passe est utilisé, les données authentiques des utilisateurs sont utilisées comme des données d'imposture des autres utilisateurs. Cette information peut être spécifiée en utilisant (a) un booléen, (b) une chaîne de caractères, ou (c) une liste de chaînes de caractères représentant les mots de passe.

3.2.2.1.6. Système d'exploitation utilisé Le système d'exploitation (SE) joue un rôle dans la précision des informations de capture [Narainsamy *et al.*, 2010]. Cette information peut être spécifiée (a) avec une chaîne de caractères donnant le nom du SE, (b) un nombre symbolisant une valeur de SE dans une liste prédéfinie, ou (c) un histogramme si chaque utilisateur a effectué la capture sur un SE différent.

3.2.2.1.7. Type de clavier utilisé La forme des claviers varie grandement, aussi bien que la position des touches. Cette information peut avoir un impact sur la performance de reconnaissance, car le mouvement des doigts peut être différent. Cette information peut être spécifiée à l'aide de mots ou d'une image.

3.2.2.1.8. Taux d'erreur d'acquisition Nous pensons que le FTAR donne une idée de la difficulté du mot de passe. Une erreur d'acquisition intervient lorsque l'utilisateur effectue une erreur de frappe et doit ressaisir le mot de passe. Un FTAR trop élevé est dérangeant pour l'utilisateur en augmentant le taux de faux rejets (*false rejection rate*) (FRR).

3.2.2.1.9. Résolution de l'horloge La résolution de l'horloge utilisée pour capturer les temps a un impact conséquent sur les performances [Killourhy et Maxion, 2008; Narainsamy *et al.*, 2010]. C'est pourquoi, il est important de connaître cette information qui peut être fournie avec une précision en millisecondes.

3.2.2.1.10. Informations disponibles La plupart des jeux de données publiques ne fournissent pas les données brutes : seules les données extraites sont disponibles. Les données sélectionnées sont différentes d'une base à l'autre. Comme les performances peuvent varier d'un type de données à un autre, il est important de préciser quels types de données extraites sont disponibles dans le jeu de données, ou quelles données sont utilisées dans l'étude. Cette information peut être fournie à l'aide d'une liste de données disponibles.

3.2.2.2. Distribution de la population et des exemples

Les bases sont construites grâce à des volontaires acceptant de participer au protocole. La population sélectionnée peut influer sur la variabilité des données collectées.

3.2.2.2.1. Distribution du genre des individus Les hommes et les femmes peuvent être différenciés selon leur façon de taper au clavier, ce que nous verrons en 4.4. Nous pouvons également nous demander s'il y a des différences de performances de vérification entre les hommes et les femmes (c'est le cas pour la reconnaissance faciale). Il nous semble important de connaître cette information. Cette information peut être fournie en utilisant un ratio d'hommes présents dans la base.

3.2.2.2.2. Maîtrise du clavier Tout le monde ne maîtrise pas de la même façon la frappe au clavier. Les personnes n'étant pas habituées à utiliser un clavier peuvent avoir de moins bonnes performances de reconnaissance. Cette information peut être proposée par une liste de performance de frappe par utilisateur (bien que cette information soit relativement subjective) (*c.-à-d.*, tous les utilisateurs maîtrisent la saisie au clavier). De plus, le nombre de doigts utilisés, ou la présence ou absence de coordination entre les deux mains peut également être intéressante.

3.2.2.2.3. Individus droitiers et gauchers À notre connaissance, il n'existe pas d'étude sur ce sujet, mais nous pensons que les droitiers et gauchers peuvent taper différemment : la difficulté de saisie peut être différente. C'est pourquoi cette information peut être intéressante et fournie à l'aide d'un ratio de personnes droitières contenues dans la base.

3.2.2.2.4. Âge des individus Les personnes âgées peuvent taper différemment des plus jeunes car elles ont appris à utiliser un ordinateur plus tardivement ou peuvent avoir des problèmes de santé impactant le mouvement des doigts, ou peuvent avoir suivit des formations de dactylographie. Une étude de 1984 (à ce moment, les personnes plus âgées devaient moins maitriser l'informatique qu'aujourd'hui) montre que le comportement de frappe est différent entre les jeunes et les personnes âgées [Salthouse, 1984]. Nous pensons donc que cet indice peut impacter les performances. Cette information peut être fournie en utilisant (a) la liste d'âge des personnes impliquées, (b) un histogramme, (c) l'âge moyen.

3.2.2.2.5. Nombre d'individus dans la base Plus la population est importante, plus les résultats sont fiables. Bien sûr, la performance peut également se dégrader lorsque le nombre d'individus augmente, mais la confiance des résultats est plus importante. Cette information peut être fournie à l'aide d'un nombre entier.

3.2.2.2.6. Nombre de sessions La DDF étant une modalité comportementale, la façon de taper dépend de paramètres externes non contrôlés comme l'état émotionnel de l'individu [Epp, 2010], ou le clavier. Utiliser une seule session réduit radicalement la variabilité des données, et, augmente injustement les performances. Il est préférable d'utiliser au minimum trois sessions

différentes [Cherifi *et al.*, 2009]. Bien sûr, nous pouvons nous attendre à une diminution des performances en fonction du nombre de sessions. Cette information peut être fournie avec (a) un entier, ou (b) un réel.

3.2.2.2.7. Nombre de saisies par utilisateurs Si nous n'avons pas suffisamment de données, nous ne pouvons pas tester suffisamment la variabilité intra-classe des individus. Bien sûr, les performances diminuent lorsque le nombre d'exemples par utilisateurs croit. Cette information peut être fournie avec (a) un entier, ou (b) un réel.

3.2.2.3. Complexité du mot de passe

Les mots de passe complexes peuvent être difficiles à retenir [Yan *et al.*, 2000]. Nous devons gérer cette information.

3.2.2.3.1. Mots de passe imposés ou choisis Si le mot de passe est choisi par l'utilisateur, nous nous attendons à ce qu'il en ait une meilleure imprégnation, ainsi qu'une saisie plus stable et rapide que pour un mot de passe imposé. Cette information peut être fournie par un booléen.

3.2.2.3.2. Complexité du mot de passe La complexité du mot de passe est un bon indicateur de la sécurité de l'authentification. Un mot de passe complexe est difficile à casser, mais il peut être difficile à se rappeler ou saisir. Cette information peut être fournie à l'aide d'un score. Nous avons choisi d'utiliser une méthode standard utilisée dans les applications WEB pour connaitre la complexité du mot de passe.

3.2.2.3.3. Entropie du mot de passe L'entropie du mot de passe donne des informations intéressantes sur la quantité d'information fournie par celui-ci. Avec une faible entropie, les mêmes lettres sont présentes plusieurs fois, alors qu'avec une entropie plus importante, il y a un plus grand nombre de lettres différentes. Il s'agit également d'une métrique utilisée pour tester la force d'un mot de passe. Cette information peut être fournie à l'aide d'un score.

3.2.2.3.4. Difficulté de saisie Même s'il n'existe pas d'études à ce sujet dans la littérature, nous pouvons nous attendre à obtenir de moins bonnes performances lorsque la complexité de saisie est importante. Il n'existe pas de méthode standard pour tester la difficulté de saisie de mot de passe, et, nous pensons qu'il s'agit d'un problème ouvert en DDF. La difficulté peut être différente entre les différents claviers en raison de leur forme ou langue. Cette information peut être fournie à l'aide d'un score. Il serait intéressant de vérifier si la complexité du mot de passe est corrélée avec la difficulté de la saisie. La méthode que nous avons utilisée pour calculer la difficulté est plutôt naïve et mériterait d'être validée plus en détails. Chaque lettre du clavier est représentée par un nœud dans un graphe, tandis que les arcs correspondent à l'adjacence entre les touches (poids de 1 verticalement et horizontalement et $\sqrt{2}$ en diagonale. Les poids sont plus élevés entre les touches alphabétiques de droite et le pavé numérique). Le score de difficulté consiste à la somme des plus courts chemins entre chaque touche successive.

3.2.2.4. Performance de la base de données

Pour calculer les performances d'un jeu de données, il est nécessaire de sélectionner une certaine quantité d'exemples pour l'enregistrement et une autre pour la vérification. Comme tous les jeux de données sont différents, il est presque impossible d'utiliser la même quantité de données pour l'apprentissage et la vérification entre les jeux de données. Une bonne méthode serait d'utiliser la

première session pour l'enregistrement et les sessions suivantes pour la validation. Si le nombre d'exemples par session est très différent entre deux jeux de données, la comparaison pourra être biaisée. Les informations fournies dans ce sous-ensemble sont liées à la difficulté de reconnaissance du jeu de données.

3.2.2.4.1. Respect des propriétés biométriques Nous avons vu (section 3.1) qu'il existe différents indicateurs qui permettent de mesurer l'unicité, la discriminabilité et la consistance de l'ensemble des saisies de chaque utilisateur [Hwang *et al.*, 2006]. Même si ces indicateurs ont été publiés en 2006, ils n'ont jamais été utilisés dans des études effectuées par d'autres laboratoires, à notre connaissance. Ces informations peuvent être fournies avec (a) une liste de scores, (b) son histogramme, ou (c) une moyenne selon les utilisateurs. Elles peuvent dépendre du temps (cela n'a pas été vérifié). Il peut donc être utile de calculer les valeurs session par session, puis de calculer la moyenne finale.

3.2.2.4.2. Performance sur un classifieur de référence Le but des jeux de données en DDF est de les utiliser pour comparer les performances de différents classifieurs selon différentes conditions. Il est important de donner les performances d'un jeu de données avec un classifieur de référence. Les indicateurs de performance peuvent être l'EER ou l'aire sous la courbe ROC (*Area Under the ROC Curve*) (AUC). L'EER donne le point de fonctionnement souvent utilisé dans la littérature pour comparer les classifieurs. Étant donné que nous caractérisons une base, et pas un algorithme de reconnaissance, nous pensons qu'il n'est pas nécessaire d'utiliser d'autres métriques. Nous avons choisi le calcul de distance présenté dans l'équation (3.6), page 26.

3.2.2.4.3. Dégradation des performances au cours du temps Comme nous le verrons au chapitre 5, la DDF souffre de dégradation de performances au cours du temps. Il est utile de vérifier si un jeu de données est sensible à de telles perturbations en:

1. calculant la performance par session, en utilisant les données de la première session pour l'enregistrement.
2. calculant la pente de la droite de régression du EER au cours du temps.

Plus la pente est grande, plus la dégradation au cours du temps est importante.

3.2.3. Illustration de la caractérisation sur les jeux de données publiques

Nous illustrons les indices présentés précédemment sur les jeux de données publiques recensés en section 3.2.1. Certaines bases sont séparées en sous bases en fonction de leur différence de protocole. Les descriptions des bases montrent leur grande variabilité et jeunesse en comparaison des bases des autres modalités. Plusieurs jeux de données ont été capturés sans notion de session. Dans ce cas, nous les avons créées artificiellement. D'autres proposent différents mots de passes avec différents utilisateurs. Dans ce cas, nous avons créé des sous-bases pour chaque mot de passe. Pour réduire la taille des tableaux, nous utilisons un symbole par jeu de données, dont voici la liste (les sous-bases sont post-fixés avec leur code):

A	GREYC	**D**	PRESSURE
B	WEBGREYC	**E**	BIOCHAVES
C	DSL2009	**F**	KEYSTROKE100

TAB. 3.5: Informations d'acquisition pour les différents jeux de données. «???» indique un point non indiqué dans le papier présentant la base

Base	A	BA	BB	C	D	EA	EB	EC	F
Durée de l'acquisition	3 mois	17 mois	17 mois	> 8 jours	???	1 mois	1 semaine	???	???
Erreurs de saisies autorisées	✗	✗	✗	✗	???	✗	✗	✗	✗
Acquisition calme et stable	✗	✗	???	???	???	???	???	???	???
Acquisition contrôlée	✓	✗	✗	✓	???	???	???	???	???
Mot de passe unique	✓	✓	✗	✓	3	✓	✓	✓	✓
Système d'exploitation	win. xp	navigateurs	navigateurs	windows	???	???	???	???	???
Type de clavier	[1]	aucun contrôle	aucun contrôle	[2]	???	[3]	variés	aucun contrôle	???
Résolution d'horloge		???	???	200 μsecondes	???	???	???	???	???
Failure To Acquire Rate	$\approx 20\%$	???	???	???	???	???	???	???	???
Information capturées	PP,PR,RP,RR	PP,PR,RP,RR	PP,PR,RP,RR	PP,PR,RP,RR	PP	PP	PP	PP	PP

[1] 2 AZERTY (portable + USB), [2] 1 QWERTY portable [3] Clavier brésilien

3.2.3.1. Collecte des méta-données

Le tableau 3.5 présente les informations d'acquisition de chaque jeu de données. Plusieurs indices sont inconnus et référencés par «???». Nous observons les points suivants :

- Aucun jeu de données, excepté WEBGREYCB, ne propose un unique mot de passe par individu.
- Aucun jeu de données ne permet de corriger les erreurs de frappes et tous imposent une re-saisie systématique en cas d'erreur.
- La résolution de l'horloge n'est pas une information couramment présentée, alors qu'il est connu qu'elle est importante.
- Un seul jeu de données présente le FTAR. Il est ainsi difficile de savoir s'il s'agit d'un comportement anormal, ou de quelque chose en rapport avec le scénario.

Le tableau 3.6 présente la distribution des utilisateurs ; même si certains articles donnent des indications sur la distribution des individus [Giot *et al.*, 2009; Killourhy et Maxion, 2009], elles apparaissent rarement dans les méta-données des bases et ne sont pas disponibles pour les chercheurs voulant utiliser ces bases. Nous observons les points suivants :

- Peu de jeu de données proposent des informations à propos des utilisateurs. Si tel est le cas, la population n'est pas vraiment équilibrée. C'est un problème, car aucun jeu de données ne présente une population réaliste.
- La plupart du temps, le nombre d'utilisateurs impliqués dans l'étude est trop faible (moins de 50 utilisateurs).
- La quantité de données fournie par un jeu de données varie énormément d'un jeu à l'autre.

3.2.3.2. Caractérisation des bases

Après avoir analysé les méta-données des différentes bases, il est intéressant d'analyser leur contenu. Le tableau 3.7 présente les informations des différents mots de passe. Nous observons les points suivants :

- Tous les jeux de données utilisent un mot de passe imposé par le protocole, sauf WEBGREYCB. Ce point est compréhensible en raison de la complexité de création d'un jeu de données. Malheureusement, ils ne sont pas réalistes.
- Le coefficient de corrélation de Pearson (annexe B.3) entre l'entropie et la complexité du texte est de 0.90 sans WEBREGREYCB et 0.29 avec. L'entropie semble être une information suffisante car sa définition est plus claire que celle de la complexité du texte.

TAB. 3.6: Distribution des individus

Base	A	BA	BB	C	D0	D1	D2	EA	EB	EC	F
Genre (% d'hommes)	73	80	80	59	???	???	???	???	???	???	???
Maîtrise du clavier	???	???	???	???	???	???	???	[1]	[2]	[1]	???
Gaucher/droitier	???	???	???	???	???	???	???	???	???	???	???
Age	19-56	20-40	20-40	18-70	???	???	???	???	???	???	???
Nombre d'individus	97	32	32	51	13	14	14	10	8	14	100
Nombre de sessions	5	10	10	8	3	3	3	2	2	???	1
Nombre d'exemples/utilisateur	60	94.81	89	400	15	15	15	10	10	10	10
Nombre d'exemples d'apprentissage[3]	12	10	10	50	5	5	5	5	5	5	5
Nombre d'exemples de validation[3]	48	84.81	79	350	10	10	10	5	5	5	5

[1] pas tous familiers [2] étudiants en génie électrique et informatique [3] ces valeurs ont majoritairement été choisies par nous, pour permettre une analyse la plus objective précise, car les bases proposent rarement une séparation des exemples d'apprentissage et de validation

TAB. 3.7: Complexité des mots de passe

Base	A	BA	BB	C	D0	D1	D2	EA	EB	EC	F
Imposé par le protocole	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓
Valeur	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[8]	[9]	[10]
Entropie	3,33	3,91	0,34 (0,25)	3,32	2,52	3,03	2,58	3,67	3,67	3,31	3,00
Complexité du texte	120	246	184,00 (71,78)	187	7	102	40	324	324	132	126,00
Difficulté de saisie	5	4,21	4,00 (0,79)	6,50	5,50	3,46	10,29	3,77	3,77	4,50	6,67

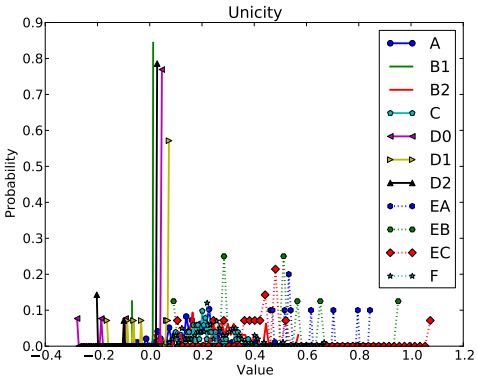
[1]"greyc laboratory", [2]"laboratoire greyc" + "sésame", [3]Chaque utilisateur en utilise un différent, [4]".tie5Roanl", [5]"drizzle", [6]"jeffrey allen" , [7]"pr7q1z", [8]"chocolate, zebra, banana, taxi", [9]"computador calcula", [11]"try4-mbs"

– Les mots de passes sont relativement différents d'une base à l'autre.

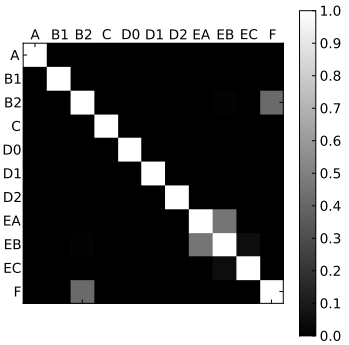
Le tableau 3.8 donne les indices de performances des différentes bases. Nous observons les points suivants :

– La réduction de performance au cours du temps est liée au nombre de sessions (coefficient de Pearson de 0.73, en utilisant plusieurs sessions). Cela indique différents aspects :
  – La dégradation de la façon de taper au cours du temps est importante.
  – Il est obligatoire d'évaluer la DDF avec des bases capturées sur plusieurs sessions.
– Le coefficient de Pearson entre la consistance et la dégradation du EER est de 0.901 sans WEBGREYCB et 0.55 avec. Cela montre l'intérêt de cette mesure afin de contrôler les performances de reconnaissances individuelles d'un utilisateur (rappelez-vous que cette indice a été créé afin d'aider les individus à générer des motifs stables).

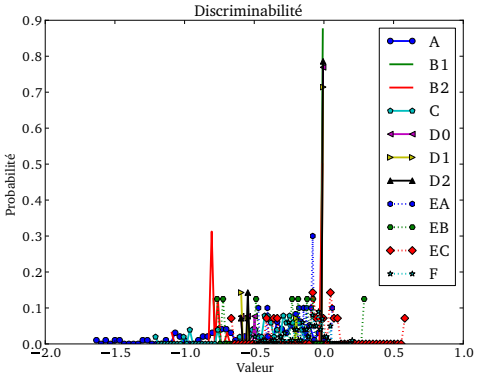
La figure 3.5 présente les distributions d'unicité, discriminabilité et consistance des utilisateurs de chaque base. Elles ont été calculées en normalisant par z-score (équation (4.2), page 70) les données extraites des différentes bases pour avoir des chiffres plus facilement comparables. Il serait intéressant de comparer ces valeurs avec celles fournies par leurs créateurs Hwang *et al.* [2006], malheureusement leur base n'est pas publique. La matrice de confusion, calculée en utilisant la p-value du test de Kolmogorov-Smirnof pour comparer les valeurs de deux bases, est également présentée. Le test de Kolmogorov-Smirnof permet de vérifier si deux distributions suivent la même loi. La Figure 3.6 présente la courbe ROC de chaque système, tandis que les



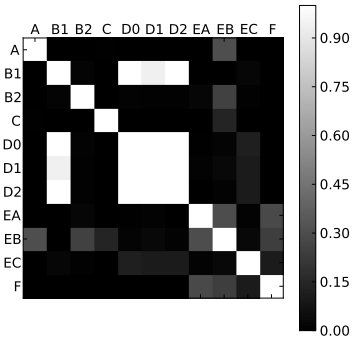
(a) Distribution de l'unicité



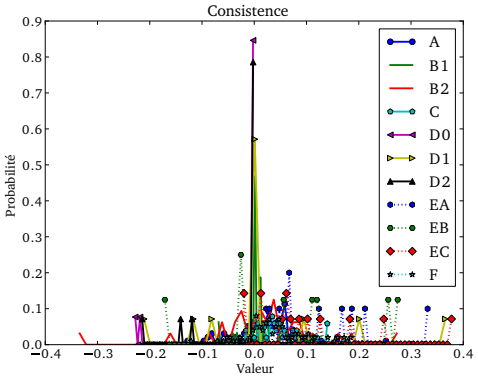
(b) KS pvalue de l'unicité



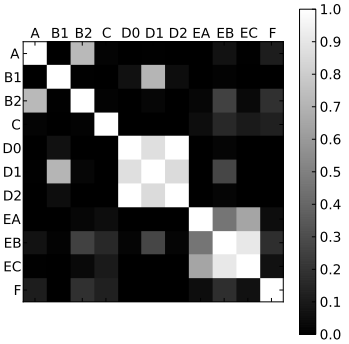
(c) Distribution de la discriminabilité



(d) KS pvalue de la discriminabilité



(e) Distribution de la consistence



(f) KS pvalue de la consistence

FIG. 3.5: Distributions de l'unicité, de la discriminabilité et de la consistence et comparaison entre les bases

TAB. 3.8: Évaluation des performances

Base	A	BA	BB	C	D0	D1	D2	EA	EB	EC	F
Unicité (μ)	0,17	0,00	0,26	0,22	−0,00	0,03	−0,01	0,60	0,48	0,44	0,25
Unicité (σ)	0,09	0,03	0,12	0,06	0,11	0,08	0,08	0,13	0,25	0,21	0,10
Consistance (μ)	0,02	0,00	0,02	0,05	−0,03	0,02	−0,03	0,13	0,08	0,10	0,04
Consistance (σ)	0,06	0,03	0,10	0,04	0,08	0,13	0,07	0,10	0,14	0,11	0,06
Discriminabilité (μ)	0,17	0,00	0,26	0,22	−0,00	0,03	−0,01	0,60	0,48	0,44	0,25
Discriminabilité (σ)	0,09	0,03	0,12	0,06	0,11	0,08	0,08	0,13	0,25	0,21	0,10
EER du classifeur	0,18	0,14	0,18	0,24	0,25	0,18	0,31	0,36	0,42	0,34	0,20
AUC du classifieur	0,90	0,93	0,89	0,84	0,82	0,90	0,74	0,69	0,61	0,70	0,89
Dégradation du EER	74,86	93,34	56,93	49,98	−26,90	28,00	17,17	x	x	x	x
Dégradation de l'AUC	−71,60	−109,34	−50,94	−44,67	36,15	−57,75	−30,20	x	x	x	x

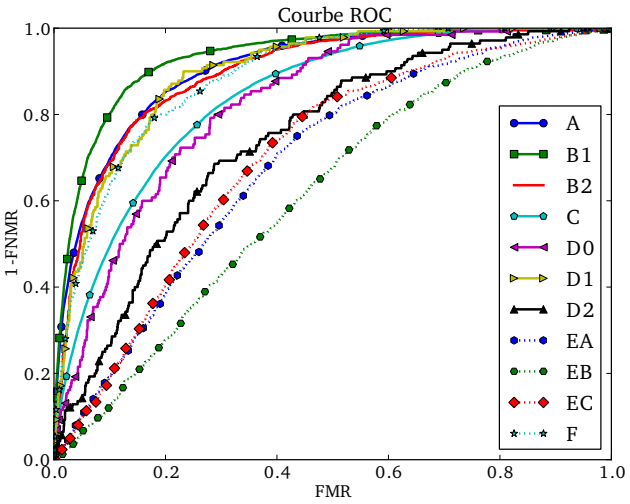
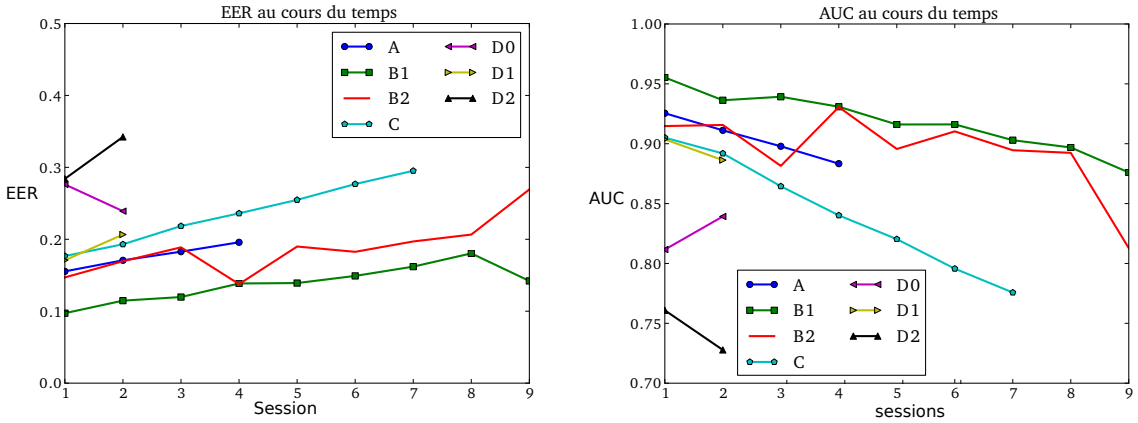


FIG. 3.6: Courbe ROC des différents systèmes

figures 3.7b et 3.7a présentent respectivement la AUC et l'EER par session pour chaque base. Nous faisons les observations suivantes :

- En utilisant le classifieur de référence, le EER en DDF est vraiment différent d'une base à l'autre. Sa valeur varie de 14% à 42%. Il y a une différence de 28% entre la base donnant de meilleures performances (A), et la base donnant les moins bonnes performances (EA), en utilisant exactement la même méthode d'authentification. Cela montre l'importance de valider ses algorithmes avec plusieurs jeux de données.

- Toutes les bases, exceptée D2 (mot de passe « pr4q12z »), voient une diminution des performances au cours du temps, même lorsque le nombre d'exemples est faible. Cela montre la nécessité de prendre en compte le vieillissement de la donnée biométrique ou la stabilisation au cours du temps.



(a) Évolution du EER au cours des sessions

(b) Évolution de l'AUC au cours des sessions

FIG. 3.7: Évolution du EER et de l'AUC au cours des sessions

3.2.4. Discussion

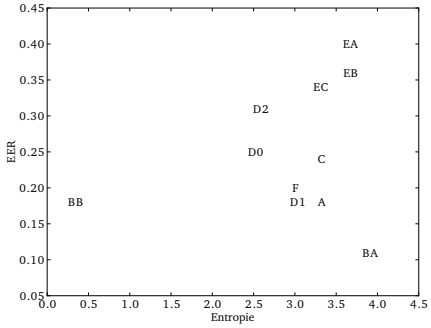
Nous avons analysé et comparé les propriétés de toutes les bases publiques de DDF. Cette analyse montre que les bases sont relativement différentes et partagent peu de propriétés communes. Ce manque de propriétés communes ne facilite pas leur comparaison. Le tableau 3.5 montre que la majorité des bases ne fournissent pas d'informations précises quant à leur collecte. Le tableau 3.6 montre que peu d'informations sont présentes sur les propriétés des utilisateurs. De plus, la majeure partie des bases possède peu d'individus et d'exemples.

Le tableau 3.9 présente une comparaison des différentes bases. Le but de ce tableau est de voir rapidement les différentes propriétés des jeux de données, afin de sélectionner ceux qui semblent intéressants pour une tâche particulière. Cependant, il n'y a pas de méthode générique et objective pour construire une telle table. Nous l'avons construite au regard de nos contraintes et d'idéal de base de bonne qualité pour la DDF :

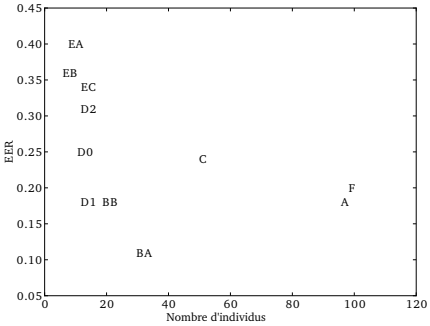
- Le nombre d'utilisateurs doit être supérieur ou égal à 50 (même si cela reste un nombre faible). Avoir beaucoup d'utilisateurs permet de généraliser plus facilement les résultats en leur garantissant une validité statistique.
- Le nombre d'exemples par utilisateur doit être supérieur à 30. Il est nécessaire de disposer de suffisamment d'exemples pour générer la référence biométrique et pour tester le système sans utiliser de méthodes de validation croisée qui ne permettent pas de respecter la chronologie d'acquisition.
- Le nombre de sessions d'acquisition doit être supérieur ou égal à 3 [Cherifi *et al.*, 2009] : une session pour calculer les références biométriques, et deux sessions pour calculer les différentes métriques d'erreur. Il est insensé de travailler avec des exemples capturés sur une période courte, étant donné qu'il est connu que le vieillissement de la donnée biométrique est important en DDF.

En analysant les scores obtenus par chaque jeu de données, nous voyons que les jeux PRESSURE, BIOCHAVES et KEYSTROKE100 ne sont pas adaptés pour les études en DDF, selon nos contraintes. WEBGREYC non plus, mais elle pourrait être utilisée en tant qu'alternative. Ainsi, les deux seuls jeux de données qui peuvent être utilisés dans notre étude sont GREYC et DSL2009.

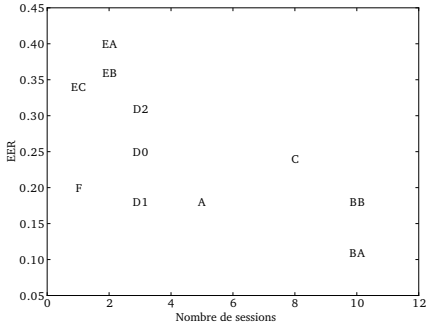
Bien sûr, ces différentes bases peuvent ne pas correspondre à de futures études. Dans ce cas, il est nécessaire pour les chercheurs d'acquérir de nouveaux jeux de données. Ces nouveaux jeux



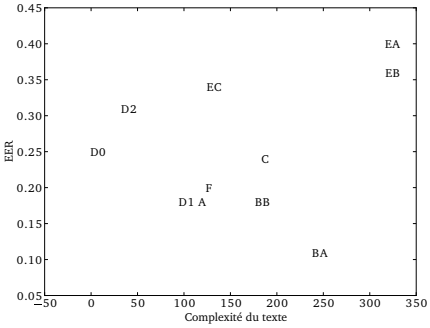
(a) EER vs entropie



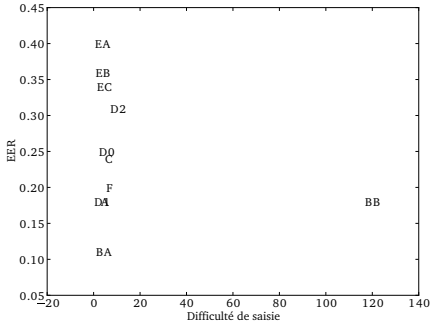
(b) EER vs nombre d'individus



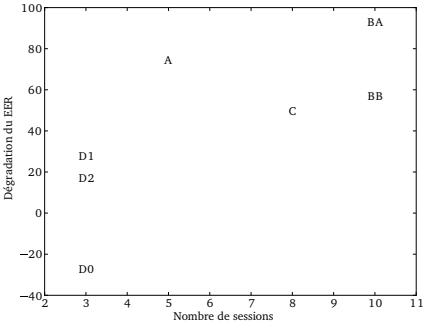
(c) EER vs nombre de sessions



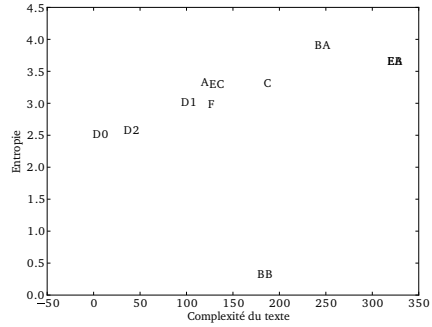
(d) EER vs complexité du texte



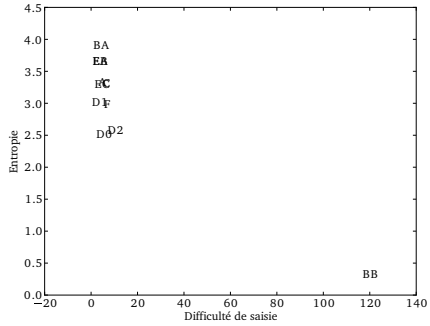
(e) EER vs difficulté de frappe



(f) EER dégradation vs nombre de sessions



(g) Entropie vs complexité du texte



(h) Entropie vs difficulté de frappe

FIG. 3.8: Comparaison de différentes propriétés pour chaque base

TAB. 3.9: Comparaison succincte des différents jeux de données afin d'aider à la décision de sélection. Chaque ligne présente le meilleur jeu en gras

Keu	A	B	C	D	E	F
Nb utilisateurs $\geq$ 50	✓	✗	✓	✗	✗	✓
Nb d'exemples/utilisateurs $\geq$ 30	✓	✓	✓	✗	✗	✗
Nb de sessions $\geq$ 3	✓	✓	✓	✓	✗	✗
Score	**3**	2	**3**	1	0	1
I = Rang par score	**1**	3	**1**	4	6	4
II = Rang par nombre d'utilisateurs valides	2	4	3	6	5	**1**
III = Rang par nombre d'exemples/utilisateurs	3	2	**1**	4	5	5
IV = Rang par nombre de sessions	3	**1**	2	4	5	5
V = Somme des rangs (II+III+IV)	8	7	**6**	14	15	11
VI = Somme totale des rangs (I + V)	9	10	**7**	18	21	15

de données doivent correspondre aux attentes du tableau 3.9, aussi bien qu'à des conditions spécifiques à la nouvelle étude.

3.3. Proposition d'une approche à l'aide d'un mot de passe partagé

3.3.1. Méthode développée

Nous avons développé une méthode qui prend en compte (i) le fait que nous demandons peu de captures aux utilisateurs, pour des questions évidentes d'usage tout en maintenant de bonnes performances, (ii) le fait que tous les utilisateurs saisissent le même texte. L'originalité de la méthode est due à l'utilisation conjointe des points suivants (bien qu'ils aient pu être rencontrés individuellement dans d'autres travaux de la littérature) :

- l'utilisation d'une méthode de discrétisation pendant une étape de pré-traitement.
- le calcul d'un score à partir de la décision d'un SVM (afin de prendre en compte les erreurs de classification du SVM).

Les SVM ont déjà été appliqués avec succès dans la littérature sur une base de 10 individus, ce qui nous semble trop peu pour valider les résultats [Sang *et al.*, 2005]. La discrétisation des données a également été utilisée par [Revett, 2009] afin de simuler une représentation sous formes de chaines d'acides aminées et d'appliquer des méthodes communes en bio-informatique. Les sections suivantes présentent le fonctionnement du système, tandis que la figure 3.9 résume le fonctionnement global.

3.3.1.1. Enregistrement

Cinq saisies sont nécessaires, au minimum, pour effectuer l'enregistrement. Chacune des données biométriques collectées est discrétisée en un alphabet de cinq valeurs (section 3.3.2.3.4). Ensuite, une machine à vecteur support est utilisée pour l'apprentissage de la référence biométrique de l'utilisateur (il y a donc un SVM par utilisateur). Nous utilisons un SVM à deux classes (annexe E). L'apprentissage du SVM se fait en utilisant les saisies d'enregistrement de l'utilisateur en les étiquetant 1 et les saisies d'enregistrement des imposteurs (les autres utilisateurs) en les étiquetant -1. La référence biométrique contient donc deux types d'informations : les données nécessaires à la discrétisation des requêtes ; et le SVM déjà entrainé.

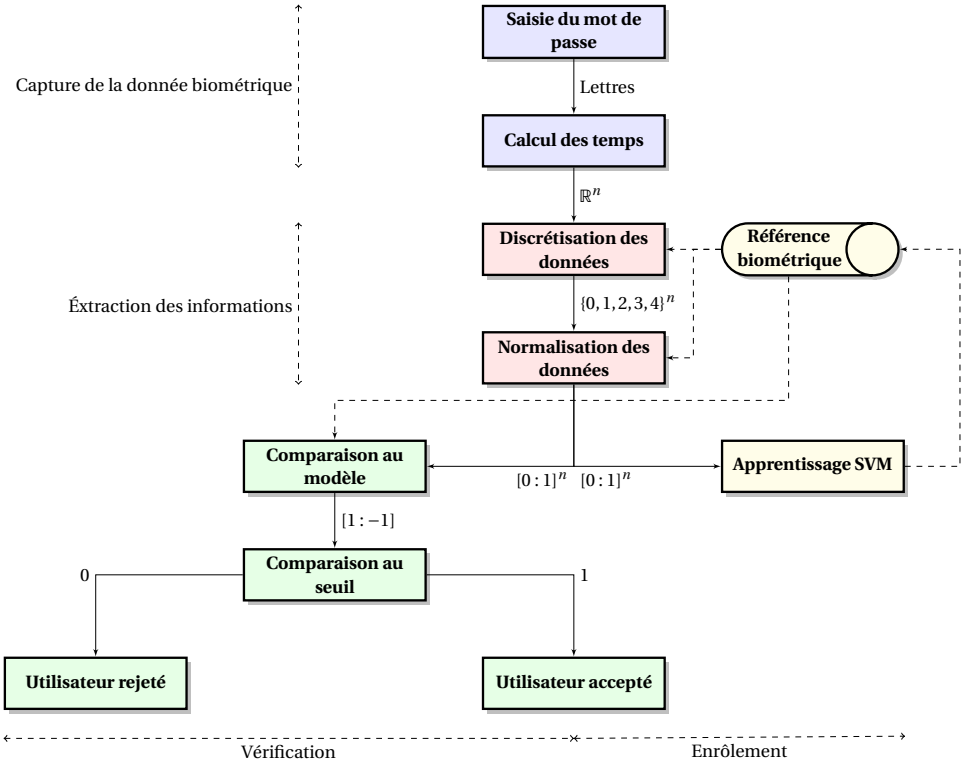


FIG. 3.9: Vue globale du système

Comme nous utilisons des données d'imposteurs (données des autres utilisateurs du système, qui ne cherchent pas à imiter le comportement de l'utilisateur authentique) durant l'étape d'apprentissage, la définition de la référence biométrique nécessite l'utilisation des références des autres utilisateurs. Quand les données de tous les utilisateurs (m est le nombre d'utilisateurs) sont prises en compte (c'est le scénario choisi dans cette expérience), il y a $5 * m$ données d'apprentissage (5 appartenant à l'utilisateur et $5*(m-1)$ appartenant aux imposteurs). Si un nouvel utilisateur est ajouté au système plus tard, différents scénarios peuvent être appliqués :

- Nous re-calculons la référence biométrique de tous les utilisateurs. Dans ce cas, $m+1$ références doivent être calculées en utilisant $5 * (m+1)$ exemples à chaque fois. Cette méthode peut être longue si beaucoup d'utilisateurs sont enregistrés dans le système. De plus, les performances de reconnaissance pour un utilisateur peuvent ne pas s'améliorer en ajoutant ces 5 nouvelles données aux exemples d'impostures. Ces nouvelles données peuvent être insignifiantes (en terme d'apport d'information) face au $5 * (m-1)$ données déjà présentes. Donc, le ratio entre la perte de temps et l'évolution des performances peut ne pas être un bon compromis.
- Nous calculons seulement le modèle du nouvel utilisateur en conservant les autres identiques. C'est plus efficace car seule une référence doit être générée.

Cependant, nous n'avons exploré aucune de ces pistes, car nous n'avons pas étudié l'inclusion des utilisateurs durant la vie du système.

3.3.1.2. Vérification

L'étape de vérification consiste à effectuer la procédure de reconnaissance de la requête par le SVM. Un score est généré, puis comparé à un seuil pour décider s'il s'agit d'une donnée authentique ou d'imposture. Si la vérification est réussie, la nouvelle capture peut être utilisée pour

mettre à jour de façon supervisée la référence afin de prendre en compte l'évolution de la façon de taper au clavier.La requête est discrétisée à l'aide des informations sur les bornes de discrétisation stockées dans la référence biométrique de l'utilisateur. Il s'agit donc d'un changement de repère spécifique à l'utilisateur. Elle est ensuite classée à l'aide du SVM, et la confiance de l'étiquette trouvée est également calculée. Il est ensuite nécessaire de générer un score afin d'obtenir une courbe ROC avec plusieurs points, et permettre une meilleure configuration du système (autrement, la courbe n'aurait qu'un unique point de fonctionnement). Le score est calculé de la façon suivante[11] :

$$score = -prb * prd \tag{3.8}$$

avec *prb* représentant la probabilité accordée au résultat du SVM et *prd* correspondant à la classe prédite du résultat (-1 pour un imposteur, 1 pour un authentique). Le seuil de décision peut être choisi à l'aide des deux approches suivantes:

- en utilisant le même seuil pour tous les utilisateurs ;
- en utilisant un seuil spécifique à chaque utilisateur.

Il est connu que les performances mesurées du système sont différentes en fonction du choix effectué [Hocquet *et al.*, 2006; Hosseinzadeh et Krishnan, 2008] (l'approche par seuil individuel étant plus avantageuse en termes de pourcentage d'erreur calculé). Les deux approches sont comparées à la section 3.3.3.4. Le but du travail n'est pas de présenter comment choisir ces seuils. Leur configuration dépend du niveau de sécurité à atteindre et peut être définie empiriquement ou automatiquement.

3.3.2. Étude comparative avec les systèmes de l'état de l'art

3.3.2.1. De l'absence d'études comparatives

Comme cela a été précisé dans plusieurs études [Crawford, 2010; Giot *et al.*, 2009; Karnan *et al.*, 2011; Killourhy et Maxion, 2011], il est difficile, voire impossible, de comparer les études entre elles. Plusieurs raisons expliquent ce problème :

- Les études sont toutes faites avec des bases de données différentes et rarement publiques, et les algorithmes efficaces sur une base particulière ne le sont pas autant sur d'autres (section 3.2). Il est donc impossible de comparer objectivement un algorithme évalué sur une base avec un autre algorithme évalué sur une autre base. Par exemple, un système utilisant des réseaux de neurones montre un FRR de 1% dans l'article initial, alors que son FRR est de 85.9% en utilisant une autre base de données [Killourhy et Maxion, 2011]. On peut s'attendre à ce que ce type de phénomène soit courant dans la littérature.
- À peine 54% (en étant relativement souple sur les critères) des articles sur la DDF effectuent des études comparatives. Même les auteurs des articles publiés ne savent pas si leur méthode est plus efficace que d'autres de l'état de l'art.
- À peine 8% (toujours en étant relativement souple sur les critères) des articles sur la DDF effectuent des analyses statistiques de leurs résultats. Les conclusions ne sont donc pas rigoureuses et peuvent être interprétées différemment en fonction du point de vue du lecteur.

Étant donné qu'il n'existe pas d'étude comparative dans la littérature, nous avons dû en réaliser une nous-même à l'aide d'une base créée pour l'occasion, et diffusée par la suite[12].

11. On aurait aussi pu calculer un score utilisant directement la probabilité que l'exemple soit dans la classe $+1$

12. La base DSN2009 n'était pas encore diffusée au moment de ce travail.

3.3.2.2. Base de données utilisée

3.3.2.2.1. Caractéristique de la base Nous utilisons la base rapidement décrite en 3.2.1.3, plus de précisions concernant son acquisition sont présentées ici. Dans [Cherifi *et al.*, 2009], nous argumentons qu'il est nécessaire d'avoir au minimum trois sessions dans une base de données biométrique comportementale. Ces sessions devant être espacées dans le temps avec une population large et diversifiée. Nous avons tenté de correspondre au maximum à ces remarques lors de la création de notre base de données. Il faut noter que c'est une condition préalable qui a rarement été respectée dans la littérature ; le nombre d'utilisateurs est souvent très faible, et la capture est souvent effectuée lors d'une session unique. *GREYC-Keystroke* [Giot *et al.*, 2009] est le logiciel qui nous a permis de créer la base de données[13]. Une capture de l'application est disponible en figure 3.10. Cette application a également été créée dans le but d'être utilisée par la communauté scientifique pour produire d'autres bases de données désirant répondre à des questions différentes. Les données sont stockées dans un fichier sqlite qui permet d'accéder rapidement, et facilement, à des informations spécifiques à l'aide de requêtes SQL (Structured Query Language).

La plupart des participants sont des chercheurs en informatique, des secrétaires, des étudiants en informatique et chimie. Il y a différents types de « tapeurs » : rapide, lent, à deux doigts, avec tous les doigts, etc... Cependant, nous n'avons pas capturé cette information. Un total de 133 individus ont participé à la création de la base en saisissant la phrase « greyc laboratory » entre 5 et 107 fois, du 18/03/2009 au 05/07/2009. 7555 sont disponibles, et le nombre moyen d'acquisitions par utilisateur est 51 avec 100 possédant plus de 60 captures. Ainsi, la plupart des utilisateurs ont participé à 5 sessions. Nous avons choisi cette phrase pour plusieurs raisons :

- c'est le nom de notre laboratoire, et, l'utiliser peut nous aider à le rendre plus visible ;
- il s'agit d'un mot de passe suffisamment long, possédant une bonne distribution des touches sur le clavier, ce qui aide à la discriminabilité [Revett *et al.*, 2006].

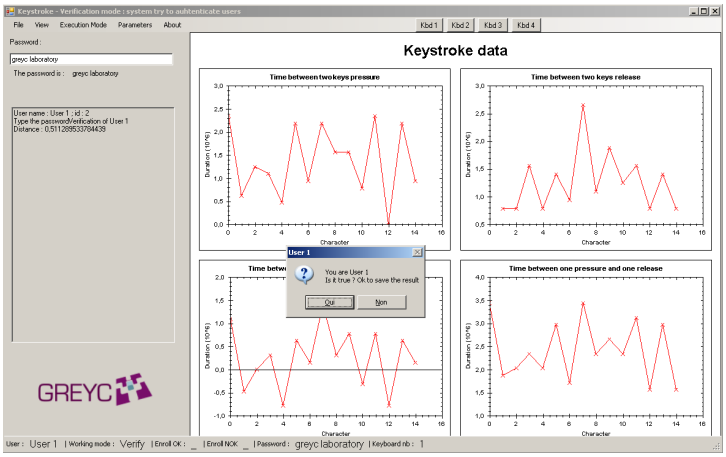


FIG. 3.10: Capture de l'application de collecte de données

La position des touches sur un clavier AZERTY est présentée en figure 3.11. Nous n'avons pas capturé d'autres mots de passe en raison du temps nécessaire pour créer une base de données. L'investissement étant trop important, nous avons choisi de distribuer l'application de collection de base.

En comparaison avec les bases utilisées dans les articles souvent cités (tableau 3.3, 31), notre base est plutôt une base importante, collectée sur une longue période. Nous avons demandé aux participants de contribuer à une session par semaine (une faible quantité d'entre eux a

13. http://www.ecole.ensicaen.fr/~rosenber/keystroke.html

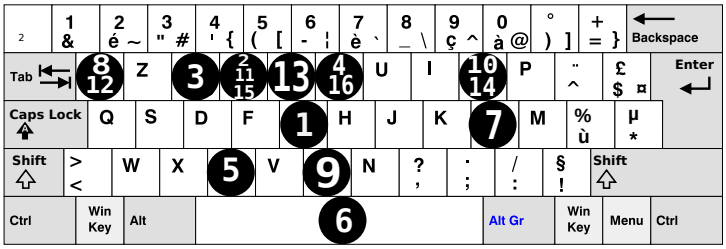


FIG. 3.11: Position des touches utilisées sur un clavier AZERTY. Les touches marquées sont utilisées dans le mot de passe. Les numéros indiquent l'ordre du caractère dans le mot de passe. Le schéma original est extrait de wikipedia

participé à plusieurs sessions dans une semaine pour des contraintes de calendrier). Chaque session consiste en la saisie correcte du même mot de passe douze fois. Lors de la première session, les participants étaient autorisés à s'entrainer à taper le mot de passe sur une courte période. Deux claviers différents ont été utilisés : (i) le clavier original de l'ordinateur portable, et (ii) un clavier USB branché sur l'ordinateur portable. Nous avons fait ce choix pour vérifier s'il y a une dépendance entre les performances et le clavier. Ainsi, durant chaque session, les utilisateurs doivent saisir six fois le mot de passe sur chaque clavier en alternant les claviers à chaque fois. Cela permet également d'éviter les saisies qui sont non réalistes et mécaniques pour les utilisateurs ne levant pas les mains du clavier. Nous ne réduisons donc pas la variabilité intra-classe de façon artificielle. La figure 3.12 présente les deux claviers différents utilisés pour l'expérience. Nous pouvons observer que les formes sont différentes. La sensation de pression sur les touches est également différente, et, la présence du curseur rouge au milieu du clavier du portable est perturbante pour la majorité des utilisateurs.

(a) Portable (b) USB

FIG. 3.12: Différence entre les deux claviers de l'expérience

Lors de la première session, les participants peuvent s'entrainer à saisir le mot de passe, sans que celui-ci ne soit capturé, aussi longtemps qu'ils le souhaitent. Nous n'avons pas enregistré le nombre d'essais par utilisateur, mais, la plupart d'entre eux ne l'ont pas fait plus de cinq fois. Cet entrainement est nécessaire car il s'agit d'un mot de passe imposé, les utilisateurs n'ont pas nécessairement l'habitude de le saisir (de plus, il est écrit dans une langue étrangère). C'est une étape nécessaire car la variabilité intra-classe serait trop importante sans cet apprentissage. Les participants savaient lorsqu'ils étaient en mode d'entrainement ou de capture. La phase d'entrainement n'était pas autorisée lors des autres sessions. Un résumé du sous ensemble de la base utilisée pour l'étude comparative est présenté en table 3.10. La base fait partie de la famille des bases utilisant un seul mot de passe [Killourhy et Maxion, 2009; Sheng *et al.*, 2005].

Nous utilisons toutes les données extraites lors de cette évaluation. Ce qui signifie que pour chaque mot de passe, nous disposons des types de temps suivants :

PP Il s'agit du temps entre chaque pression de touches successives.

TAB. 3.10: Résumé de la sous-base utilisée pour la comparaison renseignée à l'aide d'un questionnaire. Les utilisateurs ayant répondu au questionnaire ne sont pas forcément ceux ayant participé à la comparaison

Information	Description
Utilisateurs	100 utilisateurs
Exemples	6000 phrases (60 exemples par utilisateur)
Taille d'un exemple	16 caractères ('greyc laboratory')
Erreurs de frappe	non autorisé
Acquisition contrôlée	oui
Distribution des âges	entre 19 et 56 (répartition présentée en table Table 3.11)
Distribution des genres	environ 73% d'hommes et 27% de femmes
Fréquence d'utilisation d'un PC	inconnue
Profession de l'utilisateur	étudiants, chercheurs, secrétaires, employés (répartition inconnue)
Clavier	2 claviers AZERTY (1 sur le portable, 1 USB)
Plateforme d'acquisition	Windows XP/Greyc keystroke

PR Il s'agit de la durée de pression d'une touche.

RR Il s'agit de l'intervalle de temps entre le relâchement de chaque touche.

RP Il s'agit du délai entre le relâchement d'une touche et la pression de la suivante.

Ainsi, une donnée $\boldsymbol{v}$ est représentée de la façon suivante :

$$\boldsymbol{v} = \{RR_0,PP_0,RP_0,PR_0,RR_1,PP_1,RP_1,PR_1...\} \tag{3.9}$$

La taille des données extraites dépend naturellement de celle du mot de passe. Pour un mot de passe de n caractères, $\boldsymbol{v}$ a une dimension de $3 * (n-1) + n$.

3.3.2.2.2. Analyse de la base 100 volontaires ayant participé à la création de la base ont répondu à un questionnaire dont les résultats sont présentés en annexe A. Leur âge et genre sont présents en table 3.11.

TAB. 3.11: Diversité de la population ayant répondu au questionnaire

	Homme	**Femme**	**Total**
18-25	46	13	59
26-35	19	6	25
36-49	8	6	14
50+	0	2	2
Total	73	27	100

Étant donné qu'il est nécessaire d'effectuer une action précise (la saisie correcte du mot de passe), le risque d'effectuer une acquisition erronée est plus important que pour une modalité morphologique. Ces erreurs peuvent être dues au fait que la correction des fautes de frappe n'est pas permise lors de la saisie du mot de passe : une erreur implique de resaisir le mot de passe depuis le début. Il est donc intéressant d'analyser ces erreurs car, à notre connaissance, cela n'a jamais été fait dans la littérature. La figure 3.13 présente la quantité de captures effectuées par chaque utilisateur en affichant le nombre de saisies correctes (en gris) et de saisies erronées (en noir). Le nombre d'erreurs faites est important pour la plupart des volontaires. Le taux d'erreur

à l'acquisition est proche de 20% : une saisie sur cinq est fausse en raison d'une faute de frappe. Ces erreurs sont dues à plusieurs raisons :

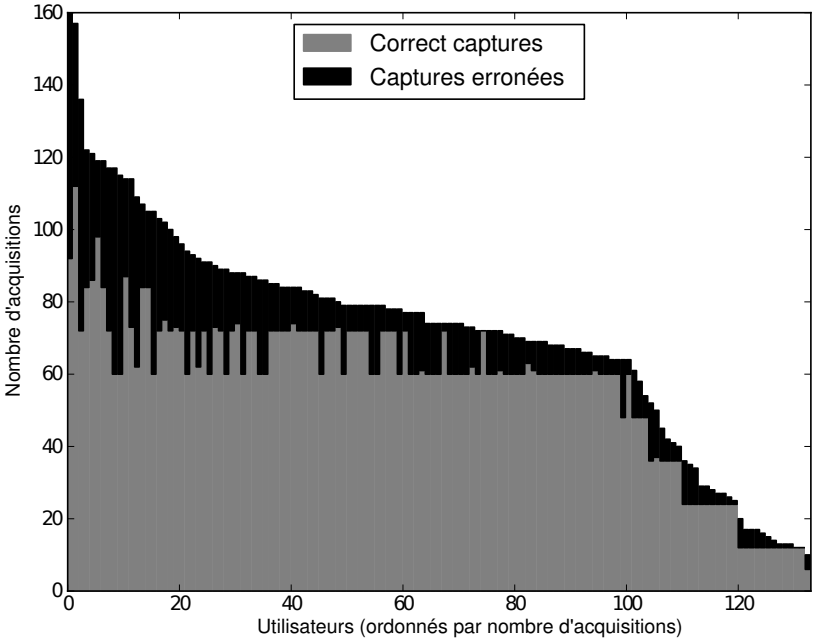


FIG. 3.13: Nombre d'acquisitions de chaque participant. Acquisitions correctes en gris et erronées en noir

- le mot de passe est relativement long à saisir (seize caractères), et, il est connu que le taux d'erreurs de frappe augmente si plus de huit caractères sont utilisés [Hosseinzadeh et Krishnan, 2008] ;
- les utilisateurs ne sont pas nécessairement habitués au clavier et peuvent hésiter en tapant (certains participants n'utilisent pas régulièrement un ordinateur, tandis que d'autres sont perturbés de saisir une phrase en anglais);
- certains utilisateurs veulent taper plus vite qu'ils ne sont réellement capables de le faire[14]... ;
- certains utilisateurs oublient le mot de passe à saisir (les sessions sont espacées de plus d'une semaine pour certains participants, et il n'y a pas de phase d'entrainement avant les sessions. D'autres utilisateurs ont également participé à la création d'autres jeux de données avec un mot de passe différent et les confondent);
- les utilisateurs peuvent être perturbés par leur environnement (discussion avec un collègue, fond sonore bruyant,...);
- les utilisateurs doivent saisir un mot de passe imposé ;

Habituellement, nous tapons plus vite un mot de passe choisi qu'un mot de passe imposé. Nous avons vérifié si le taux d'erreur d'acquisition était dépendant de la vitesse de frappe. Nous avons montré qu'il n'y a pas de corrélation significative (coefficient de corrélation de Pearson de $-0{,}28$). La figure 3.14 présente les taux d'erreur d'acquisition (lors de la création de la base) en fonction de la vitesse de frappe moyenne des utilisateurs. Comme nous pouvons voir, l'expérience ne révèle pas de dépendance entre ces facteurs : dans tous les intervalles, il y a des taux d'erreur importants.

14. ce comportement est probablement dû au fait qu'il y a quelqu'un pour contrôler l'acquisition

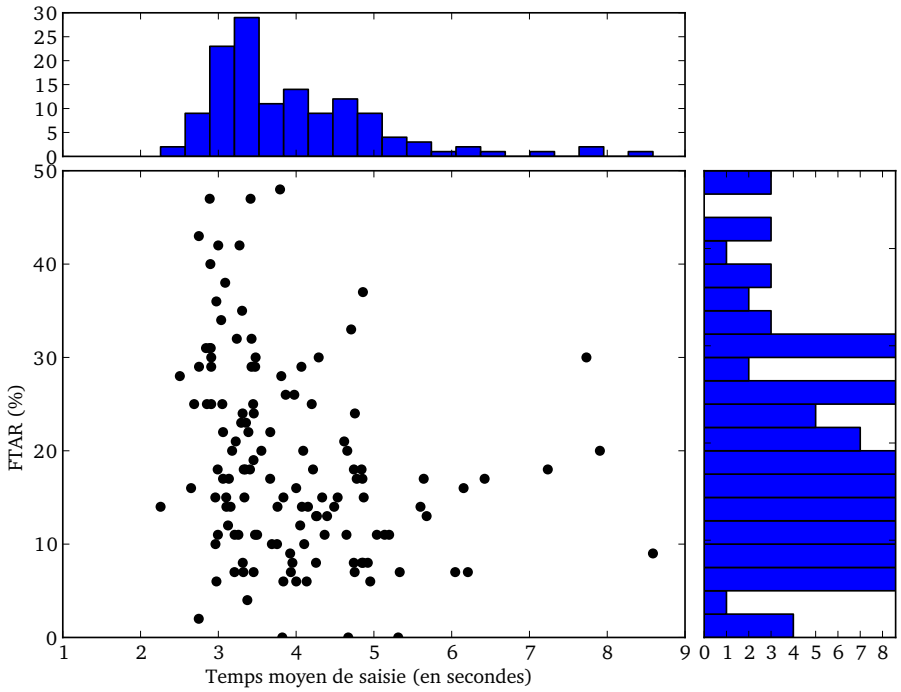


FIG. 3.14: Taux d'erreur d'acquisition en fonction de la vitesse moyenne de saisie de chaque utilisateur

3.3.2.3. Méthodes sélectionnées

Notre thème de recherche s'intéresse aux méthodes d'authentification faiblement contraintes et nous percevons le fait de devoir capturer plusieurs données lors de l'enregistrement comme une contrainte. Pour cette raison, parmi la multitude de méthodes présentées dans la littérature, nous n'avons choisi que des méthodes susceptibles de fonctionner avec peu de données d'apprentissage, ou des méthodes proches de notre proposition présentée en 3.3.

3.3.2.3.1. Méthodes sélectionnées pour l'étude comparative Dans cette partie, nous présentons les méthodes sélectionnées pour l'étude comparative. Nous avons sélectionné celles qui fonctionnent avec peu de données d'enregistrement (moins de 10). Nous notons $\mathbf{v}$ la requête, n la taille des données extraites (et des données de référence). Comme l'étude est faite en utilisant un seul mot de passe, les scores ne sont pas normalisés.

3.3.2.3.2. Algorithmes de type statistique Trois méthodes statistiques différentes sont testées. Elles diffèrent dans le contenu du modèle biométrique généré et la complexité de la méthode de calcul de distance. Dans la première méthode, la référence contient la moyenne $\boldsymbol{\mu}$ des exemples d'apprentissage [Bleha *et al.*, 1990]:

$$STAT1 = \frac{(\mathbf{v} - \boldsymbol{\mu})^t (\mathbf{v} - \boldsymbol{\mu})}{||\mathbf{v}||.||\boldsymbol{\mu}||} \tag{3.10}$$

Pour la seconde méthode, la référence contient à la fois la moyenne $\boldsymbol{\mu}$ et l'écart type $\boldsymbol{\sigma}$ [Hocquet *et al.*, 2007]:

$$STAT2 = 1 - \frac{1}{n}\sum_{i=1}^{n} e^{-\frac{|v_i - \mu_i|}{\sigma_i}} \tag{3.11}$$

La troisième méthode utilise la moyenne $\boldsymbol{\mu}$, l'écart type $\boldsymbol{\sigma}$ et la médiane $\boldsymbol{m}$ [Revett *et al.*, 2006] des exemples d'apprentissage. Alors que les deux méthodes précédentes pouvaient être représentées par une simple équation, celle-ci nécessite plusieurs étapes de calcul qui permettent de calculer un score basé sur le nombre d'éléments du vecteur de vérification inclus dans un ensemble de bornes définies par les données d'enregistrement. D'abord, nous vérifions si la requête satisfait les conditions présentées en (3.12). Il s'agit d'un calcul vectoriel effectué sur tous les éléments de la requête $\mathbf{v}$.

$$\mathbf{boolres} = min(\boldsymbol{\mu},\mathbf{m}) * \left(0.95 - \frac{\boldsymbol{\sigma}}{\boldsymbol{\mu}}\right) \leq \mathbf{v} \leq min(\boldsymbol{\mu},\mathbf{m}) * \left(1.05 + \frac{\boldsymbol{\sigma}}{\boldsymbol{\mu}}\right) \tag{3.12}$$

Dans la seconde étape, toutes les occurrences de faux sont remplacées par un 0, et chaque occurrence de vrai précédée par un 0 est remplacée par 1,5, tandis que les autres occurrences de vrai sont remplacées par 1. Nous disposons ainsi d'un tableau de nombres. La troisième étape consiste à effectuer la somme de tous les éléments du tableau, cette somme étant le score de la méthode biométrique. Nous notons $STAT3$ cette méthode.

3.3.2.3.3. Algorithme basé sur une distance Nous considérons une métrique simple basée sur la distance euclidienne [Monrose et Rubin, 1997]. Dans cette méthode, la référence biométrique contient tout simplement la liste des données d'enregistrement. La distance euclidienne est calculée entre la requête et chaque donnée d'enregistrement. Le score est alors la distance minimale comme décrit en (3.13).

$$DIST = \min\left(\forall_{\mathbf{u}\in enrol}, \sqrt{\sum_{i=1}^{n}(u_i - v_i)^2}\right) \tag{3.13}$$

3.3.2.3.4. Méthode basée sur le rythme Le principe de la méthode est de discrétiser les données de DDF selon cinq classes différentes, puis d'utiliser une distance de Hamming classique [Hocquet *et al.*, 2007]. La référence biométrique contient le dictionnaire de discrétisation (afin de pouvoir discrétiser la requête) ainsi que $\boldsymbol{\mu}^d$ la version discrétisée du vecteur moyen. Le calcul du score est décrit en (3.14).

$$RHYTHM = \frac{1}{n}\sum_{i=1}^{n} \text{abs}\left(\text{classe}\,(v_i) - \text{classe}\left(\mu_i^d\right)\right) \tag{3.14}$$

avec $classe(i)$ la fonction retournant la classe de i (*c.-à-d.*, application de la discrétisation) selon cinq classes différentes (cinq est un nombre qui semble optimal [Giot *et al.*, 2011b]). Pour calculer ces classes, nous divisons l'espace en cinq ensembles de même taille entre les valeurs minimales et maximales des composantes du vecteur $\mathbf{V}$ sur l'ensemble d'apprentissage (équation 3.15). La classe assignée à chaque dimension du vecteur est le numéro du groupe.

$$cluster\_width = \frac{max(train\_data) - min(train\_data)}{5} \tag{3.15}$$

qui est propre à chaque utilisateur.

3.3.2.3.5. Réseaux de neurones Les réseaux de neurones ont été utilisés dans différentes études de DDF [Anagun, 2006; Brown et Rogers, 1993; Cho *et al.*, 2000; Obaidat et Sadoun, 1997; Uzun et Bicakci, 2012]. Ils nécessitent habituellement une quantité d'exemples d'apprentissage pour créer la référence biométrique. Les réseaux de neurones et les SVM étant utilisés dans le même type de problèmes, nous avons choisi de comparer notre méthode (utilisant un SVM) aux réseaux de neurones. Nous utilisons un perceptron multi-couches contenant une couche cachée avec 45% de nœuds par rapport aux nombre de nœuds de la couche d'entrée. Nous avons choisi cette valeur de façon empirique afin de limiter le temps d'apprentissage tout en gardant une performance acceptable. La fonction coût est la somme des différences au carré. La méthode d'apprentissage est l'algorithme de Newton tronqué. Nous n'avons pas testé d'autres configurations de réseaux de neurones. Ainsi, dans cette méthode, la référence biométrique contient le réseau entrainé à l'aide des données d'enregistrement authentiques et d'imposteurs.

3.3.3. Résultats expérimentaux

Dans cette section, nous présentons les différents résultats expérimentaux sur notre base de données (GREYC). Une validation a également été faite sur une seconde base de l'état de l'art (DSL2009). CONTRIB se réfère à notre méthode de vérification. Les autres méthodes sont celles présentées dans la section 3.3.2.

3.3.3.1. Indépendance par rapport au clavier

Le tableau 3.12 présente l'EER en fonction du clavier utilisé pour l'enregistrement et la vérification pour notre méthode et six autres de la littérature. L'EER a été calculé en utilisant les dix premiers exemples pour l'apprentissage, et les autres pour la vérification. Le seuil de décision est global. Quand les claviers utilisés pour l'enregistrement et la vérification sont différents, les calculs sont effectués plusieurs fois en sélectionnant les captures d'enregistrement aléatoirement et en moyennant les résultats.

TAB. 3.12: EER(%) des méthodes en fonction de la configuration clavier. Le meilleur EER de chaque méthode est présenté en italique, tandis que le meilleur EER de chaque configuration est présenté en gras

Methode	EER11 (%)	EER22 (%)	EER12 (%)	EER21 (%)	EERaa (%)
STAT1	24,91	23,96	24,73	*23,51*	25,50
STAT2	17,68	*16,55*	17,10	16,65	17,58
STAT3	15,10	13,81	14,68	*13,22*	15,43
DIST	27,01	26,00	26,46	*25,07*	27,56
RHYTHM	19,40	20,09	*19,25*	19,50	19,78
NEURAL	12,65	12,03	12,15	***11,21***	13,62
CONTRIB	**10,68**	**10,37**	***10,30***	11,76	**11,96**
Moyenne	18,20%	17.54%	17.81%	17.27%	18.77%

EERab signifie que le clavier d'enregistrement est le clavier a, et le clavier de test est le clavier b. Le clavier 1 est le clavier de l'ordinateur portable, et le clavier 2 est le clavier USB branché dessus. Nous voyons que les résultats sont différents en fonction de la configuration du clavier. Curieusement, six fois sur sept, les meilleurs résultats sont obtenus quand le clavier d'enregistrement et de vérification sont différents, alors qu'on s'attendrait à obtenir de meilleures performances en utilisant le même clavier. Cinq fois sur sept, les meilleurs résultats sont obtenus lorsque le clavier d'enregistrement est le clavier 2 (clavier USB). On peut interpréter ses points en disant

que les données sont plus stables sur ce clavier contrairement à celles capturés sur le clavier de l'ordinateur portable. Les résultats les moins bons sont obtenus en utilisant le même clavier pour l'enregistrement et la vérification. La méthode proposée donne de meilleurs résultats dans la plupart des configurations.

Cependant, en utilisant un test de Kruskal-Wallis (annexe B.4) sur les cinq colonnes du tableau, nous obtenons une valeur p de 0.9673. Cette valeur montre qu'il n'y a pas de différences significatives entre les deux claviers.Nous avons également vérifié s'il est possible de reconnaitre le clavier utilisé pour saisir le mot de passe. En utilisant un SVM avec un mécanisme de validation croisée à dix échantillons, et en répétant la procédure cinquante fois, nous obtenons un taux de reconnaissance du clavier de 61,48% avec un écart type de 0,17. Ces résultats ne sont pas suffisamment différents du hasard pour dire que nous sommes capables de reconnaitre le clavier et expliquer ces différences.

3.3.3.2. Impact de la taille de la galerie

La quantité de données stockées dans la galerie utilisée lors de l'enregistrement diffère d'une étude à l'autre. La performance des algorithmes peut varier en fonction de cette quantité. La plupart des études utilisent plus de trente captures pour créer la référence biométrique, alors que nous pensons que cinq captures par utilisateur est vraiment le nombre maximum toléré par un utilisateur (surtout en considérant que celui-ci doit s'entrainer à saisir le mot de passe avant). La figure 3.15 présente l'EER des différents algorithmes en fonction du nombre d'exemples utilisés lors de l'enregistrement. Il est clair que les performances s'accroissent (*c.-à-d.*, l'EER diminue) en augmentant le nombre d'exemples. Pour toutes les méthodes, moins de dix captures donnent de très mauvais résultats. Afin d'obtenir les meilleurs résultats, le nombre d'exemples requis doit se situer autour de quarante (cependant, dans ce cas, le nombre de données de test est très petit, et les résultats sont moins significatifs). La méthode que nous proposons donne les meilleurs résultats, même en utilisant moins de dix exemples. Plus le nombre d'exemples d'enregistrements est important, plus notre méthode donne de meilleurs résultats que les autres méthodes. Elle peut donc être utilisée dans un environnement non critique et faiblement contraint.

3.3.3.3. Impact du pourcentage d'imposteurs

Tous les calculs précédents ont été faits en utilisant tous les imposteurs disponibles. Ici, nous voulons voir quel est l'impact sur les performances en fonction du nombre d'imposteurs conservé. Pour différents ratios d'imposteurs, nous avons calculé le modèle en utilisant toutes les données clientes et toutes les données des imposteurs sélectionnés, puis nous avons calculé l'EER. L'expérience a été réalisée 20 fois et les résultats moyennés. La figure 3.16 présente le gain en fonction du nombre d'imposteurs sélectionnés comparé à la méthode utilisant tous les imposteurs. Les valeurs négatives indiquent que les performances sont moins bonnes qu'en utilisant 100% d'utilisateurs, tandis que les valeurs positives indiquent que les résultats sont meilleurs. On voit clairement qu'utiliser tous les imposteurs donne de moins bonnes performances qu'utiliser 20% d'entre eux. Cela peut s'expliquer par un phénomène de manque de généralisation. Utiliser peu d'imposteurs donne de très mauvaises performances en comparaison de l'utilisation de tous. Cela peut s'expliquer par un manque de spécialisation. Dans les deux cas, un facteur important doit être le déséquilibre entre le nombre d'exemples de chaque classe.

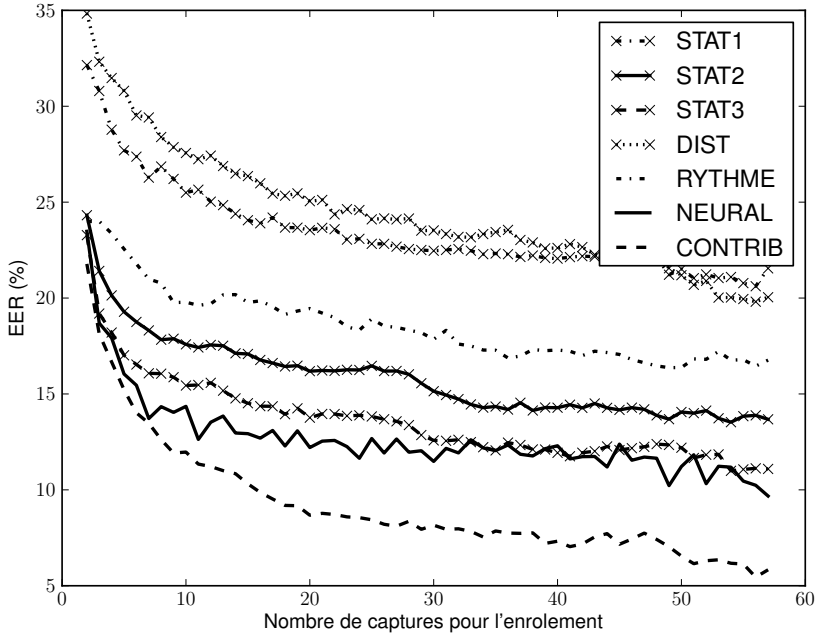


FIG. 3.15: Évolution de l'EER sur les différents algorithmes en considérant le nombre d'exemples d'apprentissage utilisé pour chaque individu client et imposteur

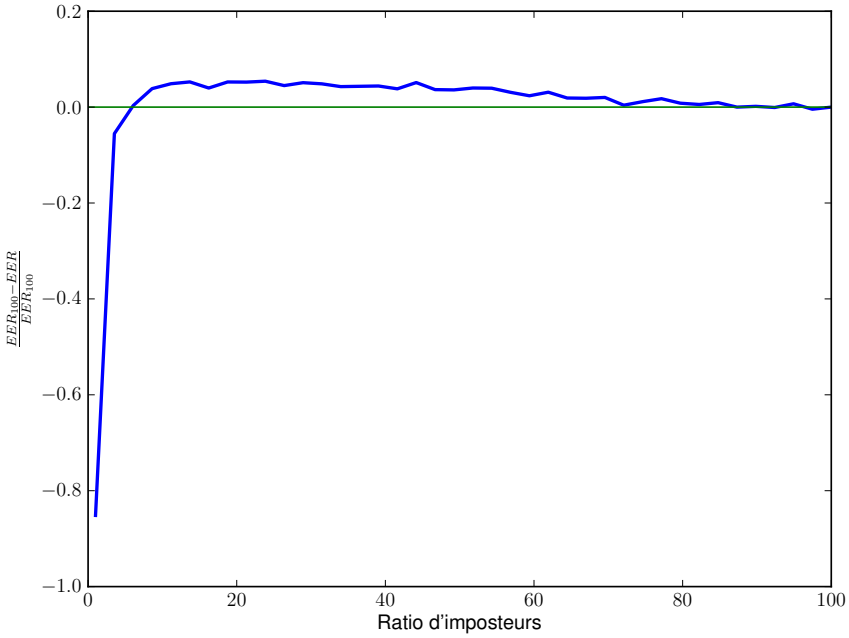


FIG. 3.16: Gain en fonction du nombre d'imposteurs conservés

3.3.3.4. Indépendance du seuil

Utiliser des seuils individuels, plutôt qu'un unique seuil global augmente les performances. Le tableau 3.13 présente le gain, en termes d'EER, à utiliser un seuil individuel comparé à un seuil global. Les EER sont calculés en utilisant cinq données pour l'enregistrement, la méthode de mise à jour intermédiaire et des données des deux claviers. En utilisant un test de Kruskal-Wallis (annexe B.4), nous avons obtenu une valeur p de 0,2774 qui indique que le gain à utiliser un seuil individuel par rapport à un seuil global est acceptable mais pas forcément significatif (valeur p inférieure à 0,05) sur ce jeu de données. Il faut noter que la configuration automatique d'un seuil individuel avec un système utilisant un mot de passe partagé est possible, mais elle ne peut pas être appliquée à un système où chaque utilisateur possède un mot de passe différent (aucun utilisateur n'accepterait de fournir son mot de passe pour permettre à des imposteurs de le saisir). Hocquet *et al.* [2006] proposent une solution à ce problème en classifiant, de façon automatique, les utilisateurs dans différents groupes en fonction de différents facteurs. Ces groupes sont initialement créés à l'aide d'une base de test. Chaque membre d'un même groupe possède les mêmes paramètres de configuration de la méthode de vérification. De cette façon, le seuil de décision individuel d'un utilisateur est obtenu en utilisant celui du groupe auquel il appartient (ce compromis permet d'obtenir des résultats proches de l'utilisation d'un seuil individuel).

TAB. 3.13: EER pour chacunes des méthodes en utilisant un seuil global et seuil individuel. Le meilleur EER de chaque méthode de calcul est présenté en gras

Methode	EER(global) (%)	EER(individuel) (%)	Gain
STAT1	20,94	*19,54*	1,4
STAT2	10,75	*9,22*	1,53
STAT3	9,78	*8,64*	1,14
DIST	24,65	*21,53*	3,12
RHYTHM	13,21	*10,02*	**3,18**
NEURAL	10,3	*8,75*	1,55
CONTRIB	**6,96**	***6,95***	0,01
Moyenne	13,8	**12,1**	1,7

3.3.3.5. Temps de calcul

Le temps de calcul nécessaire pour vérifier une requête à une référence biométrique est relativement similaire pour toutes les méthodes, cependant ce n'est pas le cas pour la génération de la référence biométrique. Le temps de calcul nécessaire pour générer la référence biométrique de tous les utilisateurs de la base (incluant le temps de lecture de la base) est présenté dans le tableau 3.14. Les temps sont calculés en utilisant 5 et 10 données pour générer le modèle à l'aide d'un script *Python* exécuté sur un ordinateur de bureau avec un Pentium IV à 3Ghz et 1Go de mémoire vive utilisant le système d'exploitation Linux. Nous pouvons voir que le temps de génération des références pour les méthodes STAT1, STAT2, STAT3, DIST, RHYTHM est équivalent. Le temps de calcul est plus important pour CONTRIB et NEURAL, mais CONTRIB reste près de sept fois plus rapide que NEURAL. Cependant il faut surement tempérer ces résultats, car le réseau de neurones n'a pas été optimisé. Tous les scripts sont écrits en Python (à la fois les algorithmes et l'évaluation) en utilisant psyco [Rigo, 2010] (un compilateur JIT qui accélère l'exécution du programme). La bibliothèque ffnet [Wojciechowski, 2007] est utilisée pour les réseaux de neurones et la bibliothèques libsvm [chung Chang et Lin, 2001] est utilisée pour les SVM.

TAB. 3.14: Temps de calcul nécessaire pour générer les références biométriques

Nb	STAT1	STAT2	STAT3	DIST	RHYTHM	NEURAL	CONTRIB
5	3s	3s	3s	3s	4s	5m 55s	54s
10	3s	3s	3s	3s	4s	30m 4s	4m 24s

3.3.3.6. Analyse statistique

La différence de performance entre les méthodes est faible. Cela signifie qu'elles peuvent ne pas être statistiquement différentes. Afin de comparer les algorithmes de façon plus juste, nous pouvons utiliser des tests d'hypothèse ou des intervalles de confiance. Nous avons calculé l'intervalle de confiance de l'EER, en utilisant cinq exemples pour l'apprentissage, pas de mécanisme d'adaptation, et un seuil global. La méthode appliquée est présentée dans [Mayoue, 2007] [15]. Les résultats obtenus sont présentés dans le tableau 3.15. En utilisant seulement cinq exemples d'enregistrement et pas de mécanisme d'adaptation, notre contribution donne de meilleurs résultats que les méthodes STAT1, STAT2, STAT3, DISTANCE et RHYTHME. Il y a un petit chevauchement avec NEURAL. Notre méthode reste intéressante, comparée à NEURAL, en raison du temps plus faible nécessaire pour générer les références.

TAB. 3.15: Intervalle de confiance des EER à 95%

Method	**EER min (%)**	**EER max (%)**	**Largeur de l'intervalle**
STAT1	27,09	28,42	1,33
STAT2	18,69	19,85	1,16
STAT3	16,64	17,71	1,07
DISTANCE	30,12	31,49	1,37
RHYTHME	21,70	22,95	1,25
NEURAL	15,32	16,40	1,08
CONTRIB	14,62	15,69	1,07

3.3.3.7. Variabilité du détecteur

Il existe une seconde base conséquente de DDF [16] citée en 3.2.1.2. Le tableau 3.16 résume cette base qui comporte quelques différences comparées à la nôtre. Notre base contient deux fois plus d'utilisateurs, nous pouvons donc obtenir des résultats sur une plus grande population. Notre base contient plus de variabilité intra-classe, car : (i) nos sessions semblent être plus espacées (une semaine au lieu d'un jour) pour la majorité des utilisateurs (ii) nous forçons une pause avant de saisir un nouveau mot de passe (*cf.*, l'utilisation de deux claviers), donc le participant ne peut pas saisir plusieurs fois à la suite le mot de passe de façon mécanique et (iii) nous utilisons deux claviers. Cependant, notre base a moins de sessions. Killourhy et Maxion [2009] ont testé 14 détecteurs d'anomalie différents sur leur base de données (se référer à leur travail pour plus d'informations). Ils présentent leurs résultats d'une façon différente de la nôtre : une courbe ROC est calculée pour chaque utilisateur, et son EER en est extrait. Ensuite, la moyenne et l'écart type des EER sont présentés dans l'article. Nous avons utilisé leur protocole sur leur base de données avec notre système afin d'observer son comportement. En utilisant un seuil global, nous obtenons un EER de 10.63%, alors qu'avec des seuils individuels, nous obtenons un EER moyen

15. en utilisant un intervalle de confiance à 95% au lieu de 90%

16. créée au même moment que la nôtre

de 9.39% (avec un écart type de 6.72%). Cela placerait notre méthode à la première position du classement de leur tableau 2 (qui présente les méthodes en fonction de leur performance et est reproduit en figure 3.17), car leur meilleure méthode (Manhattan scaled) donne un EER de 9.6% (avec un écart type de 6.9%). Nos résultats sont également meilleurs que l'utilisation du SVM à une classe. Cependant, nos meilleures performances peuvent s'expliquer par l'utilisation de données d'impostures, alors que les détecteurs d'anomalie utilisent uniquement des données authentiques.

TAB. 3.16: Résumé de la base DSN2009 [Killourhy et Maxion, 2009]

Information	Description
Utilisateurs	51
Exemples	20400 phrases (50*8 exemples par utilisateur)
Taille d'un exemple	10 caractères (« .tie5Roanl »)
Erreurs de frappe	non autorisé
Acquisition contrôlée	oui
Distribution des âges	entre 18 et 70
Distribution des genres	30 hommes et 21 femmes
Fréquence d'utilisation d'un PC	inconnu
Profession des utilisateurs	inconnu
Clavier	clavier QWERTY (portable)
Plateforme d'acquisition	Windows
Résolution de la capture	200 microsecondes

	Detector	equal-error rate			Detector	zero-miss false-alarm rate	
1	**Manhattan (scaled)**	**0.096**	**(0.069)**	**1**	**Nearest Neighbor (Mahalanobis)**	**0.468**	**(0.272)**
2	**Nearest Neighbor (Mahalanobis)**	**0.100**	**(0.064)**	**2**	**Mahalanobis**	**0.482**	**(0.273)**
3	**Outlier Count (z-score)**	**0.102**	**(0.077)**	**3**	**Mahalanobis (normed)**	**0.482**	**(0.273)**
4	SVM (one-class)	0.102	(0.065)	**4**	**SVM (one-class)**	**0.504**	**(0.316)**
5	Mahalanobis	0.110	(0.065)	5	Manhattan (scaled)	0.601	(0.337)
6	Mahalanobis (normed)	0.110	(0.065)	6	Manhattan (filter)	0.757	(0.282)
7	Manhattan (filter)	0.136	(0.083)	7	Outlier Count (z-score)	0.782	(0.306)
8	Manhattan	0.153	(0.092)	8	Manhattan	0.843	(0.242)
9	Neural Network (auto-assoc)	0.161	(0.080)	9	Neural Network (auto-assoc)	0.859	(0.220)
10	Euclidean	0.171	(0.095)	10	Euclidean	0.875	(0.200)
11	Euclidean (normed)	0.215	(0.119)	11	Euclidean (normed)	0.911	(0.148)
12	Fuzzy Logic	0.221	(0.105)	12	Fuzzy Logic	0.935	(0.108)
13	k Means	0.372	(0.139)	13	k Means	0.989	(0.040)
14	Neural Network (standard)	0.828	(0.148)	14	Neural Network (standard)	1.000	(0.000)

Table 2. The average equal-error rates (left side) and average zero-miss false-alarm rates (right side) from the evaluation of the 14 detectors are ranked from best to worst (with standard deviations in parentheses). The set of top-performing detectors is indicated in bold-face (i.e., those that are not significantly worse than the best-performing detector).

FIG. 3.17: Reproduction du tableau 2 des travaux de Killourhy et Maxion

3.4. Conclusion de la dynamique de frappe au clavier

L'authentification par Dynamique De Frappe au clavier (DDF) est une modalité intéressante, car elle ne nécessite pas de capteur supplémentaire et est bien acceptée par les utilisateurs [El-Abed *et al.*, 2010]. La performance de tels systèmes d'authentification est suffisamment importante pour pouvoir être utilisée dans des contextes de sécurité personnelle. La méthode que nous proposons, en authentification par mot de passe partagé, est plus efficace que les méthodes de l'état de l'art (en condition contrainte) même si le temps de calcul pour créer la référence reste élevé. Afin de pouvoir comparer la performance de notre méthode à celles des méthodes de l'état de l'art, nous avons dû créer un jeu de données conséquent avec plus de 100 utilisateurs ayant

participé à au moins 5 sessions. La base a été rendue disponible aux chercheurs du domaine. D'autres pistes restent à explorer pour la DDF. Cependant, il peut être nécessaire de créer une nouvelle base spécifique aux nouvelles études. Pour faciliter ce travail aux autres chercheurs du domaine, nous avons également rendu public l'outil de création de la base de données.

Notre contribution permet d'augmenter les performances de reconnaissance grâce à l'utilisation des données d'impostures pour créer la référence d'un utilisateur. Cependant, elle a majoritairement été comparée à d'autres méthodes n'étant pas capable d'utiliser de données d'imposture. Il serait intéressant d'analyser leur performance en les modifiant de telle façon à pouvoir modéliser le comportement d'un utilisateur et d'un imposteur, afin de prendre la décision d'acceptation à l'aide de ces deux informations.

Il serait intéressant d'analyser quelles seraient les meilleures stratégies à adopter en cas d'augmentation de la base de données au cours du temps, afin de recalculer les références en intégrant de nouveaux imposteurs. Il est probable que des méthodes efficaces peuvent permettre de diminuer le taux d'erreur de reconnaissance avec l'augmentation de la taille de la base.

Le fait d'utiliser des méthodes utilisant des données d'impostures (et d'utiliser des modèles génératifs) comparé à des méthodes n'utilisant pas de données d'impostures (et utilisant des modèles discriminants) à l'avantage d'améliorer les performances de reconnaissance et de simplifier l'administration du système (pas de gestion de perte de mot de passe de l'utilisateur, configuration de seuil de décision optimal aisé à l'aide d'une base de validation), cependant il ne devient plus possible d'utiliser la DDF dans un contexte d'authentification forte car le mot de passe n'est plus secret.

La discrétisation des données en cinq valeurs a permis d'améliorer les performances de reconnaissance comparée à l'utilisation des données réelles. Cela peut s'expliquer par le fait que les données de DDF sont fortement bruitées (charge processeur de la machine sur laquelle est effectuée l'acquisition, saisie non stable de l'utilisateur, délais inhérents à n'importe quelle modalité comportementale) et que cette discrétisation participe à effectuer un lissage des erreurs de temps. Il serait néanmoins intéressant de chercher d'autres méthodes de discrétisation plus efficaces.

Les deux chapitres suivants présentent des méthodes d'amélioration des performances dans des systèmes de reconnaissance par DDF ne nécessitant pas de données d'imposteurs pour créer la référence biométrique.

Rappel des contributions de cette partie

- Une méthodologie de comparaison de bases de données d'authentification par DDF, afin d'aider à sélectionner, ou créer, des bases pertinentes pour des études précises.
- Une nouvelle base de données de DDF de meilleure qualité que la majorité des bases publiques.
- La comparaison de différentes méthodes de l'état de l'art, respectueuses de nos contraintes, sur la base que nous fournissons.
- La proposition d'une méthode de reconnaissance par mot de passe partagé, plus performante que les méthodes de l'état de l'art en termes de taux d'erreur ou de temps de calcul.

Travaux de l'auteur sur ce thème de travail

Romain Giot : État de l'art de la dynamique de frappe au clavier. Rapport technique, UMR6072 - GREYC - Groupe de REcherche en Informatique, Image, Automatique et Instrumentation de Caen, novembre 2009.

Romain Giot, Mohamad El-Abed et Christophe Rosenberger : Keystroke dynamics authentication for collaborative systems. *In The IEEE International Symposium on Collaborative Technologies*

and Systems (CTS), pages 172–179, Baltimore, Maryland, USA, mai 2009a. IEEE Computer Society. URL `http://hal.archives-ouvertes.fr/hal-00432764/en`. Acceptance rate: 59/100.

Romain Giot, Mohamad El-Abed et Christophe Rosenberger : Greyc keystroke: a benchmark for keystroke dynamics biometric systems. *In IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS 2009)*, pages 1–6, Washington, District of Columbia, USA, septembre 2009b. IEEE Computer Society. URL `http://hal.archives-ouvertes.fr/hal-00432768/en/`. Acceptance rate : 56/100.

Romain Giot, Mohamad El-Abed et Christophe Rosenberger : Keystroke dynamics with low constraints svm based passphrase enrollment. *In IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS 2009)*, pages 1–6, Washington, District of Columbia, USA, septembre 2009c. IEEE Computer Society. URL `http://hal.archives-ouvertes.fr/hal-00432775/en`. Acceptance rate : 56/100.

Romain Giot et Christophe Rosenberger : Greyc keystroke : un logiciel de démonstration et de création de benchmark de dynamique de frappe au clavier. Logiciel, juin 2009.

Fouad Cherifi, Baptiste Hemery, **Romain Giot**, Marc Pasquet et Christophe Rosenberger : *Behavioral Biometrics for Human Identification: Intelligent Applications*, chapitre Performance Evaluation Of Behavioral Biometric Systems, pages 57–74. IGI Global, 2009.

Romain Giot, Mohamad El-Abed et Christophe Rosenberger : Authentification faiblement contrainte par dynamique de frappe au clavier. *In Conférence Reconnaissance des Formes et Intelligence Artificielle (RFIA 2010)*, Caen, France, janvier 2010. URL `http://hal.archives-ouvertes.fr/hal-00472618/en`. Acceptance rate (oral): 23.65/100.

Romain Giot, Mohamad El-Abed, Baptiste Hemery et Christophe Rosenberger : Unconstrained keystroke dynamics authentication with shared secret. *ELSEVIER International journal on Computers & Security*, 30(6-7):427–445, septembre 2011a. Impact Factor: 1.488.

Romain Giot, Mohamad El-Abed et Christophe Rosenberger : *Keystroke Dynamics Overview*, volume 1, chapitre 8, pages 157–182. InTech, juillet 2011b. URL `http://www.intechopen.com/articles/show/title/keystroke-dynamics-overview`.

Romain Giot : Static keystroke dynamics authentication. Poster at the 8th Summer School for Advanced Studies on Biometrics for Secure Authentication, juin 2011.

Mohamad El-Abed, Baptiste Hemery, **Romain Giot** et Christophe Rosenberger : Evaluation of biometric systems: A study of users' acceptance and satisfaction. *International Journal of Biometrics*, 4(3):265–290, 2012.

Romain Giot : Keystroke dataset denoising. Rapport technique, UMR6072 - GREYC - Groupe de REcherche en Informatique, Image, Automatique et Instrumentation de Caen, février 2012. Unpublished report.

Romain Giot, Mohamad El-Abed et Christophe Rosenberger : Web-based benchmark for keystroke dynamics biometric systems: A statistical analysis. *In The Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP 2012), Special Session 1: Advances on Biometrics*, pages 11–15, Piraeus, Greece, juillet 2012a. URL `http://hal.archives-ouvertes.fr/hal-00714251`. Acceptance rate: 40/100.

Romain Giot, Alexandre Ninassi, Mohamad El-Abed et Christophe Rosenberger : Analysis of the acquisition process for keystroke dynamics. *In Proceedings of the 11th International Conference of the Biometrics Special Interest Group*, pages 123–134, Darmstadt, Germany, septembre 2012b. GI-Edition. URL `http://hal.archives-ouvertes.fr/hal-00730381`. Acceptance rate: 29/100.

4. Multimodalité et biométrie douce

Sommaire

Présentation

Qu'est-ce que la multibiométrie et la biométrie douce ? Est-il possible d'augmenter les performances des systèmes de reconnaissance par Dynamique De Frappe au clavier (DDF) à l'aide de ces techniques ? Ce chapitre répond à ces questions et présente nos travaux de recherche liés à ces deux thématiques. Nous avons étudié différentes fonctions de fusion de scores, certaines étant générés automatiquement, ainsi que la reconnaissance du genre ou d'une catégorie d'âge d'un individu suivant sa façon de taper au clavier.

Mots clés :

Multimodalité, état de l'art, biométrie douce, fusion de scores, algorithme génétique, programmation génétique, dichotomie, évaluation, reconnaissance du genre, reconnaissance de la catégorie d'âge

Contributions de ce chapitre

- Le développement d'une méthode de calcul approximant le taux d'égales erreurs (*Equal Error Rate*) (EER) afin d'accélérer le temps de calcul des algorithmes évolutionnaires utilisant l'EER comme fonction d'évaluation. La méthode a un grand intérêt dans la configuration de systèmes multimodaux.
- L'illustration de l'intérêt d'utiliser deux biométries de performance moyenne (la DDF et la Reconnaissance Faciale (RF)) avec une forte acceptation de la part des utilisateurs afin d'obtenir un système plus performant.
- L'intérêt de l'utilisation de la programmation génétique (PG) afin de déterminer, de façon totalement automatique, la fonction de fusion des scores.
- La détection automatique du genre d'un individu selon sa façon de taper au clavier un texte fixe, ainsi que l'utilisation de cette information pour augmenter les performances de reconnaissance par DDF.
- La détection automatique du genre et d'une classification selon son âge d'un individu selon sa façon de taper un texte quelconque.

Organisation du chapitre

La section 4.1 présente l'état de l'art de la multibiométrie et des biométries douces. Les sections suivantes présentent nos travaux sur ces thématiques. La section 4.2 présente une méthode d'approximation rapide de l'EER qui a un intérêt dans les mécanismes d'optimisation de systèmes multimodaux. La section 4.3 présente nos propositions de systèmes de fusion de scores en utilisant deux types d'algorithmes évolutionnaires : les algorithmes génétiques et la programmation génétique. La section 4.4 présente nos travaux sur la reconnaissance d'information de type biométrie douce en DDF. Il s'agit de travaux originaux dans la littérature.

4.1. État de l'art de la multibiométrie et de la biométrie douce

4.1.1. Introduction

Même si les systèmes biométriques présentent des avantages certains par rapport aux autres systèmes de reconnaissance, ils possèdent néanmoins certaines limitations. On peut mentionner (pour plus de détails, voir Ross *et al.* [2006b]) : le bruit d'acquisition, une sensibilité plus ou moins importante vis-à-vis des variabilités intra- ou inter-classes. Ces systèmes peuvent également être sensibles à des attaques par usurpation d'identité, ou d'autres types d'attaques identifiées par Ratha *et al.*. L'universalité n'est pas garantie à cent pour cent par tous les systèmes, dans le sens où certaines personnes sont dans l'impossibilité d'utiliser des systèmes biométriques reposant sur les empreintes digitales (en cas de brûlure par exemple), ou sur l'iris, voire en cas de refus de présenter une modalité particulière mais acceptation d'en utiliser une autre.

L'utilisation d'un système multibiométrique permet en partie de pallier les limitations précédentes. En effet, la fusion de données (ou de signaux) issues de plusieurs sous-systèmes biométriques permet d'exploiter plus d'informations dans le processus de vérification d'identité : le bruit d'acquisition n'affecte pas tous les signaux issus des différents capteurs de la même façon, il suffit donc de sélectionner le capteur dont les signaux sont les moins bruités ; la variabilité diffère selon la modalité et impacte à des degrés plus ou moins importants chaque sous-système. On pourra ainsi réduire l'impact de la variabilité sur la décision définitive. En exploitant les points forts de chaque sous-système, les performances du système multibiométrique devraient être globalement

supérieures. Concernant la non-universalité de certaines modalités, la multibiométrie permet de changer de modalité si nécessaire. Le fait de devoir présenter plusieurs modalités rend également les attaques plus complexes à mettre en œuvre pour un intrus, qui aura plus de difficultés à fournir au système à la fois une empreinte digitale et un iris correspondant au même utilisateur légitime par exemple. Et en cas de défaillance d'un capteur, les autres capteurs peuvent prendre le relais.

Lors de la conception d'un système multibiométrique, il est nécessaire de trouver un compromis entre : le prix des capteurs supplémentaires éventuels et le gain de performance, sans oublier la contrainte supplémentaire pour l'utilisateur. Il faut également bien choisir les sous-systèmes en fonction de l'application souhaitée, ainsi que l'algorithme de fusion des données, fusion qui peut être effectuée à tous les niveaux du système global. Tous ces points sont abordés dans la suite de cette section, qui est organisée de la manière suivante : la section 4.1.2 présente les différentes structures de systèmes multibiométriques, ensuite les différents procédés de fusion de données biométriques sont décrits dans la section 4.1.3. La section 4.1.4, quant à elle, présente les informations sur la biométrie douce.

4.1.2. Les différents principes de multibiométrie

Différentes sources d'informations peuvent être utilisées par un système multibiométrique [Ross *et al.*, 2006b, chap. 2.4] ; voici une liste des principaux scénarios d'obtention de sources de systèmes multibiométriques (résumés en figure 4.1) : les systèmes (a) multi-algorithmes, (b) multicapteurs, (c) multi-instances, (d) multi-échantillons, (e) multi-caractères, et (f) hybrides. Les principaux critères de comparaison de ces types de systèmes sont le coût (financier ou en termes de temps de calcul) et la dégradation de l'expérience utilisateur par rapport à un sous-système considéré séparément.

4.1.2.1. Les systèmes multi-algorithmes

La même donnée biométrique est vérifiée à l'aide de plusieurs algorithmes de reconnaissance (*c.-à-d.*, reconnaissance d'empreinte digitale à partir de la comparaison de minuties et de texture). Le coût d'une telle solution est peu élevé en raison de l'utilisation d'un seul capteur. Elle est peu contraignante pour l'utilisateur en raison de l'absence d'interactions supplémentaires. L'utilisation de plusieurs algorithmes peut cependant entraîner une augmentation du coût et du temps de calcul.

4.1.2.2. Les systèmes multi-capteurs

Le même caractère biométrique est capturé à l'aide de plusieurs capteurs différents, afin d'acquérir le plus d'informations différentes possibles (*c.-à-d.*, utilisation d'un lecteur d'empreinte digitale capacitif et d'un lecteur d'empreinte digitale résistif). Le coût du système est plus élevé en raison de l'utilisation de plusieurs capteurs. L'expérience utilisateur peut être dégradée si ces capteurs doivent être utilisés séquentiellement comme le montre Allano *et al.* [2010].

4.1.2.3. Les systèmes multi-instances

Plusieurs instances du même caractère biométrique sont capturées (*c.-à-d.*, capture des iris droit et gauche, ou capture du pouce et de l'index de la main droite). Le même capteur peut être utilisé pour capturer toutes les données biométriques, dans ce cas, le coût de la solution n'est pas plus élevé, mais l'utilisateur doit effectuer la capture de toutes les instances, ce qui

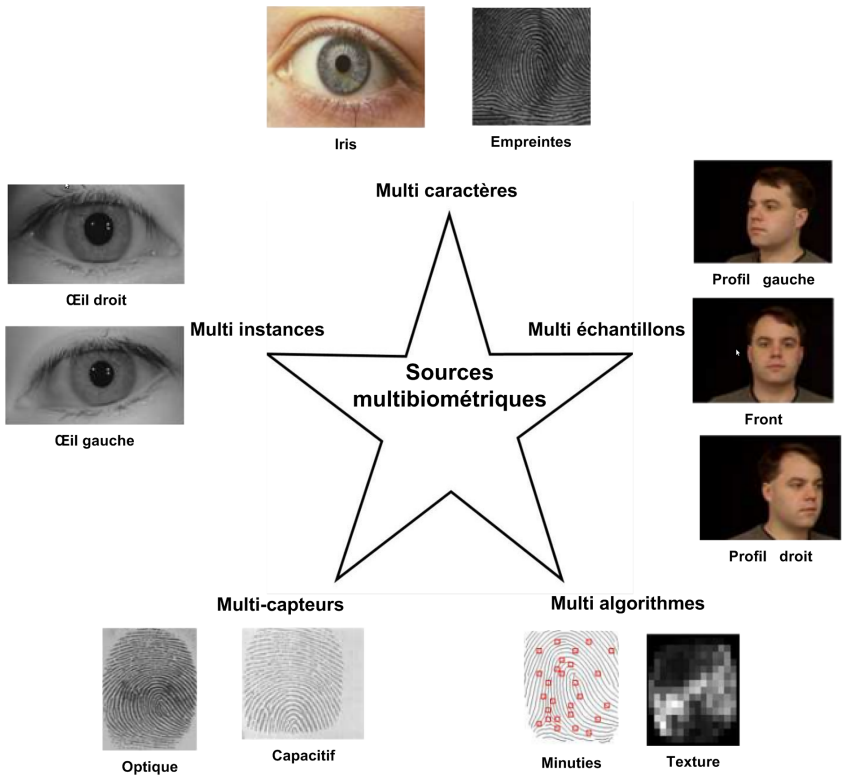


FIG. 4.1: Différentes sources multibiométriques. Image inspirée de Ross *et al.* [2006b]

peut être contraignant. Il est également possible d'utiliser un nouveau capteur capable d'acquérir toutes les données simultanément. Dans ce dernier cas, le coût du système peut être plus élevé, mais en contrepartie, l'expérience utilisateur n'est pas plus contraignante quelque soit le nombre d'instances.

4.1.2.4. Les systèmes multi-échantillons

Plusieurs acquisitions de la même donnée biométrique sont effectuées. Il s'agit donc d'une variante d'un système *multi-instances*. L'intérêt d'un tel système est d'augmenter la robustesse au bruit en augmentant le nombre de captures de la donnée. En contrepartie, l'expérience utilisateur est fortement dégradée, sauf s'il s'agit de systèmes exploitant la vidéo. Une vérification peut être effectuée sur chacune des captures, ou une super capture peut être générée à l'aide des différentes captures.

4.1.2.5. Les systèmes multi-caractères

Les systèmes multi-caractères utilisent l'information de plusieurs caractères biométriques différents pour authentifier les individus (par exemple, reconnaissance faciale associée à la reconnaissance de la parole). Le coût d'un tel système est nécessairement plus important en raison de la nécessité de disposer d'un capteur spécifique par caractère.

4.1.2.6. Les systèmes hybrides

Les systèmes hybrides concernent les autres types de systèmes : ils sont composés de plusieurs scénarios parmi ceux présentés précédemment. Les systèmes hybrides disposent donc de plus d'information que les systèmes précédents.

4.1.3. Niveaux de fusion

En fonction des sous-systèmes retenus dans le scénario, il est possible d'effectuer une fusion des données à différents niveaux de l'architecture du système de multibiométrie (figure 4.2).

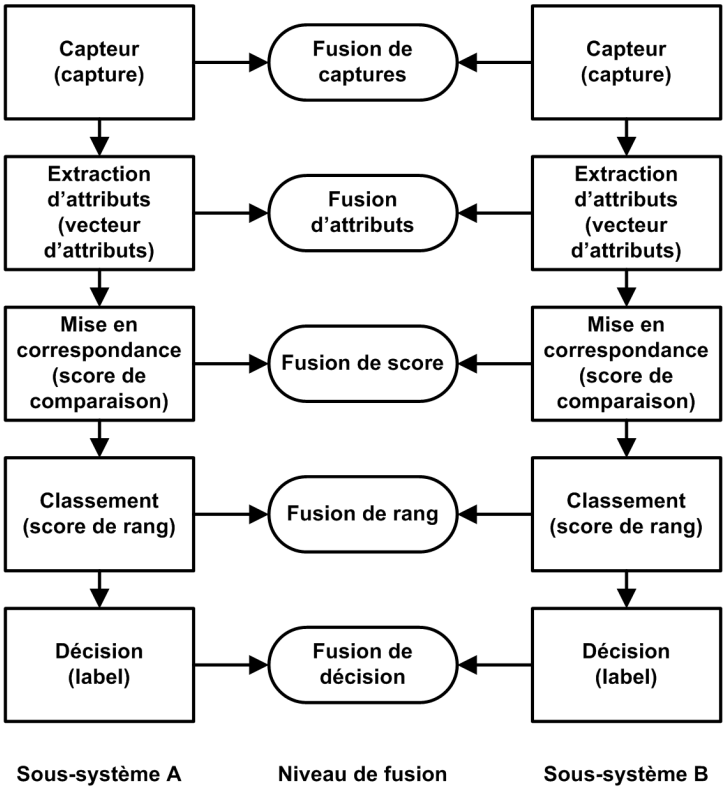


FIG. 4.2: Liste des différents niveaux de fusion

Plus la fusion est réalisée à un niveau éloigné de la capture, plus la quantité d'informations disponibles pour prendre une décision est faible ; la figure 4.3 illustre ce point par la diminution de la taille des signaux traités.

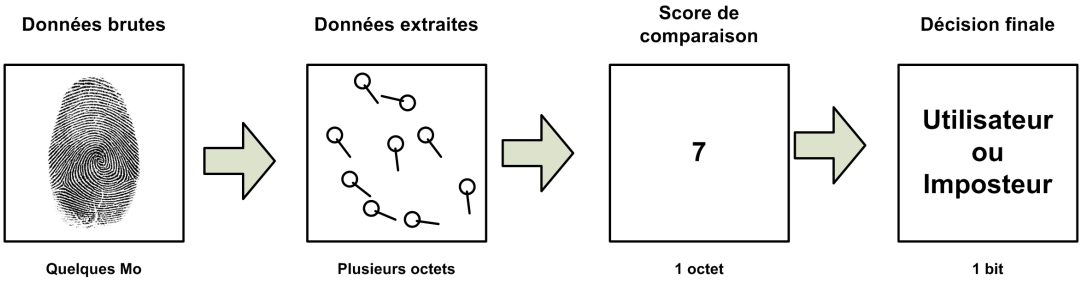


FIG. 4.3: Illustration de la perte d'information au cours de la progression dans le système de vérification

4.1.3.1. Fusion de captures

Le premier niveau de fusion est au niveau des capteurs. L'objectif est de générer une nouvelle capture, de meilleure qualité que les captures sources, à traiter avant l'extraction d'attributs. Il s'agit d'une technique, appelée *fusion d'images* ou *fusion de pixels* dans le domaine du traitement d'images. La figure 4.4 schématise le procédé de fusion de captures.

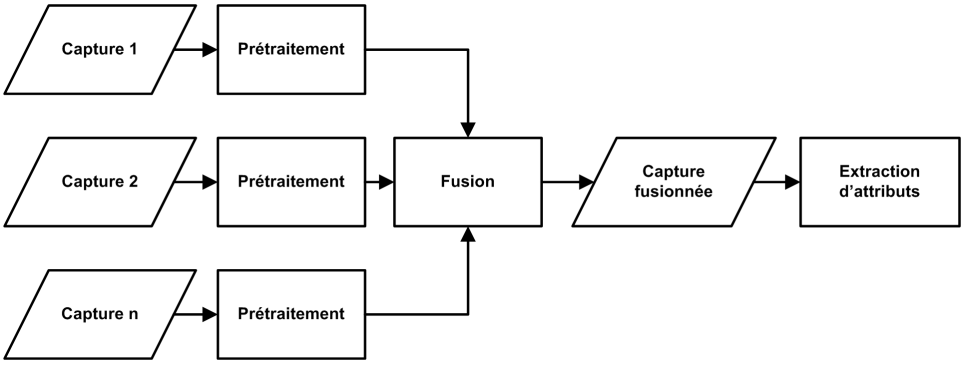
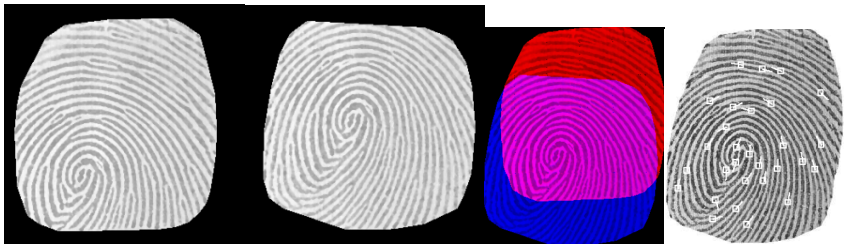


FIG. 4.4: Schéma de fusion au niveau du capteur

Les travaux présentés dans [Choi *et al.*, 2005; Jain et Ross, 2002; Ross *et al.*, 2006a] visent à créer une seule image d'empreinte digitale à partir de plusieurs images. Cette technique est utilisée dans le cadre de l'empreinte digitale car certains capteurs d'empreintes fournissent une image partielle du doigt. Ainsi, plusieurs acquisitions d'un même doigt peuvent présenter des informations avec peu de recouvrement. Pouvoir créer une image de plus grande taille, à l'enregistrement par exemple, permet par la suite d'avoir un recouvrement idéal lors de la vérification.

Le prétraitement nécessaire à la fusion d'empreintes consiste principalement à les aligner et à les déformer. En effet, le doigt étant élastique, deux captures peuvent avoir une légère déformation (rotation, mise à l'échelle,...) qu'il faut compenser avant de les fusionner. Nous voyons en figure 4.5 (a) et (b) deux captures à fusionner. On remarque qu'il y a un important décalage entre les deux captures. La figure 4.5 (c) montre les deux captures après déformation et recalage tandis que la figure 4.5 (d) montre que des minuties sont détectées aussi bien dans les zones présentes uniquement dans la capture 1 que dans la capture 2. On obtient une capture de meilleure qualité et ainsi une image contenant plus d'informations.



(a) Capture 1 (b) Capture 2 (c) Capture fusionnée (d) Minuties extraites de la capture fusionnée

FIG. 4.5: Fusion de captures d'empreintes digitales Jain et Ross [2002]

Des travaux similaires sont également présents pour d'autres modalités telles que le visage dans [Singh *et al.*, 2007; Yang *et al.*, 2006; Zhang et Gao, 2009]. L'objectif est de reconstruire

une image panoramique du visage d'un individu afin d'améliorer la robustesse aux problèmes liés à la pose lors de l'acquisition de l'image du visage. En effet, contrairement à la biométrie de l'empreinte, pour laquelle le doigt est obligatoirement posé sur un capteur, l'image du visage est prise sans contact avec le capteur. Elle peut donc être prise de face, de profil, en plongée ou contre-plongée.

La reconstruction panoramique peut se faire soit en utilisant un ensemble d'images acquises successivement avec le même capteur, soit avec un ensemble de capteurs différents. Yang *et al.* [2006] montrent un dispositif d'acquisition contenant 5 caméras qui permettent d'acquérir en même temps 5 vues différentes du même visage. Ce dispositif est présenté en figure 4.6.

(a) Dispositif d'acquisition (b) Logiciel d'acquisition

FIG. 4.6: Système d'acquisition pour de la fusion de captures du visage [Yang *et al.*, 2006]

La création de l'image panoramique du visage consiste à altérer les images afin de les aligner deux à deux, puis de sélectionner dans chaque image la partie qui sera gardée. En découpant et recollant les différentes parties des images, on arrive à obtenir l'image panoramique souhaitée. On voit à la figure 4.7 plusieurs visages panoramiques reconstruits à partir de séquences d'images de visages.

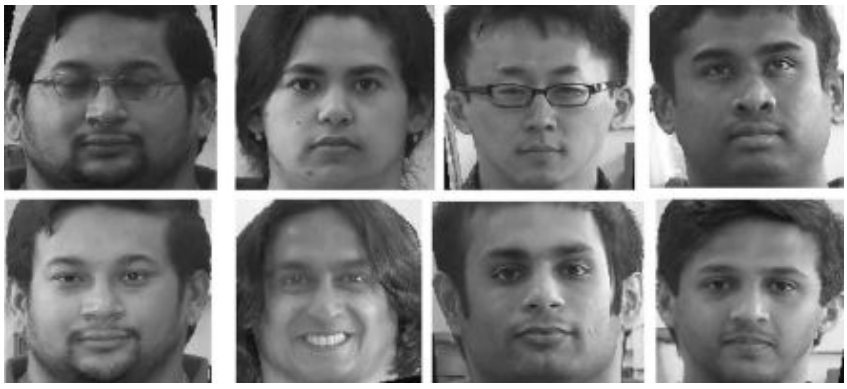


FIG. 4.7: Visages panoramiques [Singh *et al.*, 2007]

Raghavendra *et al.* [2011] effectuent quant à eux une fusion d'image visible et proche infra-rouge. Un mécanisme d'optimisation du type Particle Swarm Optimisation (PSO) est utilisé dans une première méthode pour sélectionner les poids pour une combinaison linéaire entre les deux images, et dans une deuxième méthode pour sélectionner le sous-ensemble optimal des données à conserver entre l'image proche infra-rouge et l'image visible. Les résultats obtenus sont meilleurs que ceux obtenus avec une fusion des scores (cf. section 4.1.3.3).

4.1.3.2. Fusion d'attributs

La fusion peut également avoir lieu après le traitement de la donnée provenant du capteur, au niveau des attributs extraits. On l'effectue dans des systèmes multi-capteurs ou multi-échantillons comme pour la fusion de capteurs, mais aussi multi-instances ou multi-modalités. L'objectif est d'obtenir des attributs qui sont soit plus robustes dans le cadre de données homogènes (*c.-à-d.*, provenant de la même modalité avec le même algorithme d'extraction d'attributs), soit contenant plus d'informations dans le cadre de données hétérogènes (*c.-à-d.*, provenant de modalités différentes ou de méthodes d'extraction d'attributs différentes).

La figure 4.8 présente le schéma global de la fusion d'attributs. Nous voyons que les données sont préalablement homogénéisées (cf. normalisation des données de la même façon que les scores, section 4.1.3.3.1) avant d'être fusionnées. Cette étape n'est nécessaire que dans le cas de données hétérogènes. La fusion intervient ensuite pour créer un nouveau vecteur d'attributs. Pour cela, les attributs peuvent être simplement concaténés, avec une réduction de la dimension, ou bien une sélection des attributs à concaténer peut être réalisée. Le vecteur d'attributs obtenu est utilisé pour la mise en correspondance.

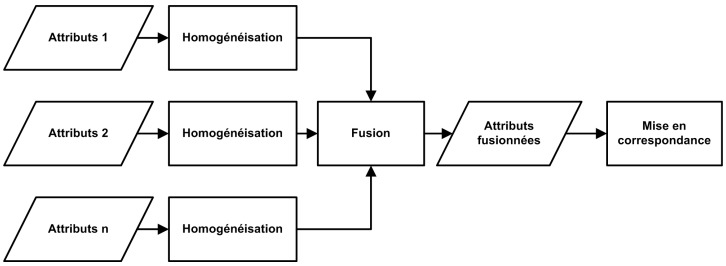


Fig. 4.8: Schéma de fusion au niveau des vecteurs d'attributs

Dans le cas de données hétérogènes, la façon la plus simple de fusionner des attributs est de les concaténer. On obtient alors un vecteur de plus grande taille, qui contient plus d'informations. La figure 4.9 illustre cette approche. Chang *et al.* [2003] présentent cette méthode de fusion appliquée sur des vecteurs d'attributs provenant d'une part d'une image de visage et d'autre part d'une image de l'oreille de l'utilisateur. Ils affirment que cette simple méthode de fusion permet d'améliorer les performances de l'une ou l'autre des modalités prises individuellement.

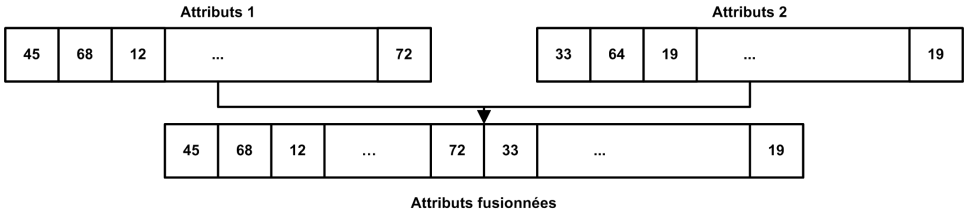


Fig. 4.9: Concaténation de vecteurs d'attributs

Pour lutter contre la « malédiction de la dimension », liée à la difficulté d'apprentissage lorsque la taille du vecteur obtenu est plus grande que le nombre d'exemplaires d'apprentissage, il est souvent intéressant de réduire la taille du vecteur d'attributs obtenu après la fusion. Plusieurs méthodes sont possibles : soit une réduction de la dimension est effectuée par une analyse statistique sur le vecteur concaténé, soit une sélection des attributs les plus pertinents est réalisée avant la concaténation. La réduction de la dimension peut se faire facilement avec une analyse en composantes principales, ce qui permet de garder une part importante de l'information pertinente contenue dans le vecteur concaténé, ou bien avec une analyse discriminante linéaire. Pour le cas de la sélection d'attributs, les travaux de [Raghavendra *et al.*, 2009] comparent l'utilisation de

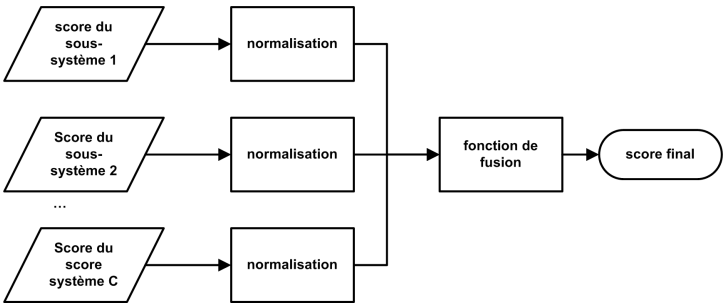


FIG. 4.10: Résumé du fonctionnement de la fusion de scores à l'aide d'une fonction de fusion

l'algorithme AdaBoost et l'optimisation par essaims particulaires. Dans les deux cas, l'objectif est de filtrer le vecteur concaténé pour en supprimer certains attributs. Ces travaux montrent que la sélection d'attributs permet de garder l'efficacité de la fusion d'attributs, tout en ayant des vecteurs d'attributs plus petits de 36% et 45% pour, respectivement, AdaBoost et l'optimisation par essaims particulaires.

Pour les données homogènes, deux ou plusieurs vecteurs d'attributs de même taille sont disponibles. Bien évidement, les méthodes de fusion pour les données hétérogènes peuvent être utilisées pour des données homogènes. Par exemple, dans le cas de minuties de l'empreinte digitale, on peut concaténer l'ensemble des minuties extraites depuis plusieurs images du même doigt, et faire une sélection pour supprimer les minuties redondantes. Cependant, il est préférable de fusionner ces vecteurs de données homogènes afin d'en obtenir un nouveau de même taille. En effet, les vecteurs d'attributs représentent la même donnée. Chacun des attributs des vecteurs de données peut être fusionné individuellement. On peut utiliser la moyenne ou une moyenne pondérée sur chaque attribut des vecteurs pour obtenir le vecteur fusionné.

4.1.3.3. Fusion des scores

Les différents sous-systèmes du système multibiométrique produisent des scores après l'étape de comparaison. Le mécanisme de fusion des scores permet de générer un nouveau score ou une classe finale à partir de ces scores.

4.1.3.3.1. Classification à base de transformation de scores Deux étapes sont nécessaires pour effectuer une fusion des scores (figure 4.10) : (i) la normalisation des scores, et (ii) la combinaison des scores.

Fonctions de normalisation des scores Comme la distribution des scores issus des différents sous-systèmes est rarement compatible (*c.-à-d.*, il est inutile de faire la somme des scores du système A avec ceux du système B si la distribution des scores du système A est [0,1] et celle du système B est [1000,10000]), la normalisation est une étape nécessaire [Jain *et al.*, 2005].

Une des méthodes de normalisation les plus simples est la normalisation *min-max*. Elle est utilisée lorsque les bornes de la distribution des scores sont connues (on peut en général les trouver sur une base de dévelopement). En utilisant cette technique, les scores sont normalisés entre 0 et 1. À partir d'un ensemble de scores $\{s_k\}$,$k = 1,2,..,n$, les scores normalisés sont obtenus de la façon suivante :

$$s'_k = \frac{s_k - min}{max - min} \tag{4.1}$$

avec min et max respectivement les scores minimum et maximum.

Une normalisation communément utilisée est la normalisation par *z-score* en utilisant la moyenne arithmétique μ et l'écart-type σ des données. Il est donc nécessaire de connaître ou de disposer de données pour estimer ces valeurs. Les scores normalisés sont obtenus de la façon suivante :

$$s'_k = \frac{s_k - \mu}{\sigma} \tag{4.2}$$

Les deux méthodes précédentes sont sensibles au bruit en raison de l'utilisation de min, max, μ et σ. En revanche, la *mediane* et l'écart absolu à la médiane MAD ne sont pas sensibles au bruit. Une normalisation utilisant ces données est obtenue de la façon suivante :

$$s'_k = \frac{s_k - mediane}{MAD} \tag{4.3}$$

avec $MAD = \text{mediane}(|s_k - mediane|)$

Une double sigmoïde peut également être utilisée :

$$s'_k = \begin{cases} \dfrac{1}{1 + exp(-2((s_k - t)/r_1)} & \text{si } s_k < t, \\ \dfrac{1}{1 + exp(-2((s_k - t)/r_2)} & \text{autrement} \end{cases} \tag{4.4}$$

avec t un point de référence et r_1 et r_2 les régions gauche et droite de t sur lesquelles la sigmoïde est linéaire. Il est donc nécessaire de spécifier t, r_1 et r_2.

Une normalisation robuste et efficace est la normalisation *tanh* basée sur les estimateurs de Hampel. La normalisation est obtenue de la façon suivante :

$$s'_k = \frac{1}{2}\left\{tanh\left(0.001\left(\frac{s_k - \mu_{client}}{\sigma_{client}}\right)\right) + 1\right\} \tag{4.5}$$

avec μ_{client} et σ_{client} la moyenne et l'écart-type des scores authentiques donnés par les estimateurs de Hampel. Les estimateurs de Hampel sont basés sur la fonction d'influence ψ :

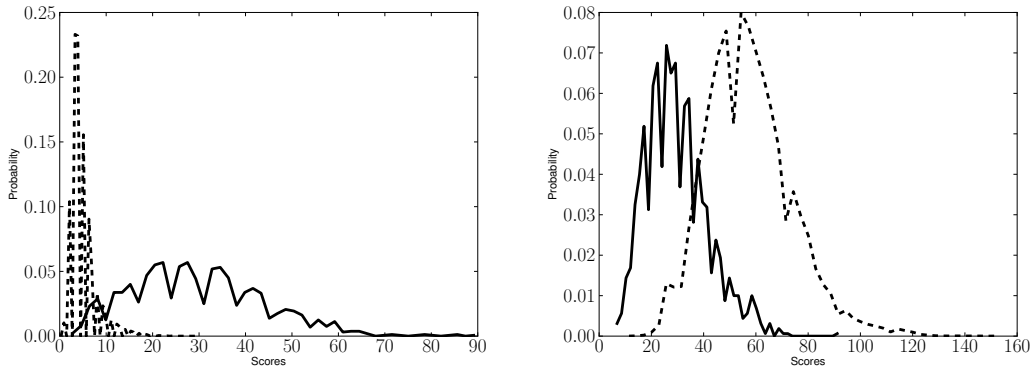
$$\psi(u) = \begin{cases} u & 0 \leq |u| < a, \\ asign(u) & a \leq |u| < b, \\ asign(u)\left(\frac{c-|u|}{c-b}\right) & b \leq |u| < c, \\ 0 & |u| \geq c. \end{cases} \tag{4.6}$$

La fonction d'influence réduit l'influence des points de la queue de la distribution (identifiés par a, b, et c). Ainsi, la méthode n'est pas sensible au bruit, mais les valeurs de a, b, c doivent être choisies avec précaution en fonction de l'estimation du bruit dans les données. Il semble que cette normalisation ait majoritairement été appliquée sans effectuer le filtrage des scores par ψ. La figure 4.11 illustre certaines de ces méthodes.

Fonctions de combinaison des scores Un ensemble de règles de combinaison de scores est présenté dans [Kittler *et al.*, 1998]. Ces travaux ne sont pas spécifiques à la multi-modalité biométrique, mais ont été largement utilisés dans ce contexte. Le but est d'obtenir un score de classification c à l'aide de plusieurs scores s_i obtenus à l'aide des C différents sous-systèmes.

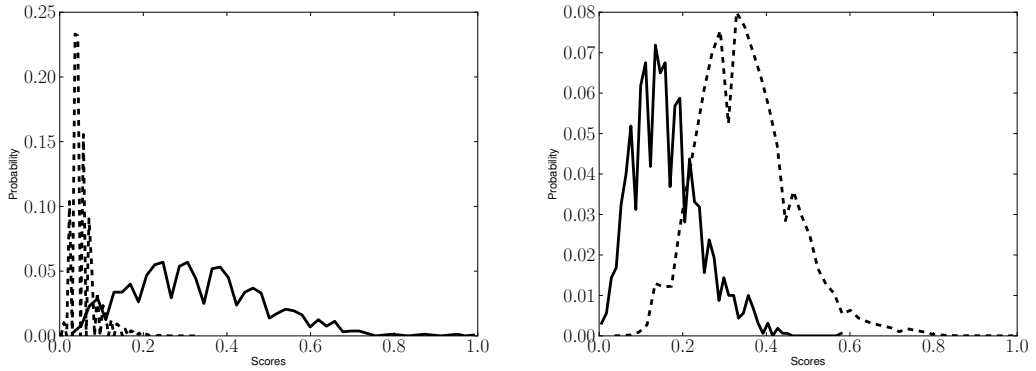
La règle *produit* effectue le produit des scores des différents sous-systèmes afin de générer le score final :

$$c = \prod_{i=1}^{C} s_i \tag{4.7}$$



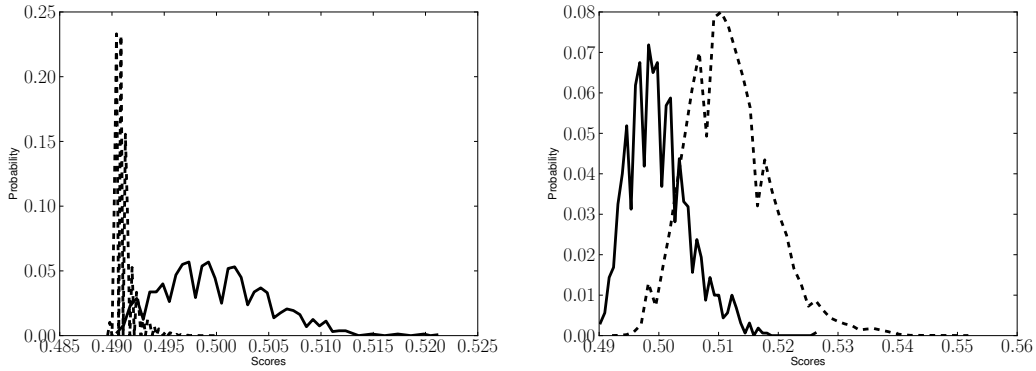
(a) visage/scores bruts

(b) dynamique de frappe/scores bruts



(c) visage/minmax

(d) dynamique de frappe/minmax



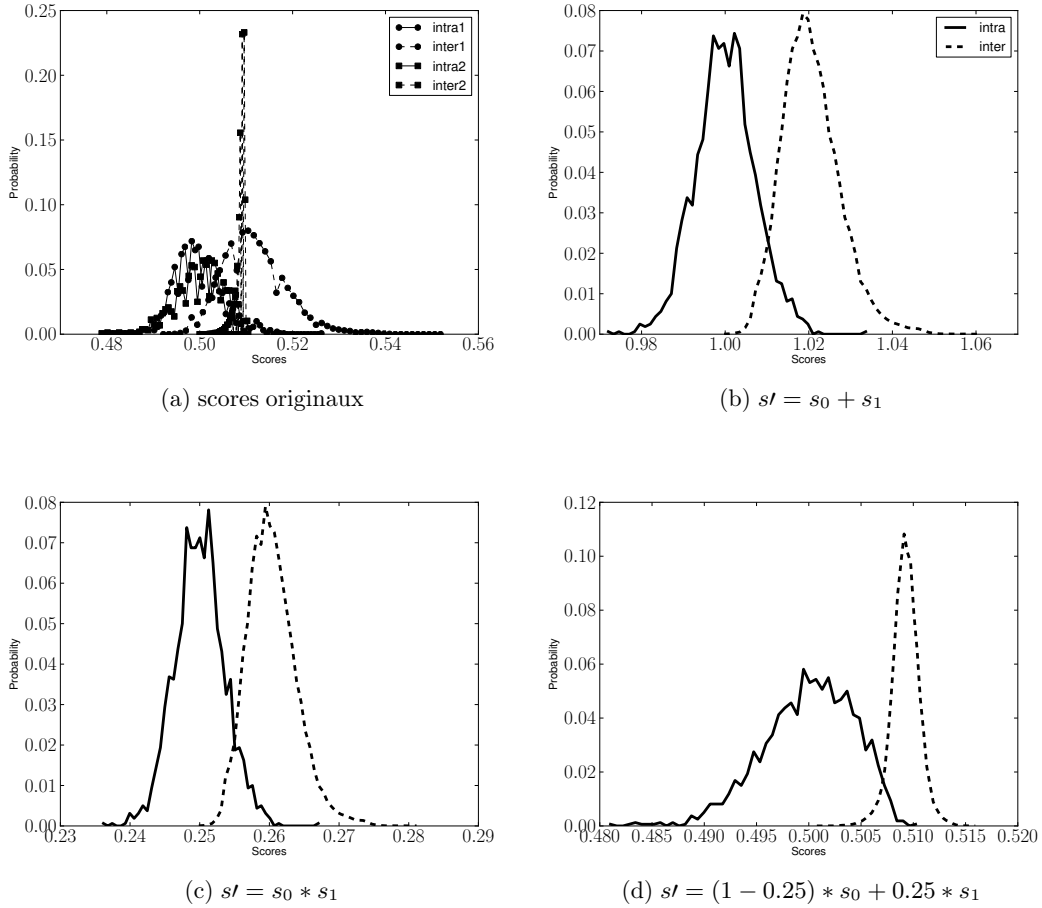
(e) visage/tanh

(f) dynamique de frappe/tanh

FIG. 4.11: Illustration de quelques méthodes de normalisation. Traits pleins : score authentiques. Traits pointillés : scores d'imposture. À gauche (resp. droite) : distribution des scores d'un système de reconnaissance faciale (resp. dynamique de frappe au clavier). a-b : scores réels. c-d : scores normalisés avec minmax. e-f : scores normalisés avec tanh

La règle *somme* effectue la somme des scores des différents sous-systèmes afin de générer le score final :

$$c = \sum_{i=1}^{C} s_i \tag{4.8}$$



(a) scores originaux

(b) $s\prime = s_0 + s_1$

(c) $s\prime = s_0 * s_1$

(d) $s\prime = (1 - 0.25) * s_0 + 0.25 * s_1$

FIG. 4.12: Illustration de quelques méthodes de combinaison entre les deux systèmes normalisés précédemment avec la technique *tanh* (un des systèmes étant basé sur un score, nous avons soustrait ce score normalisé à 1 pour obtenir une distance)

Cette fusion relativement simple donne souvent les meilleurs résultats. L'avantage par rapport au produit, est que si un score vaut zéro, la combinaison des scores ne vaudra pas zéro.

La règle *maximum* retourne comme score final le score le plus élevé de l'ensemble des sous-systèmes :

$$c = \max_i s_i \tag{4.9}$$

La règle *minimum* retourne comme score final le score le plus faible de l'ensemble des sous-systèmes :

$$c = \min_i s_i \tag{4.10}$$

Une *somme pondérée* peut également être utilisée afin de donner plus de poids à certains systèmes, soit de façon empirique en raison de leur performance, soit de façon automatique en utilisant des mécanismes d'optimisation. Il est donc nécessaire d'utiliser des exemples de validation pour configurer les poids :

$$c = \sum_{i=1}^{C} \alpha_i * s_i \tag{4.11}$$

La figure 4.12 illustre la distribution des scores après combinaison de deux sous-systèmes.

4.1.3.3.2. Fusion utilisant un classifieur La figure 4.13 résume le fonctionnement de la fusion utilisant un classifieur. Il est également possible d'utiliser des méthodes de classification permettant d'indiquer si l'ensemble des scores correspond à un intrus ou un utilisateur authentique. Cette fois-ci, nous utilisons un n-uplets de scores $\mathbf{s} = [s_1,...,s_C]$ pour représenter l'ensemble des scores d'une authentification, chaque n-uplet pouvant appartenir à la classe « Client » ou la classe « Imposteur ».

Le principe de tous les classifieurs utilisés est le même, à savoir trouver une fonction de séparation dans l'espace des scores permettant de séparer les ensembles scores clients et scores imposteurs. Des mécanismes à base de Séparateur à Vaste Marge (SVM) (annexe E) peuvent être utilisés pour classer les n-uplets selon deux classes : -1 pour les n-uplets « Imposteurs » et $+1$ pour les n-uplets « Clients ». L'apprentissage consiste à trouver la fonction $y : \mathcal{S} \to \mathbb{R}$ à l'aide de l'ensemble d'apprentissage. y est la fonction de décision qui permet d'effectuer une séparation non linéaire des deux ensembles en utilisant un système de maximisation de la marge. Cependant, il a été montré que bien que cette méthode soit performante, une moyenne arithmétique peut être suffisante [Garcia-Salicetti *et al.*, 2005]. D'autres types de classifieurs peuvent être utilisés, comme les réseaux de neurones, etc...

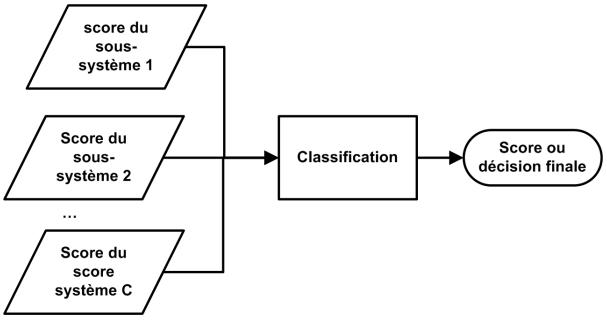


FIG. 4.13: Résumé du fonctionnement de la fusion de scores à l'aide d'un classifieur

4.1.3.3.3. Fusion utilisant une modélisation de densités Une autre méthode de fusion consiste à modéliser la distribution (multidimensionnelle) des scores « Clients » et « Imposteur », afin de classer les utilisateurs en fonction de leur probabilité d'appartenance. Nandakumar *et al.* [2008] modélisent les distributions clientes $\hat{f}_{client}(\cdot)$ et imposteurs $\hat{f}_{imposteur}(\cdot)$ en utilisant des mélanges de gaussiennes. Un test de vraisemblance (likelihood ratio) est calculé $LR(\mathbf{s}) = \frac{\hat{f}_{client}(\mathbf{s})}{\hat{f}_{imposteur}(\mathbf{s})}$. Il est comparé à un seuil η (déterminé sur le taux de fausses correspondances (*false match rate*) (FMR) spécifié) pour vérifier si $\mathbf{s}$ est un ensemble de score client : $LR(\mathbf{s}) \geq \eta$.

4.1.3.3.4. Améliorations Une amélioration possible est la normalisation des scores de façon individuelle pour chaque utilisateur [Poh *et al.*, 2010c] (cependant, nous disposons de peu de données) ou en fonction de catégories [Poh *et al.*, 2010b] (où nous disposons de plus de données). Il est également possible d'utiliser des méthodes de normalisation adaptatives [Indovina *et al.*, 2003].

4.1.3.4. Fusion de décision et rang

Les fusions de décision et de rang interviennent au niveau le plus abstrait, après la comparaison, et utilisent uniquement les résultats donnés par les différents sous-systèmes (*accepté*, *rejeté* ou un identifiant avec un indice de confiance). Nous avons vu que la quantité d'information disponible pour le processus de fusion diminue lorsqu'on s'éloigne du capteur. Il semble donc plus efficace

d'effectuer la fusion justement avant la comparaison. Cependant, cela n'est pas toujours possible, dans la plupart des systèmes courants, la fusion ne peut être faite qu'au niveau des scores, des rangs ou de la décision. Par rapport aux processus de fusion présentés précédemment, la fusion de décision ou de rang utilise beaucoup moins d'informations : les seules données considérées sont les sorties des différents sous-systèmes. Cette perte d'information, notamment par rapport à la fusion de scores, peut expliquer que la fusion de décision et la fusion de rang soient moins étudiées. Néanmoins, les méthodes mises en œuvre sont généralement plus simples et évitent le problème de normalisation des données.

4.1.3.4.1. Fusion de rang La fusion de rang (figure 4.14) concerne l'identification d'un individu parmi tous les individus autorisés. Le processus d'identification de chaque sous-système renvoie une liste comportant plusieurs identifiants classés par ordre (décroissant) de confiance. Ces classements peuvent être comparés directement, même s'ils proviennent de modalités différentes, dans le sens où aucune normalisation préalable des données n'est requise. Par conséquent, les schémas de fusion de rang sont plutôt simples à implémenter. Plusieurs méthodes peuvent être utilisées pour combiner les différents rangs comme le montrent Ho *et al.* [1994].

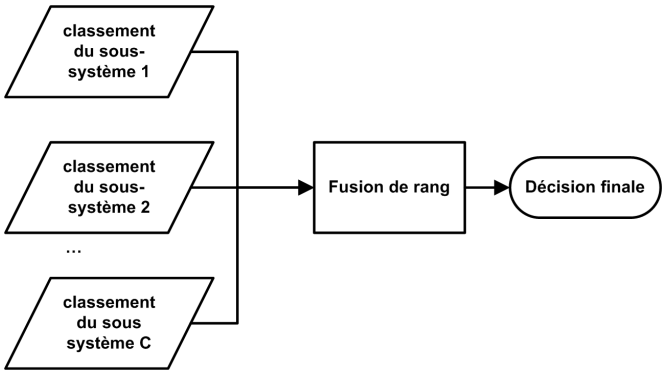


Fig. 4.14: Résumé du fonctionnement de la fusion de rang

La *méthode de plus haut rang* sélectionne le meilleur rang pour chaque sous-système ; elle peut être considérée comme un vote majoritaire. Cette méthode est intéressante lorsque le nombre d'utilisateurs répertoriés dans la base de données est très supérieur au nombre de sous systèmes: c'est souvent le cas des systèmes d'identification. En pratique, même si seulement un sous-système classe correctement le client à identifier, avec un bon indice de confiance, le système global devrait donner à ce client un rang assez haut.

La *méthode* « *Borda count* » repose sur un processus de vote pondéré à partir de la somme des rangs donnés par chaque sous-système. Une hypothèse d'indépendance statistique entre les différents modules est nécessaire pour appliquer cette méthode.

La *méthode de régression logistique* utilise des connaissances statistiques sur les performances des sous-systèmes et généralise la méthode « Borda count ». Elle affecte un poids différent à chaque sous-système, en fonction de son efficacité. Ces poids sont calculés pendant une phase d'apprentissage, par une régression logistique. Ils dépendent donc des données traitées. Cette méthode présente donc l'avantage de tenir compte de différences au niveau de l'efficacité de chaque sous-système. D'autres méthodes moins utilisées sont détaillées par Saranli et Demirekler qui présentent une autre méthode de fusion de rang statistique englobant les trois méthodes précédentes, et Nandakumar *et al.* [2009] qui proposent une approche bayésienne de fusion de rang.

4.1.3.4.2. Fusion de décision La fusion de décision (figure 4.15) intervient dans une problématique d'authentification ou de vérification d'identité. Elle utilise moins d'information que la fusion de rang, dans le sens où seules les décisions définitives (*accepté* ou *rejeté*) de chaque sous-système sont considérées : un seul identifiant est conservé. Il s'agit donc du niveau le plus abstrait de décision dans un système de multibiométrie.

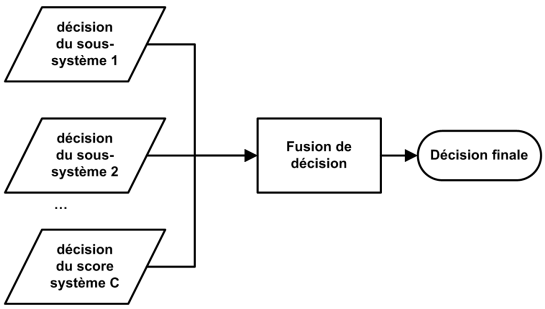


FIG. 4.15: Résumé du fonctionnement de la fusion de décision.

Un utilisateur se présente au système de multibiométrie, il donne son identifiant ; le système effectue ensuite toutes les captures nécessaires à la vérification d'identité ; chaque sous-système produit donc une décision, soit accepté, soit rejeté. Il existe plusieurs méthodes pour fusionner ces décisions en une décision définitive ; seules les méthodes les plus utilisées sont présentées ici.

Daugman [2000] introduit une méthode utilisant l'opérateur logique ET qui consiste à accepter le client si et seulement si tous les sous-systèmes ont reconnu l'utilisateur. Ce schéma de fusion de décision induit un taux de fausses acceptations (*false acceptance rate*) (FAR) global plus bas que celui de chaque sous-système considéré séparément. Au contraire, le taux de faux rejets (*false rejection rate*) (FRR) global est plus haut que celui de chaque sous-système considéré séparément. Daugman [2000] introduit une autre méthode utilisant l'opérateur logique OU qui consiste à accepter le client si au moins un sous-système a reconnu l'utilisateur. Ce schéma de fusion de décision induit un FAR global plus haut que celui de chaque sous-système considéré séparément. Au contraire, le FRR global est plus bas que celui de chaque sous-système considéré séparément. Les deux méthodes précédentes sont très simples, mais rarement utilisées. En effet, elles dégradent les performances du système multibiométrique en terme d'EER dès lors qu'un des sous-systèmes possède un EER beaucoup plus élevé que les autres.

Lam et Suen [1997] proposent une méthode utilisant le *vote majoritaire*. L'utilisateur est authentifié par le système global si la moitié au moins des sous-systèmes l'authentifient. Différentes configurations ont été testées dans l'article [Kuncheva *et al.*, 2003]. Cette méthode ne nécessite aucune phase d'apprentissage, ni aucune connaissance a priori sur les différents sous-systèmes. Elle s'applique cependant dans le cas où tous les sous-systèmes ont des performances comparables. Dans le cas contraire, s'il existe un déséquilibre entre les différentes performances, une solution possible consiste à pondérer la décision de chaque sous-système, en donnant un poids plus important aux sous-systèmes les plus fiables.

D'autres méthodes plus complexes, largement détaillées dans la thèse de Verlinde, s'apparentent aux classifieurs paramétriques développés dans le domaine de la reconnaissance des formes. Elles utilisent des informations *a priori* sur les performances des différents sous-systèmes biométriques et comportent une phase d'apprentissage. On peut citer les méthodes utilisant la théorie de Bayes [Domingos et Pazzani, 1997; Xu *et al.*, 1992], la théorie de l'évidence de Dempster-Shafer [Xu *et al.*, 1992], la méthode BKS (Behavior Knowledge Space) [Huang et Suen, 1995; Lam et Suen, 1995], ou encore IBC (Iterative Boolean Combination) [Granger *et al.*, 2012].

4.1.4. Biométrie douce

Bien qu'il ne s'agisse pas de biométrie à proprement parler, la biométrie douce (*soft biometrics* en anglais) peut être utilisée comme une information utile à prendre en compte en complément d'une biométrie classique ; c'est pour cette raison qu'elle est présentée ici. Elle est qualifiée par [Jain *et al.*, 2004a] comme : *un caractère fournissant de l'information sur l'individu, mais manquant d'unicité et de permanence pour différencier suffisamment deux individus.* Dantcheva *et al.* proposent quant à eux une autre définition : *les biométries douces sont des caractéristiques physiques, comportementales ou propres à l'homme qui peuvent être classées dans différentes catégories humaines. Ces catégories sont, à l'inverse de la biométrie classique, établies et éprouvées par les hommes afin de différencier les individus. En d'autres termes, les instances de biométrie douce sont créées de façon naturelle et utilisées par les humains pour distinguer leurs pairs.*

Dantcheva *et al.* [2011] listent les biométries douces suivantes : la couleur de la peau, la couleur des cheveux, la couleur des yeux, la présence d'une barbe, la présence d'une moustache, les mesures du visage, la forme du visage, des informations extraites depuis le visage, le maquillage, l'ethnie, des cicatrices, le genre, l'âge, la taille, le poids, la démarche, différentes mesures du corps, la forme du corps, la couleur des vêtements ou la présence de lunettes.

Les informations peuvent être mesurées grâce à un capteur spécifique, déterminées de façon automatique à l'aide d'un système de reconnaissance de formes ou définies par un opérateur. Nous présentons ces trois catégories.

4.1.4.1. Utilisation de capteurs spécifiques

Jain *et al.* présentent un cadre générique permettant d'utiliser la biométrie douce dans un système de reconnaissance biométrique à l'aide de probabilités bayésiennes [Jain *et al.*, 2004a]. Il s'agit de travaux précurseurs dans le domaine. Ils obtiennent une amélioration de 5% pour les systèmes de reconnaissance d'empreinte digitale en utilisant des informations sur l'ethnie (indien ou pas), l'âge et la taille (artificiellement générés à l'aide d'une distribution gaussienne). Le problème principal de cette approche est la nécessité de disposer d'un matériel spécifique pour mesurer la taille de l'individu lors du processus d'authentification[1].

Ailisto *et al.* utilisent le poids et le pourcentage de graisse dans le corps comme information supplémentaire d'un système de reconnaissance par empreinte digitale [Ailisto *et al.*, 2006]. Ils réduisent l'erreur totale de 3,9% à 1,5%. Le score lié aux poids est calculé en effectuant la différence absolue entre le poids de référence et celui mesuré (une procédure identique est employée pour la graisse). Leur expérience montre que le poids peut être une mesure discriminante et utilisé comme une biométrie douce (utilisé seul, l'erreur est de 11%), mais ça n'est pas le cas de la mesure de graisse. Une fois de plus, le choix de la modalité sélectionnée peut être problématique en raison de la nécessité d'utiliser une balance spécifique qui n'est pas ergonomique.

4.1.4.2. Détermination sans capteur spécifique

Dans les travaux suivants, les informations sont prédites par des systèmes d'apprentissage au lieu d'être mesurées par des capteurs spécifiques. Jain et Park utilisent des marques sur la peau du visage (tâche de rousseur, grain de beauté, cicatrice) [Jain et Park, 2009]. L'intérêt de ces informations est multiple : (i) ajouter des données supplémentaires au système de reconnaissance faciale ; (ii) permettre une recherche plus rapide dans les grandes bases de visage ; (iii) faciliter la reconnaissance depuis des images partielles avec des marques. La performance d'identification au rang 1 a été améliorée sur deux jeux de données significatifs.

1. Problème contourné lors des expériences en générant les informations de façon automatique

Les biométries douces peuvent également être utilisées pour localiser des individus dans des bases de vidéo-surveillance [Vaquero *et al.*, 2009]. Dans ce contexte, Vaquero *et al.* ne sont pas intéressés par le système de reconnaissance faciale, mais ils veulent pouvoir trouver des individus en fonction d'une requête spécifique (par exemple: trouver les individus chauves, étant entrés dans le bâtiment samedi dernier, et portant un pull rouge et des lunettes de soleil). Les informations utilisées sont des attributs visuels tels que le type de chevelure, la présence ou non de lunettes (et leur type), le type de barbe, la couleur des habits. Certains de ces attributs peuvent être classés en tant que biométrie douce.

Le genre d'un individu peut être prédit depuis des images de visage [Alexandre, 2010]. Alexandre présente un système fusionnant différents classifieurs utilisant des descripteurs de forme, texture et couleur d'images faciales à différentes échelles. Le système a été validé sur deux jeux de données significatifs et le taux de reconnaissance avoisine les 90%.

Hwang *et al.* proposent un système de reconnaissance faciale utilisant des informations de genre afin d'améliorer les performances de reconnaissance [Hwang *et al.*, 2009]. Dans leur travail, ils partent du principe que le système de reconnaissance de genre fonctionne sans erreur. Ils utilisent trois différents modèles de visages (en utilisant Linear Discriminent Algorithm (LDA) [Belhumeur *et al.*, 1997]) : un pour les hommes, un pour les femmes, un pour les deux genres. Le système de reconnaissance utilise le modèle spécifique au genre et le modèle général. Ces travaux indiquent que la reconnaissance des femmes est plus difficile que la reconnaissance des hommes et qu'utiliser l'information du genre permet d'augmenter les performances. Les travaux sont validés sur la base FRGC.

Plutôt que d'utiliser une approche ensembliste pour prédire le genre d'un individu à partir de son visage, Li *et al.* utilisent une approche basée sur des informations locales et globales en utilisant un modèle de mélange de gaussiennes spatiales (Spatial Gaussian Mixture Models) [Li *et al.*, 2009]. Cette approche permet d'obtenir de meilleures performances (amélioration entre 40% et 50%) qu'en utilisant des GMM et SVM sur la base de visage YGA.

Dantcheva *et al.* [2011] proposent un état de l'art complet des différents types de biométrie douce. Ils proposent également un système nommé « sac de biométrie douce » qui permet d'identifier un individu avec un ensemble de biométrie douce. Le système part du principe que même si une information de type biométrie douce ne permet pas d'identifier un individu, un ensemble d'informations peut le faire. Les auteurs ont analysé la probabilité d'interférence (deux personnes décrites par le même ensemble de biométrie douce) sur une base chimérique.

Les travaux précédents ont montré qu'utiliser des informations de type biométrie douce permet d'augmenter le taux de reconnaissance des systèmes biométriques. Cette nouvelle information peut être mesurée à l'aide d'un capteur ou extraite depuis la donnée biométrique. La plupart des articles de l'état de l'art ont travaillé sur la reconnaissance du genre sur des visages 2D. Une telle solution n'est pas utilisable dans tous les contextes (ordinateur sans webcam, environnement inapproprié, ...). Il peut donc être intéressant de voir si de telles informations sont également utiles dans le cas de la DDF.

4.1.5. Discussion

Cette section a permis d'aborder la multibiométrie et la biométrie douce. Bien que cet état de l'art ne soit pas complet, nous avons vu qu'il existe différents types d'architectures, ainsi que différents niveaux pour effectuer la fusion d'informations. Celle-ci peut être effectuée depuis la capture de la donnée biométrique jusqu'à la décision d'accepter ou de rejeter un individu, en fonction des informations disponibles.

La multibiométrie permet d'améliorer les performances de reconnaissance de plusieurs systèmes biométriques en effectuant la fusion de différentes informations. Cependant, contrairement

à un système biométrique reposant sur une seule capture, il est nécessaire d'acquérir des informations supplémentaires, lors des phases d'enregistrement ou de vérification. Sauf dans le cas des systèmes multi-algorithmes, cela signifie que l'utilisateur doit présenter une ou plusieurs informations biométriques supplémentaires au système. Cette acquisition de données biométriques supplémentaires peut être contraignante pour l'utilisateur, à la fois en terme d'usage et de temps de réponse du système.

L'état de l'art de la recherche en multibiométrie montre que de tels systèmes sont efficaces. Les nouvelles perspectives de recherche devront donc se porter à la fois sur la poursuite de l'amélioration des performances de tels systèmes, et sur l'amélioration de l'expérience utilisateur. Une possibilité est l'utilisation de stratégies séquentielles de fusion de scores [Allano *et al.*, 2010]. Généralement, les systèmes multibiométriques développent une stratégie dite *en série*, dans le sens où toutes les données nécessaires au système de vérification sont capturées et fusionnées. En revanche, avec une stratégie séquentielle, on fait appel aux différents sous-systèmes les uns après les autres, et on s'arrête dès que l'un d'entre eux donne un résultat satisfaisant. Cette stratégie peut se révéler moins contraignante que la stratégie en série classique, dans le cas de capteurs séparés.

Dans le chapitre précédent, nous avons vu que la DDF est une modalité très intéressante, mais ayant des performances plutôt moyennes comparées aux modalités morphologiques. La multimodalité peut être une alternative avantageuse permettant d'améliorer ces performances. Hocquet *et al.* ont montré que fusionner plusieurs méthodes de DDF à partir des mêmes données biométriques (fusion d'algorithmes) permet d'améliorer de façon significative les performances de reconnaissance [Hocquet *et al.*, 2007]. Cependant, à notre connaissance, deux points n'ont pas été traités dans la littérature :

- La fusion de la DDF et de la RF (qui peuvent être vues comme deux modalités à bas coût sur les ordinateurs modernes : tous les ordinateurs possèdent un clavier, et, presque tous les ordinateurs possèdent une webcam).
- L'utilisation de techniques de biométrie douce à l'aide de la DDF, soit pour acquérir des informations particulières sur la personne tapant sur le clavier, soit pour augmenter les performances de reconnaissance.

La suite de ce chapitre présente nos contributions sur ces deux points, ainsi que sur une méthode d'approximation rapide d'EER qui permet d'accélérer la configuration de tels systèmes.

4.2. Approximation rapide de l'EER

4.2.1. Motivation

Différentes méthodes d'optimisation des paramètres d'un système biométrique ou d'un système multi-biométrique sont basées sur la minimisation de l'EER. Une telle fonction d'évaluation peut facilement être très coûteuse en temps de calcul lorsque le nombre de scores en jeu est relativement important. C'est pourquoi, nous avons préféré utiliser une méthode d'approximation de l'EER, plutôt qu'une méthode de calcul exact de l'EER. Curieusement, bien que cette approximation soit une méthode simple, nous n'avons jamais rencontré dans la littérature de travaux équivalents.

4.2.2. Petits rappels sur l'EER

L'International Organization for Standardization (ISO/IEC 19795-1) ISO [2006] propose différentes métriques pour caractériser les systèmes biométriques. Nous nous intéressons à certaines d'entre elles permettant de noter les performances de reconnaissance des systèmes : le FMR, le

taux de fausses non-correspondances (*false non-match rate*) (FNMR) et l'EER. Le calcul du FMR et du FNMR est basé sur la comparaison des scores par rapport à un seuil — le sens de la comparaison est à inverser si les scores sont des similarités au lieu d'être des distances. Le FMR et le FNMR sont respectivement calculés — dans le cas de distances — suivant les équations (4.12) et (4.13), avec $\boldsymbol{intra}$ (respectivement $\boldsymbol{inter}$) les scores de comparaison intraclasse (respectivement interclasse) et $Card(\boldsymbol{ensemble})$ le nombre de scores dans $\boldsymbol{ensemble}$. thr est le seuil de décision et $\mathbb{1}$ est la fonction indicatrice.

$$FMR = \frac{\sum_{score \in \boldsymbol{inter}} \mathbb{1}\{score \leq thr\}}{Card(\boldsymbol{intra})} \tag{4.12}$$

$$FNMR = \frac{\sum_{score \in \boldsymbol{intra}} \mathbb{1}\{score > thr\}}{Card(\boldsymbol{intra})} \tag{4.13}$$

La courbe caractéristique de performance (*Receiver Operating Characteristic*) (ROC) est obtenue en calculant le couple (FMR_τ,$FNMR_\tau$) pour chaque seuil de décision τ testé. Elle affiche le FNMR en fonction du FNMR (ou 1-FNMR en fonction du FMR). L'intérêt de cette courbe est de présenter le compromis entre FMR et FNMR et d'avoir une vision rapide des performances globales du système en fonction de sa configuration. L'EER est la valeur lorsque le FMR et le FNMR sont tous les deux égaux. Il constitue l'indicateur le plus couramment utilisé pour évaluer et comparer les systèmes biométriques. Plus l'EER est faible, meilleures sont les performances du système pour ce point de fonctionnement. À l'aide de la courbe ROC, on peut calculer l'EER en selectionnant le couple ($FMR(\tau)$,$FNMR(\tau)$) ayant la plus faible différence en valeur absolue pour le seuil τ :

$$\tau = \underset{\tau}{\operatorname{argmin}} abs(FMR(\tau) - FNMR(\tau)) \tag{4.14}$$

puis en retournant leur moyenne (4.15) :

$$EER = HTER(\tau) = \frac{FMR(\tau) + FNMR(\tau)}{2} \tag{4.15}$$

τ est couramment choisit parmi un ensemble de valeur linéairement répartie entre les valeurs maximum et minimum obtenues. De cette façon, nous obtenons la meilleure valeur approchée de l'EER avec l'erreur de précision la plus faible. Il s'agit de la méthode classique de calcul d'EER qui est présentée en figure 4.16[2]. La figure 4.16 montre que la complexité est en $O((Card(\boldsymbol{intra}) + Card(\boldsymbol{inter})) * Card(\boldsymbol{\Omega})) = O(n * m)$ avec n le nombre de seuils de décision utilisés et m le nombre de scores à comparer aux seuils. Étant donné qu'il est impossible de réduire m, et que celui-ci est généralement très grand, nous devons trouver un moyen de réduire n afin d'obtenir le résultat plus rapidement.

4.2.3. Méthode développée

Le temps de calcul de l'EER peut être relativement important, et lorsqu'il est nécessaire de le calculer beaucoup de fois (notamment pour optimiser des paramètres d'une fusion par somme pondérée), il est nécessaire de trouver une méthode de calcul plus rapide, quitte à perdre de la précision.

2. une autre façon, plus lente, mais plus précise, de calcul serait de tester tous les scores de comparaison des ensembles intra et inter, mais, cela produirait un nombre d'itération trop important

FIG. 4.16: Méthode standard de calcul de l'EER

Entrées : Intrascores et interscores, N (le nombre de seuils à tester)
Sorties : L'EER et les informations de la courbe ROC
roc ← [] ;
eer ← 1.0;
diff ← 1.0;
start ← $min(scores)$;
end ← $max(scores)$;
pour τ *de* start *à* end *en* N *pas* **faire**
    far ← calcul du FMR avec le seuil τ;
    frr ← calcul du FNMR avec le seuil τ;
    ajoute (far,frr) à roc ;
    si $abs(FMR - FNMR) < DIFF$ **alors**
        diff ← abs(far − frr);
        eer ← (far + frr)/2;
retourner *EER, ROC*

Dans la communauté de la biométrie, la forme de la courbe ROC suit toujours la même structure : il s'agit d'une fonction monotone décroissante (en présentant le FNMR en fonction du FMR) et la valeur de l'EER correspond au point de la courbe vérifiant $x_{ROC} = y_{ROC}$ (ou FMR = FNMR). Grâce à cette propriété, la courbe symbolisant la différence entre y_{ROC} et x_{ROC} est également une fonction monotone décroissante sur [−1,1], avec $EER = x_{DIFF}$ lorsque $y_{DIFF} = 0$. Grâce à cette information, nous savons que pour obtenir le EER, il suffit de trouver x_{DIFF} pour lequel y_{DIFF} est le plus proche possible de zéro. Une analogie avec la version classique de calcul de EER serait de calculer incrémentalement y_{DIFF} par chaque seuil de décision par ordre croissant et d'arrêter le calcul lorsque y_{DIFF} change de signe. De cette façon, on peut espérer effectuer deux fois moins de comparaisons qu'avec la version classique si les scores sont distribués convenablement. Une façon plus efficace est d'utiliser quelque chose d'équivalent à l'algorithme « diviser pour mieux régner » (comme l'utilisation d'un arbre de recherche binaire) afin d'obtenir une complexité plus proche de $O(\log(n))$. C'est pourquoi nous avons mis au point une version polytomique de calcul d'EER approximatif. La méthode est la suivante :

1. Choisir ω seuils distribués linéairement le long de l'ensemble des scores.
2. Pour chaque seuil τ parmi les ω seuils selectionnés, nous calculons les valeurs ($FMR(\tau)$,$FNMR(\tau)$).
3. Prendre les deux seuils successifs τ_1 et τ_2 tels que $signe\,(FNMR(\tau_1) - FMR(\tau_1))$ soit différent de $signe\,(FNMR(\tau_2) - FMR(\tau_2))$.
4. Répéter l'étape 2 en sélectionnant ω seuils entre τ_1 et τ_2 inclus, tant que $FNMR(\tau_1) - FMR(\tau_1)$ n'atteint pas la précision attendue.

Ainsi, le nombre de comparaisons de seuils est largement inférieur à celui de la méthode classique, et, bien que cela ne soit pas le but recherché, le résultat peut être plus précis car l'espace de recherche est estimé à nouveau à chaque itération. L'analyse de complexité de l'algorithme n'est pas une chose aisée car elle dépend à la fois de la précision attendue et du choix de ω. Elle peut être estimée à $O(\log(n))$. La figure 4.17 décrit l'algorithme, tandis que la figure 4.18 illustre son fonctionnement en montrant les différentes itérations afin d'obtenir l'EER d'un vrai jeu de données. Dans cet exemple, nous avons choisi $\omega = 5$ points de calcul pour chaque itération et la valeur approchée de l'EER est obtenue en cinq itérations. Les cercles présentent les points calculés à l'itération actuelle, les triangles présentent les points calculés aux itérations précédentes et la

courbe en pointillés représente la vraie courbe ROC si tous les seuils sont utilisés. Très peu de points sont donc calculés pour obtenir l'estimation de l'EER. La figure 4.18f présente la courbe ROC complète, ainsi que la courbe ROC estimée par la méthode proposée. Nous pouvons voir que la courbe ROC estimée est approximative mais relativement similaire à la vraie courbe autour de la valeur de l'EER.

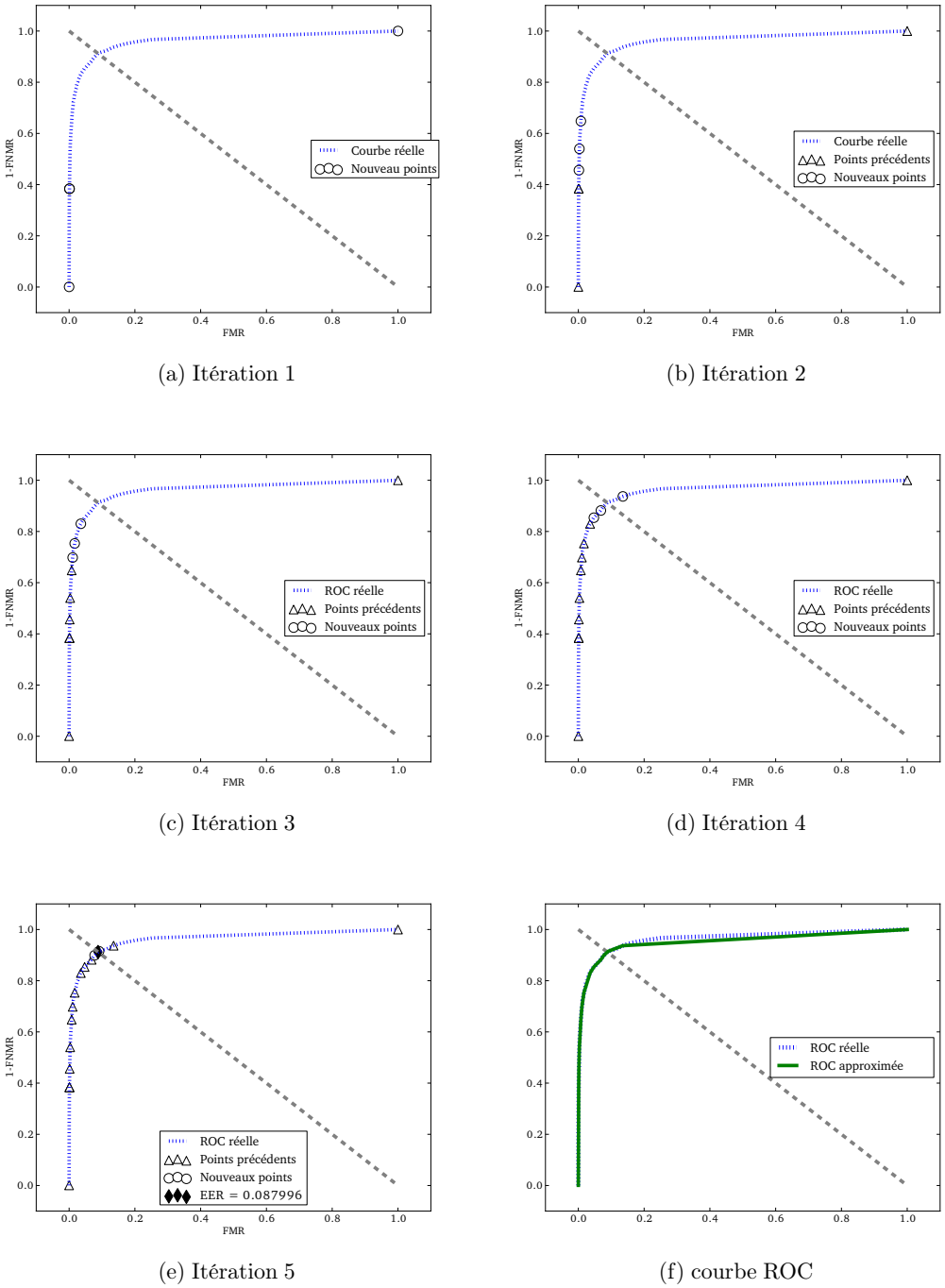
FIG. 4.17: Algorithme d'approximation rapide de l'EER

Entrées : Intrascores et interscores, ω, la précision
Sorties : L'EER et les informations de la courbe ROC

```
roc ← [] ;
cache ← {} ;
debut ← min(scores) ;
fin ← max(scores) ;
tant que Vrai faire
    pour chaque seuil de debut à fin en ω étapes faire
        sdiff ← [] ;
        seuils ← [] ;
        si cache[seuil] existe alors
            /* Évite de recalculer les bornes */
            (far,frr) ← cache[seuil] ;
        sinon
            far ← calcul du FMR avec le seuil seuil ;
            frr ← calcul du FNMR avec le seuil seuil ;
            ajoute le couple (far,frr) à roc ;
            cache[seuil] ← (far, frr) ;
        si abs(far − frr) ≤ precision alors
            eer ← (far + frr)/2 ;
            retourner (eer, roc);
        ajoute far-frr à sdiff ;
        ajoute seuil à seuils ;
    pdebut ← −1 ;
    pfin ← −1 ;
    pour pivot de 0 à steps-1 faire
        si signe(sdiff[pivot]) ≠ signe(sdiff[pivot + 1]) alors
            pdebut ← pivot ;
            pfin ← pivot + 1 ;
            Quitter la boucle ;
    /* pdebut et pfin sont configurés */
    debut ← seuils[pdebut];
    fin ← seuils[pfin];
```

4.2.4. Validation de la méthode

Nous avons utilisé trois bases ayant des configurations différentes pour valider la proposition. Il s'agit des bases de scores de BANCA [Poh], BSSR1 [NIST, 2006] et une base privée utilisant des utilisateurs chimériques sur les modalités DDF et RF (notre base de DDF (section 3.2.1.3)



(a) Itération 1

(b) Itération 2

(c) Itération 3

(d) Itération 4

(e) Itération 5

(f) courbe ROC

FIG. 4.18: Points calculés par l'algorithme proposé lorsque $\omega = 5$. Dans cette configuration, l'EER est calculé en cinq itérations. Chaque image représente une itération, avec la courbe ROC complète, les points calculés à l'itération actuelle et les points calculés aux itérations précédentes (différents seuils peuvent produire le même point)

et AR [Martinez et Benavente, 1998]). Pour la base chimérique cinq jeux de scores sont calculés, trois pour la reconnaissance par DDF et deux pour la reconnaissance par RF :

– a : calcul des scores à l'aide de notre contribution, section 3.3 (utilisation de données d'imposteurs) ;

- b : calcul des scores à l'aide de la méthode nommée STAT3, section 3.3 ;
- c : calcul des scores à l'aide de la méthode nommée RHYTHM, section 3.3 ;
- d : calcul des scores à l'aide d'un système de double association SIFT ;
- e : calcul des scores en utilisant EIGENFACE.

Les méthodes ont été sélectionnées car elles sont basées sur des techniques totalement différentes, et nous espérons ainsi réduire au maximum la corrélation entre les scores générés. Pour plus de détails concernant l'évaluation de la méthode proposée, se reporter à [Giot *et al.*, 2010a, 2012a]. La méthode proposée (notée polyto) a été testée avec différents paramètres :

- Nombre d'itérations de 3 à 7
- Précision de calcul de 0.01, 0.005 et 0.003

et comparée avec la méthode usuelle présentée précédemment (notée classique) et une variante plus précise qui utilise tous les scores et pas d'échantillonnage (notée tout). La méthode *classique* est utilisée avec un nombre de seuils valant 50, 100, 500 et 1000. Le tableau 4.1 présente : les paramètres des méthodes ; l'erreur obtenue sur la précision du calcul de la durée en millisecondes pour obtenir le résultat ; et le nombre de seuils utilisés pour la comparaison.

TAB. 4.1: Comparaison des méthodes de calcul de l'EER sur le premier jeu de données (qui comporte moins de scores que dans certaines configurations de la méthode classique). Comp. est le nombre de comparaisons de seuils effectués

Méthode	**EER (%)**	**Erreur (%)**	**Temps (ms.)**	**Comp.**
tout	8.747159	0.07	3720	2383
classique_100	9.557449	4.13	160	100
classique_1000	8.764836	0.20	1530	1000
classique_5000	8.716540	0.16	7609	5000
classique_10000	8.747159	0.07	15680	10000
dicho_3_0.010	8.799558	0.30	19	11
dicho_3_0.005	8.747159	0.07	10	14
dicho_3_0.003	8.747159	0.07	20	14
dicho_4_0.010	8.726641	0.40	19	15
dicho_4_0.005	8.764836	0.20	19	16
dicho_4_0.003	8.723801	0.10	30	18
dicho_5_0.010	8.799558	0.30	29	16
dicho_5_0.005	8.747159	0.07	30	20
dicho_5_0.003	8.747159	0.07	30	20
dicho_6_0.010	8.726641	0.40	19	15
dicho_6_0.005	8.723801	0.10	19	19
dicho_6_0.003	8.723801	0.10	29	19
dicho_7_0.010	8.747159	0.07	30	21
dicho_7_0.005	8.747159	0.07	30	21
dicho_7_0.003	8.747159	0.07	39	21

En utilisant un test de Kruskal-Wallis avec un degré de confiance de 95%, la méthode proposée surpasse largement la méthode classique en terme d'erreur de précision de calcul (avec une valeur p de 0.0305) et en terme de temps de calcul (avec une valeur p de 0.002562). Notre objectif initial était de diminuer le temps de calcul quitte à perdre en précision dans les résultats. Nous voyons que nous obtenons les résultats plus rapidement, ce qui s'explique grâce à la nette diminution du nombre de comparaisons à effectuer. Cependant, pour certaines configurations, nous obtenons également une meilleure approximation de la valeur du EER qu'avec la méthode classique.

4.3. Combinaison de différents systèmes

Pour améliorer les performances des systèmes de DDF, nous avons choisi de les fusionner avec des systèmes de RF. Nous avons testé deux systèmes de fusion de scores basés sur des algorithmes évolutionnaires :

- Un système de fusion où nous avons défini manuellement des fonctions de fusion dont les paramètres sont automatiquement configurés par Algorithmes Génétiques (AG).
- Un système de fusion où la fonction de fusion est générée automatiquement par PG.

Les méthodes de fusion ont été validées sur les mêmes bases de données que celles utilisées pour évaluer la méthode de calcul de l'EER approximé. D'ailleurs l'utilisation de la méthode d'approximation d'EER a permis de grandement réduire le temps de calcul en ayant un gain moyen (sur trois bases de données différentes et trois fonctions de fusions) en temps de calcul de 55,20% sur l'ensemble du processus d'évolution. Notre thèse étant sur la DDF, nous ne présenterons que les résultats spécifiques à la DDF, pour plus d'informations, se reporter à [Giot *et al.*, 2010a,b; Giot et Rosenberger, 2012]. Le système de fusion nous intéressant utilise la RF et la DDF et est composé des sous-systèmes suivants :

- Un système de RF utilisant un système de comparaison basé sur une double association de points d'intérêts proposée par Rosenberger et Brun [2008] comme extension de la méthode SIFT définie par Lowe [2004].
- Un système de RF utilisant un système de comparaison basé sur EIGENFACE [Turk et Pentland, 1991].
- Un système de DDF basé sur le rythme.
- Un système de DDF basé sur une méthode statistique [Magalhães *et al.*, 2006].
- Un système de DDF basé sur notre méthode présentée en section 3.3. Comme elle utilise les exemples d'imposteurs, nous ne l'avons pas utilisée dans toutes les expériences.

4.3.1. Méthodes de fusion développées

La somme pondérée est notre classifieur de référence contre lequel nous voulons effectuer des comparaisons. En effet, il est relativement simple et efficace. La fonction est la suivante :

$$ga1 = \sum_{i=1}^{c} w_i * s_i \tag{4.16}$$

avec $\boldsymbol{s} = \{s_1, s_2, \cdots, s_c\}$ les c scores des c sous-systèmes à fusionner, et $\boldsymbol{w} = \{w_1, w_2, \cdots, w_c\}$ les poids de la fonction de fusion.

4.3.1.1. Fonctions paramétrées par algorithmes génétiques

Nous avons proposé deux fonctions simples de fusion. Elles ont été créées pour favoriser les scores faibles par rapport aux scores importants (nous travaillons avec des scores de dissimilarité):

$$ga2 = \prod_{i=1}^{c} s_i^{x_i} \tag{4.17}$$

avec $\boldsymbol{x} = \{x_1, x_2, \cdots, x_c\}$ les poids de la fonction de fusion.

$$ga3 = \sum_{i=1}^{c} w_i * s_i^{x_i} \tag{4.18}$$

avec $\boldsymbol{w} = \{w_1,w_2,\cdots,w_c\}$ et $\boldsymbol{x} = \{x_1,x_2,\cdots,x_c\}$ les poids de la fonction de fusion. Les trois fonctions (la somme pondérée $ga1$ et nos contributions $ga2$ et $ga3$) nécessitent une configuration de leurs poids. Nous avons choisi d'utiliser un système d'AG pour effectuer cette configuration :

- Les chromosomes sont des vecteurs à valeur réelle dans l'intervalle $[-5;5]$ et ont une dimension égale au nombre de poids de la fonction à optimiser.
- La fonction d'évaluation consiste à calculer l'EER approximé par notre méthode (section 4.2) sur le jeu de score produit par la fonction de fusion paramétrée par le chromosome à évaluer.
- Les calculs sont effectués sur une population de 5000 individus, sur 500 générations avec une probabilité de mutation de 0.9.

Le tableau 4.2 résume les paramètres du système d'optimisation.

TAB. 4.2: Résumé de la configuration du système à base d'algorithme génétique

Configuration	**Valeurs**
Objectif	Optimiser une fonction de fusion afin de produire un score multi-biométrique.
Chromosomes	Poids de la fonction dans l'intervalle $[-5;5]$
Évaluation	Calcul de l'EER approximé du système généré (section 4.2)
Terminaux	– a, b, c : scores de DDF – d, e : scores de RF
Population initiale	5000 chromosomes aléatoires
Paramètres d'évolution	– Nombre d'individus : 5000, – Nombre de générations maximum : 500, – Probabilité de croisement : 45%, – Probabilité de mutation : 50% – Probabilité de reproduction : 5% (avec élitisme), – Sélection : tournois de taille 10 avec une probabilité de sélection de 90%.
Critère de fin	Le meilleur individu a une évaluation inférieure à 0.001 ou le nombre maximal de générations est atteint.
Ensemble d'apprentissage	La première moitié des n-uplets de scores intra-classes et la première moitié des n-uplets de scores inter-classes.
Ensemble de validation	La seconde moitié des n-uplets de scores intra-classes et la seconde moitié des n-uplets de scores inter-classes.

4.3.1.2. Fonctions générées par programmation génétique

La configuration automatique des fonctions de fusion permet d'obtenir les poids optimaux afin d'obtenir les meilleures performances de fusion sur la base de validation. Cependant, il est nécessaire de définir manuellement les fonctions à optimiser. La définition de ces fonctions repose sur des heuristiques ou des idées sur la distribution des scores qui peuvent ne pas être vérifiées. Il est donc possible que les fonctions définies ne soient pas optimales. Pour cette raison, une

approche par programmation génétique peut se révéler plus intéressante : au lieu de chercher à obtenir les poids optimaux de la fonction de fusion, nous cherchons directement la fonction de fusion optimale. Le forme de la fonction obtenue n'est pas limitée à la classe des polynômes.

- Les fonctions générées par programmation génétique n'agissant pas forcément comme des fonctions de fusion, nous n'avons pas pu utiliser la méthode de calcul de l'EER approximé (la courbe ROC ne respectant pas forcément les propriétés attendues).
- Les fonctions suivantes peuvent être utilisées par l'arbre généré : $\{-, +, *, /, max, min, avg\}$.
- L'évolution est effectuée sur une population de 500 individus sur 50 générations avec une probabilité de mutation de 0.5.

Le mécanisme est résumé dans le tableau 4.3.

TAB. 4.3: Résumé de la configuration du système à base de programmation génétique. Les nombres utilisés dans les fonctions peuvent être des scores ou des constantes

Configuration	**Valeur**
Objectif	Générer une fonction de fusion produisant un score multi-biométrique.
Fonctions	– $+$: addition de deux nombres, – $-$: soustraction de deux nombres – $*$: multiplication de deux nombres – $/$: division de deux nombres – min: minimum de deux nombres – max: maximum de deux nombres – avg: moyenne de deux nombres
Évaluation	Calcul de l'EER du système généré (méthode classique)
Terminaux	– a, b, c : scores de DDF – d, e : scores de RF – 50 constantes distribuées linéairement entre 0 et 1.
Population initiale	500 arbres aléatoires avec une profondeur entre 2 et 8 générés avec la méthode « ramped half and half ».
Paramètres d'évolution	– Nombre d'individus : 500, – Nombre de generations maximal : 50, – Profondeur maximum : 8, – Probabilité de croisement : 45%, – Probabilité de mutation : 50% – Probabilité de reproduction : 5% (avec élitisme), – Sélection : tournois de taille 10 avec une probabilité de sélection de 80%.
Critère de fin	Le meilleur individu a une évaluation inférieure à 0.001, ou le nombre maximal de générations est atteint.
Ensemble d'apprentissage	La première moitié des n-uplets de scores intra-classe et la première moitié des n-uplets de scores inter-classe.
Ensemble de validation	La seconde moitié des n-uplets de scores intra-classe et la seconde moitié des n-uplets de scores inter-classe.

4.3.2. Performances obtenues

L'évaluation des méthodes développées est effectuée en utilisant un système de validation croisée à deux ensembles sur la base chimérique. La première moitié des n-uplets de scores est utilisée comme ensemble d'apprentissage, tandis que l'autre moitié des n-uplets de scores est utilisée pour valider la méthode. Le processus est ensuite répété en échangeant le rôle des deux moitiés. Les scores originaux ont tous été préalablement normalisés par la méthode *tanh*, et transformés en distances le cas échéant. Nos propositions sont comparées aux méthodes classiques de l'état de l'art, ainsi qu'à un classifieur SVM (une recherche des paramètres optimaux est effectuée en utilisant une validation croisée à trois ensembles avec l'outil *easy.py* fourni par la bibliothèque libsvm). Nous effectuons la fusion de tous les sous-systèmes ensemble ; nous n'avons pas effectué de fusion deux à deux en partant du principe que les algorithmes évolutionnaires sont capables d'évincer les systèmes inutiles en leur donnant des poids faibles (pour les fonctions de fusion configurées par AG) ou en ne les incluant pas dans l'arbre généré (pour les fonctions de fusions créées par PG). Nous avons également calculé les performances des méthodes de fusion de référence sur la DDF uniquement et la RF uniquement. Nous n'avons pas calculé d'intervalle de confiance, mais il est fort probable que l'intervalle des propositions se chevauche.

Le tableau 4.4 présente les résultats de fusion. On peut voir que nos fonctions configurées par AG donnent de meilleures performances que la somme pondérée configurée par AG qui est la méthode souvent considérée comme la plus performante de l'état de l'art. La fonction de fusion générée par PG donne de meilleurs résultats. Il faut noter que même si la méthode à base de PG permet d'obtenir une fonction de fusion plus efficace que celles configurées par AG, le temps de calcul est beaucoup plus important (même avec dix fois moins d'individus et de générations). Nous n'avons pas enregistré le temps de calcul nécessaire pour effectuer l'optimisation, mais on peut estimer qu'il faut quelques jours pour la méthode à base d'AG et quelques semaines pour la méthode à base de PG. De plus, les arbres générés peuvent être inutilement complexes. À titre d'exemple, la figure 4.19 présente un arbre relativement simple généré automatiquement. On peut observer que la branche $avg(a,a-1/12)$ pourrait être simplifiée en $a-1/24$.

La fusion des deux méthodes de RF ne donne pas de meilleures performances, cependant cela peut s'expliquer par les très faibles performances de la méthode egeinface comparé à la double association SIFT. La fusion des différentes méthodes de DDF donne de meilleurs résultats lorsque la méthode à base de SVM (notre contribution en section 3.3) n'est pas utilisée. Une fois de plus cela peut s'expliquer par la différence de performance entre cette méthode et les autres. Par contre, nous pouvons voir que la DDF et la RF sont fortement complémentaires au vu de la forte amélioration des performances en combinant les systèmes des deux modalités.

4.4. Biométrie douce pour la dynamique de frappe au clavier

La multibiométrie permet d'augmenter les performances de reconnaissance et les meilleurs systèmes biométriques sont ceux associant la DDF à une autre modalité, plutôt que ceux fusionnant plusieurs sous-systèmes de DDF (section 4.3).Nous nous sommes également intéressés dans les travaux de cette thèse à la reconnaissance du genre et de catégorie d'âge par DDF. À notre connaissance, bien que de telles études soient nombreuses en RF, ou en reconnaissance de la démarche, il s'agit des premiers travaux de ce type en DDF. Ceci permet d'ouvrir une nouvelle voie dans le domaine de recherche en biométrie douce. Utiliser le genre de la personne saisissant le mot de passe a plusieurs avantages :

– Comme toute biométrie douce, il s'agit d'un paramètre additionnel permettant de réduire le taux d'erreur de reconnaissance, sans avoir à capturer ou mesurer d'information supplé-

TAB. 4.4: Performance des méthodes de fusion sur la base chimérique (GREYC+AR). Chaque système de fusion effectue la fusion des cinq systèmes

Methode		**EER**
Sous-systèmes	DDF (section 3.3)	
	a (CONTRIB)	08,92%
	b (STAT3)	11,53%
	c (RHYTHM)	15,69%
	RF	
	d (Double association SIFT)	06,21%
	e (EIGENFACE)	31,43%
Fusion de base (tous les systèmes)	*sum*	**02,70%**
	min	13,72%
	mul	**02,67%**
	SVM	08,80%
	*ga*1	**02,31%**
Fusion de base (DDF seule)	*sum*	08,99%
	min	09,00%
	mul	10,19%
Fusion de base (DDF seule, sans SVM)	*sum*	**09,99%**
	min	**09,89%**
	mul	11,66%
Fusion de base (RF seule)	*sum*	07,75%
	min	07,83%
	mul	19,70%
Proposition (tous les systèmes)	*ga*2	**02,22%**
	*ga*3	**02,26%**
	gp	**01,57%***

mentaire (la reconnaissance du genre est effectuée avec la donnée biométrique utilisée pour la vérification). On peut donc l'utiliser avec une approche de fusion biométrique à bas coût.

– L'information peut être utilisée seule pour vérifier si le genre clamé par l'individu correspond bien à son genre réel (cette approche est relativement importante dans les environnements WEB, où les utilisateurs sont inconnus et peuvent mentir facilement).

Nos contributions principales dans ce domaine sont :

– La proposition d'un nouveau système permettant de prédire le genre d'un individu selon sa façon de saisir un texte prédéfini sur un clavier standard.
– De montrer des résultats nettement meilleurs que le hasard sur la reconnaissance du genre utilisant la DDF.
– De montrer des résultats nettement meilleurs que le hasard de vérification par DDF utilisant le genre comme information additionnelle aux informations temporelles.
– La proposition d'un nouveau système permettant de prédire le genre d'un individu selon sa façon de saisir un texte quelconque sur un clavier standard.

4.4.1. Reconnaissance du genre sur texte fixe

Nous ne disposons d'informations de genre que sur notre base de données (section 3.2.1.3), car les autres bases publiques ne diffusent pas de telles informations. Afin de réduire le biais dû à la supériorité numérique des utilisateurs masculins, nous avons seulement sélectionné n données biométriques appartenant à des hommes, avec n le nombre de données biométriques appartenant

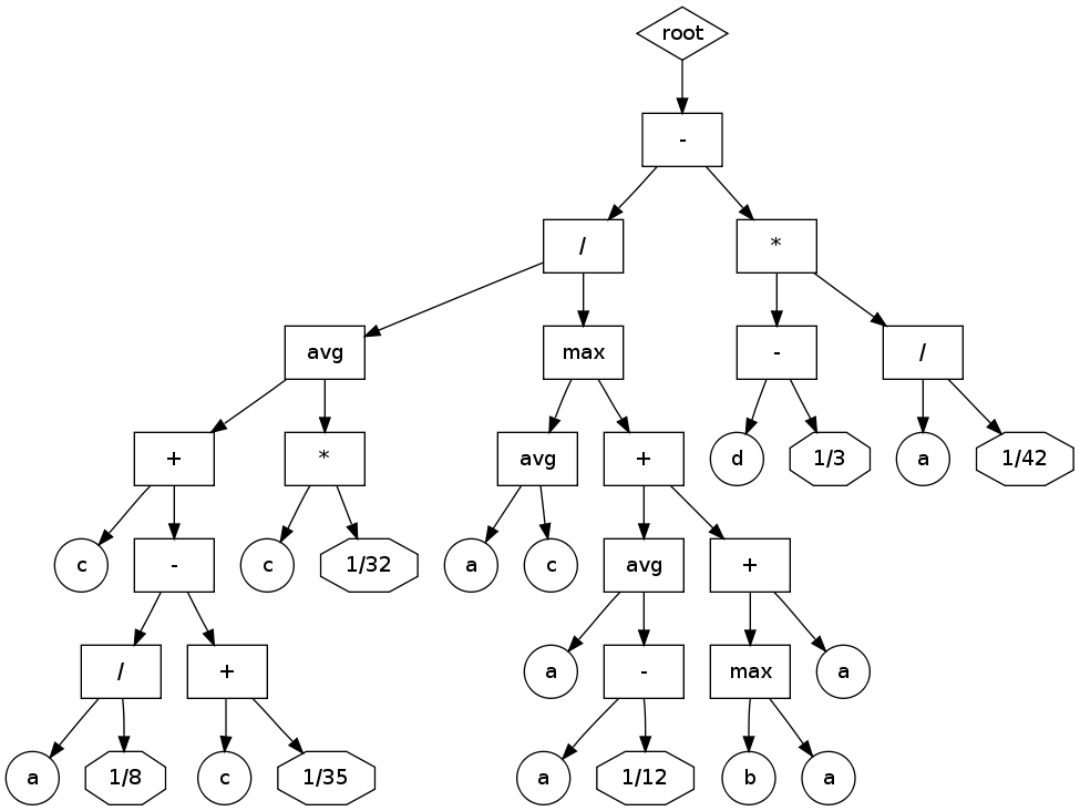


FIG. 4.19: Exemple de fonction de fusion représentée par un arbre produit par la programmation génétique

à des femmes. Bien que cela ne permette pas la généralisation des résultats, une étude plus récente sur une autre base de données privée a obtenu des résultats équivalents (travaux non encore publiés) : `http://skynet.cs.clemson.edu/ival/KEYBIO/PrelimGenderCheck.html`.

4.4.1.1. Méthode de reconnaissance sur texte fixe

La première étape concerne la sélection des informations pertinentes pour la reconnaissance du genre.

4.4.1.1.1. Informations pertinentes Nous utilisons les données de type (i) intervalle de temps entre le relâchement de touches successives (Release-Release) (RR), (ii) intervalle de temps entre le relâchement d'une touche et la pression de la suivante (Release-Press) (RP), (iii) intervalle de temps entre la pression de touches successives (Press-Press) (PP), (iv) intervalle de temps entre la pression et le relâchement d'une touche, autrement dit : durée de pression de la touche (Press-Release) (PR) et (v) le vecteur V consistant en la concaténation des informations précédentes (ces informations sont présentées en section 3.3.2.2). Ainsi, pour chaque capture, nous disposons de cinq représentations différentes. Ces vecteurs sont les données que nous utilisons tout au long de l'étude pour effectuer la classification. Comme le texte à saisir est de taille fixe, ces vecteurs sont également de taille fixe. Nous cherchons à savoir lequel est le plus approprié.

4.4.1.1.2. Apprentissage Nous utilisons un SVM [Vapnik, 1998] avec un noyau gaussien pour effectuer la classification des saisies effectuées par des hommes et des saisies effectuées par des femmes. Un SVM est donc appris pour chaque type de données extraites afin de comparer

leur performance. Les données des hommes sont représentées par l'étiquette 1, et les données des femmes sont représentées par l'étiquette -1. En utilisant le script *Python* fourni avec la bibliothèque libsvm [chung Chang et Lin, 2001], nous avons effectué une recherche des paramètres optimaux de la façon suivante (pour chaque jeu de données) :

- Le jeu de données est normalisé afin d'avoir des valeurs dans l'intervalle $[-1; 1]$.
- Plusieurs couples de (γ, C) sont testés. Pour chaque couple, une validation croisée à cinq ensembles est effectuée.
- Le couple (γ, C) générant le meilleur taux de reconnaissance est sélectionné.

Une fois la reconnaissance de genre effectuée, nous utilisons cette information pour améliorer les performances des algorithmes de DDF.

4.4.1.1.3. Score de genre Plutôt que d'utiliser les étiquettes prédites, nous calculons un score. Nous calculons ce score en utilisant l'étiquette prédite et la probabilité de confiance accordée à ce résultat. Nous générons un score compris dans l'intervalle $[0; 1]$. Les scores proches de 0 représentent les hommes, tandis que les scores proches de 1 représentent les femmes. L'équation (4.19) présente la façon de calculer ce score, avec *predict* l'étiquette prédite par le classifieur et *probability* sa probabilité :

$$genderscore = \frac{1 - predict * probability}{2} \tag{4.19}$$

4.4.1.2. Fusion de la dynamique de frappe et de la reconnaissance du genre

Nous verrons plus loin que nous sommes capables à 90% de reconnaitre le genre des individus. Il est intéressant de vérifier si cette information peut être utilisée dans les systèmes de reconnaissance par DDF, afin d'en améliorer les performances.

4.4.1.2.1. Reconnaissance par dynamique de frappe Pour cette étude, nous avons utilisé la méthode de reconnaissance statistique nommée STAT2 (équation (3.11)) qui a été choisie car elle est plus rapide à calculer que la méthode STAT3 qui est plus performante. Pour chaque utilisateur, 20 exemples sont utilisés pour l'apprentissage (le calcul des vecteurs moyens et écart type) et 10 exemples sont utilisés pour la vérification. La méthode de reconnaissance retourne une distance entre 0 et 1.

4.4.1.2.2. Fusion d'information Nous avons testé deux façons d'effectuer la fusion d'information. En effectuant une « fusion d'attributs » : nous ajoutons à chaque vecteur d'attribut le score de genre. Le mécanisme de reconnaissance de DDF est appliqué sur ces nouvelles données. En effectuant une « fusion des scores » : après le calcul du score de reconnaissance de DDF ($biometricscore$) et du score de genre ($genderscore$). Le score final est calculé suivant (4.20), avec μ_{gender} la moyenne des scores de genre ($genderscore$) des captures d'enregistrement et w un poids.

$$decisionscore = biometricscore + w * abs(\mu_{gender} - genderscore) \tag{4.20}$$

Nous avons fixé empiriquement w à 0.25 afin de donner plus de poids au score de reconnaissance par DDF. La partie $w * abs(\mu_{label} - genderscore)$ permet de pénaliser les scores trop différents de ceux qui sont attendus. Ce qui permet d'utiliser dans la même formule des données n'ayant pas le même sens Nous avons testé plusieurs scénarios pour vérifier les bénéfices de l'approche. Par abus de langage, nous employons le terme « étiqueter » pour « calculer le score de genre ».

4.4.1.2.3. Scénarios testés Quatre scénarios de génération d'information de genre ont été testés :

- *Pas d'information de genre* : c'est la façon classique de faire de la DDF. Ce système nous sert de référence.
- *Étiquetage manuel* : le genre de chaque donnée biométrique est étiqueté manuellement en utilisant le genre réel du propriétaire de la capture. Le score de genre vaut donc, soit 0, soit 1. Ce scénario peut être vu comme une *approche supervisée*, ou un opérateur informe le système de reconnaissance du genre de l'individu.
- *Étiquetage automatique* : le genre de chaque donnée biométrique est étiqueté automatiquement en utilisant le score prédit de la capture. Il s'agit d'une approche *non supervisée* où tout est automatique. Quelques erreurs peuvent survenir lors du calcul du score de genre des captures d'enregistrement ou de test.
- *Étiquetage semi-automatique* : Les captures d'enregistrement sont étiquetées manuellement, tandis que les captures de validation sont étiquetées automatiquement. C'est une approche *semi-supervisée* où un opérateur étiquette manuellement les captures d'enregistrement alors que l'information de genre est prédite lors de l'authentification.

Afin de ne jamais avoir de captures du même utilisateur à la fois dans l'ensemble d'apprentissage du classifieur de genre et dans les données de validation, lors de la reconnaissance par DDF, nous avons utilisé le mécanisme de validation croisé suivant : les utilisateurs sont séparés en deux ensembles de même taille : un SVM est appris pour chaque ensemble et utilisé pour calculer le score de genre dans l'autre ensemble. Cela nous permet d'éviter de reconnaitre les utilisateurs plutôt que le genre.

4.4.1.3. Résultats

Cette section présente les résultats expérimentaux de l'étude de reconnaissance du genre sur texte fixe. 35 femmes et 98 hommes ont participé à la création de notre base. Il y a trois fois plus d'hommes que de femmes, ce qui peut être un problème pour l'apprentissage. C'est pourquoi, nous n'avons pas utilisé toute la base de données lors des expériences et avons gardé 35 hommes et 35 femmes.

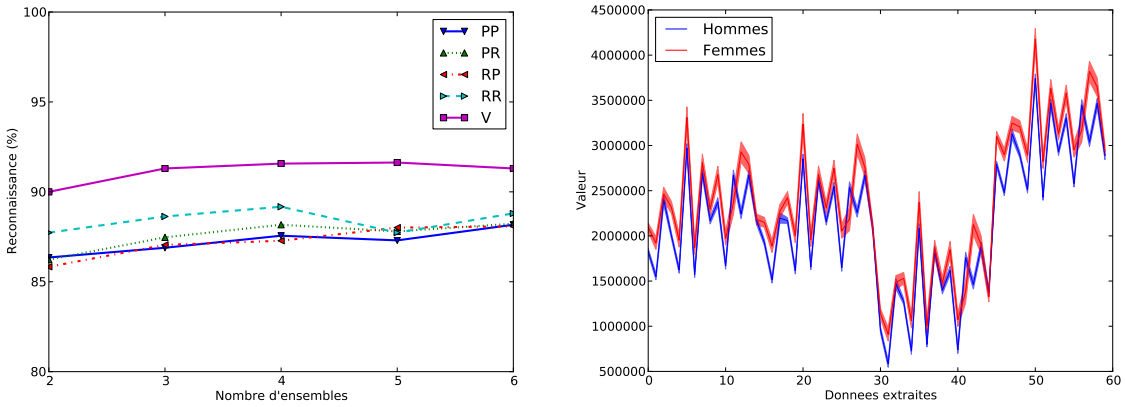
4.4.1.3.1. Performance de reconnaissance du genre Cette partie présente les résultats de la méthode de reconnaissance du genre sur texte fixe. Le tableau 4.5 présente le taux de reconnaissance avec la meilleure configuration (γ, C) pour chaque type de données extraites en utilisant une validation croisée à cinq ensembles. La performance la moins intéressante est de 87,31% tandis que la meilleure est de 91,63%. Ces résultats ne sont pas trop éloignés de ce que l'on peut rencontrer pour la reconnaissance du genre en RF. Nous pouvons affirmer que la reconnaissance du genre en DDF est possible et fonctionnelle (testé sur 70 personnes seulement).

TAB. 4.5: Taux de reconnaissance du genre en utilisant une validation croisée en cinq ensembles

Données extraites	RR	PP	RP	PR	V
Performance (%)	88,57	87,31	88,01	87,8	91,63

La figure 4.20a présente la performance de reconnaissance, pour chaque type de données extraites, en fonction du nombre d'ensembles lors de la validation croisée. L'intérêt de cette figure est de montrer que les performances diminuent lorsque le nombre d'ensembles (et donc de données utilisées pour calculer le modèle) diminue. Nous voyons également qu'utiliser le plus d'informations extraites possible augmente grandement les performances par rapport à

l'utilisation d'un seul type de données extraites. Il est donc nécessaire de disposer d'une base de configuration conséquente pour avoir un système fonctionnel. Nous pouvons observer que, quelle que soit la configuration de la validation croisée, le vecteur V donne les meilleures performances. La figure 4.20b présente la bande de confiance à 95% des données extraites catégorisants les hommes et les femmes. On peut voir que les femmes semblent taper plus lentement (les valeurs des données correspondent à des durées et sont plus grandes que celles des hommes) que les hommes au clavier, ce qui expliquerait la possibilité de les différencier.



(a) Performance du système de reconnaissance de genre en fonction de la quantité de données utilisée pour calculer le modèle

(b) Intervalle de confiance à 95% des données extraites pour les hommes et les femmes. Les bornes sont calculées avec $\mu \pm 1.96\frac{\sigma}{N}$, μ le vecteur moyen, σ le vecteur écart type et N le nombre d'exemples concernés

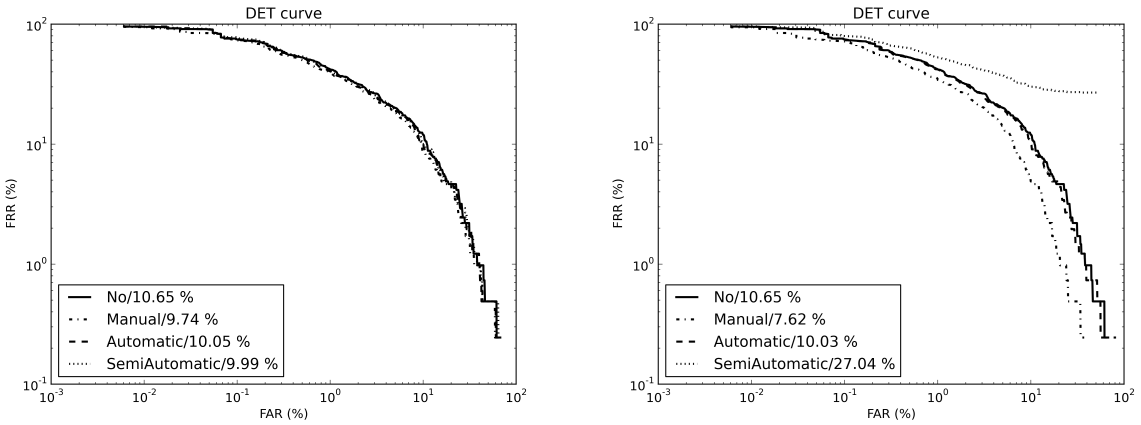
FIG. 4.20: Analyse des différences entre les hommes et les femmes sur la phrase « greyc laboratory »

4.4.1.3.2. Performance de la reconnaissance par dynamique de frappe associée à la reconnaissance de genre Les résultats précédents montrent qu'il est plus intéressant d'utiliser le vecteur V, c'est donc celui que nous avons utilisé dans les expériences suivantes. Nous avons également testé la méthode de fusion en utilisant l'étiquette $\{-1,1\}$ de genre au lieu du score. Lorsqu'une étiquette est utilisée, dans le cas de la fusion de score, nous utilisons une fusion de score logique, tandis que la fusion d'attributs ne change pas. Quand le genre prédit appartient aux labels d'enregistrement de l'individu clamé, le score final est le score de DDF, autrement c'est la distance maximale du système de vérification (pour simuler une comparaison d'imposteur).

Nous utilisons un sous-ensemble de la base afin de conserver le même nombre d'hommes et de femmes ayant 30 captures. Après suppression des individus non valides, nous avons gardé 41 utilisateurs ce qui permet de calculer 410 comparaisons intra-classe et 16400 comparaisons inter-classe. Il serait naturellement préférable de tester avec plus d'utilisateurs.

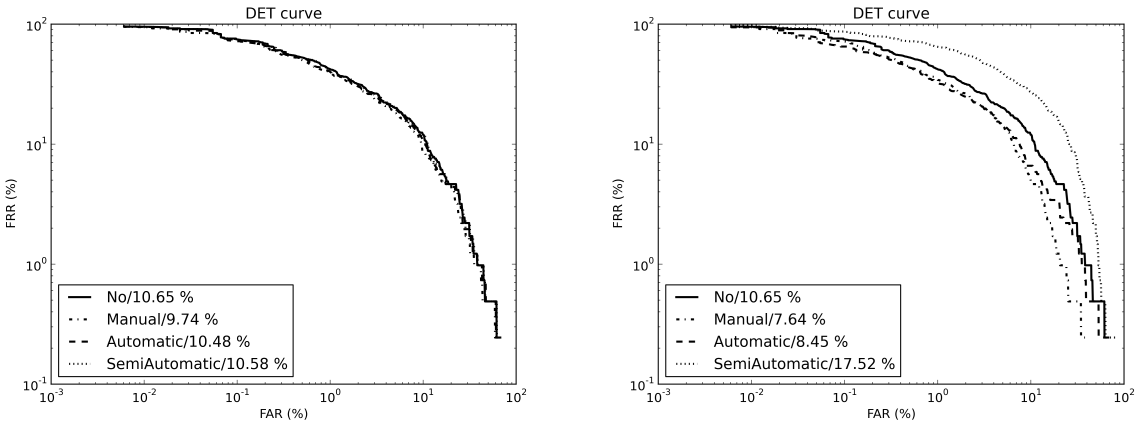
La figure 4.21 présente les courbes ROC des différents systèmes en utilisant l'étiquette de genre, tandis que la figure 4.22 présente les mêmes résultats en utilisant le score de genre. Le tableau 4.6 présente l'EER pour chaque variation et le gain de la meilleure méthode comparée à la méthode de base n'utilisant aucune information de genre.

Nous observons que l'information de genre ne permet pas toujours d'augmenter les performances de reconnaissance. Les méthodes qui utilisent un score pour représenter l'information de genre, plutôt qu'une étiquette, sont plus performantes (ce qui est facilement compréhensible en raison de la notion de confiance dans le score généré). La plupart du temps, dans les scénarios ma-



(a) Fusion d'attributs (b) Fusion de scores

FIG. 4.21: Performances globales de la reconnaissance par dynamique de frappe en utilisant le genre comme une étiquette ($\{-1,1\}$)

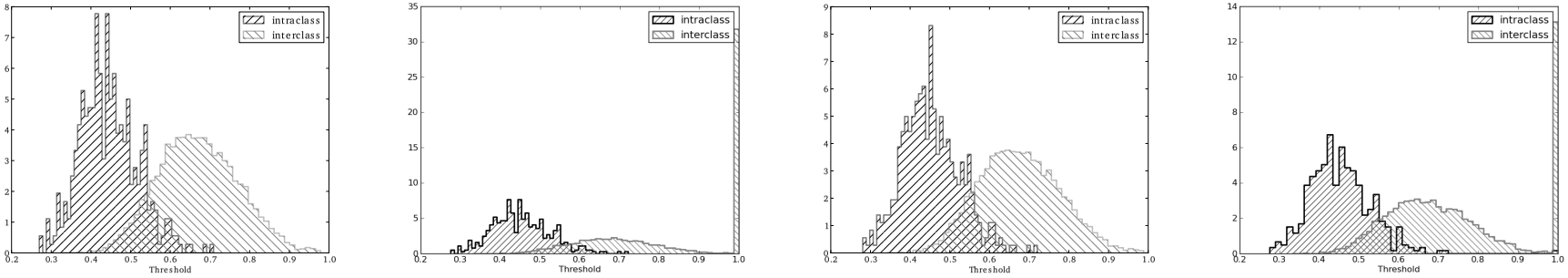


(a) Fusion d'attributs (b) Fusion de scores

FIG. 4.22: Performances globales de la reconnaissance par dynamique de frappe en utilisant le genre comme un score ($[0;1]$)

nuels ou automatiques, nous pouvons améliorer les reconnaissances du système de reconnaissance, mais (comme c'est attendu), ça n'est pas le cas pour la version semi-automatique. L'étiquetage manuel donne de meilleures performances que l'étiquetage automatique, mais cette différence est légère.

La figure 4.23 présente la distribution des scores en utilisant un score de fusion pour les deux schémas différents. En regardant la figure 4.23, nous pouvons comprendre pourquoi la version semi-automatique donne de si mauvais résultats : beaucoup de captures authentiques sont incorrectement étiquetées (sur les données de test), alors qu'elles sont correctement étiquetées manuellement (sur les données d'enregistrement). Avec la procédure automatique, ces erreurs sont lissées car la même erreur peut arriver à la fois dans les données d'enregistrement et dans les données de test.



(a) Score manuel (b) Étiquette manuelle (c) Score automatique (d) Étiquette automatique

(e) Score semi-automatique (f) Étiquette semi-automatique (g) Pas de genre

FIG. 4.23: Fréquence des genres en utilisant la fusion de scores avec la dynamique de frappe et la reconnaissance du genre

TAB. 4.6: EER du système par reconnaissance de dynamique de frappe en utilisant l'information de genre. EER (a/b) présente l'EER avec (a) la représentation du genre et, (b) le type de fusion opéré. La ligne « meilleure » représente la meilleure valeur pour chaque type de reconnaissance de genre. La ligne « gain » représente le gain par rapport au système de base. Les valeurs en gras représentent les valeurs supérieures à la non utilisation du genre. Les valeurs soulignées représentent le meilleur scénario de chaque type

Méthode	Sans	Manuel	Automatique	Semi-automatique
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%
Meilleure	10,65%	**7,64%**	**8,45%**	**9,99%**
Gain	x	28,26%	20,67%	6,20%

Pour résumer, l'étiquetage manuel (en utilisant la fusion de score) donne les meilleures performances, mais un tel système ne peut pas être utilisé dans un contexte réaliste. L'approche automatique propose de moins bons résultats, mais peut être utilisée dans un contexte opérationnel de système de reconnaissance par texte fixe. Les bénéfices de cette approche comparée à une authentification sans reconnaissance du genre sont de 20%. Cependant, il faut relativiser ces résultats par rapport au petit nombre de tests réalisés.

4.4.2. Classification par le genre et l'âge sur texte libre

Nous venons de montrer l'efficacité de la reconnaissance du genre sur texte fixe. Le nombre de cas d'utilisations d'un tel système est limité, car, il est nécessaire de créer le modèle à partir d'un texte fixe, et de reconnaitre le genre à partir du même texte. Les cas d'usage de la reconnaissance à partir d'un texte libre sont souvent plus intéressants, c'est pourquoi nous étudions la possibilité de reconnaitre le genre d'un individu selon sa façon de saisir un texte quelconque. Nous étudions également la possibilité de catégoriser les utilisateurs en fonction de leur âge (moins de 30 ans, contre plus de 30 ans comme illustration). De nombreuses applications peuvent utiliser ce type d'information pour décider si une personne a plus de 18 ans (âge de la majorité) (hélas, nous n'avons pas de données sur cette catégorie d'âge) ou de 26 ans (réduction SNCF).

4.4.2.1. Données disponibles pour la reconnaissance sur texte libre

Cette section présente la base utilisée, ainsi que les données que nous avons extraites.

4.4.2.1.1. Base de données Bello *et coll.* ont diffusé une base de données de DDF de texte fixe Bello *et al.* [2010] contenant plusieurs phrases. Plusieurs volontaires ont saisi 15 paragraphes de texte espagnol, et 15 lignes de commandes shell UNIX courantes dans une application WEB. Étant donné que cette base contient différentes phrases, nous pouvons nous en servir pour faire de la reconnaissance du genre sur texte libre. Pour chacun des évènements claviers, l'application capture les informations suivantes :

- le code de la phrase en cours de saisie ;
- l'instant de l'évènement clavier (timestamp UNIX en millisecondes) ;

– le type d'évènement (pression ou relâchement de la touche) ;
– le code ascii de la touche actionnée.

Pour des raisons techniques liées à l'utilisation de Javascript, certaines touches ne sont pas considérées. Les erreurs de saisies sont acceptées à partir du moment où la saisie d'une phrase contient entre 90% et 130% du nombre de caractères attendus. Pour plus d'informations, se rapporter à l'article original [Bello *et al.*, 2010]. Nous sélectionnons les 55 volontaires ayant participé intégralement à l'étude. Chacun d'entre eux a participé à une session, ainsi chaque utilisateur n'a saisi qu'une seule fois chaque phrase. Nous ne travaillons qu'avec les données de type texte, et pas de type ligne de commande. Plusieurs méta-données sont associées à chaque utilisateur. Dans notre cas, nous nous intéressons au genre et à l'âge de l'utilisateur. Ainsi, la base dispose de 40 hommes et 15 femmes et 29 personnes âgées de moins de 30 ans et 26 personnes âgées de plus de 30 ans.

4.4.2.1.2. Données extraites Contrairement à Bertacchini *et al.*, dans leur étude sur l'identification [Bertacchini *et al.*, 2010], nous ne filtrons pas les données erronées. Nous considérons que de telles données sont toujours présentes dans des applications déployées en environnement opérationnel, et qu'il est nécessaire que les algorithmes de reconnaissance travaillent avec. Nous avons choisi d'utiliser plusieurs types de données extraites :

– l'intervalle de temps entre la pression de deux touches successives ;
– l'intervalle de temps entre le relâchement de deux touches successives ;
– le temps de pression d'une touche ;
– le délai d'attente entre le relâchement d'une touche et la pression de la touche suivante.

Pour chaque phrase saisie, nous calculons les différentes informations liées à chaque digraphe (deux touches successives). Parmi tous les digraphes possibles, nous ne conservons que les digraphes présents au moins une fois dans chacune des phrases [Bello *et al.*, 2010], à savoir : («e », «a », « d», «de», «la», « l»). De même, nous ne conservons la durée de pression que pour les touches présentes dans les digraphes précédents (« », «a», «d», «e», «l»). Il faut noter que ces choix ont été fixés par la structure du texte de la base de données concernée. Il est possible que dans un corpus rédigé avec une langue différente, les digraphes les plus communs soient différents. Lorsqu'une même information est présente plusieurs fois dans une phrase, nous calculons sa moyenne. Ainsi, chaque phrase j de chaque utilisateur i est encodée par un vecteur $\boldsymbol{feat}_j^i$ de dimension 23 :

$$\begin{aligned}&\boldsymbol{feat}_j^i = \\ &\{PP_d, PP_l, PP_a, PP_e, PP_{de}, PP_{la}, RR_d, RR_l, RR_a, RR_e, RR_{de}, RR_{la}, \\ &PR, PR_d, PR_l, PR_a, PR_e, RP_d, RP_l, RP_a, RP_e, RP_{de}, RP_{la}\}\end{aligned} \tag{4.21}$$

avec PP_{ab}, l'intervalle de temps entre la pression de la touche a et la pression de la touche b, RR_{ab}, l'intervalle de temps entre le relâchement de la touche a et le relâchement de la touche b, PR_a, le temps de pression de la touche a, et RP_{ab}, le temps de latence entre la touche a et la touche b.

4.4.2.2. Méthode de reconnaissance sur texte libre

Pour cette étude, nous avons choisi d'utiliser une méthode similaire à celle de la reconnaissance sur texte fixe (section 4.4.1), étant donné que les descripteurs des phrases (section 4.4.2.1.2) sont relativement homogènes bien que les phrases soient hétérogènes : ils ont la même dimension et chaque dimension représente la même chose, quel que soit le texte saisi.

Différentes variantes, présentées dans la sous-section suivante, sont testées. À chaque fois, un ensemble de descripteurs étiquetés (1 pour un texte saisi par un homme, ou quelqu'un de moins de 30 ans, -1 pour un texte saisi par une femme ou quelqu'un de plus de 30 ans) est utilisé. Le nombre d'exemples pouvant être faible (54 exemples au minimum dans certains cas), nous avons utilisé le mécanisme de validation croisée Leave One Out. Avec un ensemble contenant n exemples, chaque exemple est utilisé successivement comme exemple de test, tandis que les $n-1$ autres exemples servent pour l'apprentissage du modèle. De cette façon, nous pouvons disposer d'une quantité de données plus importante pour apprendre le modèle, sans avoir à la fois la même donnée dans l'ensemble d'apprentissage et l'ensemble de validation. Nous travaillons avec des données, où chaque dimension est normalisée par la méthode zscore présentée en équation (4.2). Les vecteurs moyens et écart-type sont calculés avec les exemples d'apprentissage.

Un SVM est entrainé pour apprendre à différencier les deux ensembles de phrases normalisées (soit homme ou femme, soit moins de 30 ans ou plus de 30 ans). Nous utilisons un noyau gaussien. Nous n'avons pas cherché à optimiser les paramètres nécessaires au fonctionnement du SVM : le paramètre γ du noyau et le paramètre C. Nous utilisons les valeurs proposées par défaut par la bibliothèque scikit-learn [Pedregosa *et al.*, 2011] (à savoir $C = 1$ et $\gamma = 1/23$). Plutôt que d'utiliser l'étiquette prédite par le classifieur, nous utilisons la probabilité [chung Chang et Lin, 2001] que l'exemple testé soit saisi par un homme. De cette façon, nous obtenons un score plutôt qu'un booléen et nous pouvons calculer différents taux d'erreurs. Les résultats sont présentés à l'aide de la courbe ROC et de l'aire sous celle-ci. Nous utilisons également l'EER. La figure 4.24 résume le fonctionnement du système de reconnaissance.

4.4.2.3. Scénarios de reconnaissance

Nous testons trois différents scénarios de configuration du mécanisme de reconnaissance du genre par dynamique de frappe :

1. en apprenant à reconnaitre le genre à partir d'une phrase précise, puis en reconnaissant le genre uniquement sur d'autre saisies de la même phrase ;
2. en apprenant à reconnaitre le genre à partir d'un ensemble de phrases quelconques, puis en reconnaissant le genre uniquement à partir d'une phrase inconnue. Il s'agit du scénario qui nous intéresse le plus, le précédent étant similaire à la reconnaissance sur texte fixe ;
3. en utilisant la saisie de plusieurs phrases (au lieu d'une seule) pour reconnaitre le genre d'un individu, le modèle ayant été appris avec d'autres phrases. Ce scénario est utile lorsque la réponse n'a pas besoin d'être instantanée.

4.4.2.3.1. Reconnaissance sur phrase connue Ce scénario est proche de la reconnaissance du genre par mot de passe (voir 4.4.1) à la différence que dans ce cas :

- les descripteurs représentent des informations globales (la moyenne du temps de réalisation de quelques digraphes) plutôt que des informations locales (le temps entre chaque digraphe consécutif) ;
- la quantité d'information disponible est plus faible (descripteur de taille 23 au lieu de 60).

Le problème est donc plus compliqué. Dans ce scénario, nous lançons les calculs pour chaque phrase individuellement. Pour les 15 sous-scénarios possibles, nous disposons de 54 exemples (d'où la nécessité d'utiliser un mécanisme de validation croisée). Pour chaque exemple de test, le modèle est appris avec les données des autres utilisateurs ayant saisi la même phrase (il n'y a donc aucune donnée de l'individu de l'exemple de validation dans les exemples d'apprentissage).

Nous avons ajouté un 16^{e} sous-scénario (nommé global) qui travaille avec toutes les données possibles, soit $810 = 15 * 54$. Pour chaque exemple de test, le modèle est appris avec les données de tous les autres utilisateurs (y compris celles de la même phrase), et les données des autres

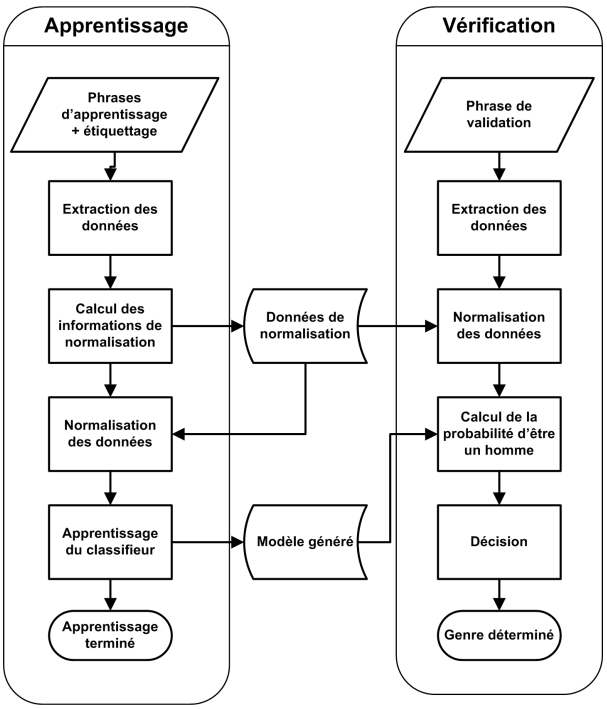


FIG. 4.24: Résumé du fonctionnement du système de reconnaissance du genre par dynamique de frappe au clavier sur texte libre

phrases pour l'utilisateur ayant saisi l'exemple. Il permet de donner des performances générales en faisant abstraction de la phrase choisie.

4.4.2.3.2. Reconnaissance sur phrase inconnue Ce scénario nous semble plus proche d'un cas d'utilisation réel. Le modèle de reconnaissance de genre est appris à partir d'un ensemble de phrases, mais le test est effectué uniquement sur des phrases inconnues du modèle. Ce scénario nous semble le plus réaliste car il ne demande pas à créer un modèle exhaustif. Pour les 15 sous-scénarios possibles, nous disposons de $756 = 14 * 54$ exemples d'apprentissage et de 54 exemples de test. Les exemples d'apprentissage étant fixés, nous n'utilisons pas de système de validation croisée. Nous avons ajouté un 16^{e} scénario (nommé global) qui consiste à calculer les performances à partir de l'ensemble des scores des 15 sous scénarios précédents. Il permet de donner des performances générales en faisant abstraction de la phrase choisie.

4.4.2.3.3. Utilisation de plusieurs phrases Utiliser une seule phrase pour effectuer la classification d'un individu peut ne pas être suffisant. Nous avons testé une méthode prenant en compte la saisie de plusieurs phrases différentes. Nous avons également testé l'évolution de la performance en fonction du nombre de phrases saisies. Cela permet de tester un scénario où il n'est pas nécessaire de disposer instantanément de l'information de classification. Pour évaluer ce scénario, nous avons utilisé le jeu de scores du scénario « Reconnaissance sur phrase inconnue ». Pour une reconnaissance à N phrases, pour chaque utilisateur, nous avons effectué la fusion de toutes les combinaisons possibles de N scores parmi les 15 de l'utilisateur. La méthode de fusion effectuée est une moyenne des scores. Le nombre de scores après fusion peut donc être différent du nombre de scores initiaux (C_N^{15} au lieu de 15 pour chaque utilisateur).

4.4.2.4. Résultats

La figure 4.25 présente les courbes ROC des différentes expériences, tandis que le tableau 4.7 présente l'EER des approches globales. À partir de ces résultats, nous pouvons faire les conclusions suivantes :

- Les performances sont grandement différentes d'une phrase à l'autre : la reconnaissance du genre fonctionne mieux sur certaines phrases (par exemple, la phrase «ks_07» donne toujours de moins bonnes performances que la phrase «ks_02»). Cela peut s'expliquer par la différence de distribution des digraphes (par exemple, la phrase «ks_07» comporte plus de doublons de digraphes que la phrase «ks_02», il y a plus de données moyennées).
- Contrairement à nos attentes, les performances obtenues avec l'utilisation des mêmes phrases en apprentissage et validation donnent de moins bons résultats que l'apprentissage et la validation sur des phrases différentes. Nous pensons que nous obtenons un tel résultat car la quantité de données est faible (à peine une cinquantaine au lieu de plusieurs centaines) lors de l'apprentissage du modèle. Les données n'étant pas suffisamment représentatives, le modèle ne se spécialise pas suffisamment et les performances sont mauvaises. Le fait que le sous-système global donne de bonnes performances appuie ce point.
- La reconnaissance du genre à partir de texte inconnu donne des résultats acceptables : aire sous la courbe de 0,84 lorsque les phrases utilisées lors de la reconnaissance sont inconnues.
- Le taux de reconnaissance de la catégorie d'âge à partir de texte inconnu donne des résultats acceptables : aire sous la courbe de 0,81 lorsque les phrases utilisées lors de la reconnaissance sont inconnues.
- Le taux de reconnaissance de la catégorie d'âge est légèrement inférieur à la reconnaissance du genre.
- Les phrases ayant un bon taux de reconnaissance du genre ne sont pas nécessairement celles ayant un bon taux de reconnaissance de la catégorie d'âge.

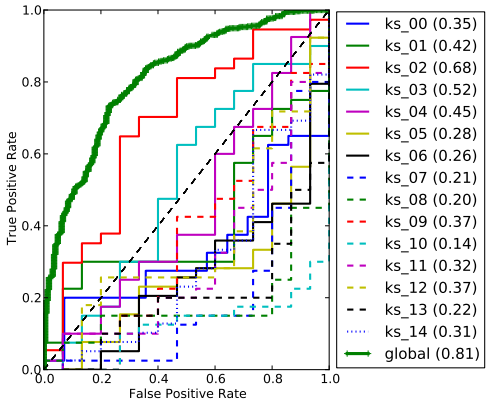
Ainsi, à partir de la saisie d'une unique phrase, nous pouvons détecter le genre d'un individu dans près de 75% des cas. Ces résultats sont inférieurs à la reconnaissance de 90% à partir de visage 2D, ou de dynamique de frappe statique, mais ils restent encourageants car nous n'avons pas cherché à optimiser les classifieurs. La reconnaissance de la catégorie d'âge est également proche de 75%. La figure 4.26 et la figure 4.27 représentent respectivement : la courbe ROC en

Tab. 4.7: EER de reconnaissance pour les deux scénarios

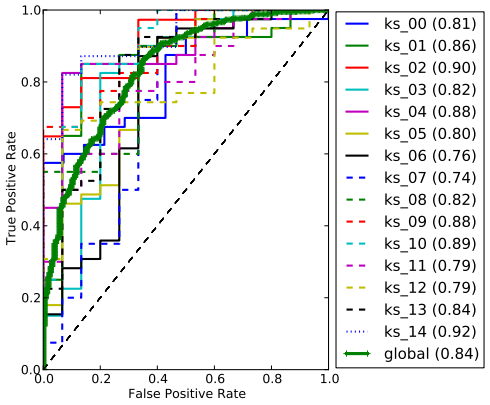
	Phrases égales	Phrases différentes
EER (%)/genre	25,54	26,80
EER (%)/âge	25,70	26,93

fonction du nombre de phrases utilisées pour reconnaitre le genre ou la catégorie d'âge, l'EER en fonction du nombre de phrases utilisées pour la reconnaissance (ainsi que la largeur de l'intervalle de confiance à 90% calculé avec une méthode de bootstrapping à 1000 ré-échantillonnages). Les résultats montrent que :

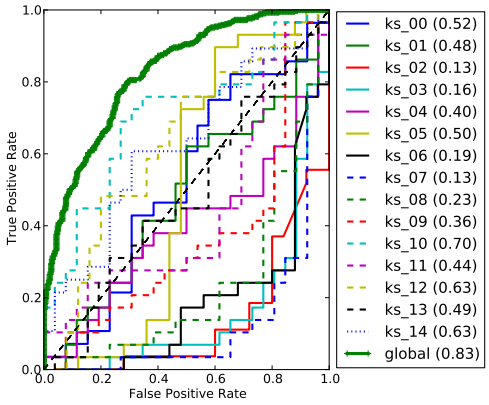
- Plus le nombre de phrases utilisées est important, plus les résultats sont fiables.
- Le gain de performance pour la reconnaissance du genre décroit exponentiellement avec le nombre de phrases et semble se stabiliser à partir de 7 phrases.
- Le gain de performance pour la reconnaissance de la catégorie d'âge décroit linéairement avec le nombre de phrases. Des phrases additionnelles pourraient augmenter encore les performances.



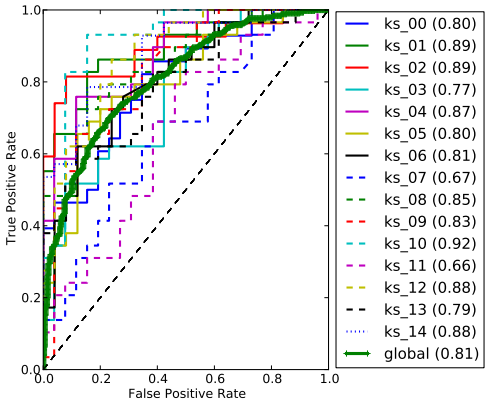
(a) Même phrase utilisée lors de l'apprentissage et de la reconnaissance du genre.



(b) Phrases différentes utilisées lors de l'apprentissage et de la reconnaissance du genre.



(c) Même phrase utilisée lors de l'apprentissage et de la reconnaissance de la classe d'âge.



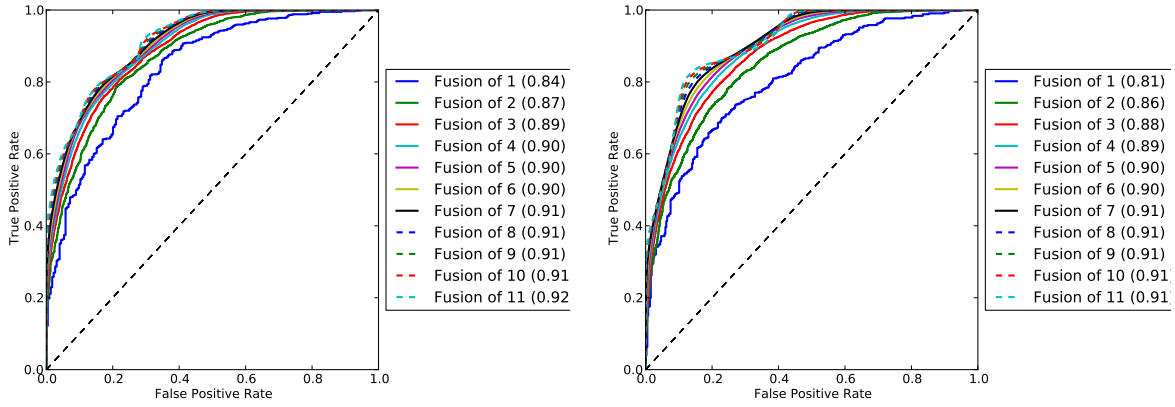
(d) Phrases différentes utilisées lors de l'apprentissage et de la reconnaissance de la classe d'âge.

FIG. 4.25: Courbe ROC de la méthode de reconnaissance en fonction de la phrase utilisée et du scénario appliqué

4.4.2.5. Discussion

Cette étude a montré la faisabilité d'un système de reconnaissance du genre et de reconnaissance de la catégorie d'âge basée sur la saisie de texte libre. L'étude comporte quelques limitations qu'il serait utile de prendre en compte dans des travaux futurs :

- La base contient peu d'utilisateurs (même si elle est conséquente par rapport à beaucoup d'autres bases de la littérature en DDF).
- Chaque utilisateur n'a saisi qu'une seule fois chaque phrase. Il est possible que plusieurs saisies apporteraient plus de variabilités (pouvant augmenter ou diminuer les performances de reconnaissance).
- Dans certains scénarios, le même utilisateur peut être présent à la fois dans la base d'apprentissage et dans la base de validation (avec des phrases différentes). Il pourrait être utile de comparer les résultats actuels à une étude similaire prenant soin de ne pas inclure de données de l'utilisateur testé dans l'ensemble d'apprentissage. Cette nouvelle étude serait



(a) Reconnaissance du genre

(b) Reconnaissance de la catégorie d'âge

FIG. 4.26: Courbe ROC en fonction du nombre de phrases utilisées pour la reconnaissance du genre ou de la catégorie d'âge

TAB. 4.8: Analyse de la différence de temps entre les catégories et vérification statistique à l'aide du Kruskal-Wallis de l'existence d'une différence

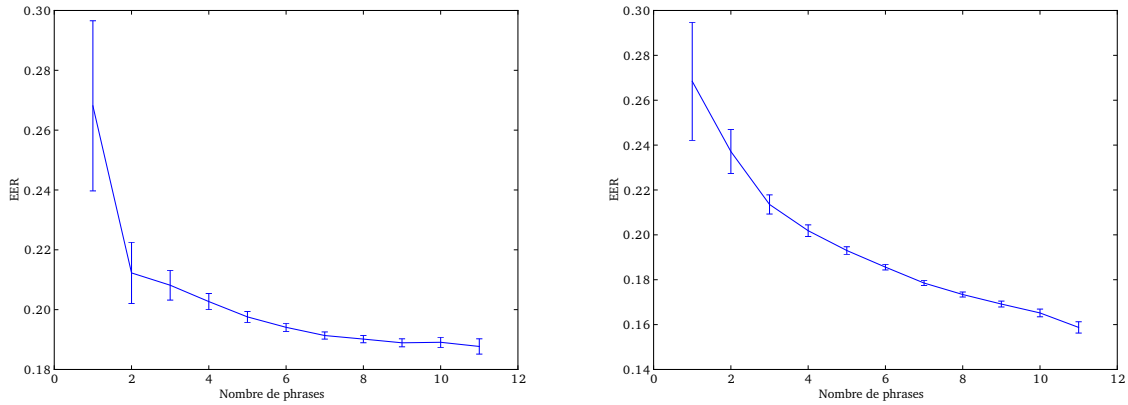
	$\mathbf{X}_{homme}$	$\mathbf{X}_{femme}$	$\mathbf{X}_{<=30}$	$\mathbf{X}_{>=30}$
Nombre d'exemples	593	224	431	386
Moyenne	131	270	184	153
P-value	1.75-08		2.99e-39	

un scénario supplémentaire, où la validation doit être réalisée sur des phrases inconnues, mais également avec des utilisateurs inconnus.

- Le texte est en espagnol. Les digraphes courants et présents dans tous les textes sont peut-être spécifiques à cette langue. Il est possible que les conclusions varient d'une langue à l'autre.
- Nous avons vu qu'il existe une différence de performance importante quel que soit le scénario en fonction de la phrase choisie. Il est intéressant d'analyser les phrases afin de comprendre pourquoi de telles différences sont constatées.

Tous ces points doivent être étudiés dans de futurs travaux. Nous avons créé les échantillons $\mathbf{X}_{homme}$, $\mathbf{X}_{femme}$, $\mathbf{X}_{<=30}$, $\mathbf{X}_{>=30}$ qui contiennent l'ensemble des valeurs de chaque vecteur correspondant aux catégories homme, femme et moins de 30 ans, plus de 30 ans. Le tableau 4.8 présente la comparaison des ensembles d'une même catégorie. On y voit que les temps des vecteurs sont majoritairement plus longs pour les femmes que pour les hommes, et qu'ils sont plus longs pour les personnes de moins de 30 ans que celles de plus de 30 ans. Ce sont ces différences de vitesse de frappe qui peuvent expliquer la possibilité d'effectuer ces catégorisations. La figure 4.28 présente la bande de confiance des vecteurs en fonction des catégories d'utilisateurs. Cependant, trois indices des données extraites ne sont pas présentés car ils ont un écart type supérieur à 500[3].

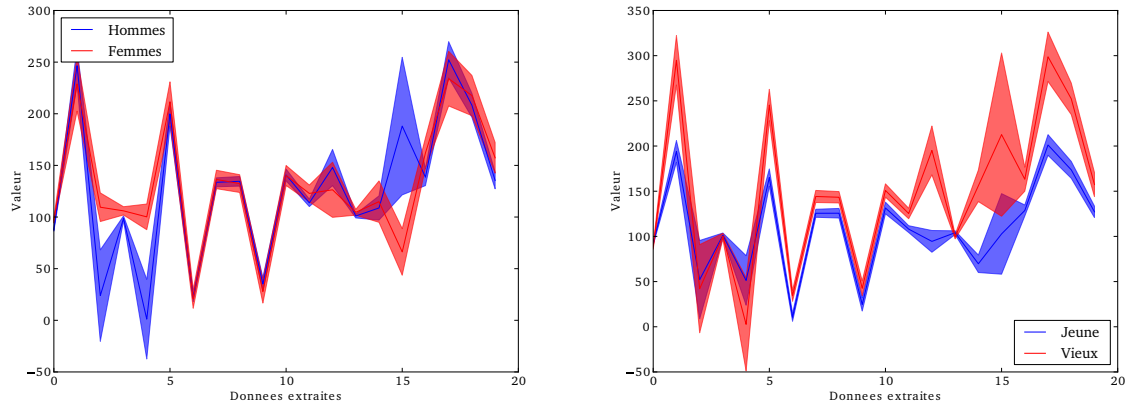
3. Il pourrait également être judicieux de retirer ces trois indices de toutes les expériences et de vérifier si de meilleures performances sont obtenues



(a) Reconnaissance du genre

(b) Reconnaissance de la catégorie d'âge

FIG. 4.27: EER (et son intervalle de confiance) en fonction du nombre de phrases utilisées pour la reconnaissance du genre



(a) Intervalle pour le genre

(b) Intervalle pour l'age

FIG. 4.28: Intervalle de confiance pour les différentes catégories d'utilisateurs (après suppression de trois indices ayant un écart type supérieur à 500)

4.4.3. Conclusion sur la biométrie douce

Nous avons montré, pour la première fois dans la littérature de DDF, qu'il est possible de détecter le genre d'un individu saisissant un texte fixe, avec un taux d'erreur acceptable. Nous n'avons testé le système que sur un seul mot de passe. Par la suite, d'autres chercheurs (sur une base considérablement plus petite) ont montré son utilité sur plusieurs mots de passe différents. Maintenant que nous savons que nous sommes capable de discriminer les hommes des femmes dans un système de DDF, il est possible d'utiliser cette information comme une biométrie douce afin de réduire le taux d'erreur de reconnaissance. L'information pourrait également être utile pour personnaliser des applications en fonction du genre de l'utilisateur, ou surveiller les réseaux sociaux. Nous avons montré qu'en utilisant cette information auxiliaire, nous pouvons augmenter les performances de reconnaissance d'un système de DDF de 20%.

Nous avons également montré qu'il est possible de reconnaitre le genre d'un individu, ainsi que sa catégorie d'âge, selon sa façon de taper un texte non connu du système. Dans cet autre cas d'utilisation, le texte est plus long qu'un mot de passe, étant donné qu'il s'agit d'une phrase. Mais, en contrepartie, le genre est reconnu à partir d'un texte quelconque : les cas d'utilisations d'un tel système se démultiplient étant donné qu'il n'est pas nécessaire de créer un modèle par texte potentiellement saisissable. Nous pouvons désormais envisager de détecter le genre d'un individu, saisissant un texte libre, à partir du moment où il a saisi suffisamment de texte (une phrase suffit pour obtenir un taux de bonne reconnaissance de près de 75%). Nous pensons qu'une telle application pourrait être utile dans des applications de détection de fraude ou de configuration automatique d'interface graphique ou de surveillance de réseaux sociaux. De futurs travaux pourraient porter sur l'amélioration des performances de reconnaissance et sur la reconnaissance d'autres caractéristiques douces afin de caractériser de façon automatique un individu saisissant du texte sur un clavier standard.

4.5. Conclusion de la multimodalité et biométrie douce

Ce chapitre a présenté nos travaux en fusion de données biométrique, en détection d'informations de biométrie douce pour la DDF, et en fusion de données et de données de biométrie douce.

Nos travaux de fusion ont porté uniquement sur la fusion des scores. Il serait intéressant de voir si de meilleures performances peuvent être obtenues avec d'autres types de fusion (attributs ou décision par exemple). Nous avons montré que des fonctions de fusion relativement simples, mais configurées convenablement à l'aide d'une méthode d'optimisation de type AG, permettent d'obtenir des performances similaires ou meilleures que la somme pondérée avec une approche coûteuse algorithmiquement. Nous avons également montré qu'il est possible de générer de meilleures fonctions de fusion à l'aide de la PG, en contrepartie d'une complexité de calcul et donc de temps d'exécution supérieurs.

De plus, nous avons montré le fonctionnement de la reconnaissance du genre à la fois sur texte fixé et sur texte libre. Nous avons aussi montré la possibilité de catégoriser des individus ayant tapé un texte quelconque selon leur catégorie d'âge. Nous avons pu améliorer les performances d'un système de reconnaissance par DDF en effectuant des fusions de scores et d'attributs avec des informations de genre. Ces nouveautés ouvrent la porte à de nouvelles applications de la DDF, dont notamment la lutte contre la pédophilie sur internet en permettant de détecter un homme se faisant passer pour une petite fille dans des applications de messagerie instantanée. On peut imaginer que des méthodes plus évoluées pourraient fonctionner sur du texte libre. Il serait intéressant, dans le futur, d'explorer de telles pistes. Comme nous avons vu que les différentes catégories d'utilisateurs peuvent avoir des temps de saisie différents, il pourrait être intéressant d'étudier des méthodes, plus simple que des classifieurs, basées sur des comparaisons de seuils.

Rappel des contributions de cette partie

- Le développement d'une méthode de calcul approximant l'EER afin d'accélérer le temps de calcul des algorithmes évolutionnaires utilisant l'EER comme fonction d'évaluation. La méthode a montré son intéret dans la configuration de systèmes multimodaux.
- L'illustration de l'intérêt d'utiliser deux biométries de performance moyenne (la DDF et la RF) avec une forte acceptation de la part des utilisateurs afin d'obtenir un système plus performant.

- L'intérêt de l'utilisation de la PG afin de déterminer, de façon totalement automatique, une fonction de fusion des scores.
- La détection automatique du genre d'un individu selon sa façon de taper au clavier un texte fixe, ainsi que l'utilisation de cette information pour augmenter les performances de reconnaissance par DDF.
- La détection automatique du genre et d'une classification selon son âge d'un individu selon sa façon de taper un texte quelconque.

Travaux de l'auteur sur ce thème de travail

Romain Giot, Mohamad El-Abed et Christophe Rosenberger : Fast learning for multibiometrics systems using genetic algorithms. *In The International Conference on High Performance Computing & Simulation (HPCS 2010)*, pages 1–8, Caen, France, juin 2010a. IEEE Computer Society. URL `http://hal.archives-ouvertes.fr/hal-00503096/en`.

Romain Giot, Baptiste Hemery et Christophe Rosenberger : Low cost and usable multimodal biometric system based on keystroke dynamics and 2d face recognition. *In IAPR International Conference on Pattern Recognition (ICPR)*, pages 1128–1131, Istanbul, Turkey, août 2010b. IAPR. URL `http://hal.archives-ouvertes.fr/hal-00503103/en`. Acceptance rate: 54/100.

Romain Giot et Christophe Rosenberger : A new soft biometric approach for keystroke dynamics based on gender recognition. *International Journal of Information Technology and Management (IJITM). Special Issue on : "Advances and Trends in Biometrics by Dr Lidong Wang*, 11(1/2):35–49, 2012a. Impact Factor : 0.727.

Romain Giot, Mohamad El-Abed et Christophe Rosenberger : Fast computation of the performance evaluation of biometric systems: application to multibiometric. *ELSEVIER International journal on Future Generation Computer Systems. HPCS 2010 special issue on Recent Developments in High Performance Computing and Security*, 2012a. URL `http://hal.archives-ouvertes.fr/hal-00674526/`. Impact Factor: 2.365.

Romain Giot, Christophe Rosenberger et Bernadette Dorizzi : Reconnaissance du genre par analyse de dynamique de frappe au clavier sur texte libre. *In* Mohammed Achemlal et Christophe Rosenberger, éditeurs : *7ème Conférence sur la Sécurité des Architectures Réseaux et Systèmes d'Information (SAR-SSI 2012)*, pages 111–118, Cabourg, France, mai 2012b. URL `http://hal.archives-ouvertes.fr/hal-00700618`. Acceptance Rate: 62/100.

Romain Giot et Christophe Rosenberger : Genetic programming for multibiometrics. *ELSEVIER International journal on Expert Systems With Applications*, 39(2):1837–1847, février 2012b. Impact factor: 1.924.

Romain Giot, Baptiste Hemery, Estelle Cherrier et Christophe Rosenberger : *La multibiométrie*, chapitre 9. Hermes Science Publications, septembre 2012c.

5. Mise à jour de la référence biométrique

Sommaire

Présentation

Qu'est-ce que le vieillissement du modèle biométrique? Comment réagit un système de DDF utilisant un système d'adaptation de la référence biométrique? Comment améliorer les performances d'un mécanisme de mise à jour du modèle biométrique? Ce sont les trois questions auxquelles nous répondons dans ce chapitre.

Mots clés :

vieillissement de la donnée biométrique, auto-apprentissage, mise à jour de la référence biométrique.

Contributions de ce chapitre

- La définition d'une procédure d'évaluation de systèmes de mise à jour de référence biométrique à l'aide d'une base de données contenant plusieurs sessions.
- L'analyse de l'efficacité de la mise à jour semi-supervisée pour la DDF.
- La proposition d'un schéma d'évaluation de la mise à jour de modèles biométriques incluant deux nouvelles métriques.
- La proposition d'un système de mise à jour hybride, plus efficace que les mécanismes d'auto-apprentissage.

Organisation du chapitre

Le chapitre est organisé de la façon suivante. Les premières sections présentent l'état de l'art de la mise à jour du modèle biométrique, en général non lié à la DDF. La section 5.1.2 présente ce qu'est un modèle biométrique. Il nous semble important de maitriser ce concept avant de chercher à le mettre à jour. La section 5.1.3 présente le fonctionnement des systèmes de mise à jour en insistant sur différents éléments : le critère de mise à jour, l'étude de la périodicité de la mise à jour, les modes de mise à jour et les stratégies de gestion du modèle. La section 5.1.4 présente l'évaluation des systèmes de mise à jour. Les sections suivantes présentent notre contribution. La section 5.2 montre l'intérêt d'utiliser un mécanisme d'auto-apprentissage en DDF, ainsi que la difficulté d'évaluer un tel système. La section 5.3 présente un nouveau système de mise à jour, plus performant que l'auto-apprentissage standard.

5.1. État de l'art de la mise à jour de la référence biométrique

5.1.1. Introduction

Tous les systèmes de reconnaissance de formes produisent des erreurs de reconnaissance. Comme les méthodes d'authentification et d'identification biométrique reposent sur de tels mécanismes, elles sont également sujettes à ces erreurs de reconnaissance. Celles-ci se traduisent par de faux rejets d'utilisateurs authentiques, et de fausses acceptations d'imposteurs. Les raisons de ces erreurs sont multiples, en voici une liste non exhaustive :

- La méthode de vérification utilisée n'est pas capable de modéliser parfaitement un utilisateur (ainsi qu'un imposteur), et donc de les différencier avec certitude.
- La modalité concernée est sujette à une très forte variabilité intra-classe qui peut être due au mécanisme bien connu de vieillissement du modèle biométrique (le modèle biométrique capturé à un instant t dérive progressivement du modèle réel à l'instant $t + n$). Drygajlo *et al.* [2009] montrent dans le cas de la reconnaissance faciale qu'il y a une corrélation non négligeable entre le score de comparaison d'une requête à un modèle et le délai écoulé entre la création de ce modèle et la capture de cette requête.
- Les conditions d'acquisition sont fortement différentes au cours du temps (luminosité, humidité, bruit, environnement mobile [Poh *et al.*, 2009], ...) et affectent les performances.
- Des capteurs différents lors de l'utilisation du système impliquent une variabilité importante. Ces variabilités peuvent être dues aux capteurs, ou à la différence d'interaction entre l'utilisateur et chacun d'entre eux.
- La coopération (ou plus exactement, le manque de coopération) de l'utilisateur est également une source génératrice de variabilité intra-classe.
- Les modalités comportementales sont sujettes à une forte variabilité intra-classe dont seul l'utilisateur est responsable :
  - Elles peuvent venir de l'acquisition d'un réflexe tout au long de l'utilisation du système : plus le temps passe, plus l'utilisateur maitrise le système et l'utilise différemment de sa première fois. On s'attend à une dérive de la donnée biométrique au cours du temps.
  - Mais, au contraire, la variabilité peut venir du fait que l'utilisateur effectue des actions qu'il est incapable de répéter exactement de la même façon. Dans ce cas, la variabilité sera toujours relativement importante et non corrélée avec le temps.

– D'autres facteurs de variabilité sont plus ou moins spécifiques aux modalités concernées, nous n'allons donc pas rentrer dans les détails. Il faut noter que les variabilités peuvent être permanentes ou temporaires.

Cette diminution des performances au cours du temps peut devenir fortement gênante pour l'utilisateur, ainsi que pour la sécurité du système. Heureusement, il existe de nombreux mécanismes pour pallier ces problèmes. Une méthode régulièrement employée en DDF est la présentation d'une nouvelle capture si la première a été rejetée [Araujo *et al.*, 2005]. Cependant, (a) ce type de méthodes n'aide pas à comparer objectivement les algorithmes utilisés, bien qu'elle semble être de plus en plus utilisée dans les papiers récents, et (b) cette solution peut être contraignante à utiliser. La *multimodalité* est également une solution intéressante (section 4.3). Elle permet d'utiliser plusieurs systèmes de vérification biométrique afin de diminuer le taux d'erreur global. Cette pratique reste efficace mais peut grandement complexifier la configuration des paramètres des différents systèmes, augmenter le coût de l'ensemble du mécanisme d'authentification, et donc être moins facile à utiliser. Nous pouvons utiliser différentes caractéristiques pour représenter le modèle biométrique. Celles-ci sont dépendantes de la modalité et du capteur utilisé. Certaines d'entre elles peuvent être redondantes, inutiles, à forte variabilité ou à faible variabilité. Rechercher et utiliser uniquement les données stables au cours du temps permet également d'améliorer les performances des systèmes de vérification biométrique [Houmani *et al.*, 2009]. Cette technique n'est pas forcément applicable pour les modalités ayant un vecteur caractéristique de faible dimension. Les erreurs peuvent être dues à une capture insuffisante de la variabilité intra-classe d'un utilisateur en raison du protocole d'enregistrement qui ne permet pas de capturer toute cette variabilité. Utiliser plusieurs sessions d'enregistrement permet de diminuer l'impact de la mauvaise représentativité de la référence biométrique mais peut être coûteux en temps et en argent. De plus, nous ne capturons pas les variabilités présentes sur une période plus courte que l'intervalle entre deux sessions, et il est difficile de savoir quand suffisamment de variabilité a été capturée pour stopper le processus d'enregistrement. Enfin, une solution qui n'a pas été intensivement testée est l'utilisation de mécanisme de mise à jour (*template update* ou *online learning* dans la littérature anglophone). Dans ces systèmes, le modèle biométrique d'un utilisateur est mis à jour au cours de l'utilisation du système. Cette mise à jour peut être faite de façon *supervisée* (ce qui revient à faire plusieurs enregistrements) ou *semi-supervisée* (c'est-à-dire de façon automatique), ce qui peut poser différents problèmes en raison de la possibilité d'intégrer des données d'imposteurs dans le modèle biométrique (et donc de diminuer les performances de reconnaissance au lieu de les améliorer). Il existe trois types majeurs de systèmes adaptatifs permettant d'augmenter les performances de reconnaissance :

– Les systèmes biométriques qui adaptent leurs paramètres en fonction de l'utilisateur ou de sa catégorie [Hocquet, 2007], ou en fonction de la qualité de la capture [Poh *et al.*, 2010a]. Il ne s'agit pas d'une modification temporelle.
– Les systèmes biométriques qui adaptent la frontière de décision d'un classifieur de scores de reconnaissance au cours du temps [Drygajlo *et al.*, 2009]. Cependant, il semble qu'une telle technique ne soit pas efficace pour la DDF [Giot *et al.*, 2012d].
– Les systèmes biométriques qui mettent à jour la référence biométrique (le modèle) de l'utilisateur au cours de son utilisation du système. C'est ce type de système qui nous intéresse dans ce chapitre.

Rattani *et al.* [2009] présentent une taxonomie de ces systèmes (figure 5.1). Nous pensons que ce schéma présente une réalité partielle des systèmes adaptatifs car il convient aux systèmes existants pour le visage et les empreintes, mais pas forcément pour d'autres modalités qui n'ont pas encore été étudiées. Comme les systèmes de mise à jour sont trop complexes pour être représentés sous la forme d'une taxonomie, nous pensons qu'il est préférable de représenter un graphe des variabilités des différents paramètres afin d'en visualiser la complexité. La figure 5.2

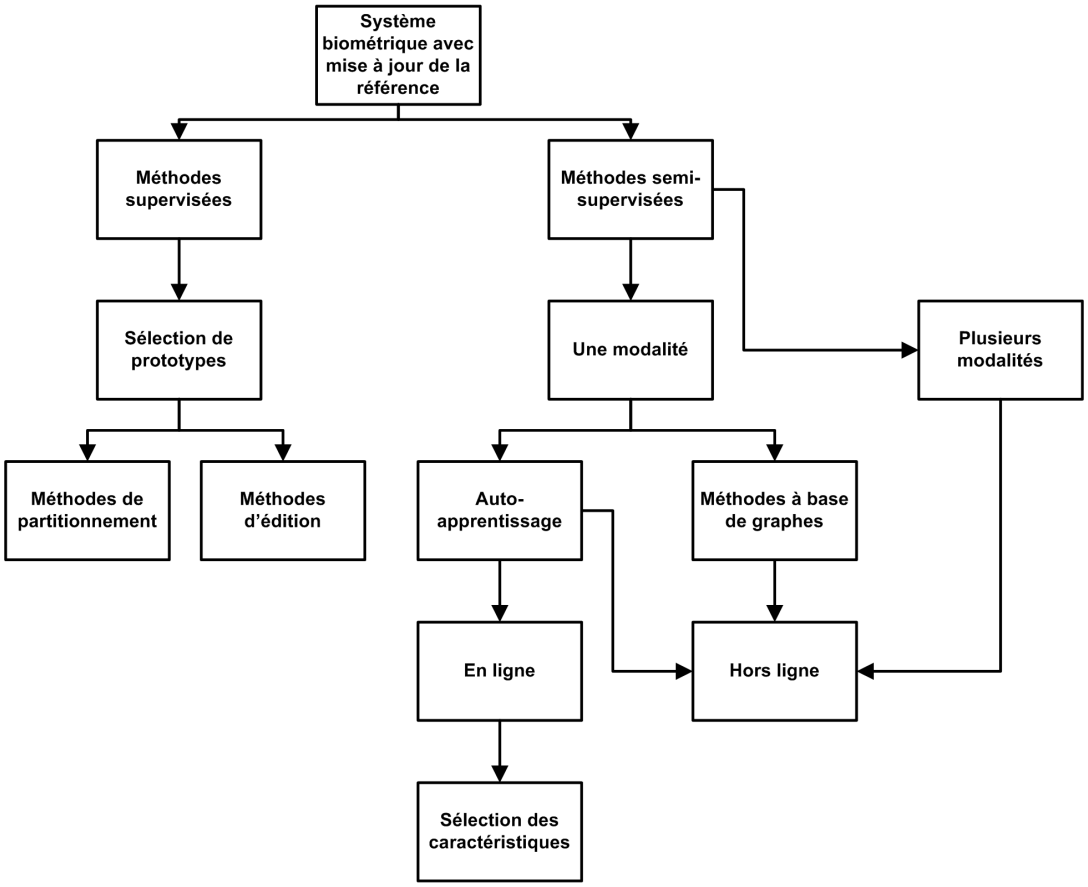


FIG. 5.1: Taxonomie des systèmes adaptatifs selon Rattani *et al.* [2009]

est une proposition pour représenter la variation de ces différents éléments que nous allons décrire tout au long de ce chapitre. Un état de l'art plus complet est également disponible [Giot, 2011].

5.1.2. Qu'est-ce qu'une référence biométrique ?

Il est nécessaire de définir le concept de référence biométrique avant de travailler sur sa mise à jour.

5.1.2.1. Définition de la référence biométrique

La référence biométrique est l'ensemble des exemples utilisés pour représenter un utilisateur. Lors d'une vérification, la capture ou *requête* (ou *query* dans la littérature anglophone) du demandeur est comparée à la référence biométrique de l'utilisateur qu'il clame être afin de déterminer si elle lui appartient. Dans la littérature, nous pouvons rencontrer les termes de *référence*, *référence biométrique* en français et *reference*, *model* et *template* (ou *master template*, si *template* désigne une requête) en anglais. La nature de la référence est complètement dépendante de la nature de la modalité concernée, ainsi que de la méthode de vérification employée. Cependant, nous pouvons dégager plusieurs variantes communes de création de référence qui peuvent être distinguées en trois catégories majeures :

– L'utilisation d'une référence à *un exemple*. Dans ce cas, une seule capture de bonne qualité est nécessaire lors de l'enregistrement. Cette unique capture sert de *référence* à l'utilisateur.

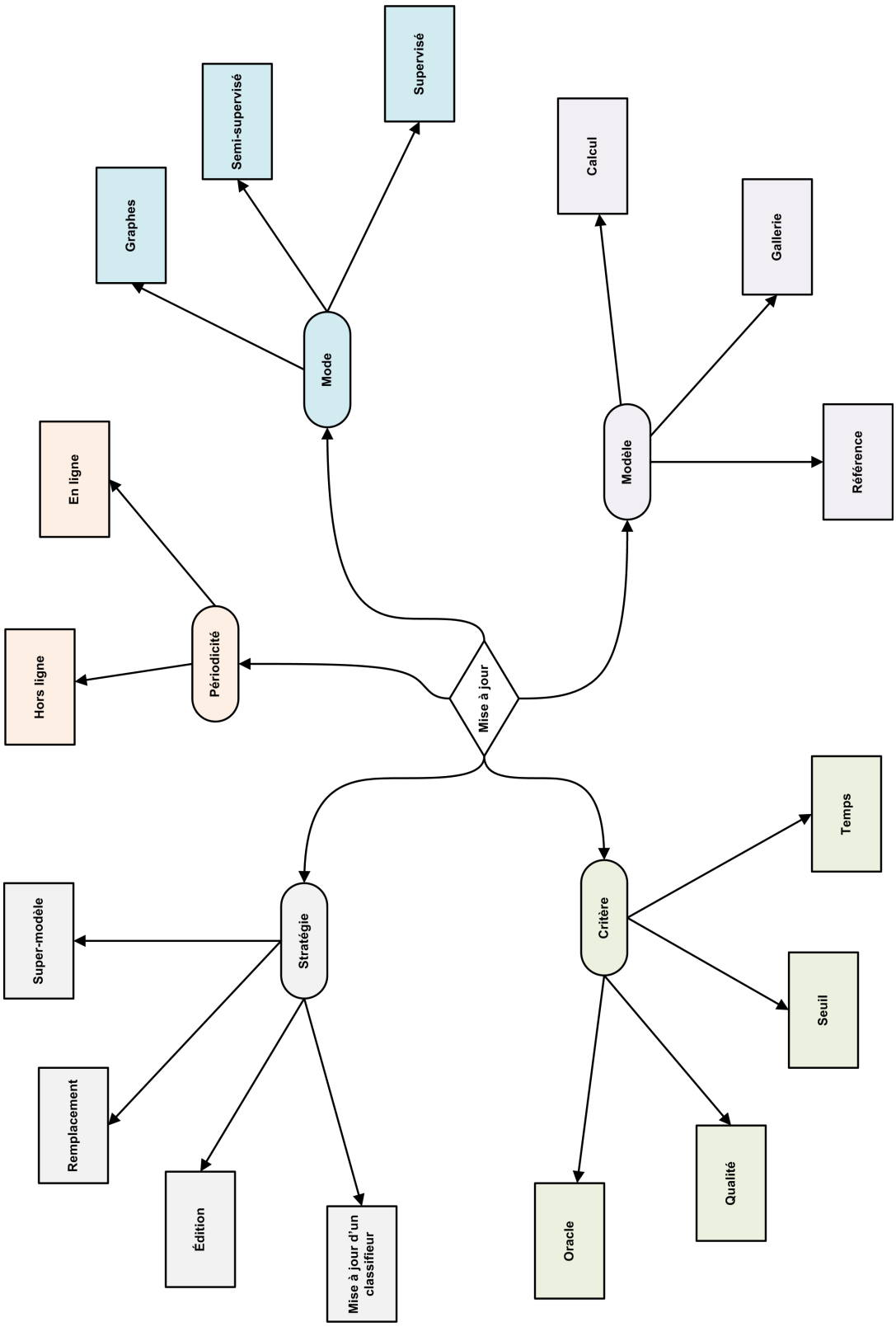


FIG. 5.2: Carte heuristique des différentes variations dans les paramètres de systèmes de mise à jour

- L'utilisation de *plusieurs exemples* dans la référence. Dans ce cas, plusieurs captures de bonne qualité sont nécessaires lors de l'enregistrement. Nous parlons de *galerie* pour désigner l'ensemble des captures de référence stockées.
- L'utilisation de *grappes de références* [Lumini et Nanni, 2006]. Il s'agit d'un cas particulier du type précédent. Les références (ou groupes de références) sont organisées sous forme hiérarchique. Chaque branche de l'arbre correspond à une contrainte particulière. Cette contrainte peut être explicite (profil, face, luminosité, ...) ou implicite (qualité de 0.1, qualité de 0.5, ...).

Cependant, cette distinction n'est pas encore totalement parfaite. Pour la plupart des méthodes d'authentification biométrique, ce ne sont pas les exemples qui sont stockés dans la référence, mais le résultat d'un calcul (*cf.*, transformée de fourrier, détection de points d'intérêt, information de texture, modèle statistique sur l'ensemble des éléments de la *galerie*...). Pour les systèmes à plusieurs références, le calcul peut être fait sur l'ensemble des références ; dans ce cas, la référence ne correspond plus à une *galerie*, mais à une *référence unique* calculée grâce aux différents éléments de la galerie (*c.-à-d.*, les modèles en dynamique de frappe au clavier sont majoritairement constitués des vecteurs moyenne et écart type).

5.1.2.2. Utilisation de la référence biométrique

Les méthodes de vérification de l'identité d'un individu sont fortement corrélées à la façon de stocker la référence. En effet, lorsque celle-ci est simple, la méthode de vérification calcule un score. Prenons par exemple le cas de la reconnaissance faciale à une référence utilisant une méthode de mise en correspondance de points d'intérêts ; le score est le nombre de correspondances entre les points d'intérêt de l'image requête et l'image de *référence* de la référence[1]. Cependant, lorsque celle-ci est multiple, la méthode de vérification a la possibilité de calculer un score pour chacun de ses éléments ; il est nécessaire de trouver un mécanisme d'agrégation de ces scores afin de n'en conserver qu'un seul (*minimum*, *moyenne*, ...). Pour l'exemple de la reconnaissance faciale à plusieurs exemples dans la galerie[2], il y a un score par référence. Le score maximal est retourné. Il est fort probable que les techniques de mise à jour soient également fortement dépendantes de la façon dont la référence doit être générée. Cependant, dans le cas avec plusieurs exemples, il peut être plus judicieux de calculer un modèle statistique basé sur celles-ci.

5.1.2.3. Discussion sur la référence biométrique

Comme nous avons pu le voir, la référence est un élément essentiel du système d'authentification biométrique, et il est nécessaire qu'elle soit de bonne qualité, afin d'avoir des performances acceptables. Comme il y a plusieurs façons de représenter la référence, la mise à jour devient difficilement générique.

5.1.3. Mise à jour du modèle biométrique

De nombreux paramètres du système de mise à jour du modèle biométrique sont différents en fonction des études et des algorithmes employés :

1. le choix du *critère de mise à jour du modèle biométrique* ;
2. la *périodicité* (en ligne ou hors ligne) de mise à jour qui est dépendante du point précédent ;
3. le *mode* (supervisé ou semi-supervisé) de fonctionnement du mécanisme de mise à jour ;

1. la référence est donc un nuage de points

2. la référence est une liste de nuages de points

4. la *stratégie* (l'algorithme utilisé) de mise à jour ;
5. la *technique* (la façon d'intégrer les modifications dans la référence) de mise à jour.

5.1.3.1. Critères de mise à jour

Avant de mettre à jour une référence, le système doit prendre la décision d'appliquer cette mise à jour. Différentes techniques de décision existent.

5.1.3.1.1. Acceptation de la requête Se baser uniquement sur l'acceptation de la capture *requête* par le système n'est pas une solution satisfaisante : une donnée d'imposteur serait trop facilement ajoutée au modèle de l'utilisateur. Il en résulterait une déviance non attendue de la nouvelle référence en comparaison de ce qu'elle devrait être (l'utilisateur sera moins facilement accepté, tandis que les imposteurs pourraient l'être plus facilement.). C'est probablement pour cette raison que nous ne rencontrons pas ce critère dans la littérature.

5.1.3.1.2. Double seuillage Le double seuillage utilise deux seuils (figure 5.3) : (i) le *premier seuil* détermine si une requête est de type client ou imposteur ; (ii) tandis que *le deuxième* décide si la requête correspond suffisamment à un client pour pouvoir être utilisée par le système de mise à jour. Ce double seuillage est souvent noté de la façon suivante dans la littérature : la requête a une *forte probabilité* d'être une donnée cliente. Cependant, la majorité du temps, il n'y a guère plus d'information sur la configuration du seuil de mise à jour. On peut s'attendre à ce que celui-ci soit configuré de telle façon à obtenir un taux de fausse acceptation (ou faux rejet) d'une certaine valeur sur une base de test [Rattani, 2010]. Dans le cas de la configuration d'un seuil automatique (*cf.*, de telle façon à avoir un TFA de 1%, par exemple), il est nécessaire que le système biométrique dispose de données d'imposteurs. Il est trivial d'en obtenir dans le cas d'une modalité morphologique, mais c'est loin d'être le cas pour les biométries demandant aux utilisateurs d'effectuer une action particulière (*cf.*, saisie d'un mot de passe, pour la dynamique de frappe au clavier).

5.1.3.1.3. Indice de qualité Une variante de la technique du double seuillage est d'utiliser un indice de qualité de la requête (figure 5.4). Si elle est de meilleure qualité que celle de la référence, elle le remplace. Noval et López [2008] présentent l'utilisation d'un tel indice de qualité : si la requête est suffisamment proche de celle de l'utilisateur, et, suffisamment éloignée de celles des imposteurs, alors elle est utilisée dans la nouvelle référence. L'intérêt d'un tel mécanisme est de pouvoir remplacer les captures de faible qualité obtenues durant l'enregistrement de l'utilisateur.

Poh *et al.* [2010a] utilisent la notion d'indice de qualité pour améliorer les performances du système. Dans ce cas, la qualité représente plus une qualité de conditions d'acquisition qui est due aux différentes modifications de l'environnement, qu'à une qualité de la donnée biométrique qui serait mauvaise car les conditions ne sont pas idéales. Dans leur papier, les auteurs ont trois conditions différentes d'acquisition, ce qui implique qu'ils ont trois types de qualité dans leurs données. Cette mesure permet d'adapter les paramètres à l'environnement d'acquisition (et non pas à l'évolution de l'utilisateur comme dans toutes les autres études de mise à jour).

5.1.3.1.4. La prédiction En analysant les scores de comparaisons authentiques produits au fil de l'utilisation du système biométrique, il est possible de détecter un vieillissement du modèle biométrique. Carls [2009] présente un framework appelé CTARP qui analyse les scores de vérification des données clientes et leur déviation au cours du temps (en fonction de la prédiction

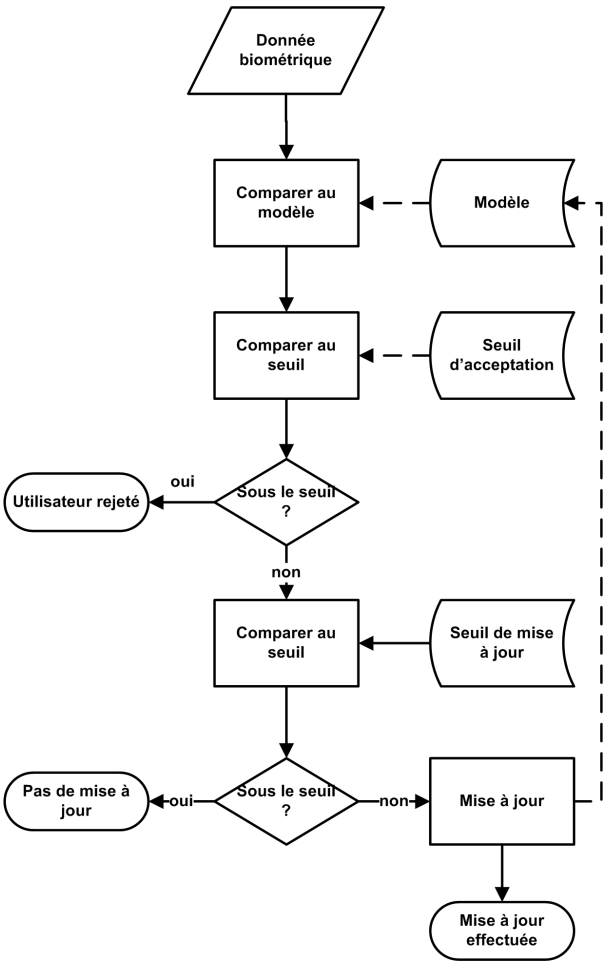


Fig. 5.3: Décision d'acceptation avec un double seuil

du score). À l'aide de ces informations (scores précédents, score prédit, et score mesuré), le système est capable de détecter le moment où il est utile de ré-enregistrer la personne [Carls *et al.*, 2008]. Un gain de performance de 20% a été observé sur des bases de reconnaissance faciale sur une longue période. Cependant, il semble que l'étude n'ait pris en compte que l'analyse de faux rejets, et pas de fausses acceptations ; elle ne donne aucune indication sur l'évolution du taux de fausse acceptation en fonction de la mise à jour (qui fonctionne de manière supervisée), ni le comportement du framework dans un environnement non supervisé.

5.1.3.1.5. L'appel à un oracle La sélection peut également être faite manuellement par un opérateur. Les critères de sélection de l'opérateur peuvent être implicites ou explicites. Nous ne nous attarderons pas non plus sur ce point, car il ne correspond pas à nos attentes (automatisation totale du processus de mise à jour souhaitée).

5.1.3.1.6. L'erreur de vérification Vandana [2007] propose un algorithme itératif pour mettre à jour l'espace propre de chaque utilisateur lorsque qu'une capture est rejetée de façon anormale. Cependant, contrairement aux autres papiers de la littérature, ce n'est pas l'administrateur qui est chargé d'initier cette mise à jour, mais les utilisateurs eux-mêmes en fonction de leur retour. Ce choix n'est pas justifié par l'auteur, mais on pourrait penser qu'il est parti du principe qu'une capture acceptée signifie qu'elle est suffisamment proche du modèle, et, que cela n'apporterait pas de variabilité supplémentaire de l'y intégrer. Tandis qu'une capture rejetée est forcément

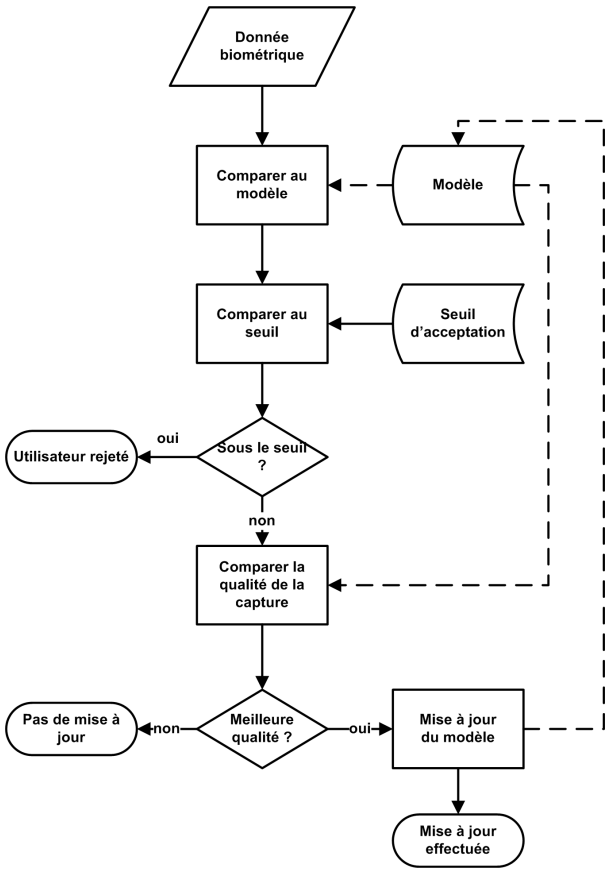


FIG. 5.4: Décision de mise à jour en utilisant une information de qualité

sujette à apporter plus de variabilité dans le modèle. Cependant, cette variabilité peut être la cause d'une erreur de capture ou de l'utilisation du système par un imposteur, et son inclusion d'office dans le modèle peut être problématique.

5.1.3.1.7. Le système de mise à jour La sélection de la donnée biométrique à intégrer dans la nouvelle référence peut être automatiquement faite par le système de mise à jour. Nous partons du principe que toutes les données capturées sont sujettes à être utilisées dans la mise à jour. Le filtrage est fait automatiquement par le système de mise à jour. Un tel exemple est présenté dans la section 5.1.3.3.2.

5.1.3.1.8. Le temps Afin de ne pas capturer une variation intra-classe locale d'un point de vue temporel (et donc qui disparaitra au bout d'un temps relativement court), il est possible de se baser sur l'utilisation d'un intervalle. Ainsi, la mise à jour peut être effectuée à intervalle régulier de façon automatique. La durée de l'intervalle est spécifique à la modalité, et plus cette modalité est sujette à variation, plus cet intervalle doit être court. Kekre et Bharadi [2009] indiquent qu'un intervalle d'un mois pour un système de reconnaissance faciale semble être couramment adopté.

5.1.3.1.9. Un système hybride On peut tout à fait imaginer que le critère de décision de mise à jour soit dépendant de plusieurs des critères présentés précédemment. Le système utilise une modalité *indépendante* du temps, n'ayant pas de variation intra-classe au cours du temps, et, une modalité *dépendante* ayant une variation intra-classe non négligeable au cours du temps.

La modalité indépendante peut être vue comme un oracle. La mise à jour n'est effectuée que si l'utilisateur est rejeté par le système multimodal alors que l'oracle l'accepte et que:

- soit le nombre de rejets par la modalité dépendante (tandis que la modalité indépendante accepte l'utilisateur) atteint un seuil;
- soit la durée d'utilisation du système sans utiliser la mise à jour a atteint son délai.

Dans tous les cas, la mise à jour ne peut être appliquée que si la modalité dépendante a fait une erreur de reconnaissance.

5.1.3.1.10. Discussion Ces méthodes de décision de mise à jour ont toutes leurs avantages et inconvénients. Pour l'utilisation du double seuillage, seules les données biométriques à forte probabilité d'appartenance à l'utilisateur sont utilisées. Cela implique fortement que peu de variabilité soit capturée (étant donné que l'on ne sélectionne que des données que l'on sait reconnaitre). Nous évitons donc au maximum d'intégrer des données d'imposteurs, mais nous limitons également l'amélioration des performances en restant dans un minimum local de la fonction de mise à jour. Utiliser un indice de qualité pour faire la mise à jour est intéressant lorsque l'on contrôle au maximum les conditions d'acquisition de la donnée biométrique. L'utilisation d'un oracle implique un coût supplémentaire (faire appel à une personne) et n'est pas nécessairement automatisable. L'utilisation d'un mécanisme d'acceptation interne au processus de mise à jour du système semble être une façon intéressante, mais ne peut pas être utilisable dans tous les cas.

5.1.3.2. Périodicité

Nous pouvons distinguer deux façons majeures d'appliquer les systèmes de mise à jour dans la littérature : la mise à jour *en ligne* ou *temps réel*, et la mise à jour *hors ligne*.

5.1.3.2.1. Mise à jour hors ligne Le mécanisme de mise à jour hors ligne (figure 5.5) est exécuté par lot ou mode « batch » lorsque le système dispose de suffisamment de nouveaux exemples (*c.-à-d.*, les données capturées lors de l'utilisation du système biométrique). L'ordre d'apparition des captures n'importe pas nécessairement, car, les données sont traitées dans leur ensemble sans nécessairement prendre en compte leur chronologie. Cette mise à jour peut être *semi-supervisée* (*c.-à-d.*, utiliser les étiquettes calculées par le classifieur), ou *supervisée* si les étiquettes des captures sont fournies par un oracle (l'administrateur du système, par exemple). Un problème ouvert est la fréquence de la mise à jour. Quelle est la meilleure stratégie à adopter ? Attendre d'avoir collecté suffisamment de données ? Ou attendre l'expiration d'un certain délai ? Dans la plupart des études, l'appel n'est fait qu'une seule fois quand suffisamment de données sont collectées : les études partagent leur base de données en trois ensembles : un pour l'apprentissage initial, un pour l'application de la mise à jour, et un pour valider les nouvelles performances (après mise à jour). La quasi-totalité des études de l'état de l'art sont en mode hors-ligne, ce n'est pas le mode qui nous intéresse car nous pensons qu'au vu des faibles performances de méthode de DDF, il est nécessaire d'effectuer la mise à jour le plus tôt possible.

5.1.3.2.2. Mise à jour en ligne Le mécanisme de mise à jour en ligne (figure 5.6) est exécuté dès la fin de la vérification, si la capture *requête* est considérée comme étant celle de l'utilisateur attendu et que le système désire la prendre en compte pour la mise à jour (voir 5.1.3.1). Il s'agit donc d'une mise à jour en temps réel, qui se déroule de façon itérative, *requête* par *requête*. Il est connu qu'avec ce type de mécanisme, l'ordre d'apparition des captures influe sur la qualité de la mise à jour du modèle [Rattani, 2010]. Un tel mécanisme est particulièrement adapté aux environnements ayant de faibles capacités de stockage ou de calculs, tels les appareils mobiles [Poh *et al.*, 2009].

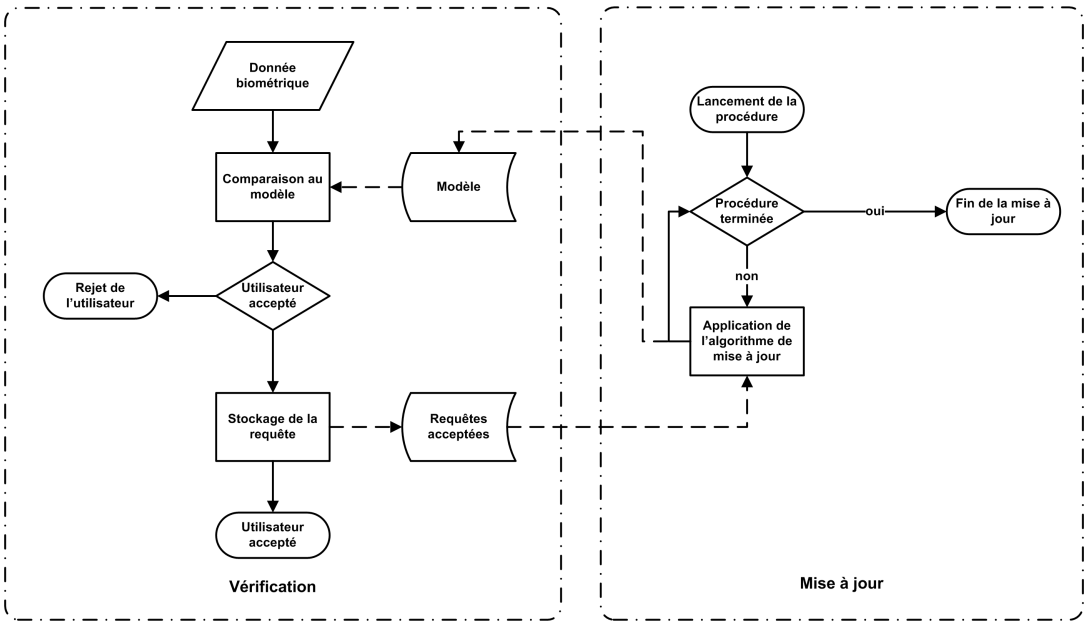


FIG. 5.5: Schéma de la mise à jour hors ligne de la référence biométrique

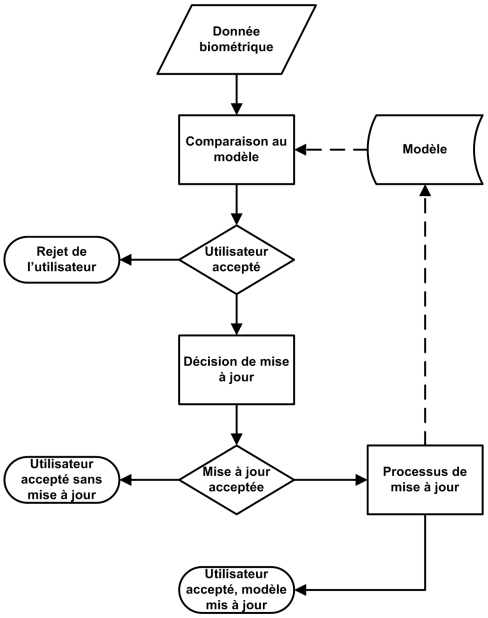


FIG. 5.6: Schéma de la mise à jour en ligne de la référence biométrique

5.1.3.2.3. Discussion La façon de gérer les mises à jour est différente en fonction de la périodicité choisie. La mise à jour en ligne ne nécessite pas de stocker l'ensemble des requêtes acceptées étant donné qu'elle effectue la mise à jour dès son acceptation. Le coût d'utilisation mémoire est donc faible. En contrepartie, il est nécessaire d'effectuer le calcul dès l'acceptation de la requête, ce qui peut être relativement coûteux en temps de calcul. La mise à jour hors ligne nécessite le stockage des requêtes acceptées au fil de l'utilisation du système. Elle est donc plus coûteuse en consommation mémoire. Cependant, la vérification est moins coûteuse en temps de calcul, car elle ne nécessite pas d'effectuer la mise à jour instantanément. Les papiers de Fabio Roli [Roli *et al.*, 2007, 2008] (en reconnaissance d'empreintes digitales ou reconnaissance de visage) insistent sur le fait que le mode en ligne est sensible à l'ordre de présentation des données et qu'il faut que

les calculs soient effectués plusieurs fois en utilisant un ordre de présentation aléatoire. Il est fort probable que cette assertion soit vraie pour de telles biométries (bien que Drygajlo *et al.* semblent montrer le contraire en analysant les scores de comparaisons produits au cours du temps par un mécanisme de reconnaissance faciale [Drygajlo *et al.*, 2009]), cependant, nous pensons que dans le cas des biométries comportementales (notamment la dynamique de frappe au clavier), l'ordre original de capture doit être impérativement conservé afin de prendre en compte l'aspect apprentissage cognitif de l'utilisateur. Il nous semble utile d'explorer ce point [3].

5.1.3.3. Modes de mise à jour

Comme nous l'avons vu précédemment, il existe deux familles principales de mécanisme de mise à jour : la *mise à jour supervisée* qui nécessite l'aide d'un opérateur humain, et la *mise à jour semi-supervisée* qui est totalement automatique, et, se sert des informations issues des classifieurs.

5.1.3.3.1. Mise à jour supervisée Nous n'allons pas nous attarder sur la gestion *supervisée* de la mise à jour, car elle a été suffisamment étudiée dans la littérature et ne présente pas de défi technique particulier. L' *opérateur* qui fournit l'étiquette des captures a pu se les procurer de deux façons distinctes :

- en gérant *plusieurs sessions d'enregistrement*. Cette étape est longue, fastidieuse et coûteuse. Elle implique de demander aux différents utilisateurs de participer à plusieurs sessions d'enregistrement. L'administrateur supervise ces enregistrements pour empêcher les erreurs (*c.-à-d.*, enregistrement d'un individu β au lieu d'un individu λ) ;
- en *étiquetant manuellement* les données ayant été capturées tout au long de l'utilisation de l'application. Dans ce cas, l'administrateur doit être un expert du domaine, et cette méthodologie n'est peut-être pas applicable pour toutes les biométries. Cette pratique est également coûteuse en temps. Sukthankar et Stockton [2005] proposent à un administrateur de modifier manuellement les étiquettes des visages reconnus en cas d'erreur du système de reconnaissance faciale d'un système de portier électronique. Ces nouvelles données sont utilisées dans le système de reconnaissance faciale afin d'en augmenter les performances. La figure 5.7 présente le fonctionnement d'un système de mise à jour supervisé.

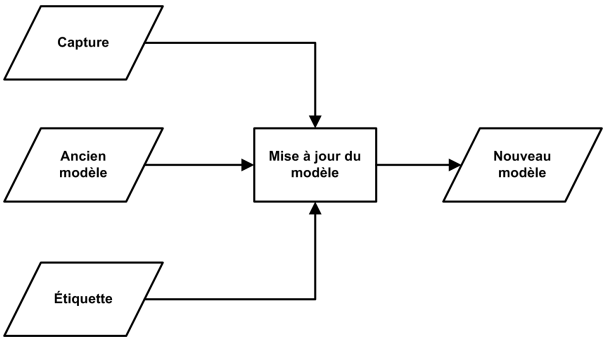


FIG. 5.7: Principe d'une mise à jour supervisée

3. Bien que nous ne présentons pas ces travaux dans le manuscrit, nous montrons dans [Giot *et al.*, 2012b] que la corrélation entre les scores et le nombre de saisies du mot de passe effectuées avant de saisir celui-ci est bien plus importante lorsque la chronologie est respectée

5.1.3.3.2. Mise à jour semi-supervisée Plusieurs techniques semi-supervisées sont référencées dans la littérature. Il n'y a pas d'étiquetage manuel, l'étiquette d'une capture est définie par l'algorithme de reconnaissance. Si l'étiquette correspond à celle de l'utilisateur, le système peut appliquer la mise à jour. Les méthodes d'apprentissage semi-supervisé utilisent ces données non étiquetées pour compléter l'apprentissage supervisé (figure 5.7). Il faut noter que ces méthodes ont majoritairement été utilisées par lots. Plusieurs méthodes d'apprentissage semi-supervisé existent. Nous allons présenter l'auto-apprentissage, le co-apprentissage, et des méthodes basées sur la propagation d'étiquettes dans un graphe.

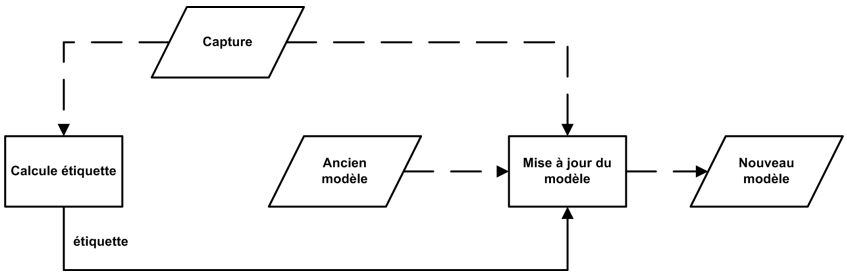


FIG. 5.8: Principe d'une mise à jour semi-supervisée

Auto-apprentissage les techniques d'*auto-apprentissage* (ou *self-training* dans la littérature anglophone) ont été beaucoup étudiées [Marcialis *et al.*, 2008; Rattani, 2010; Roli et Marcialis, 2006; Roli *et al.*, 2008]. Le terme *auto-apprentissage* est utilisé car la méthode se met à jour incrémentalement elle-même en utilisant sa propre connaissance. Dans le cas de la mise à jour, le fonctionnement hors ligne est le suivant:

1. Lors de l'enregistrement, la référence est générée en entrainant le classifieur avec les données étiquetées (D_L) récupérées de façon supervisée.
2. Un ensemble de données non étiquetées (D_U) est collecté tout au long de l'utilisation du système biométrique. Lorsque suffisamment de données sont collectées, nous pouvons passer à l'étape suivante. La difficulté réside à savoir ce qu'est « suffisamment ».
3. Le classifieur est utilisé pour étiqueter les données incomplètes (D_U).
4. Les données étiquetées avec un fort degré de confiance sont ajoutées aux données d'apprentissage (D_L). La notion « fort degré de confiance » est également subjective.
5. Le classifieur est ré-entrainé sur l'ensemble des données étiquetées (D_L) et la procédure est répétée à l'étape 3 jusqu'à satisfaire un critère d'arrêt. Ce critère d'arrêt peut être un nombre d'itérations de l'algorithme.

L'auto-apprentissage a tout de même quelques limites. Il est toujours possible d'ajouter des données d'imposteurs si la comparaison de celles-ci à la référence fait penser qu'il s'agit d'une donnée cliente à fort degré de confiance. Augmenter ce degré réduit la probabilité d'inclure des données erronées, mais seules les données clientes fortement probables sont ajoutées à la référence. Comme ce sont donc des données fortement ressemblantes à celles déjà connues, elles n'apportent pas forcément beaucoup d'information de variabilité supplémentaire. Ainsi, la configuration de ce seuil de mise à jour dépend grandement de la performance du système. Cette technique a pu être appliquée avec succès à des systèmes de reconnaissance faciale utilisant une Analyse en Composantes Principales (ACP) et de reconnaissance d'empreintes digitales utilisant la méthode de mesure *strings*. El Gayar *et al.* [2006] proposent des évolutions de l'auto-apprentissage adaptées à la multimodalité.

Co-apprentissage Le *co-apprentisage* (ou *co-training* dans la littérature anglophone) est une version de l'auto-apprentissage adaptée à l'utilisation d'au moins deux classifieurs qui vont s'entraider pour s'améliorer mutuellement. Une limite de l'auto-apprentissage est la possibilité de n'inclure que des exemples relativement ressemblants à ceux qui sont connus. Leur efficacité est donc potentiellement limitée. Le co-apprentissage permet de faire abstraction de ce problème en utilisant un second classifieur qui permet l'inclusion de données à forte variabilité qui n'auraient pas été retenues par le premier classifieur [Didaci *et al.*, 2009; Rattani, 2010; Rattani *et al.*, 2008b,c; Roli *et al.*, 2007, 2008]. Plus les modalités utilisées sont indépendantes les unes des autres, plus les performances du co-apprentissage sont bonnes. Lorsque l'environnement est fortement contrôlé, les performances sont semblables aux techniques d'auto-apprentissage, mais elles sont nettement meilleures lorsque l'environnement n'est pas contrôlé [Rattani *et al.*, 2008c]. Les données biométriques se présentent donc sous forme de couples de données : la donnée de la modalité 1, ainsi que la donnée de la modalité 2. Le fonctionnement de la mise à jour est le suivant :

1. Lors de l'enregistrement, la référence de chaque modalité est calculée à l'aide de l'ensemble D_L.
2. Un ensemble de données non étiquetées (D_U) est collecté tout au long de l'utilisation du système biométrique. Lorsque suffisamment de couples sont collectés, nous pouvons passer à l'étape suivante. La difficulté réside à savoir ce qu'est « suffisamment ».
3. Chacun des deux classifieurs est utilisé pour étiqueter les données incomplètes (D_U)
4. Les données étiquetées avec un fort degré de confiance (par un des classifieurs) sont ajoutées aux données d'apprentissage (D_L). La notion « fort degré de confiance » est également subjective ; en général, il s'agit d'un second seuil.
5. Les classifieurs sont ré-entrainés sur l'ensemble des données étiquetées (D_L) et la procédure est répétée à l'étape 3 jusqu'à satisfaire un critère d'arrêt. Ce critère d'arrêt peut être un nombre d'itérations de l'algorithme.

Rattani *et al.* [2008c] utilisent une légère variante lors de l'ajout des nouvelles données dans le système (dans le cas d'une reconnaissance d'empreintes digitales et de visage). Le système de co-apprentissage est exécuté deux fois, avec un classifieur primaire et un classifieur secondaire (les deux classifieurs intervertissent donc leur rôle au deuxième lancement). Si la requête est acceptée par les deux classifieurs, elle est considérée comme encodant peu de variabilité : elle est donc fusionnée avec le modèle le plus proche (*cf.*, techniques de super-modèle, section 5.1.3.4.1) du classifieur secondaire. Si la requête n'est acceptée que par le classifieur primaire, elle est considérée comme encodant une forte variabilité, et, est ajoutée à la galerie du classifieur secondaire. Comme le co-apprentissage permet d'intégrer des données avec plus de variabilités, les nouveaux modèles encodent donc plus de variabilité intra-classe. Rattani *et al.* [2008b] montrent que les performances individuelles de chaque systèmes de reconnaissance après mise à jour sont meilleures avec le co-apprentissage qu'avec l'auto-apprentissage, notamment car plus de données non étiquetées sont intégrées dans la référence.

El Gayar *et al.* [2006] proposent un autre mécanisme utilisant plusieurs classifieurs. Dans ce cas, l'auto-apprentissage de chaque classifieur n'est plus indépendant : les classifieurs sont mis à jour lorsqu'ils sont majoritairement d'accord sur l'étiquette à affecter aux exemples non étiquettés. Mais, contrairement à la méthode de co-apprentissage précédente, on peut s'attendre à capturer moins de variabilité en raison du consensus (le vote majoritaire) utilisé. Curieusement, ces méthodes ont été principalement utilisées dans des mécanismes hors-ligne, alors qu'elles pourraient également être adaptées à des systèmes en ligne.

Les méthodes à base de graphes L'utilisation de méthodes à base de graphes peut être une solution au manque de capture de variabilité dans le cas de l'auto-apprentissage [Rattani *et al.*,

2008a; Rattani, 2010]. Comme ces techniques nécessitent de disposer d'un certain nombre de données non étiquetées, la plupart des études travaillent hors ligne même si une méthode en ligne a été récemment proposée [Kveton *et al.*, 2010]. Contrairement aux méthodes semi-supervisées classiques, les techniques à base de graphes sont capable de capturer énormément de variabilité intra-classe. L'utilisation de ces nouvelles données va donc permettre d'améliorer la mise à jour et les modèles, en encodant plus de variabilité. Le principe des modèles semi-supervisés à base de graphes est de propager les étiquettes dans un graphe depuis les nœuds étiquetés vers les nœuds non étiquetés. Les différents mécanismes prennent en compte la structure de l'ensemble des données, dans le graphe généré à l'aide des données biométriques reliées en fonction de leur similarité. Chaque nœud représente donc une donnée biométrique, et, chaque arête relie deux nœuds ayant une certaine similarité entre eux. Celle-ci peut être pondérée en fonction de la valeur de la similarité.

La technique du flot max/coupe min Blum et Chawla [2001] présentent un système de classification des données non étiquetées en utilisant l'algorithme *flot-max/coupe-min* (annexe C). Le but est de partitionner le graphe en deux zones : une zone contenant les données authentiques, et, une zone contenant les données d'imposture. La partition appartenant à la classe authentique est utilisée comme nouveaux exemples de l'utilisateur. Les *nœuds* du graphe sont constitués des données biométriques (qu'elles soient étiquetées ou non). Les *arcs* du graphe sont pondérés par le score obtenu en comparant deux nœuds (la méthode fonctionne donc avec les systèmes capables de comparer deux captures entre elles, ce qui n'est pas le cas de la majorité des méthodes de reconnaissance par DDF qui ne manipulent pas les données biométriques mais des modèles statistiques). Le problème de mise à jour devient un problème d'optimisation de l'étiquetage. Utiliser l'algorithme de *flot-max/coupe-min* convient bien à la résolution de ce problème, car l'approche est [Blum et Chawla, 2001] :

- basée sur un binarisation des données en deux classes : les données *authentiques*, et, les données *imposteurs* ;
- l'optimisation est faite en temps polynomial ;
- la méthode est revendiquée comme étant robuste au bruit, et elle donne de bonnes performances, même avec peu de données étiquetées (ce qui est réaliste dans le cas de la mise à jour).

Champ gaussien et fonction harmonique Dans [Zhu, 2005; Zhu *et al.*, 2003], Zhu montre qu'il est possible d'étiqueter les nœuds d'un graphe, de façon semi-automatique, lorque seuls quelques nœuds sont étiquetés. Le problème est formulé sous la forme d'un champ aléatoire gaussien sur le graphe, la moyenne du champ étant caractérisée par des fonctions harmoniques qui sont obtenues à l'aide de méthodes matricielles ou de propagation de croyance. L'implémentation originale est disponible ici : `http://pages.cs.wisc.edu/~jerryzhu/pub/harmonic_function.m`. Le principe a été validé dans [Balcan *et al.*, 2005], sur un système d'identification dans un contexte de vidéo surveillance (seules les premières données sont étiquetées, les images sont de faibles qualités et rares). Les données sont constituées de plusieurs flux vidéos (0,5 images par seconde) de 10 personnes sur 3 mois. Seules les images avec une seule personne présente dans la scène sont conservées. Pour chaque image, le fond est soustrait avec un algorithme relativement simple, afin de récupérer la zone de l'image contenant l'individu (après avoir, au préalable, appliqué différents opérateurs morphologiques). Ainsi, chaque image est caractérisée par trois descripteurs différents :

- le moment d'acquisition (date et heure) ;
- un histogramme du premier plan ;
- le visage (face ou profil) extrait dans le premier plan (35% des images n'ont pas détecté de visage).

À l'aide de ces informations, il est possible de générer trois différents types d'arc dans le graphe :

- *le temps*: comme il y a une seule personne par image, deux images espacées de quelques secondes correspondent probablement à la même personne. Les images ayant une différence de temps inférieure à un seuil t_1 (de quelques secondes) sont reliées.
- *la couleur* : l'histogramme de couleur est fortement dépendant des habits de l'utilisateur. Les gens changent d'habits régulièrement, donc cette propriété est vraie sur un temps court. Pour chaque image i, les images ayant une différence de temps entre t_1 et t_2 (un autre seuil d'une demi-journée) sont sélectionnées. L'image i est ensuite reliée à ses k_c ($k_c = 3$) plus proches voisins (en termes de dissimilarité d'histogramme, avec la distance cosinus).
- *le visage* : en faisant abstraction des problèmes de pose, le visage des individus ne varie pas énormément au cours du temps. Pour chaque image i (pour laquelle un visage a été détecté), on recherche les images (avec un visage) dans un delta de temps supérieur à t_2. Les k_f ($k_f = 1$) plus proches visages (en termes de distance euclidienne) sont reliés entre eux.

Le graphe final est donc constitué de l'ensemble des images (les nœuds), avec les trois types d'arcs (qui ne sont pas différenciés, ni pondérés contrairement à l'algorithme précédent, section 5.1.3.3.2). Les expériences ont montré que la méthode semi-supervisée donne de meilleurs résultats que l'algorithme de base : un classifieur SVM linéaire (le noyau est une combinaison linéaire de noyaux linéaires sur les 3 types de descripteurs – les meilleurs poids de combinaisons sont sélectionnés avec une validation croisée). Le système est donc fait pour reconnaitre les personnes après avoir collecté un ensemble conséquent de données, il ne fonctionne donc pas en temps réel. Bien que le nombre d'images utilisées pour l'expérience soit relativement grand (>5000), le nombre d'utilisateurs concernés est relativement faible (10). Il est probable que la technique ne soit pas aussi performante avec plus de personnes. Kveton *et al.* [2010] proposent une modification de la méthode pour l'utiliser en ligne ; ils ont également validé leur algorithme avec un système de reconnaissance faciale.

5.1.3.3.3. Discussion Les techniques basées sur l'*auto-apprentissage* semblent avoir été les plus couramment utilisées dans la littérature, même si le terme n'a pas été systématiquement employé. Nous parlons de méthodes *semi-supervisées* plutôt que de méthodes *non-supervisées* car nous nous servons de la pseudo-étiquette calculée par la méthode de vérification biométrique. Il est souvent reporté dans la littérature que les méthodes semi-supervisées monomodales sont sujettes à capturer une faible variabilité, et, par conséquent à oublier d'intégrer une quantité non négligeable de données. Les méthodes à base de graphes pourraient être une solution partielle à ce problème. On peut également noter dans [Tur *et al.*, 2005] l'utilisation d'un mécanisme d'autoapprentissage associé à un mécanisme d'*apprentissage actif* [Cohn *et al.*, 1994]. Il s'agit d'une utilisation partiellement supervisée de la mise à jour du modèle : les données relativement proches du modèle initial sont intégrées au modèle à l'aide du mécanisme d'auto-apprentissage, et, les données difficiles à classifier sont étiquetées par un opérateur. Le coût est donc plus faible que d'utiliser un système entièrement supervisé, et plus de variabilité est capturée comparée à l'utilisation d'un système uniquement semi-supervisé. Le papier traite d'un problème de compréhension de phrase, et pas de biométrie, mais le concept peut facilement être adapté à un système d'identification faciale.

5.1.3.4. Stratégies de gestion des modèles

Comme il existe plusieurs façons de représenter une référence (section 5.1.2), la gestion de la mise à jour de ces modèles est donc également spécifique.

5.1.3.4.1. Utilisation d'un unique exemple Lorsque la référence est représentée par un unique exemple, trois méthodes principales peuvent être utilisées : le remplacement de cette référence, l'ajout de la nouvelle référence ou la fusion de la nouvelle référence avec l'ancienne.

Remplacement de l'ancienne référence par la nouvelle Dans ce cas, nous partons du principe que l'ancienne référence est obsolète, et que le modèle sera plus juste en utilisant uniquement la nouvelle. Cette méthode a uniquement un intérêt lorsque la place disponible pour stocker la référence est relativement limitée (et que l'on ne peut stocker qu'une seule référence), et qu'il n'est pas possible d'appliquer un traitement plus sophistiqué et coûteux en temps. Cette méthode ne semble pas avoir été expérimentée dans la littérature.

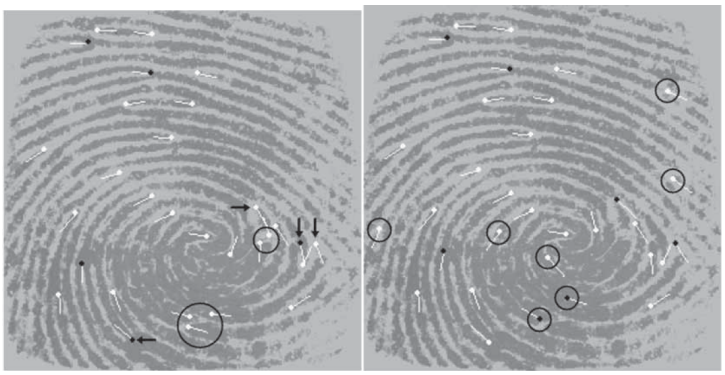
Ajout de la nouvelle capture Nous nous nous retrouvons dans le cas de la gestion des modèles par galerie. Une fois la nouvelle capture ajoutée à l'ensemble (qui était donc initialement constituée d'un seul et unique élément), les méthodes spécifiques aux galeries sont employées. Cette méthode est régulièrement employée dans la littérature.

Génération d'un super-modèle Dans ce cas, il y a toujours une seule référence dans la référence. La technique est nommée *super template generation* dans la littérature anglophone. Les super-modèles sont des modèles générés à partir de plusieurs captures (section 4.1.3.1). Ces modèles générés peuvent être considérés comme une nouvelle capture. De tels systèmes peuvent être créés pendant l'enregistrement lorsque plusieurs captures sont effectuées : un modèle est généré pour chacune de ces captures (*c.-à-d.*, la liste des minuties de la capture, pour un système d'empreintes digitales), puis un super-modèle est généré à partir des autres modèles (*c.-à-d.*, la liste des minuties jugées pertinentes pour un utilisateur, pour un système d'empreintes digitales). Ce mécanisme peut aussi être employé dans le cas d'un système adaptatif, afin de modifier le super-modèle précédent en fonction des variations nouvellement capturées. Une requête acceptée correspond donc à une capture de l'utilisateur. Cette capture peut encoder des variations par rapport à la référence actuelle (variabilité naturelle de la donnée biométrique, meilleures conditions d'acquisition ou erreurs d'acquisition). Il est intéressant d'intégrer ces variations, qui peuvent encoder une variabilité réelle qui n'a pas pu être capturée précédemment. Cette méthode a été essentiellement utilisée dans le cas de l'utilisation de systèmes d'empreintes digitales [Jiang et Ser, 2002; Ryu *et al.*, 2005, 2006] : de nouvelles minuties (qui n'ont pas été détectées lors de l'enregistrement) sont ajoutées à la référence, tandis que d'autres (considérées comme étant des erreurs) sont supprimées. Différentes familles de méthodes existent : fusion des images, fusion des minuties ou fusion des deux. Certaines méthodes utilisent un historique des captures, tandis que d'autres n'en utilisent que deux : le modèle et la requête. Les super-modèles peuvent être utilisés uniquement à l'enregistrement, sans faire de mise à jour au cours de l'utilisation Ryu *et al.* [2005].

Les super-modèles ont également été étudiés dans le cas de la reconnaissance faciale. Rattani *et al.* [2007] présentent une méthode pour générer un super-modèle, pour la reconnaissance faciale, en utilisant une image frontale, une image de profil droit, et une image de profil gauche. De cette, façon, il n'est pas nécessaire d'avoir un modèle pour les trois vues du visage. Cette technique a été utilisée pour générer le super-modèle depuis les données d'enregistrement, mais, on peut facilement imaginer son utilisation dans un système adaptatif. Rattani *et al.* [2007] montrent que les performances sont meilleures en utilisant ce super-modèle plutôt qu'une unique image frontale.

Même si le terme n'est pas employé, une technique de super-modèle a été expérimentée pour la DDF. Afin d'augmenter les performances des algorithmes, une technique est présentée dans [Bleha et Obaidat, 1991] : elle consiste à créer une capture à partir de deux en les

fusionnant (moyenne des deux vecteurs) ; ce mécanisme permet de filtrer les légères hésitations (mais nécessite à l'utilisateur d'effectuer deux saisies correctes au lieu d'une seule).



(a) modèle initial (b) super-modèle

FIG. 5.9: Illustration de la génération d'un super-modèle. Les minuties encerclées sont celles qui disparaissent ou apparaissent. Les flèches indiquent un changement de type de minuties (source : [Jiang et Ser, 2002])

5.1.3.4.2. Utilisation d'une galerie Plusieurs approches différentes existent pour la mise à jour basée sur l'utilisation de galerie. Comme les problèmes de sélection de références sont intimement liés au problème de mise à jour, nous présentons les deux types de méthodes. Le premier cas consiste à ne sélectionner que les exemples représentatifs (selon des critères spécifiques à ces méthodes) d'une galerie, tandis que le second consiste à mettre à jour le contenu de la galerie. Cette mise à jour de la galerie est une nécessité, car les individus collectés lors de l'enregistrement ont tendance à devenir de moins en moins représentatifs avec le temps.

Les méthodes de partitionnement ou de sélection Les méthodes de partitionnement ont initialement été développées pour permettre d'avoir plusieurs modèles par utilisateurs dans les méthodes utilisant une seule référence comme modèle [Uludag *et al.*, 2004]. Bien qu'il ne s'agisse pas à proprement parler de stratégies de mise à jour, elles permettent de disposer de plusieurs modèles pour un même utilisateur. Il s'agit donc d'une technique permettant d'encoder une forte variabilité intra-classe et la technique devient encore plus intéressante lorsqu'on lui ajoute un système de mise à jour. Cette méthode n'a été utilisée que dans le cas de la reconnaissance d'empreinte digitale, où une unique capture est présente dans la référence (la vérification se fait en comparant la capture requête à la capture modèle). L'intérêt de partitionner est de pouvoir récupérer K captures parmi N (avec $K < N$) de telle façon à ce que ces K captures encodent un maximum de variabilité. Uludag *et al.* [2004] présentent deux façons majeures de partitionner les données de la galerie :

– *DEND.* Les captures sont agglomérées en utilisant un classifieur hiérarchique à l'aide d'une mesure de similarité. Cette agglomération est faite à l'aide d'un dendogramme. Celui-ci est sectionné de telle façon à obtenir K partitions. La figure 5.10 présente un tel dendograme dans le cas d'un système de reconnaissance d'empreintes digitales. La capture ayant la distance moyenne minimale avec les autres membres de sa partition, est sélectionnée comme prototype de la partition. Étant donné que cette méthode cherche à capturer le plus de variabilité possible, elle est également sujette à prendre facilement des imposteurs.

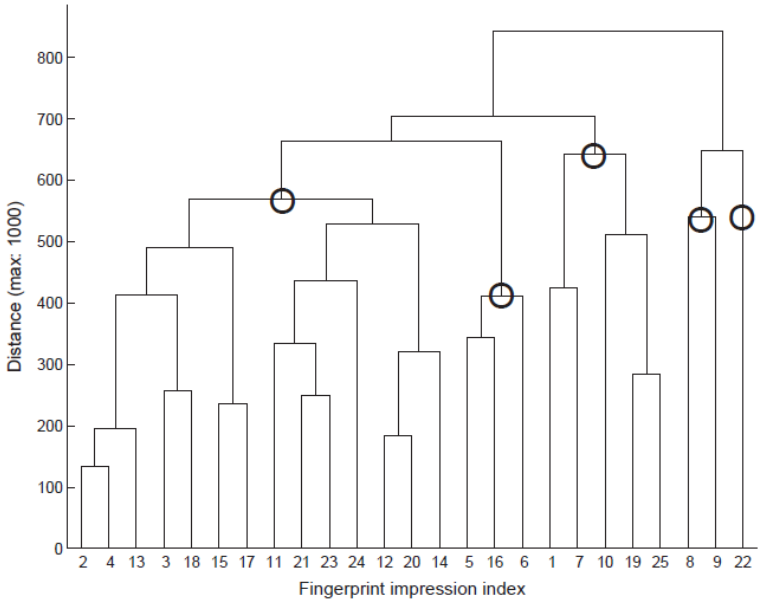


FIG. 5.10: Illustration du dendograme calculé pour un utilisateur avec 25 empreintes digitales dans sa galerie. Les cercles indiquent la coupure pour obtenir 5 partitions (source : Uludag *et al.* [2004])

– *MDIST*. Cette version trie les captures en fonction de leur distance moyenne aux autres. Les K captures donnant les K plus petites valeurs sont sélectionnées comme prototype. Les prototypes choisis sont donc ceux ayant la plus forte similarité avec l'ensemble des données.

Naturellement, le choix de K dépend donc de l'application et reste un problème ouvert. Les deux méthodes ont été validées en générant des partitions aléatoires et en comparant les performances avec ces deux types de partitionnement. Les résultats ont montré qu'un choix aléatoire donne de moins bons résultats. Ces deux techniques de partitionnement de la référence ont été utilisées avec des méthodes de mise à jour supervisée, en les exécutant plusieurs fois au cours de l'utilisation du système en intégrant de nouvelles captures. Lors d'une vérification, la requête est comparée aux K prototypes. Le score final est le score moyen des K scores. Li *et al.* [2008] proposent également une méthode de sélection de K prototypes parmi N. La méthode a également été validée dans un contexte de reconnaissance d'empreintes digitales dans le cas supervisé.

Les méthodes d'édition Le but des méthodes d'édition (annexe D), comme les méthodes de partitionnement, est de réduire le nombre d'individus représentatifs de la galerie ; elles ne sont donc pas nécessairement liées à l'utilisation de mécanisme de mise à jour. Freni [2010] a présenté dans sa thèse de doctorat, de nombreuses techniques d'édition dans le cas de systèmes de reconnaissance faciale et d'empreinte digitale. Il s'agit, une fois de plus, d'une technique intéressante pour les systèmes utilisant des classifieurs de type plus proches voisins. Le principe est : pour un ensemble d'apprentissage T, il faut trouver un sous-ensemble E qui permet d'avoir le même taux de classification que T sur lui-même. L'intérêt de la technique est d'avoir une taille de galerie dépendante de la difficulté de reconnaissance du client (et donc de réduire au maximum sa taille). Les méthodes d'édition permettent de ne conserver que les instances représentantes d'un jeu de données lors de l'utilisation des k plus proches voisins (et donc de diminuer la combinatoire

lors de la vérification, en diminuant le nombre de tests à effectuer). Deux familles de méthodes d'édition existent :

- Les méthodes *incrémentales* qui partent d'un ensemble E vide, et, le peuplent progressivement jusqu'à atteindre un critère de satisfaction.
- Les méthodes *décrémentales* qui partent d'un ensemble E complet (tous les éléments de T), et, suppriment les instances progressivement jusqu'à satisfaire un critère de décision.

Il semble que ces techniques d'édition aient été appelées sur l'ensemble des galeries en une fois (*c.-à-d.*, le processus est effectué de manière globale, et non pas individuelle pour chaque utilisateur). Ce n'est pas le cas de MDIST et DEND qui travaillent uniquement avec une galerie utilisateur à la fois. Les auteurs ont travaillé dans le cas du plus proche voisin avec la distance euclidienne comme mesure de similarité. Plusieurs méthodes d'édition existent dans la littérature et ont été testées pour la biométrie (voir annexe D).

Il semble également que ces méthodes n'aient été utilisées que dans le cas supervisé. Les auteurs montrent que les méthodes d'édition sont compétitives vis-à-vis de celles de l'état de l'art (*MDIST* et *DEND*), et, qu'elles permettent de fortement diminuer la taille de la galerie (7 représentants à la place de 50 pour la méthode la plus agressive, CNN), d'autant plus que l'administrateur n'a pas à se préoccuper de la taille des galeries. Cependant, les performances sont meilleures car les méthodes peuvent avoir un nombre conséquent d'éléments dans les galeries. Les méthodes ne peuvent être utilisées que lorsqu'un nombre suffisant de données est accessible.

Bien que ces méthodes ne soient pas des méthodes de mise à jour, elles permettent de diminuer la taille de la galerie qui est susceptible de grossir et de rendre la vérification de plus en plus lente tout au long de l'application des mises à jour. Les méthodes basées sur l'édition nécessitent d'appliquer le processus sur l'ensemble des données d'enregistrement de tous les utilisateurs. Cette méthode est donc applicable aux modalités proposant le même type de données biométriques quel que soit l'utilisateur (*c.-à-d.*, la reconnaissance faciale, où tout le monde peut donner une photo de son visage), mais pas aux modalités proposant des données biométriques n'ayant aucune relation en fonction des utilisateurs (*c.-à-d.*, la reconnaissance de la DDF, lorsque chaque utilisateur a un mot de passe différent).

Les méthodes de remplacement Le but des méthodes de remplacement est de ne pas modifier la taille de la galerie au cours des différentes mises à jour. Cependant, il s'agit bien de méthodes spécifiques à la mise à jour, car elles utilisent les nouvelles données capturées tout au long de l'utilisation du système biométrique (certaines sont spécifiques au mode en ligne, d'autres au mode hors ligne, et d'autres disposent de variantes dans les deux cas).

Remplacement aléatoire Freni *et al.* [2008] ont testé différentes méthodes de remplacement. L'une d'entre elle est le remplacement d'un exemple, de l'ensemble d'apprentissage, sélectionné aléatoirement, par la nouvelle requête. Naturellement, il ne s'agit pas de la méthode la plus performante. Cependant, Freni *et al.* ont montré que plus la taille de la galerie est importante[4], plus les autres méthodes (plus complexes, et censées être plus efficaces) se rapprochent de la méthode de remplacement aléatoire. Des méthodes plus complexes n'ont un sens que si la taille de la galerie est relativement faible.

Fenêtre glissante Une des principales techniques de remplacement est basée sur le remplacement des captures les plus anciennes. En mode *hors ligne*, l'ensemble des données d'apprentissage est remplacé par l'ensemble des nouvelles données. Cette technique est appelée *BATCH-UPDATE* dans [Uludag *et al.*, 2004] et s'est révélée efficace dans le cas de la mise à jour d'un système basé sur les empreintes digitales (EER de 7.69% au lieu de 10.32%). En mode

4. Il faut noter, que pour cette étude « relativement importante » n'a pas forcément de sens, étant donné que les utilisateurs ont réalisé 8 captures

en ligne, la donnée la plus ancienne est remplacée par la nouvelle donnée acceptée. Cette technique est appelée *moving window* dans [Kang *et al.*, 2007] et *First In First Out (FIFO)* dans [Freni *et al.*, 2008; Kekre et Bharadi, 2009; Scheidat *et al.*, 2007].

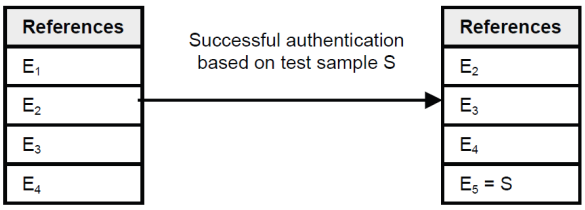


FIG. 5.11: Mécanisme de remplacement en utilisant une fenêtre glissante. (source : Scheidat *et al.* [2007])

Moins fréquemment utilisé Cette technique est présentée dans [Freni *et al.*, 2008; Scheidat *et al.*, 2007] sous le nom de *Least Frequently Used (LFU)*. Le principe est de remplacer l'exemple le moins souvent utilisé pour la vérification de l'utilisateur. Il est donc nécessaire de maintenir le nombre d'utilisations de chaque exemple de la galerie en tant que prototype ayant authentifié l'utilisateur. Le problème majeur de cette technique est le fait que les captures présentes depuis le plus longtemps dans la référence ont nécessairement été les captures les plus proches le plus souvent. Elles ont donc un poids plus important que des captures plus récentes potentiellement de meilleures qualités.

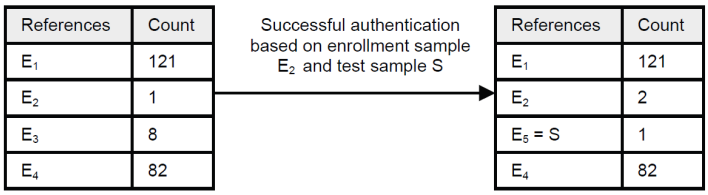
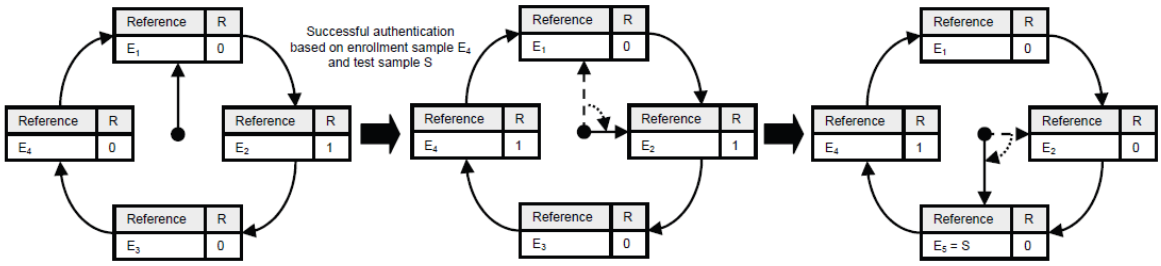


FIG. 5.12: Mécanisme de remplacement en remplaçant l'exemple le moins souvent utilisé (source : Scheidat *et al.* [2007])

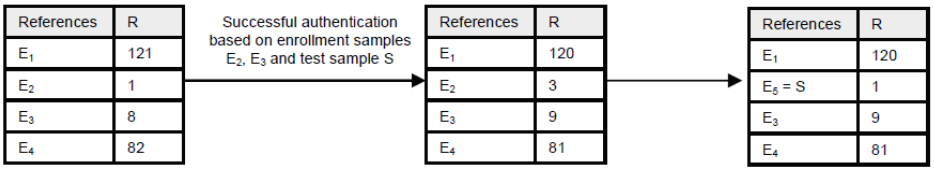
Moins récemment utilisé Cette technique est également présentée dans [Scheidat *et al.*, 2007] sous le nom de *Least Recently Used (LRU)*. La méthode considère qu'il est préférable que la nouvelle donnée remplace celle qui a été utilisée le moins récemment lors des précédentes vérifications. Une première méthode est d'utiliser un *timestamp* avec chaque exemple de la galerie, ce qui peut être trop coûteux [Scheidat *et al.*, 2007]. Les auteurs proposent deux algorithmes pour contourner ce problème : la méthode de la seconde chance (*Second Chance (clock) algorithm* dans la littérature anglophone), et une méthode étendue.

Les techniques de remplacement permettent d'obtenir une galerie dont la taille ne croit pas au cours du temps. Elles peuvent donc être utilisées dans les systèmes ne disposant pas de beaucoup d'espace pour stocker la référence d'un utilisateur.

5.1.3.4.3. Les méthodes d'ajout Le principe d'utiliser des méthodes d'ajout, est d'augmenter progressivement la taille de la galerie de la référence. Cela permet de commencer à utiliser le système biométrique après une période d'enregistrement relativement courte (n'encodant pas forcément une grande variabilité) et d'augmenter la taille de celle-ci au cours du fonctionnement du système biométrique (et potentiellement encoder une plus grande variabilité de l'utilisateur). Uludag *et al.* [2004] présentent une méthode nommée *AUGMENT-UPDATE* qui ajoute en mode



(a) algorithme d'horloge



(b) algorithme étendu

FIG. 5.13: Mécanisme de remplacement en utilisant le principe du moins récemment utilisé (source : Scheidat *et al.* [2007])

hors ligne l'ensemble des nouvelles données dans la galerie de l'utilisateur. La taille de la galerie croit donc au fil du temps. D'autres variantes ont été proposées dans la littérature: les nouvelles données sont progressivement ajoutées à la référence jusqu'à obtenir un nombre maximal de données dans la galerie. Une fois ce nombre maximal atteint, la galerie est modifiée en utilisant des systèmes de remplacement. Grabham et White [2008] ont implémenté avec succès cette méthode dans un système d'authentification par dynamique de frappe au clavier sur un clavier avec des capteurs de pression.

5.1.3.4.4. La référence est un classifieur Dans le cas de l'utilisation d'un classifieur, le modèle n'est plus constitué d'une galerie ou d'une capture unique, mais des paramètres de ce classifieur calculés grâce à la galerie. Le classifieur a été entrainé avec les données d'enregistrement. Au lieu de mettre à jour le classifieur, en effectuant à nouveau l'apprentissage en mode batch avec l'ensemble des données d'apprentissage augmenté de la nouvelle capture, le classifieur doit être entrainé en ligne afin d'en adapter les paramètres. L'intérêt d'utiliser une telle technique, plutôt qu'une méthode à base de galerie qui recalcule les paramètres du classifieur, après chaque ajout de nouvelles données, est de grandement diminuer les temps de calcul et de pouvoir être utilisée avec d'importantes bases de données.

Séparateurs à vaste marge Singh *et al.* [2010] présentent un système de mise à jour en ligne supervisé pour la reconnaissance faciale 2D proche infrarouge. Le classifieur est un « 2ν-Online Granular Soft Support Vector Machine ». Le calcul granulaire permet de s'adapter aux variations globales et locales dans la distribution des données, et, les étiquettes souples permettent d'être résistantes au bruit. Le fonctionnement d'un SVM classique est présenté en annexe E. Les paramètres optimaux (le C et les paramètres de la fonction noyau) sont obtenus en testant manuellement plusieurs jeux de paramètres, jusqu'à obtenir un taux d'erreur optimal. Le fonctionnement d'un double ν-SVM (2ν-SVM) est présenté en annexe E.2 et son adaptation avec des étiquettes souples est présentée en annexe E.3. Son intérêt est d'être calculatoirement plus efficace qu'un SVM et plus flexible lorsque les nombres de données positives et négatives

sont mal équilibrées. Même avec un 2ν-SSVM, l'apprentissage d'une grande base de données est consommateur de temps. Le calcul granulaire [Bargiela et Pedrycz, 2003] est basé sur un principe de *diviser pour régner*, et, permet de réduire le temps de calcul. Le 2ν-Granular SSVM a été entrainé en mode batch lors de l'enregistrement, puis en ligne au cours de l'utilisation du système biométrique. La mise à jour consiste à ajouter de nouveaux vecteurs supports (qui sont linéairement indépendants), et à supprimer les anciens vecteurs supports qui n'augmentent pas les performances du classifieur. De cette façon, le nombre de vecteurs supports n'augmente pas énormément au fil de l'utilisation du système de mise à jour. Il s'agit donc à la fois d'un algorithme incrémental et décrémental. L'algorithme du 2ν-Online GSSVM est le suivant :

1. Le 2ν-GSSVM est entrainé en utilisant l'ensemble initial d'apprentissage et un hyperplan de décision est obtenu avec les m vecteurs supports.
2. Pour chaque nouvel exemple $\bar{\mathbf{x}}_i$,
   a) $\bar{\mathbf{x}}_i$ est classé en utilisant le 2ν-GSSVM.
   b) Le résultat de classification est comparé avec l'étiquette $\bar{\mathbf{z}}_i$; Si la classification est correcte, alors il n'est pas nécessaire d'effectuer le réapprentissage.
   c) Sinon,
      i. L'hyperplan de décision est recalculé en utilisant les m vecteurs supports et $\{\bar{\mathbf{x}}_i,\bar{\mathbf{z}}_i\}$ à l'aide de l'apprentissage en mode batch. Cette fois-ci l'apprentissage est plus rapide, il y a moins de données d'apprentissage.
      ii. Après avoir recalculé l'hyperplan, le nombre de vecteurs supports s'accroit. Si le nombre de vecteurs supports est supérieur à $m+\lambda$ (où λ est un seuil qui contrôle le nombre de vecteurs supports), alors un vecteur support le plus loin de la frontière de décision est sélectionné.
      iii. Le vecteur support sélectionné est supprimé de la liste des vecteurs supports et stocké dans la liste l. Le classifieur avec $m + \lambda - 1$ vecteurs supports est utilisé pour la validation et le test.
3. Les vecteurs supports dans la liste l sont utilisés pour tester le nouveau classifieur. S'il y a des erreurs de classification, l'étape 2(c) est répétée pour minimiser l'erreur de classification.
4. Les vecteurs supports les moins récemment inclus sont supprimés de la liste, l, dans le classifieur final.

Le système a été utilisé dans un système à deux classes (+1 pour les clients, -1 pour les imposteurs), les attributs des données étant constitués des scores de différents systèmes de reconnaissance faciale. L'apprentissage en ligne s'est révélé calculatoirement plus efficace que l'apprentissage hors ligne, aussi bien pour l'apprentissage que le test. L'article original utilise le mécanisme pour une mise à jour supervisée, mais le système a été validé pour une mise à jour semi-supervisée également [Bhatt *et al.*, 2011].

Cartes auto-adaptives Dozono *et al.* [2007] utilisent une carte auto-adaptive [Kohonen, 1997] pour faire de la classification d'utilisateurs dans un système de reconnaissance de dynamique de frappe (temps de pression, intervalle entre deux touches et niveau sonore maximum lors de la pression de la touche). Le système qui est appris en mode batch est donc constitué d'une seule et unique référence pour l'ensemble des utilisateurs. Une fois les vecteurs représentatifs de chaque neurone correctement configurés, une étiquette leur est associé. Cette étiquette (numéro de l'utilisateur) correspond à la donnée la plus proche. La vérification se fait en récupérant l'étiquette du neurone pour lequel le vecteur représentatif est le plus proche du vecteur requête. L'étude présente une méthode de mise à jour supervisée et une méthode semi-supervisée. Dans le premier cas, le vecteur représentatif du neurone sélectionné est légèrement modifié, que la réponse soit correcte (ajout d'une portion de la différence) ou incorrecte (soustraction d'une

portion de la différence entre le neurone et l'exemple). Dans le cas semi-supervisé, la mise à jour ne peut être faite que lorsque la donnée est acceptée. Dozono *et al.* [2007] ont nommé leur carte auto-adaptive : Multi-Winner SOM (MW-SOM). Dozono *et al.* [2007] proposent une méthode incrémentale de mise à jour de la carte. Celle-ci est appelée toutes les trois authentifications, car des mises à jour trop nombreuses ont des effets négatifs sur la carte. L'étude montre que l'apprentissage semi-supervisé est relativement sensible au bruit (plus les données sont bruitées, plus les performances se dégradent), tandis que l'apprentissage supervisé est plus stable. Il faut noter que l'étude ne travaille qu'avec 10 utilisateurs et 10 captures par utilisateur (les données supplémentaires sont générées artificiellement en modifiant aléatoirement les données originales).

Analyses en composantes El Gayar *et al.* [2006] utilisent une ACP pour réduire la dimension des données manipulées d'un système de reconnaissance faciale. Le système utilise des mécanismes d'auto-apprentissage et de co-apprentissage sur un jeu de données non étiqueté. Ce qui nous importe dans cette section, n'est pas la façon de collecter ces données, mais la façon dont le modèle est mis à jour. 2/3 des données de test sont utilisées pour la mise à jour, et 1/3 est utilisé pour la validation. La mise à jour consiste simplement à recalculer l'ACP sur le nouveau jeu de données (qui n'intègre que les cinq meilleurs nouveaux exemples pour chaque classe).

Vandana [2007] propose la mise à jour itérative de l'espace propre de chaque utilisateur. La référence d'un utilisateur intègre les différentes informations nécessaires pour effectuer le changement de repère de la capture requête afin de la projeter dans cet espace propre (calculé à l'aide des données d'enregistrement de l'utilisateur (classe 0), et à l'aide des données d'enregistrement des autres utilisateurs (classe 1)). La vérification utilise un algorithme de *kppv*. La méthode employée en mode batch pour calculer l'espace propre initial est baptisée : Incremental Biased Discriminant Analysis (IBDA).

Liu *et al.* [2003] ont également utilisé un mécanisme de mise à jour de vecteurs propres de façon itérative. Leur étude porte sur la reconnaissance faciale, où l'ACP est effectuée pour chaque utilisateur. Contrairement au cas précédent, ils partent du principe que plus les données sont anciennes, plus elles sont obsolètes, et moins il est nécessaire de les prendre en compte dans le modèle. Ainsi, la moyenne et la matrice de covariance sont mises à jour en attribuant des poids aux exemples. Plus les exemples sont anciens, plus le poids est faible. Afin de profiter de l'intérêt de cette mise à jour en ligne, les deux calculs sont effectués de façon récursive. Ainsi, la moyenne $\hat{\mathbf{m}}_n$ à l'instant n est calculée de la façon suivante :

$$\hat{\mathbf{m}}_n = \alpha_m \hat{\mathbf{m}}_{n-1} + (1 - \alpha_m)\mathbf{x}_n \tag{5.1}$$

avec $\mathbf{x}_n$ le nouvel exemple à intégrer dans le modèle et α_m le facteur de pondération indiquant la vitesse d'oubli des anciens exemples. La valeur de α_m dépend principalement de la connaissance de l'évolution du processus aléatoire (le visage dans ce cas). Le même principe est utilisé pour la matrice de covariance.

Généralisation Poh *et al.* [2009] généralisent les exemples précédents à l'aide d'une interprétation bayésienne de la mise à jour de la référence.. Ils énumèrent trois opérations basiques essentielles dans la vie de la référence :

– La création du modèle :

$$\mathit{Créer} : \mathit{données} \rightarrow \mathit{nouveau\ modèle} \tag{5.2}$$

– L'adaptation du modèle :

$$\mathit{Adapter} : \mathit{modèle}, \mathit{données} \rightarrow \mathit{modèle\ adapté} \tag{5.3}$$

– La suppression du modèle:

$$Supprimer : modèle \rightarrow \emptyset \tag{5.4}$$

Ajouter une référence et la supprimer est relativement facile, ce qui est loin d'être le cas pour la mise à jour qui nécessite la combinaison de plusieurs exemples. Comme nous avons pu le voir, les classifieurs basés sur des modèles statistiques peuvent souvent être implantés dans des formes en ligne (*c.-à-d.*, les paramètres de l'ancien modèle peuvent être mis à jour avec les nouveaux exemples obtenus après sa création) et il n'est pas forcément nécessaire de stocker tout l'historique pour mettre à jour la référence. Poh *et al.* [2009] ont aussi modélisé la mise à jour pour des modèles non statistiques. Ils présentent les techniques de super-modèles (voir la 5.1.3.4.1) comme un modèle statistique simplifié sous la forme d'une distribution gaussienne multivariée avec une covariance isotrope.

5.1.3.4.5. Discussion Nous avons vu que de nombreuses techniques de gestion d'ajout d'une capture dans la référence existent. Chacune d'entre elles est plus ou moins adaptée à telle ou telle modalité. Cependant, nous pouvons remarquer que la plupart des méthodes sont vouées à des systèmes biométriques dont la vérification est basée sur le calcul d'une distance à une autre capture ou à ses k plus proches voisins. Il est fort probable qu'une majorité de ces techniques ne soient pas utilisables dans le cas de la DDF.

5.1.4. Évaluation des systèmes de mise à jour

De nombreuses mesures existent dans la littérature pour évaluer de façon statistique les performances des systèmes biométriques, cependant elles ne prennent pas en compte l'aspect temporel (qui nous semble être un point primordial du cas de la mise à jour). Il est donc nécessaire de les utiliser dans un cadre spécifique, en suivant un protocole propre à l'évaluation des systèmes de mise à jour. À notre avis, donner une valeur globale d'une métrique particulière d'un système de mise à jour n'a aucun sens car nous n'avons pas d'information sur son évolution au cours du temps. De plus, nous ne savons pas s'il est plus judicieux de tester la mise à jour de façon hors ligne ou en temps réel. Quelques travaux de la littérature ont essayé d'apporter des solutions à ces différents problèmes. Nous présentons donc les métriques qui doivent être utilisées pour l'évaluation.

5.1.4.1. Calcul des performances

5.1.4.1.1. Métriques utilisables

Rappels sur les métriques génériques Il est possible d'utiliser différentes métriques pour l'évaluation de la mise à jour. Les trois métriques principales sont le FMR, le FNMR et l'EER (section 2.2.1). Ryu *et al.* [2006] présentent l'EER, le FMR lorsque le FNMR est nul, ainsi que le FNMR lorsque le FMR est nul. L'article ne précise pas comment les seuils sont calculés, mais cela suppose de disposer d'un jeu de données suffisamment grand (ce qui n'est pas forcément problématique) et de données d'imposteurs pour le cas du *0FMR*[5] (ce qui n'est pas forcément applicable à la DDF). En règle générale, les autres études présentent les résultats avec l'EER et le *1FMR*. Ces métriques donnent les performances de reconnaissance du système ; plus les valeurs sont faibles, meilleur est le système, cependant elles ne sont pas suffisantes dans notre cas.

5. nFMR est le FMR lorsque le FNMR vaut n

Métriques spécifiques à la mise à jour Une information pertinente pour la mise à jour est le ratio d'imposteurs ajoutés dans la référence [Marcialis *et al.*, 2008]. Cette information semble être rarement présentée dans la littérature. Nous nous attendons donc à avoir le moins d'imposteurs possibles pour les meilleurs algorithmes de mise à jour. Nous allons nommer cette métrique *Taux d'Imposteurs Ajoutés aux Modèles (TIAM)*, elle se calcule de la façon suivante, et est calculée une seule fois :

$$TIAM = \frac{\sum \mathbb{1}\{étiquette(galerie) \in imposteurs\}}{|galerie|} \tag{5.5}$$

avec *galerie* l'ensemble des données collectées lors de la mise à jour. L'intérêt d'utiliser un système de mise à jour est d'augmenter les performances du système de reconnaissance (ou du moins, de faire en sorte qu'elles ne chutent pas au cours du temps). Il est donc probablement nécessaire de présenter également le pourcentage d'amélioration du système utilisant la mise à jour comparée au système n'employant pas de tel mécanisme :

$$gain = \frac{perf_{normal} - perf_{update}}{perf_{normal}} \tag{5.6}$$

5.1.4.1.2. Fréquence d'utilisation des métriques employées Il est intéressant de se demander à quelle fréquence nous devons employer les métriques de calcul d'erreur. Nous avons vu que l'aspect temporel est important dans la gestion de la mise à jour, il doit donc également être important de le prendre en compte dans la gestion du calcul des erreurs. La première façon de prendre ce point en compte est de toujours présenter les données à tester par ordre chronologique afin de suivre l'évolution temporelle de la *donnée biométrique*. La plupart des études présentent les données de façon aléatoire et ne semblent pas avoir respecté ce principe. Il est possible que cela n'ait pas d'implications particulières sur l'évolution du modèle des modalités biométriques à faible variabilité (*cf.*, empreintes digitales), mais nous pensons que dans le cas des modalités à forte variabilité temporelle (*cf.*, dynamique de frappe au clavier), il est plus que nécessaire de garder ce point à l'esprit. Toujours pour cette raison, il nous semble que ne fournir qu'un seul taux d'erreur global ne soit pas la meilleure solution, alors que c'est ce qui est fait dans la quasi-totalité des études car elles ne comprennent pas forcément de données sur une longue période. Pourquoi ne pas calculer ces métriques à chaque authentification ? À chaque journée ? Ou à chaque session de capture de la base de donnée utilisée ?

5.1.4.1.3. Calcul des performances en-ligne ou hors-ligne ? Quelle que soit la métrique d'évaluation utilisée, il existe deux façons principales d'effectuer cette évaluation :

- *Hors ligne* : la référence est mise à jour un certain nombre de fois à l'aide d'un ensemble de données de test. Au bout d'un certain temps, qui dépend de la fréquence de calcul définie plus haut, un autre jeu de test est utilisé afin d'être vérifié sur la référence mise à jour. La métrique d'évaluation est ensuite calculée sur ces scores de façon hors ligne sans faire de mise à jour du modèle. Poh *et al.* [2009] nomment ce mécanisme : *separate adapt-and-test strategy.*
- *En ligne* : la référence est évaluée à la volée tout au long de sa mise à jour. Chaque donnée sert à la fois à l'évaluation (en produisant un score de comparaison entre la donnée *requête* et la référence qui sera stockée avec les scores intra-classe ou inter-classe) et à la mise à jour (en étant intégré, le cas échéant dans le nouveau modèle). Au bout d'un certain temps, qui dépend de la fréquence de calcul définie plus haut, la métrique d'évaluation est calculée sur l'ensemble des scores précédemment produits. Poh *et al.* [2009] nomment cette méthode : *joint adapt-and-test strategy.*

Nous n'avons pas trouvé de papier qui évalue les performances en ligne dans la littérature (excepté dans [Poh *et al.*, 2009]), alors que cette procédure semble plus réaliste (*c.-à-d.*, dans un système

réel, les données utilisées pendant la vérification sont celles qui servent pour la mise à jour) et permet de disposer de plus de données pour les calculs (*c.-à-d.*, la même requête sert à la fois pour la vérification, et, peut potentiellement être utilisée pour la mise à jour). Comme expliqué précédemment, dans la littérature toutes les évaluations sont faites hors ligne avec un unique sous-ensemble de test.

5.1.4.2. Respect d'un protocole

Il est nécessaire de respecter un protocole particulier afin de pouvoir comparer les études entre elles. L'idéal étant d'avoir un protocole commun entre les études [Rattani, 2010]. Cependant, respecter un tel protocole n'est pas toujours possible ou suffisant. Dans tous les cas, il faut préciser toutes les informations nécessaires afin que l'étude puisse être reproductible facilement. Suffisamment d'informations doivent être données concernant la base de données utilisée pour l'étude. Les points les plus importants sont:

- Le nombre total d'utilisateurs concernés.
- Le nombre de captures fournies par chaque utilisateur.
- L'intervalle de temps pris pour la création de la base de données, voire la date de chacune des captures.
- Le nombre de sessions, le cas échéant, constituant la base de données.
- Le partitionnement de la base de données :
  - la partition de données utilisées pour l'enregistrement ;
  - la partition de données utilisée pour la mise à jour (données non étiquetées) ;
  - le cas échéant, la partition de données utilisée pour la validation hors ligne (données de test).

D'après la Table 1 de Rattani *et al.* [2009] (synthétisant 14 études), la quantité de données utilisée pour la mise à jour (27 en moyenne) est nettement plus faible que la quantité de données utilisée pour la validation (62 en moyenne). Il n'existe pas encore de consensus sur le ratio à garder (dans le cas d'une évaluation hors ligne), mais il parait évident que ce ratio influe sur les résultats. Pour chaque évaluation des performances, il peut être utile de connaitre le nombre de scores intra-classe et inter-classe impliqués. Ces informations permettent de donner une indication sur le ratio de données d'imposteurs utilisées dans le calcul des performances. Le protocole doit également présenter l'ordre de présentation des données de test aux modèles biométriques à faire évoluer. Il est notamment intéressant de savoir si:

- Toutes les données clientes sont présentées en premier.
- Toutes les données d'imposteurs sont présentées en premier.
- Les données sont présentées aléatoirement [Marcialis *et al.*, 2008].

Ryu et Kim [2005]; Ryu *et al.* [2006] présentent les résultats pour les trois types de présentation en indiquant que cela permet de mieux évaluer les performances des systèmes commerciaux. Ces papiers traitent de la reconnaissance d'empreintes digitales, et, les performances sont sensiblement identiques dans les trois cas. Nous ne nous attendons pas à obtenir le même résultat avec des modalités biométriques plus faibles. Les trois façons ont été utilisées dans la littérature. Il est également important de voir si les données sont également présentées de façon chronologique (*c.-à-d.*, $date(presentation_t) < date(presentation_{t+i}), \quad \forall i$). Ce point semble n'avoir quasiment jamais été pris en compte dans les études. À notre connaissance, ce point n'est respecté que dans [Liu *et al.*, 2003] (et encore, leur base de données est constituée de séquences vidéos sur une courte période, avec donc une variabilité quasiment nulle sur la forme du visage ou sa texture, et ils ont présenté les données aléatoirement pour le test), où l'espace propre est calculé récursivement tout au long du temps, en donnant moins de poids aux plus anciens exemples.

C'est surprenant car le but de ces études est de capturer la variabilité au cours du temps, ce qui implique donc de respecter un minimum de chronologie. Il est fort probable que ce non-respect de chronologie ne soit pas réellement problématique dans le cas des modalités morphologiques, mais, ce point n'est pas à négliger dans le cas des modalités comportementales, où l'utilisateur acquière un réflexe, au cours de l'utilisation du système, qui implique une évolution chronologique. Le ratio de données d'imposteur comparé aux données de test est également un point à prendre en compte. En effet, plus il y a d'imposteurs dans le jeu de données, plus la probabilité d'ajouter un imposteur dans la référence du client est grande. Le problème est donc plus difficile dans les études où la quantité de données d'imposteur est grande. Cette information n'est pas toujours présentée dans la littérature. Marcialis *et al.* [2008] utilisent le même nombre de données d'imposteurs et de clients pour la mise à jour, et le même nombre de données d'imposteurs et de clients pour le test. Les derniers points importants à vérifier sont ceux ayant été présentés précédemment : les évaluations sont-elles faites en ligne ou hors-ligne ? À quelle fréquence ? Avec quelle mesure ?

5.1.4.3. Application à la dynamique de frappe au clavier

Nous n'avons trouvé qu'une seule référence dans la littérature fixant le cadre d'utilisation des systèmes de mise à jour dans le cadre de la dynamique de frappe au clavier [Seeger et Bours, 2011][6]. Nous pouvons noter que ces points ne s'appliquent que dans un système de mise à jour en ligne lorsque la méthode de reconnaissance biométrique utilise une galerie. Dans ce papier, les auteurs expliquent les différents paramètres à spécifier dans les études afin de pouvoir les comparer et reproduire facilement. Voici les différents points énumérés (en utilisant les mêmes termes et numérotations):

1. Les trois phases d'évaluation.
   - Une phase d'initialisation (sélection des données, choix de la méthode de reconnaissance, séparation de la base, ...).
   - Calcul de la méthode de base : calcul des performances sans utiliser de système de mise à jour (procédure classique).
   - Calcul de la méthode de mise à jour en utilisant les mêmes métriques de performance que précédemment.
2. Le jeu de données.
   Les auteurs conseillent d'utiliser une base publique afin de pouvoir comparer les résultats (GREYC ou DSL2009).
3. Le seuil d'acceptation.
   Il faut définir si le seuil utilisé est un seuil global, ou un seuil individuel.
4. La taille du jeu de données.
   Avec N utilisateurs ayant k données de tests, le nombre de comparaison intra-classe est $N * k$, tandis que le nombre de comparaisons inter-classe est $N * (N - 1) * k$. Il y a donc un profond déséquilibre qui peut perturber la méthode de mise à jour en privilégiant les données d'imposteurs. Plusieurs méthodes de gestion des données d'imposteurs peuvent être appliquées:
   - Surcharge d'imposteurs : dans ce cas, tout ou une partie des imposteurs est utilisée. Le nombre de données d'imposteurs restant largement plus grand que le nombre de données authentiques.

6. Nous pouvons observer que ces points sont également facilement transposables à d'autres modalités dans le cas de la mise à jour en ligne

- Taille équivalente : autant de données d'imposteurs que de données authentiques sont utilisées.
- Surcharge de données authentiques : dans ce cas, la quantité de requêtes authentiques est nettement plus importante que la quantité de requêtes imposteur.

5. L'ordre global d'entrée.
   L'ordre de présentation global des requêtes est un facteur important sur le calcul des performances. Nous pouvons observer quatre ordres différents:
   - Les requêtes authentiques d'abord : toutes les requêtes authentiques sont présentées d'abord, suivies des requêtes d'imposteur. Il s'agit d'un scénario dans lequel l'utilisateur a utilisé le système un certain temps, puis se fait attaquer par un ou plusieurs imposteurs.
   - Les imposteurs d'abord : toutes les requêtes d'imposteur sont présentées d'abord, suivies des requêtes authentiques. Ce scénario permet de vérifier comment le modèle a fortement dévié de la façon normale de taper de l'utilisateur.
   - Présentation aléatoire : Les données authentiques et d'impostures sont sélectionnées de façon aléatoire.
   - Présentation à base de règle. Dans le cas du respect d'un scénario particulier, il peut être utile de préciser des règles de présentation particulières.
6. L'ordre local d'entrée.
   L'ordre local correspond à l'ordre des données d'un individu. Dans tous les cas, la dynamique de frappe étant une biométrie comportementale, il est nécessaire de respecter la chronologie des données des individus. Concernant les données d'attaque, il est possible de les sélectionner de la façon suivante:
   - Sélection aléatoire.
   - Plus près : sur toutes les requêtes possibles, celle qui est la plus proche du modèle est sélectionnée.
   - Personne aléatoire : un imposteur est sélectionné aléatoirement, et, toutes ses requêtes sont présentées de façon chronologique jusqu'à la sélection d'un autre imposteur.
   - Personne plus proche : la méthode est équivalente au point précédent à l'exception que l'individu choisi est celui le plus proche du modèle.
7. La décision de mise à jour.
   Il est également nécessaire de spécifier le facteur de décision de la mise à jour. Plusieurs méthodes peuvent être observées:
   - Succès de l'authentification : à partir du moment où une requête est acceptée par le système, elle est utilisée par le système de mise à jour.
   - Comparaison à un seuil : il s'agit du mécanisme de double authentification présenté précédemment. Le seuil peut être choisi selon les façons suivantes :
     - Indépendante
     - Fraction fixée : le seuil de mise à jour est un pourcentage du seuil d'acceptation
     - Fraction variable : idem que précédemment, mais la fraction peut évoluer au court du temps
   - Basé sur le contenu : La décision de mise à jour est basée sur les exemples d'apprentissage et ceux intégrés dans le système de mise à jour.

8. Le mécanisme de mise à jour.
   Les méthodes de gestions présentées concernent des systèmes à base de galeries. Seeger et Bours proposent:
   - Taille fixe : un exemple du modèle doit être remplacé par la nouvelle requête. Le nombre d'exemples reste donc constant. La méthode de remplacement peut être basée sur :
     - Le plus vieux : le plus vieil exemple est supprimé
     - La distance : l'exemple qui est, en moyenne, le plus éloigné des autres est supprimé.
   - Taille croissante : la nouvelle requête est ajoutée au modèle qui grossi à chaque mise à jour.
   - Taille maximale : les nouveaux exemples sont ajoutés progressivement jusqu'à obtenir une taille fixe.
9. La représentation des performances.

   - Minimisation du FMR
   - Minimisation du FNMR
   - Minimisation de l'EER

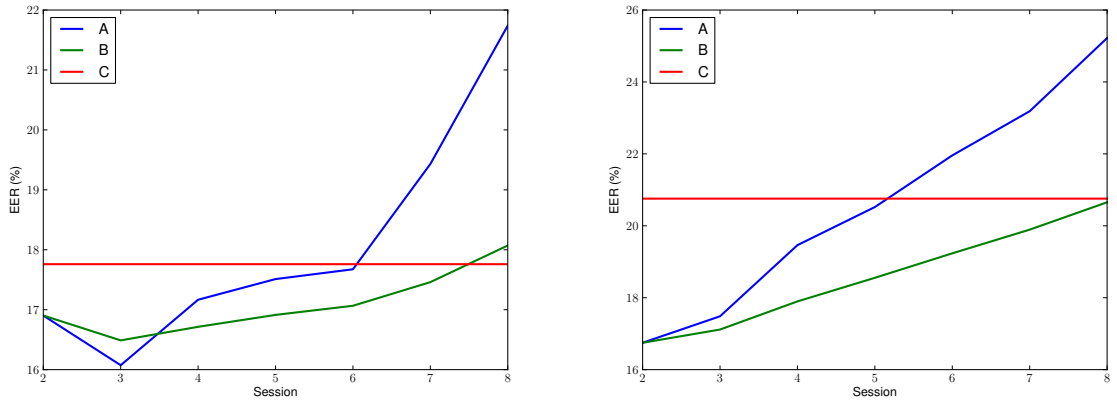
Notre travail consistant majoritairement à l'étude de la mise à jour pour la dynamique de frappe au clavier, nous nous efforcerons de conserver ces termes. Cependant, même si tous les points du protocole sont expliqués convenablement et que l'étude est reproductible, l'évaluation est souvent faite de façon différente. Ainsi, à partir des mêmes jeux de scores différents, il est possible de calculer les métriques d'évaluation différement, d'obtenir des résultats différents et d'avoir des conclusions différentes[7]. La figure 5.14 illustre, sur deux systèmes de mise à jour, qu'utiliser trois façons de calculer la même métrique de performance à partir du même jeu de scores amène une interprétation différente. Se référer à [Giot *et al.*, 2012e] pour plus d'informations.

5.2. Stratégies de mise à jour pour la dynamique de frappe au clavier

5.2.1. Introduction

Effectuer une mise à jour supervisée en ligne augmente considérablement les performances de reconnaissance [Giot *et al.*, 2011b; Kang *et al.*, 2007]. Ce comportement indique la nécessité de prendre en compte l'évolution de la façon de taper afin d'améliorer les performances. Cependant, la mise à jour supervisée est une stratégie inapplicable dans un système opérationnel. Il est nécessaire que la mise à jour s'effectue de façon automatique. C'est pourquoi, dans cette section, nous traitons de la mise à jour semi-supervisée de la référence biométrique en DDF. La mise à jour semi-supervisée ne nécessite aucune contribution d'un opérateur, le traitement est effectué de façon totalement automatique et utilise pour seule information la sortie du classifieur utilisé. Pour comprendre pourquoi il est important d'effectuer cette mise à jour, nous avons analysé la variation des caractéristiques extraites de DDF au cours du temps, sur la base publique possédant le plus grand nombre de données [Killourhy et Maxion, 2009]. Pour chaque utilisateur, nous avons calculé la mesure de dispersion de la donnée biométrique au cours du temps. Nous avons ensuite affiché en figure 5.15 la moyenne de cette mesure, ainsi que son écart type, sur l'ensemble des utilisateurs de la base de donnée. À chaque instant i, l'ensemble de référence est composé des N captures précédentes. Nous avons configuré N à 10, 25 et 50 dans les figures 5.15a, 5.15b,

7. Plus d'informations sont disponibles dans [Giot *et al.*, 2012e]



(a) Système 1

(b) Système 2

FIG. 5.14: À partir du même jeu de scores, il est possible d'utiliser différentes méthodes d'évaluation qui amènent à des conclusions contradictoires. Méthode A : calcul du EER avec les scores de la session courante [Giot *et al.*, 2011a], méthode B : calcul de la moyenne des EER jusqu'à la session courante [Rattani *et al.*, 2011], calcul du EER avec toutes les sessions [Seeger et Bours, 2011]

5.15c. Lors de l'itération m (la première itération commence à la valeur N), nous calculons la valeur moyenne de l'ensemble de données de la fenêtre de taille N : $\boldsymbol{\mu}_m = \frac{1}{N} \sum_{i=m-N}^{m-1} \boldsymbol{s}_i$ ($\boldsymbol{s}_i$ est le i^{e} exemple). Ensuite, pour chaque exemple de la fenêtre, nous calculons sa différence par rapport à la moyenne : $\boldsymbol{d}_i = \boldsymbol{\mu} - \boldsymbol{s}_i, \forall i \in [m-N; m-1]$. La mesure de dispersion est la moyenne de la distance euclidienne de $\boldsymbol{d}_i$: $dispersion = \frac{1}{N} \sum \|\boldsymbol{d}_i\|_2$. Depuis ces figures, nous pouvons conclure que :

- plus un individu effectue de saisies de son mot de passe, plus il le saisit de façon identique (la moyenne de la mesure de dispersion diminue progressivement pour toutes les tailles de fenêtres).
- Ce comportement est généralisable sur l'ensemble des utilisateurs (la valeur de l'écart type diminue également).

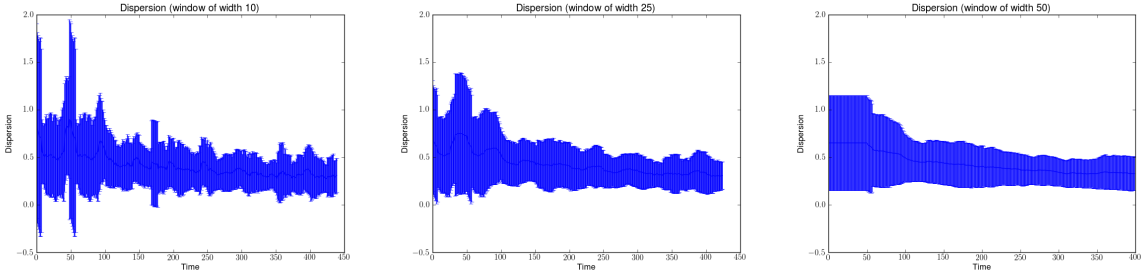
Ces points permettent de confirmer l'intérêt d'utiliser des systèmes de mise à jour de la référence biométrique afin de substituer les anciennes saisies instables par des plus récentes et stables.

5.2.2. Méthode proposée

Cette partie présente notre contribution au problème de mise à jour de la référence biométrique en DDF. La première contribution consiste à présenter un protocole d'évaluation des systèmes de mise à jour lorsque le jeu de données possède suffisamment de données sur différentes sessions. Ce protocole permet d'analyser la performance au cours du temps, contrairement à avoir un indicateur unique comme dans la plupart des études.

5.2.2.1. Méthodologie d'évaluation

Nous proposons différents scénarios pour la procédure de mise à jour.



(a) Fenêtre de taille 10 (b) Fenêtre de taille 25 (c) Fenêtre de taille 50

FIG. 5.15: Évolution de la mesure de dispersion entre la donnée à l'instant i et les données dans l'intervalle $[i-N;i-1]$ avec $N \in \{10,25,50\}$

5.2.2.1.1. Présentation des requêtes En fonction des bases choisies, il est possible d'avoir beaucoup plus de données d'imposteurs que de données de l'utilisateur. Nous supposons que l'imposteur n'a pas en sa possession un jeu de données de la dynamique de frappe de l'utilisateur (obtenu par un keylogger). L'imposteur réalise une attaque connue sous le nom de zero effort ou il utilise sa propre facon de taper au clavier le mot de passe de la victime. Afin de calculer les performances de mise à jour, il est nécessaire de connaitre plusieurs choses :

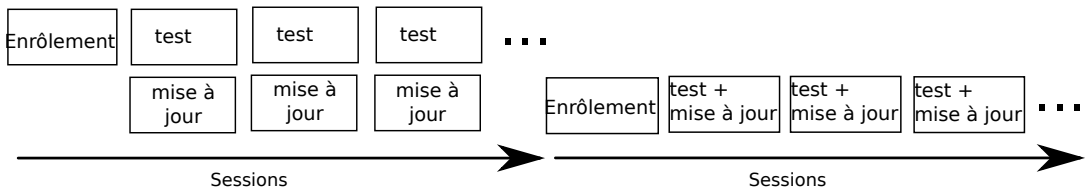
- Quel est l'ordre de présentation des requêtes lors de la validation ?
- Quel est le pourcentage de données d'imposteurs présents ? Plus ce nombre est important, plus la probabilité de faire des mises à jour erronées est importante.
- Quel est le niveau de compétence des imposteurs ? Cette information est particulièrement importante pour les modalités pour lesquelles les requêtes d'imposteurs peuvent être des tentatives d'imitation précises (en signature manuscrite par exemple). Dans le cas de la DDF, toutes les données d'imposture sont des contrefaçons aléatoires (*c.-à-d.*, tout le monde tape la même chose sans chercher à imiter la façon de saisir de la personne à attaquer).

Pour ces raisons, nous considérons qu'il est intéressant de tester différents scénarios par ordre de difficulté. Pour chaque session et utilisateur, nous construisons un jeu de données ordonné, qui est utilisé pour tester la référence biométrique de l'utilisateur lors de cette session. La difficulté du problème est proportionnelle au ratio de données clientes/impostures dans le jeu de données. Chaque jeu de données est généré de la façon suivante :

- le choix d'utiliser une donnée cliente ou d'imposture dépend d'un taux d'imposture fixé ;
- l'ordre de présentation entre une donnée cliente et une donnée d'imposture est totalement aléatoire et lié à la probabilité d'imposture (*c.-à-d.*, avec une probabilité d'imposture de α, pour chaque emplacement dans le jeu de données, la probabilité d'avoir une donnée d'imposture est α et la probabilité d'avoir une donnée authentique est $1-\alpha$.)
- les données sont ordonnées suivant un ordre chronologique afin de conserver le biais dû à l'apprentissage par l'utilisateur de la saisie du mot de passe.

Ainsi, nous ne favorisons aucun type de données (authentique ou d'imposture) et cela permet des scénarios plus réalistes.

5.2.2.1.2. Implication des sessions La première session permet de calculer la référence biométrique de chaque utilisateur (plusieurs captures sont nécessaires pour calculer la référence biométrique en DDF). Comme les bases de DDF proposent plusieurs sessions, nous voulons garder la chronologie des sessions dans les résultats. Nous avons adapté les schémas d'évaluation [Poh *et al.*, 2009] sur deux sessions (une pour générer les références, une pour effectuer la mise à jour,



(a) Stratégie de séparation de la mise à jour et de son évaluation, approche hors-ligne

(b) Stratégie de mise à jour et évaluation jointe, approche en ligne

FIG. 5.16: Stratégies d'évaluation avec plusieurs sessions. La première session sert à l'enregistrement. Il s'agit d'une amélioration des stratégies proposées dans [Poh *et al.*, 2009] en utilisant plusieurs sessions

sans aspect chronologique dans les sessions entre elles) afin qu'ils fonctionnent sur un nombre plus important de sessions. La figure 5.16 présente les stratégies d'évaluation adoptées. La procédure d'évaluation est effectuée une session après l'autre. Lors d'une session, pour chaque utilisateur, nous testons progressivement les données biométriques de son jeu de données (spécifique à cette session) contre sa référence biométrique. Pour chaque comparaison, nous obtenons un score qui est ajouté : soit à la liste des scores intra-classe (si la requête est une requête client), soit à la liste des scores inter-classe (si la requête est une requête d'imposture). L'intérêt de ces deux jeux de scores est de pouvoir calculer les performances du système de mise à jour. En fonction de la valeur de ce score, le mécanisme de mise à jour est effectué (ainsi une donnée d'imposture peut servir pendant la mise à jour). Deux types d'erreurs sont capturées durant l'expérience :

- Les erreurs *en ligne* qui sont calculées à la volée dans une session. Ce taux d'erreur est réaliste, car il correspond à une utilisation réelle du système.
- Les erreurs *hors ligne* qui sont calculées en utilisant les requêtes de la session suivante sur le modèle mis à jour avec toutes les requêtes de la session actuelle. Il s'agit de la méthode la plus couramment utilisée dans la littérature.

Nous avons choisi d'analyser l'évolution du FMR et du FNMR pour chaque session, en fonction d'un seuil d'acceptation donné, ainsi que l'EER. La figure 5.17 présente un résumé des stratégies d'évaluation lorsque plusieurs sessions sont considérées. Comme nous l'avons déjà dit, la difficulté d'un scénario réside principalement dans la proportion de requêtes d'impostures testées contre la référence biométrique (il semble que ce point n'ait pas été considéré dans la littérature) :

- Peu de requêtes d'impostures sont vérifiées contre la référence biométrique : il s'agit d'un scénario facile avec peu d'attaques.
- La même quantité de requêtes d'imposture que de requêtes authentiques sont testées contre la référence biométrique : il s'agit d'un scénario de difficulté moyenne.
- Plus de requêtes d'impostures que de requêtes authentiques sont testées contre la référence : il s'agit d'un scénario difficile avec de nombreuses attaques.

L'ordre de présentation des requêtes étant totalement aléatoire, il est nécessaire d'effectuer l'opération plusieurs fois. Ainsi, les résultats finaux sont constitués d'une moyenne des résultats de plusieurs exécutions.

5.2.3. Base et méthode utilisées

Nous avons choisi d'utiliser la base DSN2009 [Killourhy et Maxion, 2009] car c'est elle qui comporte le plus de sessions. Nous avons utilisé la méthode que nous nommons STAT3 [de Magalhaes *et al.*, 2005] comme système de vérification.

Entrées : les jeux d'exemples sont générés

Entrées : seuil de décision de mise à jour

`/* Calcul des références initiales */`

pour chaque *utilisateur* **faire**
  Calculer la référence de *utilisateur* avec ses exemples de la première session ;

pour *session* = 1 *jusqu'à l'avant dernière session* **faire**
  $intra1 \leftarrow []$;

  $inter1 \leftarrow []$;

  `/* Lance la procédure de mise à jour et calcule les performances en ligne */`

  pour chaque *utilisateur* **faire**
    tant que *le jeu d'exemples de utilisateur n'est pas vide* **faire**
      exemple ← Récupère le prochain exemple de jeu de *utilisateur* ;

      score ← Compare *exemple* au modèle de *utilisateur* ;

      si *exemple appartient utilisateur* **alors**
        Stocke *score* dans $intra1$;

      sinon
        Stocke *score* dans $inter1$;

      si $score \leq threshold$ **alors**
        Applique la procédure de mise à jour à *utilisateur* avec *exemple* ;

  $FMR1$,$FNMR1$ ← Calcule le taux d'erreur en ligne ;

  `/* Calcule les performances hors ligne */`

  $intra2 \leftarrow []$;

  $inter2 \leftarrow []$;

  pour chaque *utilisateur* **faire**
    Utilise $session + 1$ comme session de test ;

    Calcule les scores *intra* et ajoute les à $intra2$;

    Calcule les scores *inter* et ajoute les à $inter2$;

  $FMR2$,$FNMR2$ ← Calcul les taux d'erreurs ;

Affiche FMR,$FNMR$ en fonction des sessions ;

FIG. 5.17: Mécanisme d'évaluation du système de mise à jour

5.2.4. Mise à jour semi-supervisée du modèle

Quelques techniques de mise à jour de la référence biométrique ont été proposées dans l'état de l'art de la DDF. Cependant, elles ont été utilisées dans des cadres supervisés uniquement [Kang *et al.*, 2007]. Notre travail consiste à analyser leur comportement dans un cadre semi-supervisé. Dans notre contexte d'utilisation, des données d'impostures peuvent donc être intégrées à la référence. La procédure de mise à jour n'est appliquée que si la distance de correspondance entre la référence et la requête est inférieure à un *seuil de mise à jour* (un précédent *seuil d'authentification* ayant déclaré la requête comme étant une requête cliente). Un résumé de la procédure d'authentification est présenté en figure 5.18. Deux méthodes de mise à jour de la référence sont utilisées :

- *La fenêtre glissante*. Elle consiste à ajouter la nouvelle donnée biométrique (la requête sélectionnée par le seuil de mise à jour) à la galerie de l'utilisateur, tout en supprimant la donnée la plus ancienne. Le nombre d'exemples dans la galerie reste donc fixe.
- *La fenêtre croissante*. Elle consiste à ajouter le nouvel exemple à la galerie de l'utilisateur. Le nombre d'exemples dans la galerie ne cesse d'augmenter.

score ← Compare *sample* avec *template* ;

si *score* <= *decision_threshold* **alors**
  L'individu est accepté ;

  si *score* ≤ *update_threshold* **alors**
    Met à jour *template* avec *sample* ;

FIG. 5.18: Résumé de la procédure de mise à jour semi-supervisée

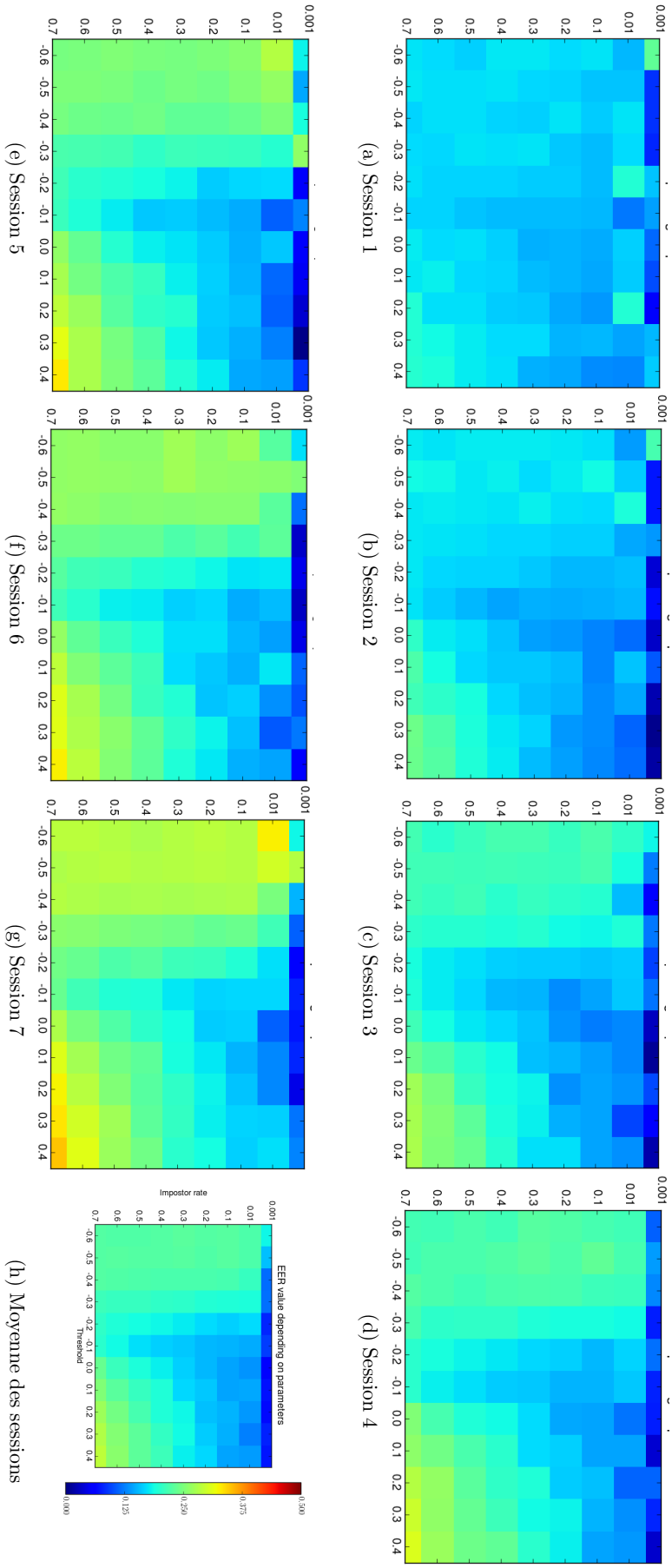
Naturellement, lorsque la galerie est modifiée, il est nécessaire de calculer à nouveau le modèle de l'utilisateur (*c.-à-d.*,, moyenne, écart type, ...).

5.2.5. Résultats

Nous avons calculé l'EER des deux méthodes de mise à jour pour différents scénarios qui correspondent à différents ratios d'imposteurs et différents seuils de décision. Les figures 5.19 et 5.20 présentent respectivement l'EER pour la fenêtre croissante et la glissante. Les couleurs plus proches du bleu foncé représentent un EER de 0%, tandis que les couleurs les plus proches du rouge correspondent à un EER de 50%. Ces figures confirment que :

- Plus le nombre d'imposteurs testés est important, moins les performances sont bonnes (en effet, la probabilité d'inclure des données d'imposteurs est importante). Cela se confirme en analysant la partie inférieure des figures.
- Utiliser un seuil élevé de mise à jour donne de moins bonnes performances (nous utilisons une distance) quand il y a une forte probabilité d'inclure des imposteurs. Cela se confirme en regardant la partie droite inférieure des figures.
- La performance se dégrade avec le temps lorsque le seuil de décision de mise à jour est trop bas (*c.-à-d.*, lorsque la mise à jour de la référence est rarement effectuée). Cela se confirme en analysant la partie gauche de la figure de la dernière session.

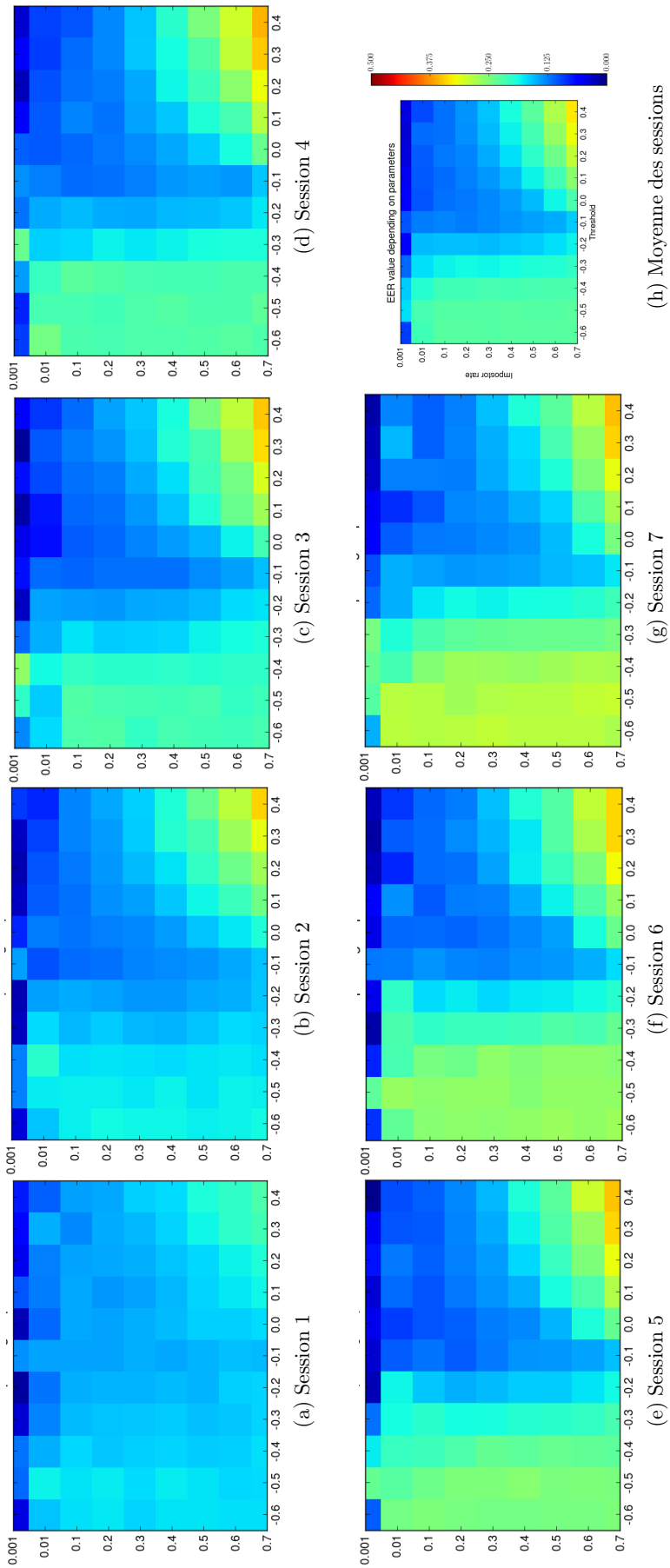
Pour l'expérience suivante, nous avons choisi un scénario avec 20% de données d'impostures, ce qui nous semble raisonnable ; un seuil de mise à jour de 0.0 et un seuil d'acceptation de 0.2. Il s'agit d'un scénario sécurisé, où nous voulons principalement réduire le FMR. Le système de base



(a) Session 1 (b) Session 2 (c) Session 3 (d) Session 4

(e) Session 5 (f) Session 6 (g) Session 7 (h) Moyenne des sessions

FIG. 5.19: EER pour les différentes configurations, pour chaque session avec la fenêtre croissante. La couleure bleue symbolise un EER de 0%. La couleur rouge symbolise un EER de 50%. L'axe des abcisses est le seuil de décision. L'axe des ordonnées est le Ratio d'imposteurs



(a) Session 1 (b) Session 2 (c) Session 3 (d) Session 4

(e) Session 5 (f) Session 6 (g) Session 7 (h) Moyenne des sessions

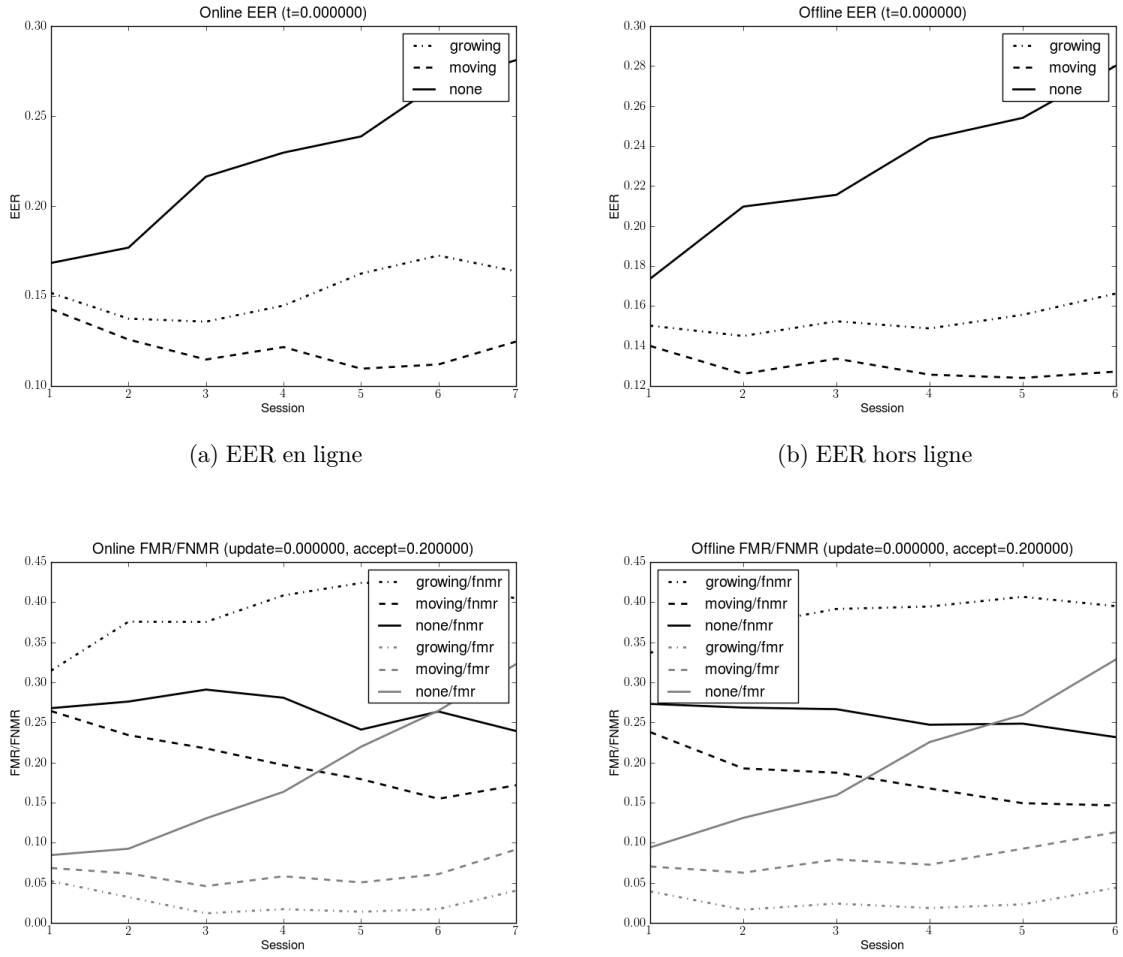
FIG. 5.20: EER pour les différentes configurations, pour chaque session avec la fenêtre glissante.La couleur rouge symbolise un EER de 50%. L'axe des abcisses est le seuil de décision. L'axe des ordonnées est le Ratio d'imposteurs

est un système sans mise à jour de la référence (nous l'avons nommé « sans »). La figure 5.21 présente l'EER, FMR et FNMR de chaque session évaluée de façon en ligne et hors ligne. La session 0 est utilisée pour configurer les références initiales. Comme calculer les performances hors ligne de la session 7 est impossible (nous n'avons pas de session 8), nous avons une session de moins de résultats. Nous pouvons faire les observations suivantes grâce à la figure 5.21 (du moins pour ces paramètres de configuration) :

- Sans système de mise à jour, l'EER se dégrade avec le temps. Cela implique de toujours valider les méthodes de reconnaissance par DDF sur des bases de données capturées sur plusieurs sessions, ainsi que de disposer de mécanisme de mise à jour performant. Si ces conditions préalables ne sont pas établies, comme dans la majorité des études de DDF, les résultats présentés seront meilleurs que dans un cas réel d'utilisation.
- En choisissant comme mesure d'erreur l'EER, la fenêtre croissante donne toujours de meilleurs résultats que ne pas utiliser d'adaptation. La fenêtre glissante donne de meilleurs résultats que la fenêtre croissante. Cela prouve qu'il est impératif d'utiliser des mécanismes de mise à jour en DDF, mais que les stratégies doivent oublier les données trop anciennes. Les données anciennes deviennent rapidement non représentatives de l'utilisateur et les conserver décroit les performances.
- Les résultats peuvent être interprétés différemment en utilisant d'autres types d'erreur que l'EER telles que le couple de FNMR/FMR. Dans le cas de l'EER, nous présentons des résultats calculés en utilisant différents seuils d'acceptation pour chaque session (ce qui est impossible sans intervention humaine dans un environnement réel), tandis que dans le second cas, tous les paramètres sont fixés au début de l'expérience. Les résultats peuvent sembler meilleurs dans le premier cas que dans le second (*cf.*, la différence de comportement de la fenêtre croissante entre les types de taux d'erreurs). Cette différence peut être expliquée par le nombre de scores d'impostures et authentiques différents.
- Avec les seuils sectionnés, le FNMR n'évolue pas beaucoup en n'utilisant pas de mise à jour (il a même tendance à réduire). Le FNMR réduit également en utilisant la fenêtre glissante, tandis qu'il augmente en utilisant la fenêtre croissante. Cela signifie que, dans le cas de la DDF, ne pas utiliser des données sur une trop longue période semble un bon choix.
- Avec les seuils sélectionnés, le FMR augmente considérablement en n'utilisant pas de mise à jour, tandis qu'il diminue considérablement avec n'importe lequel des systèmes de mise à jour.
- Les calculs hors-ligne et en-ligne de EER ne donnent pas de différences fondamentales. La méthode en-ligne semble plus intéressante car elle permet d'avoir une session supplémentaire de résultats, et diminue également le temps de calcul.
- Même si la mise à jour de la référence dans le cas de la DDF est une nécessité, cela reste un problème difficile, principalement en raison des faibles performances de cette modalité. Les distributions des scores authentiques et d'impostures se chevauchent et n'importe quel choix de seuil, implique des erreurs.

5.2.6. Discussion

Certains des résultats sont calculés en utilisant l'EER. Cette valeur donne une bonne idée des performances du système de mise à jour, mais n'est pas réaliste en raison de la configuration des deux seuils. En environnement opérationnel qui correspond à un niveau de sécurité fixé, un autre point de fonctionnement peut être choisi. En considérant les résultats de cette étude, nous pouvons affirmer que l'EER n'est pas une bonne mesure d'erreur pour l'évaluation des systèmes de mise à jour de la référence, car il donne des performances pouvant ne pas être atteintes en



(a) EER en ligne

(b) EER hors ligne

(c) FMR/FNMR en ligne

(d) FMR/FNMR hors ligne

FIG. 5.21: Différentes métriques d'erreur pour un seuil de mise à jour de 0.0 (et un seuil de décision de 0.2 le cas échéant) évaluées en ligne et hors ligne

environnement opérationnel (en raison de la configuration des seuils). Un couple de FNMR, FMR selon une configuration de seuils semble plus réaliste.

Une autre question intéressante est la façon de calculer les résultats. Nous avons choisi de représenter un indice de performance par session. Nous pensons que c'est une meilleure solution plutôt que d'afficher un résultat global pour l'ensemble du jeu de données, car les données biométriques sont acquises à des moments différents et varient avec le temps. Cette information temporelle (les erreurs au cours du temps) permet de voir si les performances augmentent ou diminuent au cours de l'usage du système. De plus, en calculant les performances avec les données d'une session, nous sommes sûrs de ne pas altérer les résultats avec les informations des sessions précédentes. Cependant, une session peut être considérée comme quelque chose d'acquis sur une trop longue ou trop courte période.

Les travaux futurs devraient analyser les façons de quantifier les erreurs des systèmes de mise à jour. La sélection aléatoire des données pour construire le jeu de test d'une session peut compliquer une telle analyse. Pour le moment, présenter les résultats d'évaluation reste un sujet ouvert.

5.3. Mise à jour hybride

5.3.1. Introduction

Nous avons vu que les systèmes d'auto-apprentissage ont un intérêt pour la DDF (section 5.2). Tandis que la section précédente a utilisé l'auto-apprentissage de base pour illustrer l'utilité de la mise à jour pour la DDF, cette section propose un nouveau schéma de mise à jour de la référence biométrique. Nous considérons que notre nouvel algorithme de mise à jour est un mécanisme hybride car il utilise à la fois des mécanismes de l'auto-apprentissage et du co-apprentissage sans pour autant être une méthode d'auto-apprentissage ou co-apprentissage. Au contraire, il casse le schéma de fonctionnement habituel et nous pouvons le voir à la fois comme une modification de la façon de représenter la référence biométrique, et de la façon de mettre à jour la galerie de l'utilisateur (l'ensemble de ses données utilisées pour créer son modèle). Ainsi, la référence biométrique d'un utilisateur est représentée par plusieurs sous-références biométriques qui évoluent en parallèle à l'aide de méthodes de mise à jour différentes. Les contributions à la mise jour de la référence biométrique sont les suivantes :

- Nous proposons un mécanisme original de mise à jour de la référence biométrique qui donne de meilleurs résultats que le système classique d'auto-apprentissage couramment utilisé dans la littérature. Il s'agit d'un système hybride, car :
  - la décision de mise à jour ce fait à l'aide d'un seul et unique score, comme pour les systèmes d'auto-apprentissage ;
  - il fonctionne à l'aide d'un mécanisme de fusion, comme pour les systèmes de co-apprentissage ;
  - la référence biométrique d'un utilisateur est composée de plusieurs sous-références biométriques.
- Nous proposons deux nouvelles métriques qui évaluent l'efficacité des systèmes de mise à jour de la référence biométrique évalués sur des bases de plusieurs sessions.
- Nous évaluons la méthode avec un jeu de données qui contient plus d'exemples par utilisateurs que la plupart des études de l'état de l'art en mise à jour de la référence biométrique.

5.3.2. Proposition d'un système semi-supervisé de mise à jour de la référence biométrique et d'un mécanisme d'évaluation

Cette sous-section présente la méthode de mise à jour proposée, ainsi que les métriques d'évaluation permettant de la comparer aux autres.

5.3.2.1. Mise à jour basée sur la co-évolution de différentes galeries

Alors que nous utilisons un sytème mono-modal, notre contribution est inspirée des systèmes de co-apprentissage [Bhatt *et al.*, 2011; Roli *et al.*, 2007] et des travaux sur les mises à jour de galeries d'utilisateur [Kang *et al.*, 2007; Scheidat *et al.*, 2007]. Dans les travaux précédents de la littérature, la référence biométrique d'un utilisateur est unique : un utilisateur est représenté par une galerie, un exemple, un modèle, ... Dans nos travaux, la référence biométrique est composite : cette méta-référence biométrique contient plusieurs sous-références biométriques qui évoluent à l'aide de différentes méthodes de mise à jour. Cependant, l'authentification est faite à l'aide d'une seule méthode d'authentification. La figure 5.22 résume le système proposé (zone verte) pour une configuration à deux sous-références biométriques par utilisateur (*c.-à-d.*, deux méthodes de mise à jour de références biométrique évoluent en parallèle) et le tableau 5.1 présente la différence entre le système hybride, l'auto-apprentissage et le co-apprentissage.

TAB. 5.1: Différences entre l'auto-apprentissage, le co-apprentissage et le système hybride

	auto-apprentissage	**Co-apprentissage**	**apprentissage hybride**
Une seule modalité	oui	non	oui
Plusieurs types de classifieurs	non	oui (un par modalité)	choix d'implémentation (un par sous-référence ou le même pour tous)
Source de décision de mise à jour	score du classifieur	désaccord entre les deux classifieurs	Score agrégé à partir des scores produits par la comparaison aux sous-références

Le système défini est indépendant des autres composants du système de mise à jour de la référence biométrique (zones rose et bleue de la figure 5.22). Lorsqu'une requête est comparée à la méta-référence biométrique du demandeur, elle est en fait comparée à chacune des sous-références biométriques, puis les scores sont fusionnés afin d'obtenir un score agrégé. Nous n'assumons rien sur la méthode de décision de mise à jour ; elle peut être basée sur un double seuillage, un indice de qualité, ou n'importe quoi d'autre (section 5.1.3.1). Avec une décision basée sur un double seuillage, la décision est prise sur le score agrégé : nous ne savons pas laquelle des sous-références biométriques est responsable de la décision de mise à jour (contrairement aux mécanismes de co-apprentissage pour lesquels c'est une obligation afin de sélectionner la référence à mettre à jour). Lorsque la méta-référence biométrique est mise à jour, nous effectuons la mise à jour de toutes les sous-références, et pas seulement de la moins influente comme dans les systèmes de co-apprentissage, en utilisant la requête acceptée et la méthode de mise à jour spécifique à la sous-référence. Nous nous attendons à ce que les erreurs de mise à jour diminuent. Chaque couple de gestion de la galerie et de calcul de la référence biométrique peut être remplacé par un système de mise à jour en ligne [Bhatt *et al.*, 2011] si les méthodes de mise à jour sont différentes pour chaque sous-référence biométrique, afin de ne pas faire évoluer deux sous-références biométriques identiques. Différentes règles de fusion et différentes techniques de mise à jour de galeries peuvent être utilisées dans cette nouvelle procédure de mise à jour.

5.3.2.2. Proposition de deux nouvelles métriques d'évaluations

Il y a un manque de métriques d'évaluation des mécanismes de mise à jour de la référence biométrique dans la littérature.Rattani, Marcialis, et Roli utilisent le ratio d'exemples d'imposteurs inclus dans la galerie après la mise à jour [Rattani *et al.*, 2008a]. Ces auteurs l'utilisent dans une procédure de mise à jour hors-ligne, alors que nous sommes intéressés par des systèmes en ligne (*c.-à-d.*, le ratio d'imposteurs inclus dans la référence biométrique après la mise à jour peut évoluer après chaque nouvelle présentation de requête). Poh, Kittler, Smith, et Tena expliquent comment estimer les performances d'authentification au cours du temps [Poh *et al.*, 2007]. Cette procédure requiert un jeu de données où les exemples sont uniformément distribués au cours d'un grand intervalle de temps, ce qui n'est jamais le cas lorsque les exemples sont acquis durant plusieurs sessions (beaucoup d'exemples sur une courte période et aucun exemple sur une plus longue période). Pour contourner ces problèmes, nous proposons deux métriques d'évaluation :

– le taux de sélection d'imposteurs pour la mise à jour (*Impostor Update Selection Rate*) (IUSR) qui correspond au ratio d'exemples d'imposteurs inclus dans le processus de mise à jour, par rapport à tous les exemples d'impostures testés ;

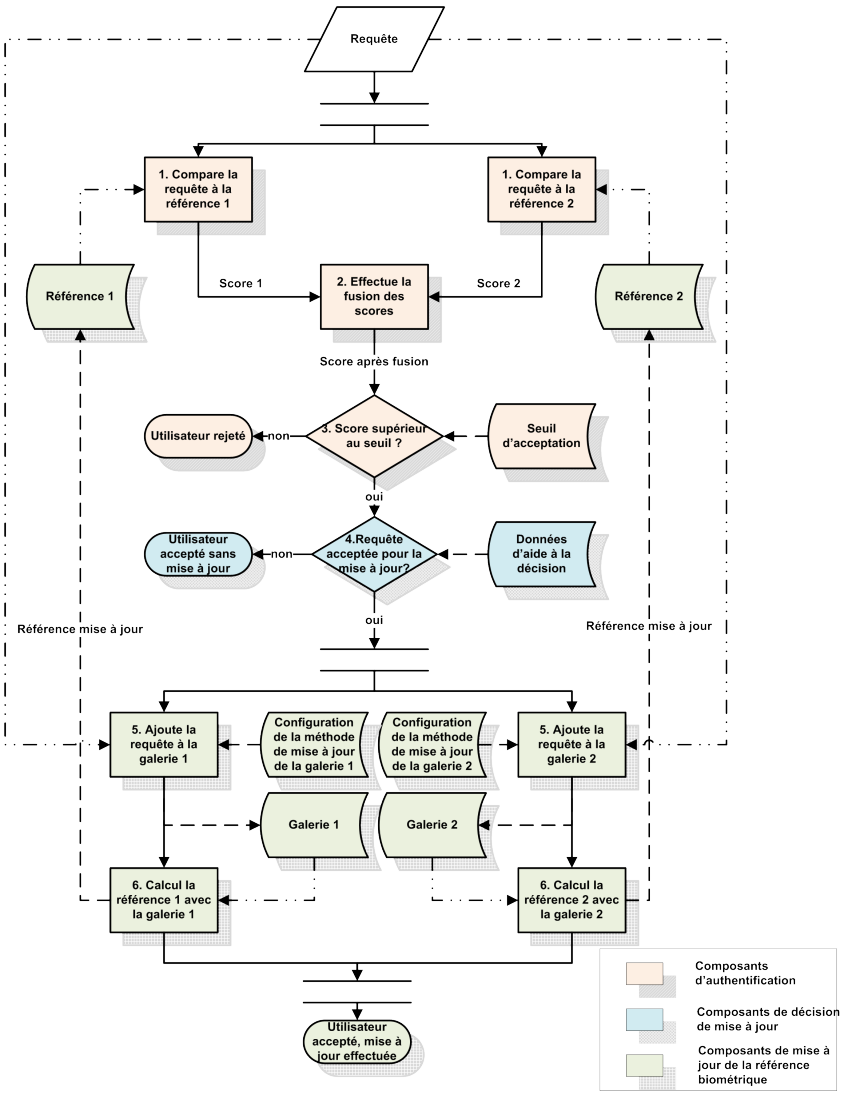


FIG. 5.22: Fonctionnement du système de mise à jour hybride, lorsque deux systèmes de mise à jour sont utilisés, dans un scénario en ligne

– le taux d'oublis de clients pour la mise à jour (*Genuine Update Miss Rate*) (GUMR) qui correspond au ratio d'exemples authentiques jamais utilisés dans le processus de mise à jour, par rapport à tous les exemples authentiques testés.

Notons N_t,N_i,N_c respectivement le nombre total d'exemples testés, le nombre d'exemples d'imposteurs testés et le nombre d'exemples authentiques testés ($N_t = N_i + N_c$). Notons U_i,U_c respectivement le nombre total d'exemples d'impostures sélectionnés dans le processus de mise à jour, et le nombre d'exemples authentiques sélectionnés dans le processus de mise à jour. Les deux taux d'erreurs peuvent être estimés de la façon suivante :

$$\widehat{IUSR} = \frac{U_i}{N_i} \tag{5.7}$$

$$\widehat{GUMR} = \frac{N_c - U_c}{N_c} \tag{5.8}$$

Dans un système sans mécanisme de mise à jour de la référence, $IUSR = 0$ et $GUMR = 1$. Le meilleur système de mise à jour de la référence biométrique tend à avoir l'IUSR aussi proche que

TAB. 5.2: Paramètres de l'expérience

Paramètre	**Valeur**
Modalité	Dynamique de frappe au clavier
Méthode d'authentification	Calcul de distance [de Magalhaes *et al.*, 2005]
Décision de mise à jour	Double seuillage en ligne et semi-supervisée
Seuil de mise à jour	Fixé empiriquement
Systèmes de mise à jour (des sous-références)	Aucun, fenêtre glissante, fenêtre croissante
Nombre de sous-références	2
Fusion des distances issues des comparaisons aux sous-références	Valeur moyenne, valeur minimum (comme nous travaillons avec des distances)
Combinaison des agrégations	(Aucun, Glissante), (Aucun, Croissante), (Glissante, Croissante)
Nombre de sessions	8 pour DSL2009, 5 pour GREYC2009
Respect de la chronologie	Oui
Ordre de présentation	Aléatoire
Quantité d'imposteurs	30% d'exemples d'impostures
Calcul de l'évaluation	En ligne (*c.-à-d.*, joint adapt-and-test strategy [Poh *et al.*, 2009] par session)
Métriques d'évaluation	EER, FNMR, FMR, IUSR, GUMR (scores de la session actuelle, pas de moyenne avec les sessions précédentes)

possible de 0 (inclure les données d'imposture est un gros problème car les références biométriques attireraient plus facilement les imposteurs) et le GUMR le plus faible possible (oublier des données clientes peut être utile si elles sont trop bruitées).

5.3.3. Protocole d'évaluation de notre mécanisme hybride de mise à jour

Cette sous-section présente la configuration définie pour évaluer notre nouveau système de mise à jour de la référence biométrique.

5.3.3.1. Paramètres d'évaluation

Différents paramètres doivent être configurés pour évaluer le système de mise à jour et permettre la reproductibilité de l'étude [Giot *et al.*, 2012e; Seeger et Bours, 2011]. Le tableau 5.2 résume ces différents paramètres. Nous avons évalué ce système sur la DDF car il s'agit à la fois de notre sujet de thèse principal, d'une des modalités ayant la plus grande variabilité intra-classe, et de l'une des modalités ayant des jeux de données avec le plus de sessions. Cependant, la méthode de mise à jour peut fonctionner avec d'autres modalités. Le système est évalué pour chaque session en utilisant seulement les scores calculés durant cette session, cependant, afin de ne pas donner de résultats trop optimistes, nous ne calculons pas la moyenne des performances de chaque session avec les sessions précédentes. Comme l'ensemble des requêtes à tester contre une référence biométrique est construit aléatoirement, les résultats peuvent varier en fonction des exécutions (section 5.2.2). Pour limiter un biais dans les résultats, nous lançons l'expérience 100 fois et moyennons les résultats. Deux jeux de données différents sont utilisés afin de valider notre proposition. Nous utilisons DSL2009 (51 utilisateurs, 400 exemples par utilisateur, 8 sessions) et GREYC2009 (100 utilisateurs, 60 exemples par utilisateur, 5 sessions). La première session est utilisée pour l'enregistrement (*c.-à-d.*, pour générer les sous-références biométriques initiales), et les autres servent pour tester et mettre à jour.

TAB. 5.3: Classement manuel de chaque méthode selon différents critères. Les trois meilleures méthodes sont en gras. * signale une méthode originale

Méthode	FMR	FNMR	IUSR	GMNR	Score	Rang
Double-min-parallèle*	**1**	8	8	**1**	15	**1**
Double-parallèle*	5	**3**	4	4	16	**2**
Croissant	**3**	7	7	**3**	19	**3**
Glissant	7	**1**	**2**	8	21	4
Parallèle-glissant*	8	**2**	**3**	7	21	4
Parallèle-min-glissant*	4	6	6	5	21	4
Parallèle-croissant*	6	4	5	6	21	4
Parallele-min-croissant*	**2**	9	9	**2**	22	8
Aucun	9	5	**1**	9	24	9

5.3.3.2. Configuration du système hybride de mise à jour

Les méthodes de mises à jour sont celles utilisées en section 5.2.4, à savoir : aucune, fenêtre glissante, fenêtre croissante. Trois méthodes d'agrégation sont utilisées :

a *Parallèle-glissante*, avec une référence biométrique jamais mise à jour, et l'autre mise à jour avec la fenêtre glissante.

b *Parallèle-croissante*, avec une référence biométrique jamais mise à jour, et l'autre mise à jour avec la fenêtre croissante.

c *Double-parallèle*, avec une référence biométrique mise à jour avec la fenêtre glissante et l'autre mise à jour avec la fenêtre croissante.

Les méthodes d'agrégation (a) et (b) produisent deux sous références biométriques qui suivent la règle suivante : dans le meilleur des cas, une des sous-références représente la façon initiale de saisir du texte, tandis que la seconde représente la façon la plus récente. Les deux méthodes de fusion de score utilisées sont :

1. la moyenne des scores ;
2. la valeur minimale des scores.

Chaque méthode d'agrégation de galerie est testée avec chaque méthode de fusion de scores.

5.3.3.3. Évaluation du système hybride de mise à jour

Il est important de savoir si le système de mise à jour de la référence biométrique se comporte bien. Nous avons vu qu'en n'utilisant pas de mise à jour, les systèmes de DDF voient leur FNMR diminuer avec le temps, et le FMR augmenter avec le temps. Un bon système de mise à jour est un système où le FNMR et le FMR diminuent ou restent stables au cours du temps. En plus de ces mesures, nous présentons aussi le IUSR et le GUMR qui donnent des informations sur les erreurs de mise à jour. La décision de mise à jour est basée sur la similarité des scores, ainsi FNMR/IUSR et FMR/GUMR peuvent être corrélés. L'EER est également utilisé car il est simple à lire, mais nous ne l'utiliserons pas pour classer les algorithmes.

5.3.4. Résultats expérimentaux

Le scénario de référence sans mise à jour est appelé « aucun » et les scénarios de référence qui utilisent un système de mise à jour sont les gestions de galeries avec fenêtre « glissante » et « croissante ». Nos contributions qui utilisent le système hybride sont « parallèle-glissante », « parallèle-croissante », « double-parallèle », « parallèle-min-glissante », « parallèle-min-croissante », « double-min-parallèle ». Il est bien connu que la décroissance du FNMR d'un

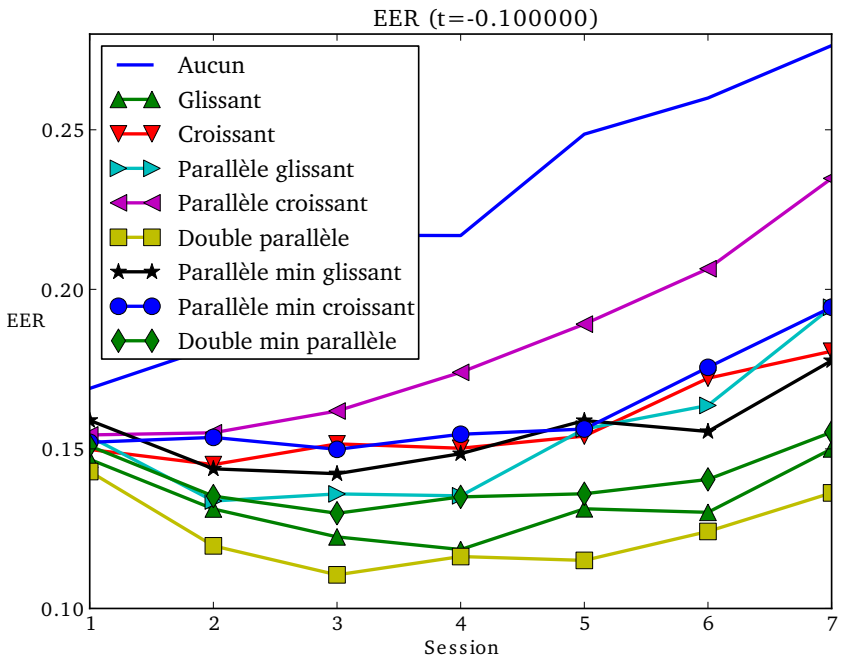
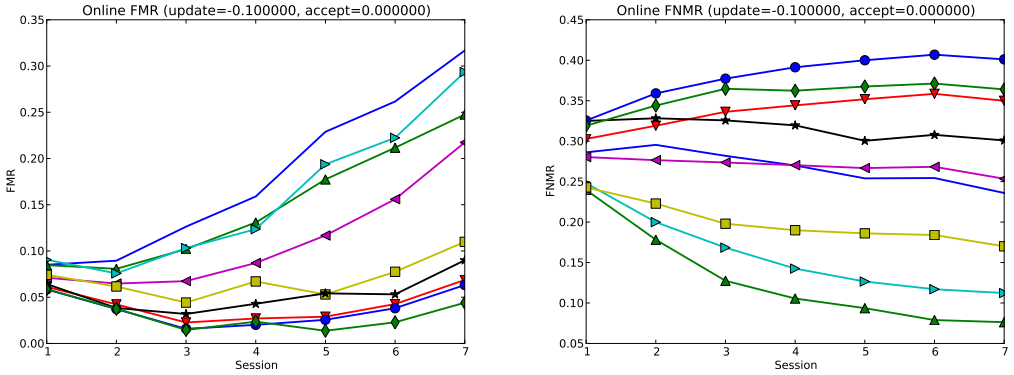


FIG. 5.23: EER au cours des sessions pour chaque système de mise à jour



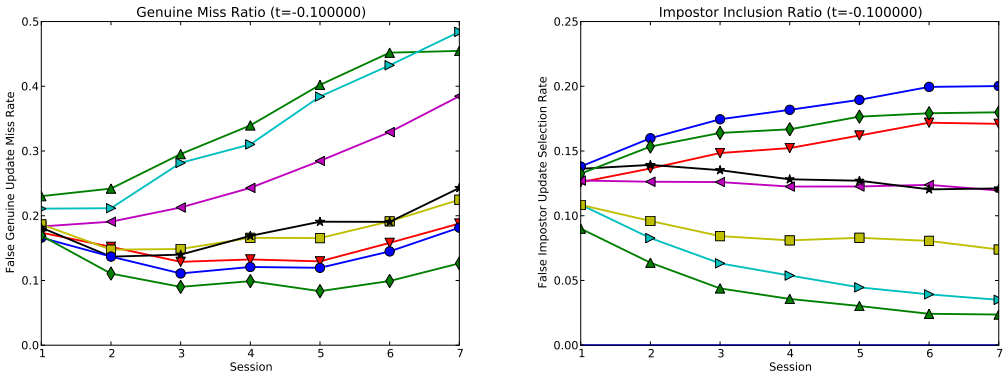
(a) Évolution du taux de faux rejets

(b) Évolution du taux de fausses acceptations

FIG. 5.24: FNMR et FMR au cours des sessions pour chaque système de mise à jour. La légende est identique à celle de la figure 5.23

système biométrique correspond à une augmentation du FMR (et vice versa). Nous pouvons observer un comportement similaire, mais lié avec le temps, en figure 5.24 : les méthodes qui font décroitre le FNMR au cours du temps, tendent à augmenter le FMR au cours du temps. Comme l'évolution de l'EER ne peut pas nous donner une telle information (figure 5.23), nous pensons qu'il est préférable de fournir uniquement le FNMR et FMR afin de voir leur différence d'évolution. De plus, en utilisant un mécanisme à double seuillage, le seuil pour obtenir l'EER peut être incompatible avec le seuil de mise à jour.

La figure 5.24 montre que la méthode « double-parallèle » est la plus appropriée. Il ne s'agit pas de la meilleure méthode en termes de FMR ou FNMR, mais c'est la seule méthode présente dans



(a) Évolution du taux d'oublis de clients pour la mise à jour

(b) Évolution taux de sélection d'imposteurs pour la mise à jour

FIG. 5.25: Erreurs de mise à jour au cours du temps. La légende est identique à celle de la figure 5.23

les meilleures méthodes à chaque fois. Elle doit donc être un bon compromis. Pour confirmer ce point, nous avons ordonné manuellement chaque méthode de mise à jour (triée par performance globale) suivant les taux suivants : FMR, FNMR, IUSR, GUMR. L'EER n'est pas pris en compte en raison des problèmes mentionnés précédement. Pour chaque méthode de mise à jour, nous sommons les rangs des différents critères et les trions suivant ce rang. Les résultats sont présentés en table 5.3. Les deux meilleures méthodes sont « double-parallèle » et « double-min-parallèle ». Il s'agit de notre proposition lorsque nous faisons évoluer en parallèle deux sous-références biométriques en utilisant la fenêtre glissante et la fenêtre croissante. Cela montre le bénéfice de notre méthode, comparée aux méthodes ne faisant pas une évolution parallèle de plusieurs références. Les deux méthodes les moins performantes sont «parallèle-min-croissante» et « aucune ». C'est relativement aisé à comprendre ; dans le premier cas, il y a deux sous-références biométriques : l'initiale qui devient rapidement non représentative et aide à rejeter les exemples authentiques, et la fenêtre croissante qui contient et garde les exemples d'impostures capturés. Ce comportement est expliqué en figure 5.25 où nous voyons que cette méthode est celle attirant le plus grand nombre d'imposteurs. Dans le second cas, aucune mise à jour n'est effectuée.

La figure 5.25 montre que, pour la plupart des méthodes, avoir un haut IUSR implique un faible GUMR (et vice versa), excepté pour « parallèle-min » qui n'est jamais la meilleure méthode, mais est toujours dans les meilleures méthodes. C'est la seule méthode n'attirant pas trop d'imposteurs et ne rejetant pas trop de données authentiques. Le même comportement est observé en figure 5.24. Ainsi, étant la méthode n'ayant pas trop de FMR et FNMR, l'EER est faible comparée aux autres méthodes. La figure 5.25 et la figure 5.24 montrent qu'il y a une forte relation entre FNMR et GUMR, et FMR et IUSR. Cela montre qu'il est nécessaire de réduire le GUMR (respectivement IUSR) pour réduire le FNMR (respectivement FMR). Il y a une limite avec la procédure d'évaluation qui est liée avec la méthode de sélection pour la mise à jour. Étant donné que nous utilisons deux seuils, il est nécessaire de les spécifier. Cela peut être un problème dans un scénario opérationnel (les seuils optimaux peuvent être difficiles à obtenir). Une bonne pratique serait de calculer le seuil d'un point de fonctionnement en utilisant les données d'enregistrement ou des données de validation, et, de calculer le seuil de mise à jour à partir de celui-ci. Nous avons calculé le seuil pour obtenir l'EER avec la première session, utilisé ce seuil pour l'acceptation, et sa valeur divisée par deux pour la mise à jour. Ensuite, nous avons appliqué le mécanisme de mise à jour hybride sur les sessions suivantes. Les résultats obtenus sont similaires.

5.3.5. Discussion

Nous avons proposé un mécanisme hybride de mise à jour de la référence biométrique qui permet de faire évoluer plusieurs références biométriques en parallèle. L'évolution parallèle des sous-références biométriques permet de réduire les erreurs de mise à jour, ainsi que la réduction de performance au cours du temps. La méthode a été validée sur deux bases, pour un système de mise à jour appliqué à la DDF. Bien que la méthode ait été évaluée dans un contexte semi-supervisé en ligne, elle pourrait également être utilisée dans un scénario supervisé en ligne. Il pourrait être intéressant de valider cette proposition dans d'autres contextes et d'autres modalités (la signature par exemple), ainsi qu'avec des classifieurs en ligne à la place des méthodes utilisant une galerie.

5.4. Conclusion de la mise à jour de la référence biométrique

La plupart des modalités biométriques souffrent du problème de vieillissement de la donnée biométrique. Les modalités comportementales sont naturellement plus sensibles à ce phénomène que les modalités morphologiques. Un moyen de prendre en compte ce vieillissement, afin de ne pas avoir de réduction de performances au cours du temps, est d'utiliser des systèmes de mise à jour de la référence biométrique.

La mise à jour de la référence biométrique et la DDF sont deux thématiques totalement différentes, intéressantes mais étudiées par peu de chercheurs. Pour cette raison, il reste encore de nombreuses choses à faire dans ces domaines. Il faut également noter qu'il y a assez peu de papiers de la littérature de la DDF traitant de la mise à jour de la référence biométrique. Les méthodes actuelles restent simplistes et évaluées de façon biaisée. Quant aux papiers sur la mise à jour du modèle, ils traitent généralement le cas des modalités morphologique avec une faible variabilité intra-classe et les méthodes développées ne sont pas applicables à la dynamique de frappe. Il reste donc encore de nombreux travaux à faire dans ce domaine de recherche.

Rappel des contributions de ce chapitre

- La définition d'une procédure d'évaluation de systèmes de mise à jour de référence biométrique à l'aide d'une base séparée en plusieurs sessions.
- L'analyse de l'efficacité de la mise à jour semi-supervisée pour la DDF.
- La proposition d'un schéma d'évaluation de la mise à jour de modèles biométriques incluant deux nouvelles métriques.
- La proposition d'un système de mise à jour hybride, plus efficace que les mécanismes d'auto-apprentissage.

Travaux de l'auteur sur ce thème de travail

Romain Giot : État de l'art de la mise à jour du modèle biométrique. Rapport technique, UMR6072 - GREYC - Groupe de REcherche en Informatique, Image, Automatique et Instrumentation de Caen, janvier 2011. URL `http://hal.archives-ouvertes.fr/hal-00581700/fr`.

Romain Giot, Bernadette Dorizzi et Christophe Rosenberger : Analysis of template update strategies for keystroke dynamics. *In IEEE Symposium Series in Computational Intelligence 2011 (SSCI 2011). Workshop on Computational Intelligence in Biometrics and Identity Management (CIBIM). Special Session on Adaptive Classification Systems for Biometric Recognition.*, pages 21–28, Paris, France, avril 2011. URL `http://hal.archives-ouvertes.fr/hal-00587106/`.

Romain Giot, Christophe Rosenberger et Bernadette Dorizzi : Performance evaluation of biometric template update. *In International Biometric Performance Testing Conference (IBPC 2012)*, pages 1–4, Gaithersburg, MD, USA, mars 2012a. URL `http://biometrics.nist.gov/cs_links/ibpc2012/presentations/Day3/342_goit_supporting_paper.pdf`.

Romain Giot, Christophe Rosenberger et Bernadette Dorizzi : Can chronological information be used as a soft biometric in keystroke dynamics? *In The Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP 2012), Special Session 1: Advances on Biometrics*, pages 7–10, Piraeus, Greece, juillet 2012b. URL `http://hal.archives-ouvertes.fr/hal-00714261`. Acceptance rate: 40/100.

Romain Giot, Christophe Rosenberger et Bernadette Dorizzi : Hybrid template update system for unimodal biometric systems. *In IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS 2012)*, pages 1–7, Washington, District of Columbia, USA, septembre 2012c. IEEE Computer Society. URL `http://hal.archives-ouvertes.fr/hal-00714241`. Acceptance rate : 36/100.

6. Conclusion

Dans cette thèse, nous avons contribué à différents éléments qui permettent d'utiliser la biométrie à *bas coût*, de manière *non intrusive*, *permanente* et *faiblement contrainte pour l'utilisateur* . La Dynamique De Frappe au clavier (DDF) étant une modalité remplissant partiellement ces contraintes, nous l'avons donc sélectionnée et tenté de mettre au point des systèmes basés sur la DDF ayant des performances acceptables et d'apporter différentes contributions à ce thème de recherche grâce à :

1. la création et la diffusion de deux des bases de DDF les plus conséquentes au niveau international, la création d'une méthode de reconnaissance par mot de passe commun, la définition d'une méthode de caractérisation des bases de DDF qui a été appliquée à celles de l'état de l'art ;
2. la création de fonctions de fusion de scores et l'utilisation de la programmation génétique pour augmenter la performance de mécanismes de fusion de scores, la démonstration pour la première fois dans la littérature de la possibilité de reconnaitre le genre et la catégorie d'âge d'un individu selon sa façon de taper au clavier ;
3. la démonstration de l'intérêt d'utiliser des mécanismes de mise à jour semi-supervisée dans les systèmes de DDF, et la création d'une méthode de mise à jour semi-supervisée plus efficace que l'auto-apprentissage classique.

Bilan des contributions majeures

Les premiers travaux présentés concernent essentiellement la DDF et ont été présentés dans le chapitre 3 (*Dynamique de frappe au clavier*). Nous avons effectué l'acquisition de deux des bases biométriques de DDF les plus conséquentes (100 utilisateurs ayant fourni 60 captures sur 5 sessions, et 118 utilisateurs ayant fourni chacun un mot de passe différent). Afin d'harmoniser les évaluations des travaux en DDF, nous avons rendu publique ces bases pour la communauté scientifique. Grâce à la première base, nous avons comparé les travaux de l'état de l'art qui sont capables de travailler avec peu de données d'enregistrement. Nous avons proposé une méthodologie de caractérisation des bases de données biométriques que nous avons appliquée sur toutes les bases publiques. Les résultats montrent que peu de bases sont de qualité suffisante et qu'il est nécessaire d'utiliser cette caractérisation pour créer de nouvelles bases pertinentes. Nous avons proposé une méthode d'authentification par mot de passe partagé qui est plus performante, soit en erreur de reconnaissance, soit en temps de calcul. Ce gain de performance est dû à l'utilisation des données d'impostures lors de la création de la référence biométrique d'un utilisateur.

Les seconds travaux présentés sont consacrés à la multimodalité ainsi qu'à la biométrie douce et ont été présentés dans le chapitre 4 (*Multimodalité et biométrie douce*). Nous avons proposé une méthode d'approximation rapide du taux d'égales erreurs (*Equal Error Rate*) (EER). Le fait que l'EER est obtenu rapidement permet d'accélérer le temps de calcul d'algorithmes d'optimisation qui l'utilisent comme fonction d'évaluation. Cette méthode de calcul a donc un intérêt pour configurer les systèmes de fusion multimodaux le plus rapidement possible. Nous avons montré l'intérêt de fusionner la Reconnaissance Faciale (RF) avec la DDF afin d'obtenir un système d'authentification plus performant tout en étant à bas coût (les ordinateurs portables récents disposent tous d'un clavier et d'une webcam) et non intrusif (capturer la dynamique de frappe au

clavier et une image du visage ne requiert aucune action différente de celle de la saisie d'un mot de passe). Nous avons proposé deux fonctions de fusion simples dont les poids sont générés par Algorithmes Génétiques (AG). Nous avons également proposé l'utilisation de la programmation génétique (PG) pour générer des fonctions de fusion complexes. Les méthodes de fusion que nous avons proposées ont été validées sur des bases multimodales de l'état de l'art, ainsi que sur une base chimérique de DDF et RF et ont obtenu des performances meilleures que la somme pondérée. Nous avons montré qu'il est possible de reconnaitre le genre d'un individu selon sa façon de saisir un texte fixé à l'avance, ainsi que d'un texte quelconque. Les performances sont meilleures dans le cas du texte fixé, mais les résultats restent encourageants pour le texte libre. Nous avons également montré qu'il est possible de reconnaitre la catégorie d'âge d'un individu saisissant un texte quelconque. Ces résultats restent préliminaires et améliorables, mais on peut s'attendre à de nouvelles applications à l'aide de ces techniques.

Pour finir, les travaux sur la mise à jour de la référence biométrique appliquée à la DDF sont présentés dans le chapitre 5 (*Mise à jour de la référence biométrique*). Nous avons montré l'intérêt d'utiliser des mécanismes de mise à jour semi-supervisée pour la DDF en adaptant des techniques de mise à jour supervisée utilisées dans la littérature. Nous avons également effectué une méta-analyse, non présentée dans le chapitre, de l'évaluation de systèmes de mise à jour biométrique. En effet, la façon de calculer les performances d'un système de mise à jour sur un même jeu de score influe la perception du système. Or, des méthodes différentes et incompatibles sont utilisées dans la littérature. Nous avons créé une méthode de mise à jour semi-supervisée hybride qui utilise plusieurs références pour représenter un individu. Elle est plus performante que l'auto-apprentissage couramment utilisé et permet de réduire à la fois les erreurs de reconnaissance et les erreurs de mise à jour. Elle est illustrée dans un contexte de DDF, mais peut être utilisée avec d'autres modalités.

Perspectives

Les travaux effectués lors de cette thèse peuvent encore être poursuivis afin de les améliorer et les étendre.

Nous avons vu que les performances de reconnaissance de méthodes d'authentification par DDF sont plus faibles que celles des modalités majeures. Nous avons montré dans cette thèse différentes méthodes permettant de diminuer le taux d'erreurs, mais nous n'avons pas montré les raisons de ces erreurs. Dans [Giot *et al.*, 2012b] nous montrons que les performances de reconnaissance peuvent dépendre de la taille du mot de passe, ainsi que de son entropie, tandis que [Giot *et al.*, 2012c] laisse supposer que le taux d'échec à l'acquisition (*failure-to-acquire rate*) (FTAR) dépend de la longueur du mot de passe. Il est nécessaire d'analyser plus en profondeur les causes de ces faibles performances, car il est probable que nous pourrions obtenir de meilleures performances en éduquant les utilisateurs de telle façon à ce qu'ils choisissent des mots de passe améliorant la performance en DDF.

Les travaux de biométrie douce étant des travaux préliminaires, il reste de nombreux points à améliorer. Tout d'abord, il serait intéressant de s'intéresser uniquement au texte libre, étant donné que cela permet un plus grand nombre d'utilisations, et d'améliorer les algorithmes d'identification à l'aide de la biométrie douce. Ensuite, nos travaux ont été réalisés avec les paramètres par défaut du Séparateur à Vaste Marge (SVM). On peut s'attendre à obtenir de meilleurs résultats de reconnaissance en cherchant à les optimiser. Cependant, nous pensons qu'il est nécessaire d'effectuer l'acquisition d'une base de données plus conséquente (plus d'utilisateurs, plus de phrases) pour généraliser les résultats de façon plus fiable. Nous n'avons testé que le genre et la catégorie d'âge, cependant la base sur laquelle nous avons travaillé [Bello *et al.*, 2010] propose d'autres informations pour chaque individu : son emploi, sa thématique, s'il est gaucher ou

droitier, la configuration du clavier, le navigateur utilisé, mais la distribution de ses classes n'est pas équilibrée. Il serait donc intéressant d'analyser si ces informations sont pertinentes, puis de voir s'il est possible de les utiliser sous la forme d'un sac de biométrie douce pour identifier les individus [Dantcheva *et al.*, 2011].

Le système de mise à jour hybride pourrait être amélioré en lui permettant de mettre à jour une galerie d'imposture. De cette façon plusieurs sous-références biométriques représentent l'utilisateur et une sous-référence représente les imposteurs. Naturellement, il serait nécessaire de modifier la fonction d'agrégation des scores. Nous pensons qu'il serait possible d'augmenter les performances du système de mise à jour en utilisant un tel procédé.

Nous pensons que les techniques utilisées dans les travaux de traitement du signal peuvent être utiles pour modéliser la déviation de la donnée biométrique au cours du temps (du moins pour la DDF). Des travaux préliminaires sur l'utilisation d'un filtre de Kalman pour filtrer les données biométriques afin de les utiliser dans la mise à jour du modèle ont montré leur intérêt. Cependant, le système est encore améliorable car il souffre de problèmes de stabilité numérique et son comportement n'est pas identique sur toutes les bases de données.

La mise à jour du modèle biométrique est également un sujet important pour les autres modalités. Une piste de travail consisterait à analyser des systèmes de mise à jour dans le cas de la reconnaissance faciale en utilisant des bases de vidéos [Drygajlo *et al.*, 2009]. Chacune de ces vidéos contient le visage d'un individu sur plusieurs mois, voire années. Nous avons déjà collecté une telle base de données d'une quarantaine d'utilisateurs à partir de vidéos postées sur youtube et expérimenté des techniques de super-template SIFT. Ces travaux mériteraient d'être poursuivis. La signature en ligne manuscrite est également sensible au vieillissement de la donnée biométrique. Il pourrait être intéressant d'analyser le comportement de systèmes de mise à jour pour cette modalité biométrique en utilisant la méthode que nous avons développée.

Pour conclure, la DDF a majoritairement été étudiée dans le cadre de l'authentification statique ou dynamique. Cependant, nous avons montré qu'il est possible d'acquérir quelques informations de type biométrie douce. Il est fort probable, qu'à l'aide de bases de données adéquates, d'autres types d'informations peuvent être collectées. Ces informations pourraient avoir une utilité dans des applications sécuritaire, mercatique ou forensics. Il serait également intéressant de découvrir quel type d'informations peuvent être recueillies grâce à la DDF et d'analyser jusqu'à quel point la vie privée de l'utilisateur saisissant du texte sur son clavier peut être mise en péril.

A. Enquête sur l'usage d'un système biométrique

Cette annexe présente le formulaire d'une enquête menée auprès de différentes personnes qui ont testé un système de Reconnaissance Faciale (RF) et un système de reconnaissance par Dynamique De Frappe au clavier (DDF). Les résultats montrent que bien que les performances du système de DDF soient moins bonnes que celles du système de RF, les utilisateurs ont plus confiance au système de DDF. Il s'agit d'une modalité bien acceptée par les utilisateurs. Pour plus d'information, se reporter à [El-Abed, 2011]

Partie I. Caractéristiques socio-démographiques

Date de naissance (année)	...
Genre	☐ masculin ☐ féminin
Dans quel continent habitez-vous?	☐ asie ☐ afrique ☐ europe ☐ amérique du nord ☐ amérique du sud ☐ autre
Niveau d'éducation	☐ universitaire ou grande école ☐ lycée ☐ autre
Statut professionnel	☐ étudiant ☐ salarié ☐ retraité ☐ autre

Partie II. Perception générale des systèmes biométriques

Q_1. Avez-vous déjà entendu parler de systèmes biométriques **(avant notre étude)**?	☐ oui ☐ non
Q_2. Avez-vous déjà utilisé un système biométrique **(avant notre étude)**?	☐ oui ☐ non
Q_3. Avez-vous déjà été personnellement victime d'une fraude d'identité?	☐ oui ☐ non
Q_4. Comment évaluez-vous vos connaissances sur la technologie biométrique?	☐ pas du tout importantes ☐ plutôt pas importantes ☐ plutôt importantes ☐ tout à fait importantes ☐ je ne sais pas
Q_5. Comment évaluez-vous vos connaissances concernant le vol d'identité?	☐ pas du tout importantes ☐ plutôt pas importantes ☐ plutôt importantes ☐ tout à fait importantes ☐ je ne sais pas
Q_6. Selon vous, est-ce que les solutions basées sur un secret (*ex.*, mot de passe) sont une solution pertinente contre la fraude (par ex. commerce électronique)?	☐ pas du tout d'accord ☐ plutôt pas d'accord ☐ plutôt d'accord ☐ tout à fait d'accord ☐ je ne sais pas
Q_7. Selon vous, est-ce que la biométrie peut être une solution pertinente contre la fraude (par ex. commerce électronique)?	☐ pas du tout d'accord ☐ plutôt pas d'accord ☐ plutôt d'accord ☐ tout à fait d'accord ☐ je ne sais pas

Partie III. Perception du système biométrique testé

Q_8. Avez-vous déjà utilisé cette modalité **(avant notre étude)**?	☐ oui ☐ non
Q_9. Ressentez-vous une gêne à utiliser ce système biométrique particulier?	☐ pas du tout gênant ☐ plutôt pas gênant ☐ plutôt gênant ☐ tout à fait gênant ☐ je ne sais pas
Q_{10}. Trouvez-vous que l'utilisation de cette technologie est une atteinte à votre vie privée?	☐ pas du tout intrusive ☐ plutôt pas intrusive ☐ plutôt intrusive ☐ tout à fait intrusive ☐ je ne sais pas
Q_{11}. Trouvez-vous l'utilisation de ce système biométrique facile et agréable?	☐ pas du tout facile ☐ plutôt pas facile ☐ plutôt facile ☐ tout à fait facile ☐ je ne sais pas
Q_{12}. La vérification est-elle rapide?	☐ pas du tout rapide ☐ plutôt pas rapide ☐ plutôt rapide ☐ tout à fait rapide ☐ je ne sais pas
Q_{13}. La réponse pour la vérification est-elle correcte?	☐ jamais ☐ rarement ☐ parfois ☐ toujours ☐ je ne sais pas
Q_{14}. Selon vous, est-ce que le système utilisé peut-il être facilement contourné?	☐ pas du tout d'accord ☐ plutôt pas d'accord ☐ plutôt d'accord ☐ tout à fait d'accord ☐ je ne sais pas
Q_{15}. Seriez-vous prêt à utiliser ce système biométrique dans le futur?	☐ pas du tout d'accord ☐ plutôt pas d'accord ☐ plutôt d'accord ☐ tout à fait d'accord ☐ je ne sais pas
Q_{16}. Si vous êtes prêt à utiliser le système dans le futur, seriez-vous prêt l'utiliser pour gérer le contrôle d'accès logique (accès à un ordinateur) ou physique (accès à un lieu)?	☐ physique ☐ logique
Q_{17}. Avez-vous confiance dans ce système?	☐ non ☐ pas vraiment ☐ plutôt ☐ oui ☐ je ne sais pas
Q_{18}. Quelle est votre appréciation générale de ce système?	☐ pas du tout satisfait ☐ plutôt pas satisfait ☐ plutôt satisfait ☐ tout à fait satisfait ☐ je ne sais pas

B. Rappels statistiques

Le but de cette annexe est de faire un rappel de quelques outils statistiques utiles pour cette thèse. Notons $\mathbf{X}$ une série statistique prenant les valeurs x_1,x_2,...,x_p avec les effectifs n_1,n_2,...n_p, et $n = n_1 + n_2 + ... + n_p$ l'effectif total de cette série.

B.1. Moyenne

La moyenne exprime la grandeur qu'auraient chacun des membres de l'ensemble s'ils étaient tous identiques sans changer la dimension globale de l'ensemble. Elle est donnée par :

$$\mu = \overline{x} = \frac{\sum_{i=1}^{p} n_i x_i}{\sum_{i=1}^{p} n_i} \tag{B.1}$$

B.2. Variance et écart type

La variance caractérise la dispersion d'une distribution ou d'un échantillon. Elle est donnée par :

$$V = \frac{\sum_{i=1}^{p} n_i (x_i - \mu)^2}{\sum_{i=1}^{p} n_i} \tag{B.2}$$

L'écart-type est le nombre σ tel que $\sigma = \sqrt{V}$.

B.3. Coefficient de corrélation linéaire de Pearson

Le coefficient de corrélation entre deux variables aléatoires, $\mathbf{X} = (x_1, ..., x_n)$ et $\mathbf{Y} = (y_1, ..., y_n)$, permet de quantifier la relation de dépendance qui peut exister entres ces variables. Elle est donnée par :

$$\rho = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \overline{y})^2}} \tag{B.3}$$

Le coefficient de corrélation linéaire est compris entre -1 et 1. Plus le coefficient est proche des valeurs extrêmes -1 et 1, plus la corrélation linéaire entre les variables est forte. Les valeurs intermédiaires renseignent sur le degré de dépendance linéaire entre les deux variables. Une corrélation égale à 0 signifie que les variables sont linéairement indépendantes.

B.4. Test de Kruskal-Wallis (KW)

C'est un test non paramétrique utilisé pour tester si k échantillons ($k \geq 2$) peuvent être considérés comme issus de la même distribution. Le test étant non paramétrique, il ne suppose pas que les distributions sont gaussiennes. Les deux hypothèses du test KW sont :

- H_0: la loi ayant généré les données est la même pour tous les échantillons, ils ont la même médianne M contre
- H_1: certains échantillons présentent systématiquement des valeurs plus élevées que d'autres.

$$\begin{cases} H0 : \ M_1 \ = \ M_2 \ = \cdots = \ M_k \\ H1 : \ M_i \ \neq \ M_j \quad \exists \, (i,j) \text{ avec } i \neq j \end{cases} \tag{B.4}$$

Avec des échantillons de taille respective n_1,n_2,$\cdots$,n_k, soit au total N mesures ($N = n_1+n_2+\cdots+n_k$), on calcule les rangs sur la réunion de tous les échantillons, $r_{11},\ldots,r_{n1,1},r_{21},\ldots,r_{n2,1},\ldots,r_{1,k},\ldots,r_{nk,k}$. Le test statistique de Kruskal-Wallis (KW) est ainsi défini par :

$$H \ = \ \frac{12}{N(N+1)} \sum_{i=1}^{k} \ n_i \overline{r}_{i\cdot}^{2} - 3(N+1) \tag{B.5}$$

avec

$$\overline{r}_{i\cdot} = \frac{\sum_{j=1}^{n_i} r_{ij}}{n_i} \tag{B.6}$$

Le critère de décision est défini par :

$$\begin{cases} p - valeur \ \geq \ 0{,}05 & \text{accepter } H_0 \\ sinon & \text{rejeter } H_0 \end{cases} \tag{B.7}$$

où p-valeur est la valeur estimée, pour k échantillons, en utilisant la distribution de probabilité du χ^2 avec k-1 degrés de liberté.

C. Flot maximum et coupe minimum

Les illustrations de cette annexe sont inspirées de `http://www-igm.univ-mlv.fr/~desar/Cours/imac-algo/ch7.pdf`. Dans le problème du flot maximum et de la coupe minimum, nous travaillons avec un graphe orienté et valué. Le graphe est représenté par ses nœuds et ses arêtes : $G = (V,E)$ avec une valuation $c : E \to \mathbb{N}$. Le graphe contient deux nœuds spéciaux : la source s et le puits t tels que $deg_{entrant}(s) = 0$ et $deg_{sortant}(t) = 0$.

C.1. Les réseaux de transport

Le graphe est donc vu comme un réseau de transport quelconque(figure C.1) : chaque arête est capable de transporter une certaine quantité de données[1], et en transporte une quantité inférieure ou égale.

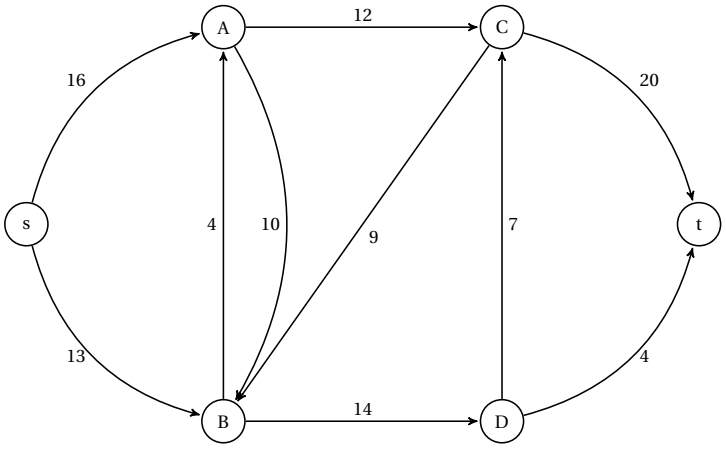


FIG. C.1: Exemple de réseau de transport

Aucune arête n'arrive sur la source s, et aucune arête ne sort du puits t. Un *flot* (figure C.2) est une application ϕ de E dans $\mathbb{N}$ telle que :

- le flot en chaque arête e est inférieur à sa capacité : $\phi(e) \leq (e)$. Il s'agit de la *contrainte de capacité* ;
- pour tout sommet de $G \setminus \{s,t\}$, la quantité de flots entrant dans ce sommet vaut la quantité de flots sortants (il n'y a aucune perte). Il s'agit de la *conservation de flot*.

La valeur $|\phi|$ du flot ϕ est le flot sortant de s (ainsi que le flot entrant de t). Un flot ϕ est dit *saturé*, si sur tout chemin de s à t, il existe une arête e tel que $\phi(e) = c(e)$.

Une *coupe* est une partition de l'ensemble des sommets de G en deux parties disjointes. L'une des parties, Y, contient la source s, tandis que l'autre, Z, contient le puits t. Les propriétés

1. qui dépend du problème concerné, mais l'unité de mesure n'est pas corrélée avec l'algorithme – dans le cas de la biométrie, il peut s'agir du score de comparaison de deux captures

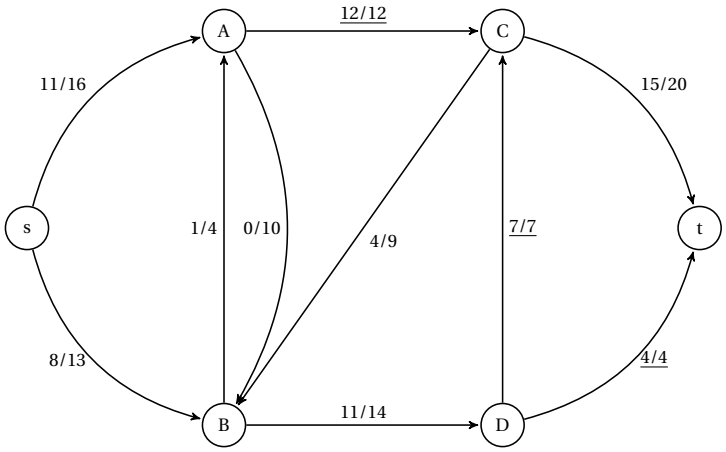


FIG. C.2: Un flot sur le réseau de transport présenté dans la figure C.1. Celui-ci est saturé (*cf.*, arêtes AC, DC et Dt)

suivantes sont vérifiées :

$$\begin{cases} Y \cup Z = G \\ Y \cap Z = \emptyset \\ s \in Y \\ t \in Z \end{cases} \tag{C.1}$$

La somme des valeurs du flot sur les arêtes de Y vers Z moins la somme des valeurs du flot sur les arêtes de Z vers Y vaut aussi $|\phi|$. Cette différence est appelée flot net. La *capacité* d'une coupe est la somme des capacités des arêtes de Y à Z.

C.2. Flot maximum et coupe minimum

L'idée est de chercher le flot maximum en améliorant progressivement le chemin en déterminant le réseau résiduel (figure C.3). Pour chaque arête $e = ij$, $\phi(e) \leq c(e)$; on peut donc augmenter le flot de $c(e) - \phi(e)$, ou le diminuer de $\phi(e)$, donc le faire passer sur l'arête $-e = ji$. Si l'arête n'existe pas, il faut la créer, sinon, on ajoute donc $\phi(e)$ à $c(-e)$.

Il faut ensuite chercher un meilleur chemin de s à t dans le réseau résiduel. Il correspond à une possibilité d'amélioration du flot, en modifiant la valeur du minimum des capacités résiduelles sur le chemin (voir la figure C.4). La figure C.5 présente le flot amélioré.

L'algorithme est effectué itérativement jusqu'à ce qu'il n'y ait plus de chemin entre s et t, voir la figure C.6 (donc pas de chemin améliorant).

Théorème 1. *Voici donc le théorème du flot maximum et de la coupe minimum : Si ϕ est un flot dans un réseau de transport, les trois conditions suivantes sont équivalentes :*

1. *ϕ est un flot maximum ;*
2. *Le réseau résiduel de ϕ ne contient aucun chemin améliorant ;*
3. *Il existe une coupe Y/Z dont la capacité vaut $|\phi|$.*

La condition 3 implique que ϕ est la valeur minimum des capacités des coupes du réseau, puisqu'on sait déjà que ϕ est inférieur à la capacité de n'importe quelle coupe.

Démonstration.

– Pour la propriété 2, si on trouve un chemin améliorant, on peut augmenter ϕ. L'ancien flot n'était donc pas maximum.

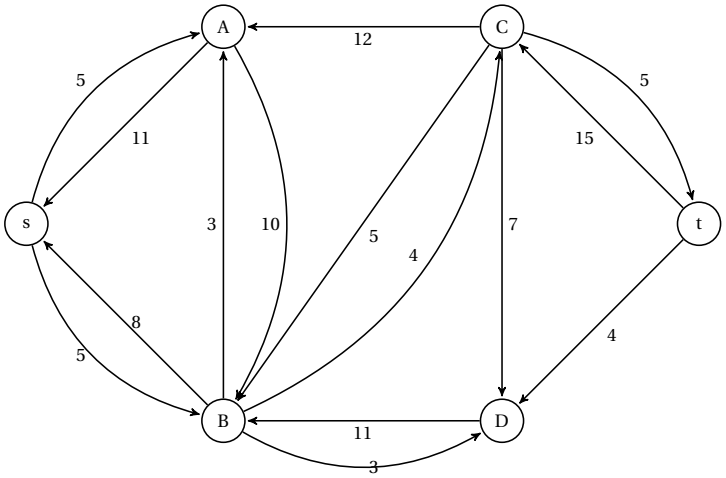


FIG. C.3: Le réseau résiduel sur le réseau de transport présenté dans la figure C.2. Notez la présence de nouveaux arcs (les arcs $-e$) et la modification de la valeur pour les arcs existants (les arcs e)

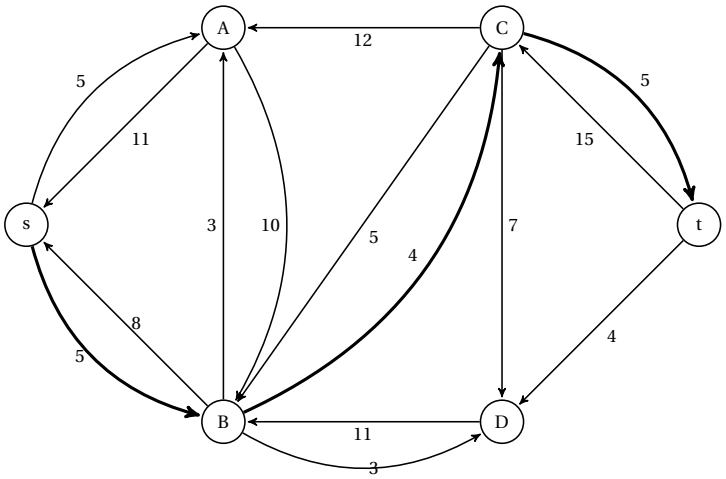


FIG. C.4: Un chemin améliorant (arêtes en gras) depuis le graphe de la figure C.3

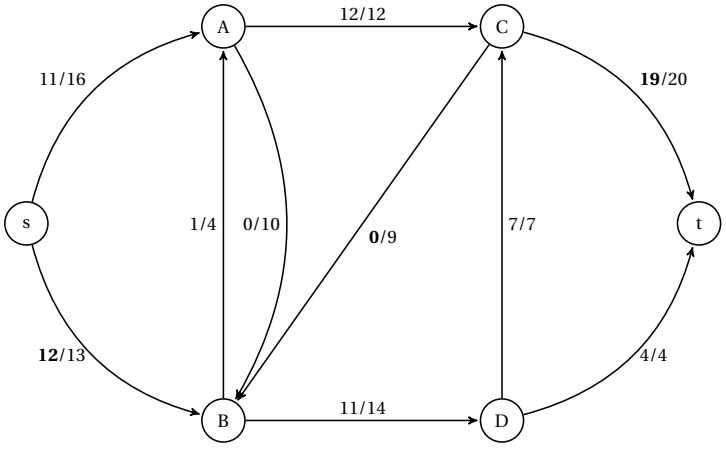


FIG. C.5: Flot après amélioration en utilisant le chemin améliorant de la figure C.4

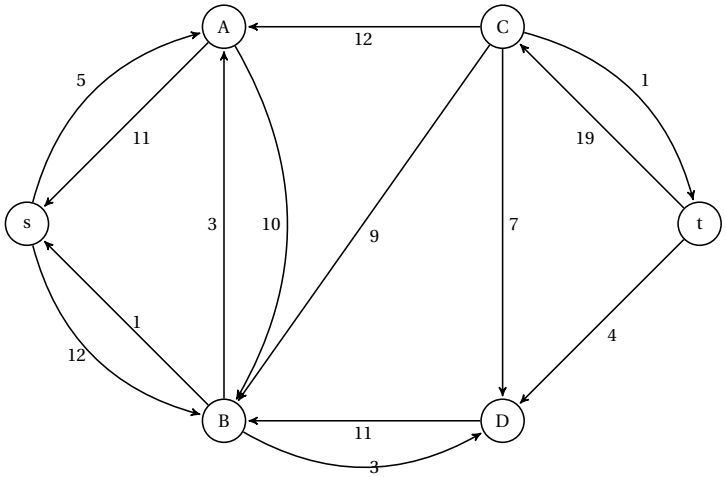


FIG. C.6: Le nouveau réseau résiduel sur le réseau de transport présenté dans la figure C.4

– Pour la propriété 3, s'il n'y a pas de chemin améliorant, on pose Y la composante fortement connexe de s dans le graphe résiduel. Le complémentaire $Z = G \setminus Y$ contient t. Toutes les arêtes entre Y et Z dans le graphe résiduel vont de Y vers Z. Donc $|\phi|$ vaut la capacité de la coupe $Y|Z$.
– Pour la propriété 1, si un flot a comme valeur la capacité d'une coupe, il est nécessairement maximum, puisque tous les flots sont inférieurs à la capacité de n'importe quelle coupe.

□

L'algorithme de Ford et Fulkerson permet de trouver la coupe.

C.3. Application à l'étiquetage de données biométriques

La méthode est la suivante : soit L l'ensemble étiqueté de données biométriques (dans certains cas, nous disposons de données authentiques, mais pas de données imposteurs ; elles sont alors représentées par les données authentiques ayant les scores de comparaison les plus faibles avec les autres [Rattani, 2010]), et d'un ensemble U de données non étiquetées (les données biométriques capturées tout au long de l'utilisation du système). Comme nous sommes dans un cas de classification binaire, nous notons L_+ l'ensemble des données positives, et L_- l'ensemble des données négatives.

1. Nous construisons un graphe pondéré $G = (V,E)$, avec l'ensemble des nœuds $V = L \cup U \cup \{v_+,v_-\}$, et l'ensemble des arêtes $E \subseteq V \times V$. Un poids $c(e)$ est associé à chaque arête $e \in E$. Les nœuds v_+ et v_- sont des *nœuds de classification*, tandis que tous les autres nœuds sont des *nœuds d'exemple*, qu'ils soient étiquetés ou pas.
2. Les nœuds de classification sont connectés par des arcs de poids infini aux exemples ayant la même étiquette que eux :

$$\begin{cases} c(v,v_+) = \infty, & \text{pour tout } v \in L_+ \\ c(v,v_-) = \infty, & \text{pour tout } v \in L_- \end{cases} \tag{C.2}$$

3. Les arcs entre les nœuds d'exemple sont assignés d'un poids basé sur une relation de similarité entre eux. Cette notion de similarité dépend donc de la nature des données

d'exemples. La fonction d'assignation des poids est désignée comme étant la *fonction de pondération des arcs* $c(\cdot)$.

4. Il faut déterminer une coupure minimum (v_+, v_-) pour le graphe. Cela signifie, trouver l'ensemble d'arêtes ayant le poids total minimum lorsque leur retrait déconnecte v_+ de v_-. Cette solution est obtenue en utilisant l'algorithme de flot maximum et coupe minimum pour lequel v_+ est la source, v_- est le puits, et les poids des arêtes sont considérés comme étant des capacités. Supprimer les arêtes au niveau de la coupe partitionne le graphe en deux ensembles de nœuds que nous appelons V_+ et V_-, avec $v_+ \in V_+$ et $v_- \in V_-$. S'il existe plusieurs coupures minimums, nous pouvons configurer l'algorithme pour qu'il choisisse le V_+ le plus petit.
5. Une étiquette positive est assignée à tous les exemples non étiquetés de V_+ (ce sont donc des données biométriques que nous considérons authentiques, et que nous pouvons utiliser dans le nouveau modèle). Une étiquette négative est assignée aux exemples de l'ensemble V_- (cependant, ces données ne nous sont d'aucun intérêt, si nous utilisons uniquement des algorithmes de détection d'anomalies).

L'algorithme part du principe que si les liens entre les nœuds similaires ont un poids important, alors, deux exemples similaires ont une forte probabilité d'être placés dans le même sous-ensemble de nœuds obtenu par la coupure minimum.

Blum et Chawla [2001] ont prouvé qu'étiqueter le graphe avec cette approche a une erreur bornée qui serait l'erreur de validation croisée en utilisant un *leave one out* pour un algorithme des k plus proches voisins. Rattani [2010] a testé ce mécanisme pour la mise à jour de modèles 2D faciaux. Différents nombres de nœuds à lier à chaque nœud ont été utilisés avec les k plus proches voisins avec $k \in \{3,5,10\}$. Le pourcentage de données d'imposteur intégré au modèle mis à jour est moins important qu'en utilisant une méthode d'auto-apprentissage classique. La variabilité des données intégrée est également plus importante. Comme il s'agit d'un problème binaire, un graphe par utilisateur est créé. Comme la galerie initiale ne dispose pas de données d'imposteurs, les données ayant les plus petits scores sont considérées comme imposteurs. La figure C.7 illustre le fonctionnement d'un tel système. Dans le cas d'un mécanisme d'auto-apprentissage classique, les données d'imposteur auraient été incluses dans le modèle. Ici, même s'ils sont les plus proches voisins de données de l'individu, ils sont filtrés par la coupure (représentée par l'arc).

FIG. C.7: Illustration du flot maximum/coupe minimum sur une mise à jour de modèle facial (source Rattani *et al.* [2008a])

D. Les méthodes d'édition

Les méthodes d'édition sont utilisés dans la thèse de Freni [2010] afin de diminuer le nombre d'exemples dans la galerie. Cette annexe présente les méthodes testées.

D.1. Méthodes incrémentales

D.1.1. Condensed NN (CNN)[Hart, 1968]

Il s'agit d'un algorithme incrémental dont le but est de trouver un ensemble édité E inclue dans T tel que l'instance $y \in T$ la plus proche de $x \in E$ possède la même étiquette que x. De cette façon, nous obtenons une performance de classification maximale sur T. Les données initiales sont ordonnées de n'importe quelle façon ; nous disposons de deux ensembles nommés E et Y et procédons comme ci-dessous :

1. Le premier élément est placé dans E.
2. Le second élément est classé par la règle des plus proches voisins en utilisant comme ensemble de référence le contenu actuel de E. Si ce second élément est classé correctement, il est placé dans Y ; autrement il est placé dans E.
3. En procédant itérativement, l'élément i est classé grâce au contenu actuel de E. S'il est classé correctement, il est placé dans Y ; autrement, il est placé dans E.
4. Après une passe complète sur l'ensemble initial, la procédure se répète en utilisant Y, jusqu'à la terminaison qui peut arriver de deux façons différentes :
   a) Y est vide, tous ces membres sont maintenant dans E (dans ce cas, l'ensemble est identique à l'ensemble original)
   b) Une passe complète est faite sur Y sans transfert sur E (les boucles suivantes amèneront au même résultat).
5. Le contenu final de E est utilisé comme nouvelle galerie.

Freni [2010] initialise E avec un exemple de chaque classe.

D.1.2. Selective NN (SNN)[Ritter *et al.*, 1975]

Cette méthode, qui est une évolution de la précédente, essaye de générer l'ensemble E le plus petit possible. Pour arriver à ce résultat, la méthode fait en sorte que les prototypes sélectionnés sont plus proches de la frontière de décision qu'avec la méthode précédente. Freni [2010] initialise E avec un exemple de chaque classe.

D.2. Méthodes décrémentales

D.2.1. Reduced NN (RNN)[Gates, 1971]

Il s'agit d'une méthode décrémentale. Les instances sont progressivement supprimées de E, tant que leur suppression n'implique pas une mauvaise classification sur T. L'algorithme s'arrête

dès qu'il ne peut plus supprimer d'instances de E. Cependant, la méthode n'est pas appliquée à partir de l'ensemble initial, mais à partir du résultat de CNN (voir D.1.1) qui n'est pas considéré comme étant minimal.

1. Copie des données résultats de CNN appliqué à T dans E.
2. Suppression du premier élément de E.
3. Utilisation de E pour classer tous les éléments de T :
   - si tous les éléments sont classés correctement, aller à 4,
   - si un élément est classé de façon incorrecte, remettre l'élément précédemment retiré dans E, puis, aller à 4.
4. Si tous les éléments de E ont été supprimé une fois (en ayant éventuellement été ré-insérés), alors la procédure s'arrête. Sinon, il faut supprimer l'élément suivant et réitérer à 3.

D.2.2. Edited NN (ENN) [Wilson, 1972]

Cette méthode est également décrémentale. Les instances sont progressivement supprimées de E si elles ne correspondent pas avec la majorité de ces k plus proches voisins (ici k=3, les méthodes précédentes utilisent k=1). L'étude montre que le risque de cette méthode approche le risque bayésien.

1. Copie de T dans E.
2. Selection de x, le premier élément de E.
3. Récupération de Y l'ensemble des éléments de T étant les k plus proches voisins de x.
4. m est la classe majoritaire de Y.
5. Si l'étiquette de x est différente de m, alors supprimer x de E.
6. Si tous les éléments de E ont été séléctionnés, alors la procédure s'arrête. Sinon, il faut sélectionner l'élément suivant et réitérer à 2.

E. Séparateurs à Vaste Marge (SVM)

E.1. Fonctionnement du SVM

Le but des machines à vecteurs de support (ou Séparateurs à Vaste Marge (SVM)) et d'effectuer une classification binaire de la donnée requête. L'étiquette de l'exemple ($\{-1,1\}$) dépend du signe de la fonction apprise. L'apprentissage consiste à trouver $y : \mathcal{X} \to \mathbb{R}$ à l'aide de l'ensemble d'apprentissage. La détermination de l'étiquette de l'exemple $\mathbf{x}$ est faite de la façon suivante : $l = sign(y(\mathbf{x}))$. Cette fonction est faite de telle façon à effectuer une séparation linéaire des données d'apprentissage en maximisant la marge. Une séparation non linéaire peut être obtenue en transformant les données d'entrée dans un autre espace des attributs $\mathcal{F}$ ($\Phi : \mathcal{X} \to \mathcal{F}$). Une fonction noyau peut être utilisée à la place du produit cartésien de deux vecteurs permet d'effectuer ce changement de repère de façon transparente.

E.1.1. Séparation (non) linéaire

La séparation (linéaire si $\phi(x) = x$) de deux ensembles s'obtient grâce à la fonction suivante :

$$y(\mathbf{x}) = \mathbf{w}^T \cdot \phi(\mathbf{x}) + b \tag{E.1}$$

Avec cette notation, si $y(\mathbf{x}) > 0$, la classe $+1$ est assignée à $\mathbf{x}$, sinon c'est la classe -1 lui est assignée. L'opération $\phi(\mathbf{x})$ correspond à la transformation de l'espace de représentation des attributs, qui permet de projeter $\mathbf{x}$ dans un espace de plus grande dimension (pouvant même être infini, dans le cas d'un noyau gaussien).

Avec N exemples d'apprentissage, $\mathbf{x}_1, \ldots, \mathbf{x}_N$ étiquetés $t_1, \ldots, t_N$, l'apprentissage consiste à trouver $\mathbf{w}$ et b tels que :

$$\forall n, t_n(\mathbf{w}^T \phi(\mathbf{x}_n) + b) \geq 1 \tag{E.2}$$

E.1.2. Maximisation de la marge

$\mathbf{w}$ et b sont choisis de telle façon que la distance entre la frontière de décision ($\mathbf{w}^T \phi(\mathbf{x}) + b = 0$) et les exemples les plus proches soit maximisée. Cette distance est calculée de la façon suivante :

$$d(\mathbf{x}) = \frac{|y(\mathbf{x})|}{||\mathbf{w}||} \tag{E.3}$$

La distance entre la frontière de décision et les exemples les plus proches est appelée la *marge*. Géométriquement, cette marge vaut $2/||\mathbf{w}||^2$, et le problème de maximisation de la marge peut être écrit de façon équivalent comme un problème de minimisation :

$$argmin_{\mathbf{w},b} \frac{1}{2} ||\mathbf{w}||^2 \tag{E.4}$$

sous la contrainte de (E.2). Pour que les calculs soient plus pratiques, en général, b est intégré à $\mathbf{w}$ et une dimension supplémentaire, toujours égale à 1, est ajoutée aux exemples.

E.1.3. Représentation duale

Étant donné qu'il s'agit d'un problème d'optimisation contraint, il est possible d'obtenir un problème dual en utilisant les multiplicateurs $a_n \geq 0$ de Lagrange [1] (il y a un multiplicateur par exemple d'apprentissage). Le problème dual est le suivant :

$$\begin{cases} \max \sum_{n=1}^{N} a_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} a_n a_m t_n t_m k(\mathbf{x}_n, \mathbf{x}_m) \\ \forall i, a_n >= 0 \\ \sum_{n=1}^{N} a_n t_n = 0 \end{cases} \tag{E.5}$$

La fonction $k(\mathbf{x}_n, \mathbf{x}_m)$ est appelée la fonction noyau et est équivalente à $\phi(\mathbf{x}_n)^T \phi(\mathbf{x}_m)$. C'est grâce à cette fonction noyau qu'il est possible d'obtenir un séparateur non linéaire.

Il s'agit d'un problème de programmation quadratique [2] de dimension N (le nombre d'exemples). Les multiplicateurs de Lagrange a_n sont obtenus en utilisant un résolveur de problèmes quadratiques. $y(\mathbf{x})$ peut également être exprimée avec seulement les multiplicateurs de Lagrange :

$$y(\mathbf{x}) = \sum_{n=1}^{N} a_n t_n k(\mathbf{x}, \mathbf{x}_n) \tag{E.6}$$

L'utilisation des multiplicateurs de Lagrange a transformé le problème d'une somme sur M dimensions (lors du produit vectoriel de (E.1) en une somme sur N points. La quantité N est souvent très largement supérieure à la quantité M, mais cette technique permet de bénéficier du *kernel trick* en ne calculant pas explicitement $\phi(\mathbf{x})$. De plus, les N points ne sont pas utilisés, car seulement quelques multiplicateurs de Lagrange a_n ne sont pas nuls. Il est donc utile de ne stocker que ceux-ci dans le modèle généré. Les exemples $\mathbf{x}_n$ correspondants aux $a_n > 0$ sont appelés *vecteurs supports*.

E.1.4. Marge souple (ou poreuse)

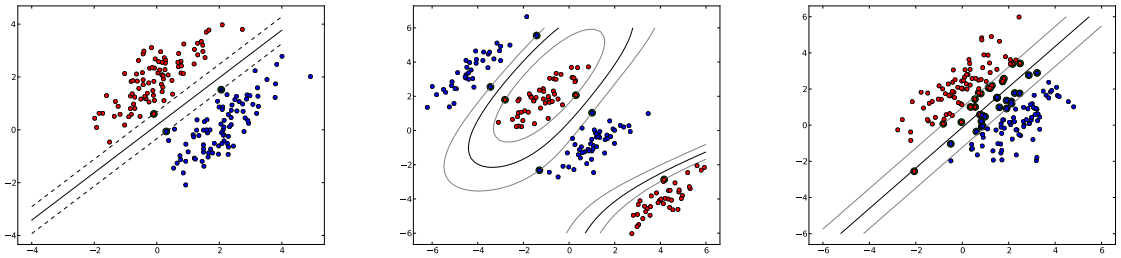
Les explications précédentes ne sont vraies que si le jeu de données d'apprentissage est séparable, ce qui est rarement vrai avec des jeux de données réels. Par conséquent, il est nécessaire de relaxer les contraintes pour avoir un mécanisme qui fonctionne également lorsque les données d'apprentissage ne sont pas séparables. La contrainte (E.2) est donc réécrite en utilisant des variables ressort ξ_n afin d'assouplir les contraintes :

$$\forall n, t_n(\mathbf{w}^T \phi(\mathbf{x}_n) + b) \geq 1 - \xi_n \tag{E.7}$$

Si $\xi_n = 0$, l'exemple d'apprentissage correspondant est classé correctement. Si $0 < \xi_n \leq 1$, l'exemple d'apprentissage est dans la marge, du bon côté de la fonction de décision. Si $\xi_n > 1$, l'exemple d'apprentissage est mal classé. L'équation (E.4) devient :

$$argmin_{\mathbf{w},b} \frac{1}{2} ||\mathbf{w}||^2 + C \sum_{n=1}^{N} \xi_n \tag{E.8}$$

1. http://en.wikipedia.org/wiki/Lagrange_multipliers
2. http://en.wikipedia.org/wiki/Quadratic_programming



(a) Séparation linéaire (b) Séparation non linéaire (c) Fonctionnement des marges poreuses

FIG. E.1: Exemple de fonctionnement de SVM sur trois jeux d'exemple. Les vecteurs de support sont encerclés en vert

$C > 0$ est le paramètre qui contrôle le lien entre la pénalité des variables ressorts et la marge. Une fois de plus, nous pouvons introduire les multiplicateurs de Lagrange, dériver la fonction Lagrangienne en fonction de $\mathbf{w}$, b et ξ_n, et injecter la solution dans la fonction Lagrangienne. On obtient le système suivant à résoudre :

$$\begin{cases} \max \sum_{n=1}^{N} a_n - \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} a_n a_m t_n t_m k(\mathbf{x}_n, \mathbf{x}_m) \\ \forall i, C \geq a_n \geq 0 \\ \sum_{n=1}^{N} a_n t_n = 0 \end{cases} \tag{E.9}$$

Les variables ressorts ont disparu, et, la seule différence avec une marge dure (le cas précédent) est que les a_n ont une borne maximum valant C.

E.1.5. Fonctions noyaux

Plusieurs fonctions noyaux sont communément utilisées, voici les principales.

Polynôme

$$k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j)^d \tag{E.10}$$

Gaussien

$$k(\mathbf{x}_i, \mathbf{x}_j) = exp(-\gamma ||\mathbf{x}_i - \mathbf{x}_j||^2), \text{ avec } \gamma > 0 \tag{E.11}$$

Tangente hyperbolique

$$k(\mathbf{x}_i, \mathbf{x}_j) = tanh(k\mathbf{x}_i \cdot \mathbf{x}_j + c) \tag{E.12}$$

E.2. 2ν-SVM

Le double ν-SVM (2ν-SVM), proposé dans Chew *et al.* [2005] est une variante du SVM calculatoirement plus efficace. Il est plus flexible pendant l'apprentissage et surmonte les problèmes lorsque les classes n'ont pas les mêmes quantités d'apprentissage. Des paramètres additionnels

(ρ, v et C_i) sont introduits, dans l'équation E.8 et l'équation E.7. La formulation devient donc :

$$\begin{cases} \min_{\mathbf{w},b} \frac{1}{2}||\mathbf{w}||^2 - \sum_{i=1}^{N} C_i(v\rho - \xi_i) \\ \forall i, t_i(\mathbf{w}^T\phi(\mathbf{x}_i) + b) \geq \rho - \xi_i, \rho, \xi_i \geq 0 \end{cases} \tag{E.13}$$

ρ est la position de la marge, et v est le paramètre d'erreur qui peut être calculé en utilisant v_+ et v_- qui sont les paramètres d'erreur en apprenant les classes positives et négatives.

$$v = \frac{2v_+v_-}{v_+ + v_-} \tag{E.14}$$

$C_i(v\rho - \xi_i)$ est le coût de l'erreur et C_i est la pénalité d'erreur pour chaque classe qui est calculée comme :

$$C_i = \begin{cases} C_+, & \text{si } y_i = +1 \\ C_-, & \text{si } y_i = -1 \end{cases} \tag{E.15}$$

avec

$$\begin{aligned} C_+ &= \frac{v}{2n_+v_+} \\ C_- &= \frac{v}{2n_-v_-} \end{aligned} \tag{E.16}$$

avec n_+ et n_- respectivement le nombre d'exemples positifs et négatifs. Ainsi, la fonction objective du 2ν-SVM peut être réécrite (formulation duale de Wolfe) :

$$L = \sum_i \alpha_i - \left\{ \frac{1}{2} \sum_{i,j} \alpha_i\alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j) \right\} \tag{E.17}$$

avec $i,j \in 1,\ldots,N$, α_i,α_j les multiplicateurs de Lagrange tels que $0 \leq \alpha_i \leq C_i$, $\sum_i \alpha_i y_i = 0$ et $\sum_i \alpha_i \geq v$. Comme pendant l'apprentissage, il est possible que certains exemples soient bruités ou mal étiquetés, le 2ν-SVM est susceptible d'effectuer des erreurs de classification. Pour contourner ce problème, il est possible d'utiliser des étiquettes souples [Tao *et al.*, 2005].

E.3. 2ν-SVM à étiquettes souples

L'utilisation d'un tel mécanisme peut diminuer les erreurs de classification, ainsi que le nombre de vecteurs supports. Notons z_i l'étiquette souple du i^{e} exemple d'apprentissage $\mathbf{x}_i$. Le 2ν-SVM peut être transformé en 2ν-Soft SVM (2ν-SSVM) comme ceci :

$$\begin{cases} \min_{\mathbf{w},b} \frac{1}{2}||\mathbf{w}||^2 - \sum_{i=1}^{N} C_i(v\rho - \xi_i) \\ \forall i, z_i(\mathbf{w}^T\phi(\mathbf{x}_i) + b) \geq z_i^2(\rho - \xi_i) \end{cases} \tag{E.18}$$

F. Publications de l'auteur

Publications de Romain Giot

4 novembre 2012

Type de publication	Nombre
Thèse	1
Chapitre de livre	4
Article de revue internationale avec comité de rédaction	5
Conférence internationale avec actes et comité de sélection	14
Conférence internationale sans comité de sélection	1
Conférence nationale avec actes et comité de sélection	4
Rapport technique	4
Affiche dans un congrès	2
Logiciel	2

1 Thèse

[1] Romain Giot. "Contributions à la dynamique de frappe au clavier : multibiométrie, biométrie douce et mise à jour de la référence". Thèse de doct. Université de Caen, oct. 2012.

2 Chapitre de livre

[1] Mohamad El-Abed, Romain Giot, Baptiste Hemery, Julien Mahier et Christophe Rosenberger. "Évaluation des performances d'un système biométrique". Dans : *Traitement du signal et de l'image pour la biométrie*. Sous la dir. d'Amine Nait-Ali et Régis Fournier. Hermes Science Publications, sept. 2012. Chap. 11.

[2] Romain Giot, Baptiste Hemery, Estelle Cherrier et Christophe Rosenberger. "La multibiométrie". Dans : *Traitement du signal et de l'image pour la biométrie*. Sous la dir. d'Amine Nait-Ali et Régis Fournier. Hermes Science Publications, sept. 2012. Chap. 9.

[3] Romain Giot, Mohamad El-Abed et Christophe Rosenberger. "Keystroke Dynamics Overview". Dans : *Biometrics / Book 1*. Sous la dir. de Dr. Jucheng Yang. T. 1. InTech, juil. 2011. Chap. 8, p. 157–182. URL : http://www.intechopen.com/articles/show/title/keystroke-dynamics-overview.

[4] Fouad Cherifi, Baptiste Hemery, Romain Giot, Marc Pasquet et Christophe Rosenberger. "Behavioral Biometrics for Human Identification : Intelligent Applications". Dans : sous la dir. de Liang Wang et Xin Geng. IGI Global, 2009. Chap. Performance Evaluation Of Behavioral Biometric Systems, p. 57–74. DOI : 10.4018/978-1-60566-725-6.ch003.

3 Article de revue internationale avec comité de rédaction

[1] Mohamad El-Abed, Baptiste Hemery, Romain Giot et Christophe Rosenberger. "Evaluation of Biometric Systems : A study of users' acceptance and satisfaction". Dans : *International Journal of Biometrics* 4.3 (2012), p. 265–290.

[2] Romain Giot, Mohamad El-Abed et Christophe Rosenberger. "Fast computation of the performance evaluation of biometric systems : application to multibiometric". Dans : *ELSEVIER International journal on Future Generation Computer Systems. HPCS 2010 special issue on Recent Developments in High Performance Computing and Security* (2012). Impact Factor : 2.365. DOI : 10.1016/j.future.2012.02.003. URL : http://hal.archives-ouvertes.fr/hal-00674526/.

[3] Romain Giot et Christophe Rosenberger. "A New Soft Biometric Approach For Keystroke Dynamics Based On Gender Recognition". Dans : *International Journal of Information Technology and Management (IJITM). Special Issue on : "Advances and Trends in Biometrics by Dr Lidong Wang* 11.1/2 (2012). Impact Factor : 0.727, p. 35–49. DOI : 10.1504/IJITM.2012.044062.

[4] Romain Giot et Christophe Rosenberger. "Genetic Programming for Multibiometrics". Dans : *ELSEVIER International journal on Expert Systems With Applications* 39.2 (fév. 2012). Impact factor : 1.924, p. 1837–1847. doi : 10.1016/j.eswa.2011.08.066.

[5] Romain Giot, Mohamad El-Abed, Baptiste Hemery et Christophe Rosenberger. "Unconstrained Keystroke Dynamics Authentication with Shared Secret". Dans : *ELSEVIER International journal on Computers & Security* 30.6-7 (sept. 2011). Impact Factor : 1.488, p. 427–445. doi : 10.1016/j.cose.2011.03.004.

4 Conférence internationale avec actes et comité de sélection

[1] Romain Giot, Christophe Charrier et Maxime Descoteaux. "Local Water Diffusion Phenomenon Clustering From High Angular Resolution Diffusion Imaging (HARDI)". Dans : *IAPR International Conference on Pattern Recognition (ICPR)*. **Oral presentation**, **IAPR Travel Stipend**. IAPR. Tsukuba, Japan, nov. 2012, p. 1–5. url : http://hal.archives-ouvertes.fr/hal-00713993.

[2] Romain Giot, Mohamad El-Abed et Christophe Rosenberger. "Web-Based Benchmark for Keystroke Dynamics Biometric Systems : A Statistical Analysis". Dans : *The Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP 2012), Special Session 1 : Advances on Biometrics*. Acceptance rate : 40/100. Piraeus, Greece, juil. 2012, p. 11–15. doi : DOI10.1109/IIHMSP.2012.10. url : http://hal.archives-ouvertes.fr/hal-00714251.

[3] Romain Giot, Alexandre Ninassi, Mohamad El-Abed et Christophe Rosenberger. "Analysis of the Acquisition Process for Keystroke Dynamics". Dans : *Proceedings of the 11th International Conference of the Biometrics Special Interest Group*. Acceptance rate : 29/100. Darmstadt, Germany : GI-Edition, sept. 2012, p. 123–134. url : http://hal.archives-ouvertes.fr/hal-00730381.

[4] Romain Giot, Christophe Rosenberger et Bernadette Dorizzi. "Can Chronological Information Be Used As A Soft Biometric In Keystroke Dynamics ?" Dans : *The Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP 2012), Special Session 1 : Advances on Biometrics*. Acceptance rate : 40/100. Piraeus, Greece, juil. 2012, p. 7–10. doi : DOI10.1109/IIHMSP.2012.9. url : http://hal.archives-ouvertes.fr/hal-00714261.

[5] Romain Giot, Christophe Rosenberger et Bernadette Dorizzi. "Hybrid Template Update System for Unimodal Biometric Systems". Dans : *IEEE International Conference on Biometrics : Theory, Applications and Systems (BTAS 2012)*. Acceptance rate : 36/100. Washington, District of Columbia, USA : IEEE Computer Society, sept. 2012, p. 1–7. url : http://hal.archives-ouvertes.fr/hal-00714241.

[6] Mohamad El-Abed, Romain Giot, Baptiste Hemery, Christophe Charrier et Christophe Rosenberger. "A SVM-Based Model for the evaluation of biometric sample quality". Dans : *IEEE Symposium Series in Computational Intelligence 2011 (SSCI 2011). CIBIM workshop*. Paris, France, avr. 2011, p. 115–122. doi : 10.1109/CIBIM.2011.5949212.

[7] Mohamad El-Abed, Romain Giot, Baptiste Hemery, Jean-Jacques Schwartzmann, Patrick Lacharme et Christophe Rosenberger. "Towards the Security Evaluation of Biometric Authentication Systems". Dans : *Proceedings of the 2011 International Conference on Security Science and Technology (ICSST 2011)*. Jan. 2011, p. 1–7.

[8] Romain Giot, Bernadette Dorizzi et Christophe Rosenberger. "Analysis of Template Update Strategies for Keystroke Dynamics". Dans : *IEEE Symposium Series in Computational Intelligence 2011 (SSCI 2011). Workshop on Computational Intelligence in Biometrics and Identity Management (CIBIM). Special Session on Adaptive Classification Systems for Biometric Recognition*. Paris, France, avr. 2011, p. 21–28. doi : 10.1109/CIBIM.2011.5949216. url : http://hal.archives-ouvertes.fr/hal-00587106/.

[9] Mohamad El-Abed, Romain Giot, Baptiste Hemery et Christophe Rosenberger. "A study of users' acceptance and satisfaction of biometric systems". Dans : *44th IEEE International Carnahan Conference on Security Technology (ICCST'10)*. San Jose, California, USA, oct. 2010, p. 1–10. doi : 10.1109/CCST.2010.5678678.

[10] Romain Giot, Mohamad El-Abed et Christophe Rosenberger. "Fast Learning For Multibiometrics Systems Using Genetic Algorithms". Dans : *The International Conference on High Performance Computing & Simulation (HPCS 2010)*. Caen, France : IEEE Computer Society, juin 2010, p. 1–8. doi : 10.1109/HPCS.2010.5547127. url : http://hal.archives-ouvertes.fr/hal-00503096/en.

[11] Romain Giot, Baptiste Hemery et Christophe Rosenberger. "Low Cost and Usable Multimodal Biometric System Based on Keystroke Dynamics and 2D Face Recognition". Dans : *IAPR International Conference on Pattern Recognition (ICPR)*. Acceptance rate : 54/100. IAPR. Istanbul, Turkey, août 2010, p. 1128–1131. DOI : 10.1109/ICPR.2010.282. URL : http://hal.archives-ouvertes.fr/hal-00503103/en.

[12] Romain Giot, Mohamad El-Abed et Christophe Rosenberger. "GREYC Keystroke : a Benchmark for Keystroke Dynamics Biometric Systems". Dans : *IEEE International Conference on Biometrics : Theory, Applications and Systems (BTAS 2009)*. Acceptance rate : 56/100. Washington, District of Columbia, USA : IEEE Computer Society, sept. 2009, p. 1–6. DOI : 10.1109/BTAS.2009.5339051. URL : http://hal.archives-ouvertes.fr/hal-00432768/en/.

[13] Romain Giot, Mohamad El-Abed et Christophe Rosenberger. "Keystroke Dynamics Authentication For Collaborative Systems". Dans : *The IEEE International Symposium on Collaborative Technologies and Systems (CTS)*. Acceptance rate : 59/100. Baltimore, Maryland, USA : IEEE Computer Society, mai 2009, p. 172–179. DOI : 10.1109/CTS.2009.5067478. URL : http://hal.archives-ouvertes.fr/hal-00432764/en.

[14] Romain Giot, Mohamad El-Abed et Christophe Rosenberger. "Keystroke Dynamics With Low Constraints SVM Based Passphrase Enrollment". Dans : *IEEE International Conference on Biometrics : Theory, Applications and Systems (BTAS 2009)*. Acceptance rate : 56/100. Washington, District of Columbia, USA : IEEE Computer Society, sept. 2009, p. 1–6. DOI : 10.1109/BTAS.2009.5339028. URL : http://hal.archives-ouvertes.fr/hal-00432775/en.

5 Conférence internationale sans comité de sélection

[1] Romain Giot, Christophe Rosenberger et Bernadette Dorizzi. "Performance Evaluation of Biometric Template Update". Dans : *International Biometric Performance Testing Conference (IBPC 2012)*. Gaithersburg, MD, USA, mar. 2012, p. 1–4. URL : http://biometrics.nist.gov/cs_links/ibpc2012/presentations/Day3/342_goit_supporting_paper.pdf.

6 Conférence nationale avec actes et comité de sélection

[1] Romain Giot, Christophe Rosenberger et Bernadette Dorizzi. "Reconnaissance du genre par analyse de dynamique de frappe au clavier sur texte libre". Dans : *7ème Conférence sur la Sécurité des Architectures Réseaux et Systèmes d'Information (SAR-SSI 2012)*. Sous la dir. de Mohammed Achemlal et Christophe Rosenberger. Acceptance Rate : 62/100. Cabourg, France, mai 2012, p. 111–118. URL : http://hal.archives-ouvertes.fr/hal-00700618.

[2] Mohamad El-Abed, Romain Giot, Christophe Charrier et Christophe Rosenberger. "Evaluation of Biometric Systems : An SVM-Based Quality index". Dans : *Norsk informasjonssikkerhetskonferanse (NISK)*. Sous la dir. de Patrick Bours. Acceptance rate : 56/100. Gjøvik, Norway, nov. 2010, p. 57–68. URL : http://tapironline.no/last-ned/355.

[3] Baptiste Hemery, Romain Giot et Christophe Rosenberger. "Sift Based Recognition of Finger Knuckle Print". Dans : *Norsk informasjonssikkerhetskonferanse (NISK)*. Acceptance rate : 56/100. Gjøvik, Norway, nov. 2010, p. 45–56. URL : http://tapironline.no/last-ned/354.

[4] Romain Giot, Mohamad El-Abed et Christophe Rosenberger. "Authentification faiblement contrainte par dynamique de frappe au clavier". Dans : *Conférence Reconnaissance des Formes et Intelligence Artificielle (RFIA 2010)*. Acceptance rate (oral) : 23.65/100. Caen, France, jan. 2010. URL : http://hal.archives-ouvertes.fr/hal-00472618/en.

7 Rapport technique

[1] Romain Giot. *Keystroke Dataset Denoising*. Rap. tech. Unpublished report. UMR6072 - GREYC - Groupe de REcherche en Informatique, Image, Automatique et Instrumentation de Caen, fév. 2012.

[2] Romain Giot. *Bilan de fin de première année de thèse*. Rap. tech. École Doctorale SIMEM, sept. 2011.

[3] Romain Giot. *État de l'art de la mise à jour du modèle biométrique*. Rap. tech. UMR6072 - GREYC - Groupe de REcherche en Informatique, Image, Automatique et Instrumentation de Caen, jan. 2011. URL : http://hal.archives-ouvertes.fr/hal-00581700/fr.

[4] Romain Giot. *État de l'art de la dynamique de frappe au clavier*. Rap. tech. UMR6072 - GREYC - Groupe de REcherche en Informatique, Image, Automatique et Instrumentation de Caen, nov. 2009.

8 Affiche dans un congrès

[1] Romain Giot. *Static Keystroke Dynamics Authentication*. Poster at the 8th Summer School for Advanced Studies on Biometrics for Secure Authentication. Juin 2011.

[2] Mohamad El-Abed, Romain Giot et Christophe Rosenberger. *Evaluation of Biometric Systems*. Poster at HPCS 2009. 2009.

9 Logiciel

[1] Romain Giot et Christophe Rosenberger. *Reconnaissance du genre par dynamique de frappe au clavier*. Logiciel. Dépôt logiciel : IDDN.FR.001.430027.000.S.P.2010.000.10800. Sept. 2010.

[2] Romain Giot et Christophe Rosenberger. *GREYC Keystroke : un logiciel de démonstration et de création de benchmark de dynamique de frappe au clavier*. Logiciel. Juin 2009.

Bibliographie

Heikki Ailisto, Elena Vildjiounaite, Mikko Lindholm, Satu-Marja Mäkelä et Johannes Peltola : Soft biometrics–combining body weight and fat measurements with fingerprint biometrics. *Pattern Recognition Letters*, 27(5):325 – 334, 2006. *Cité page 76*

M. Akila, V. Suresh Kumar, N. Anusheela et K. Sugumar : A novel feature subset selection algorithm using artificial bee colony in keystroke dynamics. *In* Kusum Deep, Atulya Nagar, Millie Pant et Jagdish Chand Bansal, éditeurs : *Proceedings of the International Conference on Soft Computing for Problem Solving (SocProS 2011) December 20-22, 2011*, volume 131 de *Advances in Intelligent and Soft Computing*, pages 813–820. Springer Berlin / Heidelberg, 2012. ISBN 978-81-322-0490-9. URL `http://dx.doi.org/10.1007/978-81-322-0491-6_74`. *Cité page 24*

Luís A. Alexandre : Gender recognition: a multiscale decision fusion approach. *Pattern Recognition Letters*, "February" 2010. *Cité page 77*

Lorene Allano, Bernadette Dorizzi et Sonia Garcia-Salicetti : Tuning cost and performance in multi-biometric systems: A novel and consistent view of fusion strategies based on the sequential probability ratio test (sprt). *Pattern Recognition Letters*, 31(9):884 – 890, 2010. ISSN 0167-8655. URL `http://www.sciencedirect.com/science/article/pii/S0167865510000437`. *2 citations pages 63 et 78*

Jeffrey D. Allen : An analysis of pressure-based keystroke dynamics algorithms. Mémoire de D.E.A., Southern Methodist University, Dallas, TX, mai 2010. *2 citations pages 21 et 32*

A.S. Anagun : Development of committee neural network for computer access security system. *Lecture Notes in Computer Science*, 3982:11, 2006. *Cité page 53*

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. *IEEE Transactions on Signal Processing*, 53(2 Part 2):851–855, 2005. *Cité page 107*

M.F. Balcan, A. Blum, P.P. Choi, J. Lafferty, B. Pantano, M.R. Rwebangira et X. Zhu : Person identification in webcam images: An application of semi-supervised learning. *In ICML 2005 Workshop on Learning with Partially Classified Training Data*, volume 2, page 6. Citeseer, 2005. *Cité page 119*

A. Bargiela et W. Pedrycz : *Granular computing: an introduction*. Springer, 2003. ISBN 1402072732. *Cité page 127*

P. N. Belhumeur, J. P. Hespanha et D. J. Kriegman : Eigenface vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997. *Cité page 77*

Luciano Bello, Maximiliano Bertacchini, Carlos Benitez, Juan Carlos, Pizzoni et Marcelo Cipriano : Collection and publication of a fixed text keystroke dynamics dataset. *In XVI Congreso Argentino de Ciencias de la Computacion (CACIC 2010)*, 2010. *3 citations pages 95, 96, et 154*

Steven S. Bender et Howard J. Postley : Key sequence rhythm recognition system and method, April 2007. URL `http://www.freepatentsonline.com/7206938.html`. *2 citations pages 19 et 20*

K.S. Benli, R. Duzagac et M.T. Eskil : Driver recognition using gaussian mixture models and decision fusion techniques. *In ISICA 2008*, 2008. *Cité page 6*

F. Bergadano, D. Gunetti et C. Picardi : User authentication through keystroke dynamics. *ACM Transactions on Information and System Security (TISSEC)*, 5(4):367–397, 2002. *7 citations pages 18, 22, 23, 26, 28, 30, et 31*

Maximiliano Bertacchini, Carlos Benitez et Pablo Fierens : User clustering based on keystroke dynamics. *In XVI Congreso Argentino de Ciencias de la Computacion (CACIC 2010)*, 2010. *Cité page 96*

H.S. Bhatt, S. Bharadwaj, R. Singh, M. Vatsa, A. Noore et A. Ross : On co-training online biometric classifiers. *In Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–7, 2011. *3 citations pages 127, 144, et 145*

Régis Bigot et Patricia Croutte : La diffusion des technologies de l'information et de la communication dans la societe franÇaise. Rapport technique, Centre de recherche pour l'étude et l'observation des conditions de vie, 2011. *Cité page 1*

S. Bleha, C. Slivinsky et B. Hussien : Computer-access security systems using keystroke dynamics. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 12 (12):1216–1222, 1990. *7 citations pages 17, 26, 27, 28, 30, 31, et 51*

SA Bleha et MS Obaidat : Dimensionality reduction and feature extraction applications inidentifying computer users. *IEEE transactions on systems, man and cybernetics*, 21(2):452–456, 1991. *3 citations pages 24, 25, et 121*

A. Blum et S Chawla : Learning from labeled and unlabeled data using graph mincuts. *In MACHINE LEARNING-INTERNATIONAL WORKSHOP THEN CONFERENCE-*, pages 19–26, 2001. *2 citations pages 119 et 165*

Marcus Brown et Samuel Joe Rogers : User identification via keystroke characteristics of typed names using neural networks. *Int. J. Man-Mach. Stud.*, 39:994–1014, 1993. *Cité page 53*

John W. Carls : *A FRAMEWORK FOR ANALYZING BIOMETRIC TEMPLATE AGING AND RENEWAL PREDICTION*. Thèse de doctorat, AIR FORCE INSTITUTE OF TECHNOLOGY, 2009. *Cité page 111*

John W. Carls, Richard Raines, Michael Grimaila et Steven Rogers : Biometric enhancements: Template aging error score analysis. *In 8th IEEE conference series on Automatic Face and Gesture Recognition (FG2008)*, 2008. *Cité page 112*

K. Chang, K.W. Bowyer, S. Sarkar et B. Victor : Comparison and combination of ear and face images in appearance-based biometrics. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(9):1160–1165, 2003. *Cité page 68*

W. Chang : Keystroke biometric system using wavelets. *In ICB 2006*, pages 647–653. Springer, 2006a. *Cité page 24*

W. Chang : Reliable keystroke biometric system based on a small number of keystroke samples. *Lecture Notes in Computer Science*, 3995:312, 2006b. *Cité page 25*

Yi-Wei Chen et Chih-Jen Lin : Combining svms with various feature selection strategies. Rapport technique, Department of Computer Science National Taiwan University Taipei 106 Taiwan, 2005. *Cité page 24*

Fouad Cherifi, Baptiste Hemery, Romain Giot, Marc Pasquet et Christophe Rosenberger : *Behavioral Biometrics for Human Identification: Intelligent Applications*, chapitre Performance Evaluation Of Behavioral Biometric Systems, pages 57–74. IGI Global, 2009. *3 citations pages 36, 42, et 47*

Hong-Gunn Chew, Cheng-Chew Lim et Robert E. Bogner : An implementation of training dual-nu support vector machines. *In* Panos M. Pardalos, Donald W. Hearn, Liqun Qi, Koklay Teo et Xiaoqi Yang, éditeurs : *Optimization and Control with Applications*, volume 96 de *Applied Optimization*, pages 157–182. Springer US, 2005. ISBN 978-0-387-24255-2. URL `http://dx.doi.org/10.1007/0-387-24255-4_7`. *Cité page 171*

S. Cho et S. Hwang : Artificial rhythms and cues for keystroke dynamics based authentication. *In IAPR International Conference on Biometrics*, volume 5, pages 626–632, 2006. *Cité page 23*

Sung-zoon Cho et Dae-hee Han : Apparatus for authenticating an individual based on a typing pattern by using a neural network system, November 2000. URL `http://www.freepatentsonline.com/6151593.html`. *Cité page 19*

Sungzoon Cho : System and method for performing user authentication based on keystroke dynamics, December 2006. URL `http://www.freepatentsonline.com/20060280339.html`. *Cité page 19*

Sungzoon Cho, Dae Hee Han, Chigeun Han et Hyung-Il Kim : Web-based keystroke dynamics identity verification using neural network. *Journal of organizational computing and electronic commerce*, 10(4):295–307, 2000. *2 citations pages 18 et 53*

Sungzoon Cho et Seong Seob Hwang : Method and system of detecting account sharing based on behavior patterns, February 2009. URL `http://www.freepatentsonline.com/20090049555.html`. *Cité page 19*

K. Choi, H. Choi et J. Kim : Fingerprint mosaicking by rolling and sliding. *In Audio-and Video-Based Biometric Person Authentication*, pages 260–269. Springer, 2005. *Cité page 66*

Chih chung Chang et Chih-Jen Lin : Libsvm: a library for support vector machines, 2001. *3 citations pages 56, 90, et 97*

N. L. Clarke et S. M. Furnell : Authenticating mobile phone users using keystroke analysis. *International Journal of Information Security*, 6:1–14, 2007. *Cité page 30*

N.L. Clarke et S.M. Furnell : Advanced user authentication for mobile devices. *computers & security*, 27:109–119, 2006. *4 citations pages 20, 24, 30, et 31*

David Cohn, Les Atlas et Richard Ladner : Improving generalization with active learning. *Machine Learning*, 15:201–221, 1994. ISSN 0885-6125. URL `http://dx.doi.org/10.1023/A:1022673506211`. 10.1023/A:1022673506211. *Cité page 120*

O. Coltell, JM Badfa et G. Torres : Biometric identification system based on keyboard filtering. *In IEEE International Carnahan Conference on Security Technology*, pages 203–209, 1999. *2 citations pages 30 et 31*

A. Conklin, G. Dietrich et D. Walz : Password-based authentication: A system perspective. *In Proceedings of the 37th Hawaii International Conference on System Sciences, Hawaii*, 2004. *Cité page 12*

H. Crawford : Keystroke dynamics: Characteristics and opportunities. *In Privacy Security and Trust (PST), 2010 Eighth Annual International Conference on*, pages 205–212. IEEE, 2010. *Cité page 46*

A. Dantcheva, C. Velardo, A. D'Angelo et J.L. Dugelay : Bag of soft biometrics for person identification. *Multimedia Tools and Applications*, 51(2):739–777, 2011. *3 citations pages 76, 77, et 155*

J. Daugman : Combining multiple biometrics, 2000. URL `http://www.cl.cam.ac.uk/~jgd1000/combine/combine.html`. *Cité page 75*

Brent L. Davis, Peeyush Gandhi, Shailesh B.and Jaiswal, James R. Lewis et Fang Wang : Method and system for establishing a biometrically enabled password, March 2009. URL `http://www.freepatentsonline.com/7506174.html`. *Cité page 19*

T. de Magalhaes, K. Revett et H. Santos : Password secured sites: stepping forward with keystroke dynamics. *In International Conference on Next Generation Web Services Practices*, 2005. *2 citations pages 137 et 147*

L. Didaci, G. Marcialis et F Roli : Modelling frr of biometric verification systems using the template co-update algorithm. *In Lecture Notes in Computer Sciences, ICB 2009*, 2009. *Cité page 118*

P. Domingos et M. Pazzani : On the optimality of the simple bayesian classifier under zero-one loss. *Machine Learning*, 29(2-3):103–130, 1997. *Cité page 75*

Hiroshi Dozono, Shinsuke Itou et Masanori Nakakuni : Comparison of the adaptive authentication systems for behavior biometrics using the variations of self organizing maps. *INTERNATIONAL JOURNAL OF COMPUTERS AND COMMUNICATIONS*, 1(4):108–116, 2007. *3 citations pages 21, 127, et 128*

A. Drygajlo, W. Li et K. Zhu : Q-stack aging model for face verification. *In Proc. 17th European Signal Processing Conference (EUSIPCO 2009)*, 2009. *4 citations pages 106, 107, 116, et 155*

Mohamad El-Abed : *Évaluation des systèmes biométriques.* Thèse de doctorat, Université de Caen, 2011. Rapport interne. *2 citations pages 9 et 157*

Mohamad El-Abed, Romain Giot, Baptiste Hemery et Christophe Rosenberger : A study of users' acceptance and satisfaction of biometric systems. *In 44th IEEE International Carnahan Conference on Security Technology (ICCST'10)*, pages 1–10, San Jose, California, USA, octobre 2010. *Cité page 58*

Mohamad El-Abed, Baptiste Hemery, Romain Giot et Christophe Rosenberger : Evaluation of biometric systems: A study of users' acceptance and satisfaction. *International Journal of Biometrics*, 4(3):265–290, 2012. *Cité page 10*

N. El Gayar, S.A. Shaban et S. Hamdy : Face recognition with semi-supervised learning and multiple classifiers. *In Proceedings of the 5th WSEAS International Conference on Computational Intelligence, Man-Machine Systems and Cybernetics*, pages 296–301. World Scientific and Engineering Academy and Society (WSEAS), 2006. ISBN 9608457564. *3 citations pages 117, 118, et 128*

W.E. Eltahir, MJE Salami, A.F. Ismail et W.K. Lai : Design and evaluation of a pressure-based typing biometric authentication system. *EURASIP Journal on Information Security, Article ID*, 345047(2008):14, 2008. *Cité page 21*

Clayton Epp : Identifying emotional states through keystroke dynamics. Mémoire de D.E.A., University of Saskatchewan, Saskatoon, CANADA, 2010. *2 citations pages 11 et 35*

Julian Fierrez et Javier Ortega-Garcia : *Handbook of Biometrics*, chapitre On-line signature, pages 189–209. Springer US, 2008. *Cité page 6*

Jugurta R. Montalvão Filho et Eduardo O. Freire : On the equalization of keystroke timing histograms. *Pattern Recognition Letters*, 27:1440–1446, 2006. *5 citations pages 26, 28, 29, 30, et 31*

B. Freni, G. Marcialis et F. Roli : Replacement algorithms for fingerprint template update. *In International Conference on Image Analysis and Recognition (ICIAR 2008)*, pages 884–893. Springer, 2008. *2 citations pages 124 et 125*

Biagio Freni : *Template Editing and Replacement: novel methods for biometric Template Selection and Update.* Thèse de doctorat, University of Cagliari, 2010. *2 citations pages 123 et 167*

R. Gaines, W. Lisowski, S. Press et N. Shapiro : Authentication by keystroke timing: some preliminary results. Rapport technique, Rand Corporation, 1980. *7 citations pages 6, 16, 17, 22, 23, 30, et 31*

John D. Garcia : Personal identification apparatus, November 1986. URL `http://www.freepatentsonline.com/4621334.html`. *Cité page 17*

S. Garcia-Salicetti, M.A. Mellakh, L. Allano et B. Dorizzi : Multimodal biometric score fusion: the mean rule vs. support vector classifiers. *In Proc. EUSIPCO*, 2005. *Cité page 73*

G. Gates : The reduced nearest neighbor rule (corresp.). *Information Theory, IEEE Transactions on*, 18(3):431–433, 1971. ISSN 0018-9448. *Cité page 167*

Michael Gilfix, Foluso Olaiya Okunseinde et Tyron Jerrod Stading : Method and apparatus for detecting password attacks using modeling techniques, June 2008. URL `http://www.freepatentsonline.com/7386892.html`. *Cité page 19*

Romain Giot : État de l'art de la mise à jour du modèle biométrique. Rapport technique, UMR6072 - GREYC - Groupe de REcherche en Informatique, Image, Automatique et Instrumentation de Caen, janvier 2011. URL `http://hal.archives-ouvertes.fr/hal-00581700/fr`. *Cité page 108*

Romain GIOT, Bernadette DORIZZI et Christophe ROSENBERGER : Analysis of template update strategies for keystroke dynamics. *In IEEE Symposium Series in Computational Intelligence 2011 (SSCI 2011). Workshop on Computational Intelligence in Biometrics and Identity Management (CIBIM). Special Session on Adaptive Classification Systems for Biometric Recognition.*, pages 21–28, Paris, France, avril 2011a. URL `http://hal.archives-ouvertes.fr/hal-00587106/`. Acceptance Rate : ??? *Cité page 135*

Romain GIOT, Mohamad EL-ABED, Baptiste HEMERY et Christophe ROSENBERGER : Unconstrained keystroke dynamics authentication with shared secret. *Computers & Security*, 30(6-7):427–445, septembre 2011b. Impact Factor: 1.488. *3 citations pages 33, 52, et 134*

Romain GIOT, Mohamad EL-ABED et Christophe ROSENBERGER : Greyc keystroke: a benchmark for keystroke dynamics biometric systems. *In IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS 2009)*, pages 1–6, Washington, District of Columbia, USA, septembre 2009. IEEE Computer Society. URL `http://hal.archives-ouvertes.fr/hal-00432768/en/`. Acceptance rate : 56/100. *4 citations pages 32, 38, 46, et 47*

Romain GIOT, Mohamad EL-ABED et Christophe ROSENBERGER : Fast learning for multibiometrics systems using genetic algorithms. *In The International Conference on High Performance Computing & Simulation (HPCS 2010)*, pages 1–8, Caen, France, juin 2010a. IEEE Computer Society. URL `http://hal.archives-ouvertes.fr/hal-00503096/en`. *2 citations pages 83 et 84*

Romain GIOT, Mohamad EL-ABED et Christophe ROSENBERGER : Keystroke dynamics overview. *In* Dr. Jucheng YANG, éditeur : *Biometrics / Book 1*, volume 1, chapitre 8, pages 157–182. InTech, juillet 2011c. URL `http://www.intechopen.com/articles/show/title/keystroke-dynamics-overview`. *Cité page 6*

Romain GIOT, Mohamad EL-ABED et Christophe ROSENBERGER : Fast computation of the performance evaluation of biometric systems: application to multibiometric. *Future Generation Computer Systems. HPCS 2010 special issue on Recent Developments in High Performance Computing and Security*, 2012a. URL `http://hal.archives-ouvertes.fr/hal-00674526/`. Impact Factor: 2.365. *Cité page 83*

Romain GIOT, Mohamad EL-ABED et Christophe ROSENBERGER : Web-Based Benchmark for Keystroke Dynamics Biometric Systems: A Statistical Analysis. *In The Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP 2012)*, 2012b. *4 citations pages 13, 33, 116, et 154*

Romain GIOT, Baptiste HEMERY et Christophe ROSENBERGER : Low cost and usable multimodal biometric system based on keystroke dynamics and 2d face recognition. *In IAPR International Conference on Pattern Recognition (ICPR)*, pages 1128–1131, Istanbul, Turkey, août 2010b. IAPR. URL `http://hal.archives-ouvertes.fr/hal-00503103/en`. Acceptance rate: 54/100. *Cité page 84*

Romain GIOT, Alexandre NINASSI, Mohamad EL-ABED et Christophe ROSENBERGER : Analysis of the acquisition process for keystroke dynamics. *In Proceedings of IEEE BIOSIG 2012*, pages 1–12, Darmstadt, Germany, 2012c. *2 citations pages 28 et 154*

Romain GIOT et Christophe ROSENBERGER : Genetic programming for multibiometrics. *Expert Systems With Applications*, 39(2):1837–1847, février 2012. Impact factor: 1.924. *Cité page 84*

Romain GIOT, Christophe ROSENBERGER et Bernadette DORIZZI : Can Chronological Information Be Used As A Soft Biometric In Keystroke Dynamics? *In The Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP 2012)*, 2012d. *Cité page 107*

Romain GIOT, Christophe ROSENBERGER et Bernadette DORIZZI : Performance evaluation of biometric template update. *In International Biometric Performance Testing Conference 2012*, pages 1–4, Gaithersburg, MD, USA, mars 2012e. URL `http://biometrics.nist.gov/cs_links/ibpc2012/presentations/Day3/342_goit_supporting_paper.pdf`. *2 citations pages 134 et 147*

N.J. GRABHAM et N.M WHITE : Use of a novel keypad biometric for enhanced user identity verification. *In Instrumentation and Measurement Technology Conference Proceedings, 2008. IMTC 2008. IEEE*, pages 12–16, 2008. *2 citations pages 21 et 126*

Eric GRANGER, Wael KHREICH, Robert SABOURIN et Dmitry O. GORODNICHY : Fusion of biometric systems using boolean combination: an application to iris–based authentication. *International Journal of Biometrics*, 4:291–315, 2012. *Cité page 75*

A. GUVEN et I. SOGUKPINAR : Understanding users' keystroke patterns for computer access security. *Computers & Security*, 22(8):695–706, 2003. *3 citations pages 26, 30, et 31*

J. HAN et B. BHANU : Individual recognition using gait energy image. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, pages 316–322, 2006. *Cité page 6*

P. HART : The condensed nearest neighbor rule (corresp.). *Information Theory, IEEE Transactions on*, 14(3):515 – 516, mai 1968. ISSN 0018-9448. *Cité page 167*

Masaki HASHIYADA : Developement of biometric dna ink for authentication security. *Tohoku J. Exp. Med.*, 204:109–117, 2004. *Cité page 6*

T. K. HO, J. J. HULL et S. N. SRIHARI : Decision combination in multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(1):66–75, 1994. *Cité page 74*

Sylvain HOCQUET : *Authentification biométrique adaptative Application à la dynamique de frappe et à la signature manuscrite*. Thèse de doctorat, Université de Tours, 2007. *2 citations pages 23 et 107*

Sylvain HOCQUET, Jean-Yves RAMEL et Hubert CARDOT : Estimation of user specific parameters in one-class problems. *In ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, pages 449–452, Washington, DC, USA, 2006. IEEE Computer Society. ISBN 0-7695-2521-0. *4 citations pages 18, 19, 46, et 56*

Sylvain HOCQUET, Jean-Yves RAMEL et Hubert CARDOT : User classification for keystroke dynamics authentication. *In The Sixth International Conference on Biometrics (ICB2007)*, pages 531–539, 2007. *7 citations pages 18, 26, 28, 30, 31, 52, et 78*

D. HOSSEINZADEH et S. KRISHNAN : Gaussian mixture modeling of keystroke patterns for biometric applications. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 38(6):816–826, 2008. *6 citations pages 19, 28, 30, 31, 46, et 50*

N. HOUMANI, S. GARCIA-SALICETTI et B. DORIZZI : On assessing the robustness of pen coordinates, pen pressure and pen inclination to time variability with personal entropy. *In Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on*, 2009. *Cité page 107*

Y.S. HUANG et C.Y. SUEN : Method of combining multiple experts for the recognition of unconstrained handwritten numerals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(1):90–94, 1995. *Cité page 75*

S. HWANG, S. CHO et S. PARK : Keystroke dynamics-based authentication for mobile devices. *Computers & Security*, 2008. *Cité page 20*

Seong-seob HWANG, Hyoung-joo LEE et Sungzoon CHO : Improving authentication accuracy of unfamiliar passwords with pauses and cues for keystroke dynamics-based authentication. *Intelligence and Security Informatics*, 3917:73–78, 2006. *7 citations pages 18, 24, 28, 30, 31, 37, et 39*

Wonjun HWANG, Haibing REN, Hyunwoo KIM, Seok-Cheol KEE et Junmo KIM : Face recognition using gender information. *In IEEE 16th International Confernce on Image Processing (ICIP 2009)*, Cairo, Egypt, novembre 2009. IEEE Signal Society. *Cité page 77*

J ILONEN : Keystroke dynamics. *Advanced Topics in Information Processing–Lecture*, 2003. *Cité page 22*

M. INDOVINA, U. ULUDAG, R. SNELICK, A. MINK et A. JAIN : Multimodal biometric authentication methods: a cots approach. *In Proc. MMUA*, pages 99–106, 2003. *Cité page 73*

ISO : Information technology biometric performance testing and reporting. Standard, 2006. ISO/IEC 19795-1. *2 citations pages 9 et 78*

A. JAIN, K. NANDAKUMAR et A. ROSS : Score normalization in multimodal biometric systems. *Pattern Recognition*, 38(12):2270–2285, 2005. *Cité page 69*

A. JAIN et A. ROSS : Fingerprint mosaicking. *In IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 4, pages IV–4064. IEEE, 2002. *Cité page 66*

A.K. JAIN, S.C. DASS et K. NANDAKUMAR : Soft biometric traits for personal recognition systems. *In Proceedings of International Conference on Biometric Authentication,*, 2004a. *2 citations pages 7 et 76*

A.K. JAIN et U. PARK : Facial marks: Soft biometric for face recognition. *In IEEE International Conference on Image Processing (ICIP)*, 2009. *Cité page 76*

A.K. JAIN, A. ROSS et S. PRABHAKAR : An introduction to biometric recognition. *Biometrics*, 14(1), 2004b. *Cité page 6*

R. JANAKIRAMAN et T. SIM : Keystroke dynamics in a general setting. *Lecture notes in computer science*, 4642:584, 2007. *3 citations pages 24, 30, et 31*

Xudong JIANG et Wee SER : Online fingerprint template improvement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:1121–1126, 2002. *2 citations pages 121 et 122*

Varun KACHOLIA et Shashank PANDIT : Biometric authentication using random distributions (bioart). *Proceedings of the 15th Canadian IT Security SymposiumOttawa, Canada*, 5:1–8, 2003. *3 citations pages 28, 30, et 31*

Pilsung KANG et Sungzoon CHO : A hybrid novelty score and its use in keystroke dynamics-based user authentication. *Pattern Recognition*, 42(11):3115–3127, 2009. *Cité page 24*

Pilsung KANG, Seong-seob HWANG et Sungzoon CHO : Continual retraining of keystroke dynamics based authenticator. *In* Seong-Whan LEE et Stan LI, éditeurs : *Proceedings of ICB 2007*, volume 4642 de *Lecture Notes in Computer Science*, pages 1203–1211. Springer Berlin / Heidelberg, 2007. URL `http://dx.doi.org/10.1007/978-3-540-74549-5_125`. *4 citations pages 125, 134, 139, et 144*

S. KARATZOUNI et N. CLARKE : *New Approaches for Security, Privacy and Trust in Complex Environments*, volume 232 de *IFIP International Federation for Information Processing*, chapitre Keystroke Analysis for Thumb-based Keyboards on Mobile Devices, pages 253–263. Springer, 2007. *Cité page 20*

M. KARNAN, M. AKILA et N. KRISHNARAJ : Biometric personal authentication using keystroke dynamics: A review. *Applied Soft Computing*, 11(2):1565 – 1573, 2011. The Impact of Soft Computing for the Progress of Artificial Intelligence. *Cité page 46*

HB Kekre et VA Bharadi : Adaptive feature set updating algorithm for multimodal biometrics. *In Proceedings of the International Conference on Advances in Computing, Communication and Control*, pages 277–282. ACM, 2009. *2 citations pages 113 et 125*

K. Killourhy et R. Maxion : The effect of clock resolution on keystroke dynamics. *In Proceedings of the 11th international symposium on Recent Advances in Intrusion Detection*, pages 331–350. Springer, 2008. *5 citations pages 19, 21, 28, 30, et 34*

Kevin S. Killourhy et Roy A. Maxion : Should security researchers experiment more and draw more inferences? *In 4th Workshop on Cyber Security Experimentation and Test (CSET'11)*, pages 1–8, août 2011. *4 citations pages 19, 29, 33, et 46*

K.S. Killourhy et R.A. Maxion : Comparing anomaly-detection algorithms for keystroke dynamics. *In IEEE/IFIP International Conference on Dependable Systems & Networks, 2009. DSN'09*, pages 125–134, 2009. *8 citations pages 32, 33, 38, 48, 57, 58, 134, et 137*

Josef Kittler, Mohamad Hatef, Robert P.W. Duin et Matas Jiri : On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence and Security Informatics*, 20(3):226–239, march 1998. *Cité page 70*

T. Kohonen : *Self-Organizing Maps*. Numéro 30 *in* Information Sciences. Springer, Heidelberg., second édition, 1997. *Cité page 127*

Z. Korotkaya : Biometric person authentication: Odor. Rapport technique, Department of Information Technology, Laboratory of Applied Mathematics, Lappeenranta University of Technology, 2003. *Cité page 6*

A. Kumar et D. Zhang : Personal recognition using hand shape and texture. *IEEE Transactions on Image Processing*, 15(8):2454, 2006. *Cité page 6*

L.I. Kuncheva, C.J. Whitaker, C.A. Shipp et R.P.W. Duin : Limits on the majority vote accuracy in classifier fusion. *Pattern Analysis and Applications*, 6(1):22–31, 2003. *Cité page 75*

B. Kveton, M. Philipose, M. Valko et L. Huang : Online semi-supervised perception: Real-time learning without explicit feedback. *In Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 15–21. IEEE, 2010. *2 citations pages 119 et 120*

L. Lam et C. Y. Suen : Optimal combinations of pattern classifiers. *Pattern Recognition Letters*, 16(9):945–954, 1995. *Cité page 75*

L. Lam et C. Y. Suen : Application of majority voting to pattern recognition: an analysis of its behavior and performance. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 27(5):553–568, 1997. *Cité page 75*

H. Lee et S. Cho : Retraining a keystroke dynamics-based authenticator with impostor patterns. *Computers & Security*, 26(4):300–310, 2007. *2 citations pages 30 et 31*

Y. Li, J. Yin, E. Zhu, C. Hu et H. Chen : Score based biometric template selection and update. *In Future Generation Communication and Networking, 2008. FGCN'08. Second International Conference on*, volume 3, pages 35–40. IEEE, 2008. *Cité page 123*

Zhen Li, Xi Zhou et Thomas S. Huang : Spatial gaussina mixture model for gender recognition. *In IEEE 16th International Confernce on Image Processing (ICIP 2009)*, Cairo, Egypt, novembre 2009. IEEE Signal Society. *Cité page 77*

X. Liu, T. Chen et S.M. Thornton : Eigenspace updating for non-stationary process and its application to face recognition. *Pattern Recognition*, 36(9):1945–1959, 2003. ISSN 0031-3203. *2 citations pages 128 et 131*

D.G. Lowe : Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004. ISSN 0920-5691. *Cité page 84*

C. C. Loy, W. K. Lai et C. P. Lim : Keystroke patterns classification using the artmap-fd neural network. *In International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 61-64, Taiwan, 2007 (IIHMSP 2007)*, pages 61–64, 2007. *Cité page 32*

Alessandra Lumini et Loris Nanni : A clustering method for automatic biometric template selection. *Pattern Recognition*, 39(3):495–497, 2006. ISSN 0031-3203. *Cité page 110*

P.S. Magalhães, K. Revett et H.D. Santos : Keystroke dynamics: stepping forward in authentication. *In Next Generation Web Services Practices, 2005. NWeSP 2005. International Conference on*, 2006. *Cité page 84*

D. Maltoni, AK Jain et S. Prabhakar : *Handbook of fingerprint recognition*. Springer, 2009. *Cité page 6*

G. Marcialis, A. Rattani et F. Roli : Biometric template update: an experimental investigation on the relationship between update errors and performance degradation in face verification. *In SSPR&SPR*, pages 684–693. Springer, 2008. *4 citations pages 117, 130, 131, et 132*

A.M. Martinez et R. Benavente : The ar face database. Rapport technique, CVC Technical report, 1998. *Cité page 82*

Aurélien Mayoue : Biosecure tool - performance evaluation of a biometric verification system, 2007. URL `http://svnext.it-sudparis.eu/svnview2-eph/ref_syst/Tools/PerformanceEvaluation/doc/howTo.pdf`. *Cité page 57*

F. Monrose, M.K. Reiter et S. Wetzel : Password hardening based on keystroke dynamics. *International Journal of Information Security*, 1(2):69–83, 2002. *2 citations pages 18 et 30*

F. Monrose et Rubin : Authentication via keystroke dynamics. *In Proceedings of the 4th ACM conference on Computer and communications security*, pages 48–56. ACM Press New York, NY, USA, 1997. *4 citations pages 13, 17, 26, et 52*

F. Monrose et A.D. Rubin : Keystroke dynamics as a biometric for authentication. *Future Generation Computer Syststems*, 16(4):351–359, 2000. *3 citations pages 13, 30, et 31*

R. Myung : Keystroke-level analysis of korean text entry methods on mobile phones. *International Journal of Human-Computer Studies*, 60(5-6):545–563, 2004. *2 citations pages 20 et 30*

K. Nandakumar, Y. Chen, S.C. Dass et A. Jain : Likelihood ratio-based biometric score fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):342, 2008. *Cité page 73*

K. Nandakumar, A. Jain et A. Ross : Fusion in multibiometric identification systems: What about the missing data. *In Proc. of the 3rd International Conference on Biometrics ICB, Sassari, Italy*, 2009. *Cité page 74*

P. Narainsamy, S. Sunjiv et N. Shrikaant : Investigating & improving the reliability and repeatability of keystroke dynamics timers. *International Journal of Network Security & Its Applications (IJNSA),*, 2(3):70–85, 2010. *3 citations pages 19, 21, et 34*

NIST : Nist biometric score set, 2006. URL `http://www.itl.nist.gov/iad/894.03/biometricscores/`. *Cité page 81*

Ricardo García Noval et Francisco Perales López : Adaptative templates in biometric authentication. *In The 16th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2008*, 2008. *Cité page 111*

MS Obaidat et B. Sadoun : Verification of computer users using keystroke dynamics. *Systems, Man and Cybernetics, Part B, IEEE Transactions on*, 27(2):261–269, 1997. *6 citations pages 18, 24, 28, 30, 31, et 53*

MS Obaidat, B. Sadoun et J. Amman : *Biometrics*, chapitre Keystroke Dynamics Based Authentication, pages 213–229. Jain, Anil K. and Bolle, Ruud and Pankanti, Sharath, 2006. *Cité page 30*

R. Palaniappan : Electroencephalogram signals from imagined activities: A novel biometric identifier for a small population. *Lecture notes in computer science*, 4224:604, 2006. *Cité page 6*

M.V.S. Paula, E.A. Kinto, E.D.M. Hernandez et T. Carvalho : User authentication based on human typing pattern with artificial neural networks and support vector machine, 2005. *Cité page 19*

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot et Duchesnay E. : Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. *Cité page 97*

D. Petrovska-Delacretaz, A. El Hannani et G. Chollet : Text-independent speaker verification: State of the art and challenges. *Lecture Notes In Computer Science*, 4391:135, 2007. *Cité page 6*

Vir V. Phoha, BabuSunil, Asok Ray et Shashi Phoha : System and method for classifying regions of keystroke density with a neural network, fev 2005. *Cité page 19*

Koksoon Phua, Jianfeng Chen, Tran Huy Dat et Louis Shue : Heart sound as a biometric. *Pattern Recognition Society.*, 41:906 – 919, 2008. *Cité page 6*

N. Poh : URL `http://info.ee.surrey.ac.uk/Personal/Norman.Poh/web/banca_multi/`. *Cité page 81*

N. Poh, J. Kittler, S. Marcel, D. Matrouf et J.F. Bonastre : Model and score adaptation for biometric systems: Coping with device interoperability and changing acquisition conditions. *In Proceedings of Internal Conference on Pattern Recogntion (ICPR 2010)*, 2010a. *2 citations pages 107 et 111*

N. Poh, J. Kittler, A. Rattani et M. Tistarelli : Group-specific score normalization for biometric systems. *In Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 38–45. IEEE, 2010b. *Cité page 73*

N. Poh, J. Kittler, R. Smith et J. Tena : A method for estimating authentication performance over time, with applications to face biometrics. *In 12th Iberoamerican Congress on Pattern Recognition CIARP*, 2007. *Cité page 145*

Norman Poh, Thirimachos Bourlai et Josef Kittler : A multimodal biometric test bed for quality-dependent, cost-sensitive and client-specific score-level fusion algorithms. *Pattern Recogn.*, 43:1094–1105, March 2010c. ISSN 0031-3203. URL `http://dl.acm.org/citation.cfm?id=1660180.1660626`. *Cité page 73*

Norman Poh, R. Wrong, J. Kittler et F. Roli : Challenges and research directions for adaptive biometric recognition systems. *In Advances in Biometrics*, pages 753–764, 2009. *8 citations pages 106, 114, 128, 129, 130, 136, 137, et 147*

Vir v Pohoa, Sashi Pohoa, Asok Ray et Shrijit Sudhakar Joshi : Hidden markov model (hmm)-based user authentication using keystroke dynamics. patent, fev 2009. *Cité page 19*

David Primeaux, Doraiswamy Sundar et Willard L. Robinson Jr. : Usage pattern based user authenticator, December 2001. URL `http://www.freepatentsonline.com/6334121.html`. *Cité page 19*

R. Raghavendra, B. Dorizzi, A. Rao et G. Hemantha Kumar : Pso versus adaboost for feature selection in multimodal biometrics. *In IEEE 3rd International Conference on Biometrics: Theory, Applications and Systems, BTAS 2009*, 2009. *Cité page 68*

R. Raghavendra, Bernadette Dorizzi, Ashok Rao et G. Hemantha Kumar : Particle swarm optimization based fusion of near infrared and visible images for improved face verification. *Pattern Recognition*, 44(2):401 – 411, 2011. ISSN 0031-3203. URL `http://www.sciencedirect.com/science/article/pii/S0031320310003924`. *Cité page 67*

Bhaskar Rao : Continuous keystroke biometric system. Mémoire de D.E.A., University of California, 2005. *4 citations pages 13, 24, 27, et 30*

N. K. Ratha, J. H. Connell et R. M. Bolle : An analysis of minutiae matching strength. *In Audio- and Video-Based Biometric Person Authentication*, pages 223–228, 2001a. *Cité page 10*

N. K. Ratha, J. H. Connell et R. M. Bolle : Enhancing security and privacy in biometrics-based authentication systems. *IBM Systems Journal*, 40(3):614–634, 2001b. *Cité page 62*

A. Rattani, B. Freni, G. Marcialis et F. Roli : Template update methods in adaptive biometric systems: A critical review. *In Internal Conference on Biometrics 2009 (ICB 2009)*, 2009. *3 citations pages 107, 108, et 131*

A. Rattani, G.L. Marcialis et F. Roli : Biometric template update using the graph mincut algorithm : A case study in face verification. *In Biometrics Symposium, 2008. BSYM '08*, pages 23 –28, septembre 2008a. *3 citations pages 118, 145, et 165*

Ajita Rattani : *Adaptive Biometric System based on Template Update Procedures.* Thèse de doctorat, Dept. of Electrical and Electronic Engineering University of Cagliari, 2010. *8 citations pages 111, 114, 117, 118, 119, 131, 164, et 165*

Ajita Rattani, D. R. Kisku, Andrea Lagorio et Massimo Tistarelli : Facial template synthesis based on sift features. *In Automatic Identification Advanced Technologies, 2007 IEEE Workshop on*, 2007. *Cité page 121*

Ajita Rattani, Gian Luca Marcialis et Fabio Roli : Boosting gallery representativeness by co-updating face and fingerprint verification systems. 5th Summer School for Advanced Studies on Biometrics for Secure Authentication, 2008b. *Cité page 118*

Ajita Rattani, Gian Luca Marcialis et Fabio Roli : Capturing large intra-class variations of biometric data by template co-updating. *In Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*, 2008c. *Cité page 118*

Ajita Rattani, Gian Luca Marcialis et Fabio Roli1 : Self adaptive systems: An experimental analysis of the performance over time. *In IEEE Symposium Series in Computational Intelligence 2011 (SSCI 2011). Workshop on Computational Intelligence in Biometrics and Identity Management (CIBIM). Special Session on Adaptive Classification Systems for Biometric Recognition*, 2011. *Cité page 135*

K. Revett, S.T. de Magalhães et H.M.D. Santos : Enhancing login security through the use of keystroke input dynamics. *Lecture notes in computer science*, 3832, 2006. *2 citations pages 47 et 52*

K. Revett, S.T. de Magalhaes et H.M.D. Santos : On the use of rough sets for user authentication via keystroke dynamics. *Lecture notes in computer science*, 4874:145, 2007a. *4 citations pages 24, 26, 30, et 31*

Kenneth Revett : Biometric security system using keystroke dynamics of a user's login attempt, october 2007. *2 citations pages 19 et 23*

Kenneth Revett : A bioinformatics based approach to user authentication via keystroke dynamics. *International Journal of Control, Automation and Systems*, 7(1):7–15, feb 2009. *3 citations pages 19, 30, et 44*

Kenneth Revett, Sérgio Tenreiro de Magalhães et Henrique M. D. Santos : Developing a keystroke dynamics based agent using rough sets. *In IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology.*, 2005. *2 citations pages 30 et 31*

Kenneth Revett, Florin Gorunescu, Marina Gorunescu, Marius Ene, Sérgio de Magalhães Tenreiro et Henrique M. Dinis Santos : A machine learning approach to keystroke dynamics based user authentication. *International Journal of Electronic Security and Digital Forensics*, 1:55–70, 2007b. *Cité page 30*

A. Riera, A. Soria-Frisch, M. Caparrini, C. Grau et G. Ruffini : Unobtrusive biometric system based on electroencephalogram analysis. *EURASIP Journal on Advances in Signal Processing*, 2008:8, 2008. *Cité page 6*

Armin Rigo : Psyco - home page, 2010. URL `http://psyco.sourceforge.net/`. *Cité page 56*

G. Ritter, H. Woodruff, S. Lowry et T. Isenhour : An algorithm for a selective nearest neighbor decision rule (corresp.). *Information Theory, IEEE Transactions on*, 21(6):665–669, 1975. ISSN 0018-9448. *Cité page 167*

JA Robinson, VW Liang, JAM Chambers et CL MacKenzie : Computer user verification using login string keystroke dynamics. *Systems, Man and Cybernetics, Part A, IEEE Transactions on*, 28(2):236–241, 1998. *3 citations pages 18, 30, et 31*

R.N. Rodrigues, G.F.G. Yared, CR do N Costa, J.B.T. Yabu-Uti, F. Violaro et L.L. Ling : Biometric access control through numerical keyboards based on keystroke dynamics. *Lecture notes in computer science*, 3832:640, 2006. *3 citations pages 24, 30, et 31*

S. J. Rogers et M.E. Brown : Method and apparatus for verification of a computer user's identification, based on keystroke characteristics, septembre 17 1996. US Patent 5,557,686. *6 citations pages 17, 21, 23, 24, 28, et 30*

F. Roli, L. Didaci et G. Marcialis : Template co-update in multimodal biometric systems. *In International Conference on Biometrics (ICB 2007)*, pages 1194–1202. Springer, 2007. *3 citations pages 115, 118, et 144*

F. Roli et G. Marcialis : Semi-supervised pca-based face recognition using self-training. *In Structural, Syntactic, and Statistical Pattern Recognition*, pages 560–568. Springer, 2006. *Cité page 117*

Fabio Roli, Luca Didaci et Gian Luca Marcialis : *Advances in Biometrics*, chapitre Adaptive Biometric Systems That Can Improve with Use, pages 447–471. SpringerLink, 2008. *3 citations pages 115, 117, et 118*

Christophe Rosenberger et Luc Brun : Similarity-based matching for face authentication. *In International Conference on Pattern Recognition (ICPR)*, 2008. *Cité page 84*

A. Ross, S. Shah et J. Shah : Image versus feature mosaicing: A case study in fingerprints. *In Proceedings of SPIE Conference on Biometric Technology for Human Identification III*, pages 620208–1. Citeseer, 2006a. *Cité page 66*

A.A. Ross, K. Nandakumar et A.K. Jain : *Handbook of Multibiometrics*. Springer, 2006b. *3 citations pages 62, 63, et 64*

Gordon Alfred Ross : Intellectual property protection and verification utilizing keystroke dynamics, February 2004. URL `http://www.freepatentsonline.com/20040034788.html`. *Cité page 19*

M. Rybnik, M. Tabedzki et K. Saeed : A keystroke dynamics based system for user identification. *Computer Information Systems and Industrial Management Applications, 2008. CISIM '08. 7th*, pages 225–230, June 2008. *2 citations pages 30 et 31*

C. Ryu, Y. Han et H. Kim : Super-template generation using successive bayesian estimation for fingerprint enrollment. *In Audio-and Video-based Biometric Person Authentication (AVBPA 2005)*, pages 710–719. Springer, 2005. *Cité page 121*

C. Ryu et H. Kim : Fingerprint verification testing scenarios for multi-impression enrollment and template adaptation. *In Proc. of Biometric Symposium*, pages 39–40, 2005. *Cité page 131*

Choonwoo Ryu, Hakil Kim et Anil K. Jain : Template adaptation based fingerprint verification. *In Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, 2006. *3 citations pages 121, 129, et 131*

T.A. Salthouse : Effects of age and skill in typing. *Journal of Experimental Psychology: General*, 113(3):345, 1984. *Cité page 35*

Y. Sang, H. Shen et P. Fan : *Parallel and Distributed Computing: Applications and Technologies*, chapitre Novel impostors detection in keystroke dynamics by support vector machine, pages 666–669. Springer, 2005. *2 citations pages 19 et 44*

Yingpeng Sang, Hong Shen et Pingzhi Fan : Novel impostors detection in keystroke dynamics by support vector machine. *In Proc. of the 5th international conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT 2004)*, 2004. *Cité page 24*

A. Saranli et M. Demirekler : A statistical unified framework for rankbased multiple classifier decision combination. *Pattern Recognition*, 34(4):865–884, 2001. *Cité page 74*

T. Scheidat, A. Makrushin et C. Vielhauer : Automatic template update strategies for biometrics. Rapport technique, Otto-von-Guericke University of Magdeburg, Magdeburg, Germany, 2007. *3 citations pages 125, 126, et 144*

Graeme G. Schreiber et Andrew R. Knox : Software method for improved password entry, December 2007. URL `http://www.freepatentsonline.com/7305559.html`. *Cité page 19*

Mark M Seeger et Patrick Bours : How to comprehensively describe a biometric update mechanisms for keystroke authentication. *In 3rd International Workshop on Security and Communication Networks (IWSCN 2011)*, pages 1–7, 2011. *4 citations pages 132, 134, 135, et 147*

Michael Lawrence Serpa : System and method for user authentication with enhanced passwords, October 2005. URL `http://www.freepatentsonline.com/6954862.html`. *Cité page 19*

D. Shanmugapriya et G. Padmavathi : An efficient feature selection technique for user authentication using keystroke dynamics. *IJCSNS International Journal of Computer Science and Network Security*, 11(10):191–195, octobre 2011. *Cité page 24*

Y. Sheng, V.V. Phoha et S.M. Rovnyak : A parallel decision tree-based method for user authentication based on keystroke patterns. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 35(4):826–833, 2005. *Cité page 48*

R. Singh, M. Vatsa, A. Ross et A. Noore : A mosaicing scheme for pose-invariant face recognition. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 37(5):1212–1225, 2007. *2 citations pages 66 et 67*

R. Singh, M. Vatsa, A. Ross et A. Noore : Biometric classifier update using online learning: A case study in near infrared face verification. *Image and Vision Computing*, 28(7):1098–1105, 2010. ISSN 0262-8856. *Cité page 126*

D. Song, P. Venable et A. Perrig : User recognition by keystroke latency pattern analysis. *Retrieved on*, 19, 1997. *3 citations pages 13, 18, et 30*

D.X. Song, D. Wagner et X. Tian : Timing analysis of keystrokes and timing attacks on ssh. *In Proceedings of the 10th conference on USENIX Security Symposium-Volume 10*, pages 25–25. USENIX Association Berkeley, CA, USA, 2001. *Cité page 13*

D. Stefan et D. Yao : Keystroke dynamics authentication and human-behavior driven bot detection. Rapport technique, Technical report, Rutgers University, 2008. *Cité page 13*

R. Sukthankar et R. Stockton : Argus: the digital doorman. *Intelligent Systems, IEEE*, 16(2):14–19, 2005. ISSN 1541-1672. *Cité page 116*

Q. Tao, G.W. Wu, F.Y. Wang et J. Wang : Posterior probability support vector machines for unbalanced data. *Neural Networks, IEEE Transactions on*, 16(6):1561–1573, 2005. ISSN 1045-9227. *Cité page 172*

P.S. Teh, A.B.J. Teoh, T.S. Ong et H.F. Neo : Statistical fusion approach on keystroke dynamics. *In Proceedings of the 2007 Third International IEEE Conference on Signal-Image Technologies and Internet-Based System-Volume 00*, pages 918–923. IEEE Computer Society, 2007. *Cité page 26*

G. Tur, D. Hakkani-Tür et R.E. Schapire : Combining active and semi-supervised learning for spoken language understanding. *Speech Communication*, 45(2):171–186, 2005. ISSN 0167-6393. *Cité page 120*

M. Turk et A. Pentland : Face recognition using eigenfaces. *In Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, volume 591, 1991. *2 citations pages 6 et 84*

U. Uludag, A. Ross et A. Jain : Biometric template selection and update: a case study in fingerprints. *Pattern Recognition*, 37(7):1533–1542, 2004. *4 citations pages 122, 123, 124, et 125*

D Umphress et G. Williams : Identity verification through keyboard characteristics. *Internat. J. Man–Machine Studies*, 23:263–273, 1985. *6 citations pages 16, 22, 23, 24, 30, et 31*

Yasin Uzun et Kemal Bicakci : A second look at the performance of neural networks for keystroke dynamics using a publicly available dataset. *Computers & Security*, 2012. In Press, Corrected Proof. *2 citations pages 31 et 53*

Kumari Vandana : Enhancing weak biometric authentication by adaptation and improved user-discrimination. Mémoire de D.E.A., International Institute of Information Technology Hyderabad, INDIA, 2007. *2 citations pages 112 et 128*

V Vapnik : *Statistical learning theory*. Wiley New York, 1998. *Cité page 89*

D.A. Vaquero, R.S. Feris, D. Tran, L. Brown, A. Hampapur et M. Turk : Attribute-based people search in surveillance environments. *In IEEE Workshop on Applications of Computer Vision (WACV)*, 2009. *Cité page 77*

P. Verlinde : *A contribution to multimodal identity verification using decision fusion*. Thèse de doctorat, Ecole Nationale Supérieure des Télécommunications, 1999. *Cité page 75*

Adam Weiss, Anil Ramapanicker, Shah Pranav, Shinese Noble et Larry Immohr : Mouse movements biometric identification: A feasibility study. *In Proceedings of Student/Faculty Research Day, CSIS, Pace University,*, 2007. *Cité page 6*

Dennis L. Wilson : Asymptotic properties of nearest neighbor rules using edited data. *Systems, Man and Cybernetics, IEEE Transactions on*, 2(3):408 –421, juillet 1972. ISSN 0018-9472. *Cité page 168*

M. Wojciechowski : *Feed-Forward neural network for python*. Technical University of Lodz (Poland), Departement of Civil Engineering, Architecture and Environmental Engineering, 2007. http://ffnet.sourceforge.net/. *Cité page 56*

L. Xu, A. Krzyzak et C.Y. Suen : Methods for combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(3):418–435, 1992. *Cité page 75*

Z.W. Xu, X.X. Guo, X.Y. Hu et X. Cheng : The blood vessel recognition of ocular fundus. *In Proceedings of the 4th International Conference on Machine Learning and Cybernetics (ICMLC'05)*, pages 4493–4498, 2005. *Cité page 6*

J. Yan, A. Blackwell, R. Anderson et A. Grant : The memorability and security of passwords-some empirical results. *Technical Report-University Of Cambridge Computer Laboratory*, 2000. *Cité page 36*

F. Yang, M. Paindavoine, H. Abdi et D. Arnoult : Fast image mosaicing for panoramic face recognition. *Journal of multimedia*, 1(2):14–20, 2006. *2 citations pages 66 et 67*

James R. Young et Robert W. Hammon : Method and apparatus for verifying an individual's identity, February 1989. URL `http://www.freepatentsonline.com/4805222.html`. *Cité page 17*

E. Yu et S. Cho : Keystroke dynamics identity verification – its problems and practical solutions. *Computers & Security*, 23(5):428–440, 2004. *4 citations pages 24, 28, 30, et 31*

X. Zhang et Y. Gao : Face recognition across pose: A review. *Pattern Recognition*, 42(11):2876–2896, 2009. *Cité page 66*

Ying Zhao : Learning user keystroke patterns for authentication. *In 14th International Enformatica Conference*, 2006. *2 citations pages 30 et 31*

Charles ZHOU : A study of keystroke dynamics as a practical form of authentication. Mémoire de D.E.A., Pomona College, march 2008. *Cité page 16*

Xiaojin ZHU : *Semi-Supervised Learning with Graphs.* Thèse de doctorat, Language Technologies Institue School of Computer Science Carnegie Mellon University, 2005. *Cité page 119*

Xiaojin ZHU, Zoubin GHAHRAMANI et John LAFFERTY : Semi-supervised learning using gaussian fields and harmonic functions. *In Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, pages 1–9, Washington DC, 2003. *Cité page 119*

Arkady G. ZILBERMAN : Security method and apparatus employing authentication by keystroke dynamics, August 2002. URL `http://www.freepatentsonline.com/6442692.html`. *Cité page 19*

La dynamique de frappe au clavier est une modalité biométrique comportementale qui permet d'authentifier des individus selon leur façon de taper au clavier. Un tel système est peu coûteux, car il ne nécessite pas de matériel d'acquisition autre que le clavier de l'ordinateur, et est facilement accepté par l'utilisateur. Nous nous sommes principalement intéressé aux systèmes statiques où le texte saisit par l'utilisateur est connu à l'avance par la machine. Malheureusement, les performances de cette modalité sont plutôt médiocres en raison de la forte variabilité de la donnée biométrique. Cette variabilité est due à l'état émotionnel de la personne, l'apprentissage de la façon de taper, ...

Nous proposons dans cette thèse différentes contributions permettant d'améliorer les performances de reconnaissance de systèmes de dynamique de frappe au clavier (DDF). Nous effectuons également une analyse des bases publiques permettant d'évaluer la performance de nouveaux systèmes de reconnaissance. Une contribution est la mise au point d'un système de DDF par mot de passe partagé. Nous étudions ensuite la fusion multibiométrique avec la dynamique de frappe au clavier et la reconnaissance faciale afin d'augmenter les performances des deux systèmes. Nous montrons, sur deux jeux de données différents, qu'il est possible de reconnaitre le genre d'un individu suivant sa façon de taper au clavier. Enfin, nous présentons une nouvelle méthode de mise à jour de la référence biométrique qui permet de prendre en compte le vieillissement de la donnée biométrique, afin de ne pas avoir une diminution des performances de reconnaissance au cours du temps.

Contribution to keystroke dynamics: mulitbiometrics, soft biometrics and template update

---

Keystroke dynamics is a behavioural biometry which allows to authenticate individuals through there way of typing on a keyboard. Such systems are cheap, as they do not need specific devices different from the keyboard of the computer. They are also well accepted by the user. We are mainly interested in static systems where the text typed by the user is known in advance by the machine. Sadly, the performance of this modality are rather mediocre because of the high variability of the biometric data which comes from emotional state of the individual, the learning of they way to type, ...

In this thesis, we propose various contributions which allow to improve the recognition performance of keystroke dynamics systems. We also do an analysis of the public datasets allowing to evaluate the performance of new recognition systems. One contribution is the creation of a system which allows the authentication of users with a shared password. Then, we study the biometric fusion with face recognition and keystroke dynamics in order to increase the performance of the two systems. We show, on two different datasets, that it is possible to guess the gender of an individual through its way of typing to a keyboard. Finally, we present a new template update method which allows to take into account the ageing of the biometric data in order to not observe a decrease of performance overtime.

Indexation Rameau : Biométrie / Systèmes informatiques – Mesures de sûreté / Reconnaissance des formes (informatique) / Classification / Calcul évolutionnaire
Indexation libre : Biométrie, Dynamique de frappe au clavier, Mise à jour de la référence, Algorithmes évolutionnaires, Fusion d'information

---

Informatique et applications

Laboratoire GREYC - UMR CNRS 6072 - Université de Caen Basse-Normandie - Ensicaen
6 Boulevard du Maréchal Juin - 14050 CAEN CEDEX