Université François Rabelais
Tours
Ecole Doctorale : Santé, Sciences et Technologies
Année Universitaire : 2004-2005

THESE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITE DE TOURS

Discipline : **Informatique**

Présentée et soutenue publiquement

par :

Matthieu WIROTIUS

le 10 novembre 2005

AUTHENTIFICATION PAR SIGNATURE MANUSCRITE SUR SUPPORT NOMADE

Sous la direction de Nicole VINCENT (co-encadrant : Jean-Yves RAMEL)

Jury

Gallinari Patrick	Professeur	Examinateur	Université Paris 6 (LIP)
Heutte Laurent	Professeur	Rapporteur	Université de Rouen (PSI)
Lorette Guy	Professeur	Rapporteur	Université Rennes 1 (IRISA)
Milgram Maurice	Professeur	Examinateur	Université Paris 6 (LISIF)
Ramel Jean-Yves	Maître de Conférences	Examinateur	Université de Tours (LI)
Vincent Nicole	Professeur	Examinateur	Université Paris 5 (CRIP5)
Barbezange Jean-Claude	Directeur de Projets	Invité	Atos WorldLine

REMERCIEMENTS

Tout d'abord, je tiens à remercier mon directeur de thèse, Nicole Vincent ainsi que Jean-Yves Ramel co-encadrant, pour leurs conseils, le temps précieux qu'ils m'ont accordé et pour leur soutient au cours de ces trois années.

Un grand merci à l'ensemble des membres du Laboratoire d'Informatique de l'Université François Rabelais de Tours, et plus particulièrement son directeur, Christian Proust, pour m'avoir accueilli au sein d'une structure de recherche dynamique. Dans ce cadre, je tiens aussi à remercier les étudiants du département informatique de l'Ecole Polytechnique de l'Université François Rabelais de Tours avec qui j'ai eu l'occasion de travailler.

Ce travail de recherche n'aurait pu exister sans le soutien de l'entreprise Atos Origin, sous forme d'un contrat Cifre. Mes plus sincères remerciements à l'équipe Recherche et Développement d'Atos Origin, et plus particulièrement à Jean-Claude Barbezange pour la confiance qu'il m'a accordé.

Je tiens également à remercier Laurent Heutte et Guy Lorette pour avoir accepté de rapporter mon travail et pour leur remarques constructives. Plus généralement, je remercie Patrick Gallinari et Maurice Milgram pour leur participation à mon jury de thèse.

Pour finir je tiens à remercier toute ma famille et plus particulièrement mes parents pour leur soutien durant ces nombreuses années d'études…

SOMMAIRE

Table des Figures

Table des Tableaux

INTRODUCTION

Depuis le 11 septembre 2001, la sécurité est devenue une préoccupation majeure au niveau international. Certaines technologies, qui étaient à l'époque encore marginales, ont alors pris un essor considérable. Actuellement, de nombreux systèmes informatiques sont envisagés dans les futurs passeports et cartes d'identité, ainsi que pour de nombreux autres usages en lien avec la sécurité. C'est dans ce contexte que la société ATOS a proposé une thèse en convention CIFRE avec le Laboratoire Informatique de Tours afin de concevoir un nouveau système d'authentification de personnes, notamment pour sécuriser les transactions financières.

Les quatre services de base en sécurité sont :

- l'authentification,
- l'intégrité,
- la confidentialité,
- la non-répudiation.

L'authentification donne l'assurance de l'identité d'un objet (une personne, un serveur, une application). Les services d'intégrité garantissent qu'un objet n'a pas été modifié depuis sa création par une personne autre que son auteur. La confidentialité assure qu'un document ne sera pas lu par un tiers n'en ayant pas le droit. Enfin, le but de la non-répudiation (ou non-désaveu) est que l'émetteur d'un message ne puisse pas nier l'avoir envoyé et le recepteur l'avoir reçu. Ces besoins existaient déjà lors de la diffusion de documents papier. Mais, la technologie évolue à grands pas et de nouvelles habitudes émergent. Les documents manipulés actuellement sont majoritairement électroniques et ils circulent en clair sur les réseaux informatiques. Il a donc été nécessaire de créer un ensemble de mécanismes pour assurer divers services de sécurité au niveau des documents et de leur transfert, c'est-à-dire réaliser les 4 services de base cités précédemment.

Le travail demandé par ATOS concerne principalement l'authentification de personnes et de documents. Pour authentifier une personne, on distingue trois types de méthodes. Les deux méthodes les plus utilisées sont basées sur ce que connaît la personne, comme un mot de

passe, ou sur ce que possède une personne comme un badge ou une carte d'identité. Un troisième type de méthodes, plus original et basé sur ce que l'on est (les empreintes digitales ou la forme de la main) ou sur ce que l'on sait faire (la signature manuscrite ou la dynamique de frappe au clavier), est aujourd'hui en plein développement. Cette troisième approche repose sur les caractéristiques de l'individu lui-même, on parle alors de biométrie.
Le terme "authentification forte" est utilisé lorsque deux des trois types de méthodes se combinent pour réaliser l'authentification. Par exemple, l'accès aux terminaux de paiement n'est permis que sur présentation de la carte bancaire (ce que l'on possède) et l'indication du code confidentiel associé (ce que l'on sait), enregistré sur la puce de la carte elle-même.

Afin de renforcer l'authentification d'un porteur de certificat numérique en utilisant la biométrie, notre étude a pour but d'améliorer les performances des systèmes d'authentification par signature manuscrite. En effet, actuellement, seul un code PIN protège les certificats. La solution envisagée pour augmenter le niveau de sécurité consiste à remplacer le code PIN par une signature manuscrite. Ainsi, l'utilisateur pourra s'authentifier en signant à l'aide d'un stylet sur une interface graphique. On peut citer deux exemples concrets d'utilisations envisagées : la vérification de l'identité d'une personne physique (authentification) et la signature d'un document électronique (non répudiation). Dans les deux cas, l'usager signe sur son interface graphique pour valider son identité.
Par ailleurs, l'arrivée de nouveaux dispositifs tels les PDA, les téléphones portables, les Tablets PC ou encore les dernières consoles de jeu munies d'un stylet rendent la signature manuscrite en ligne accessible à un public de plus en plus large. On peut donc penser que son utilisation deviendra de plus en plus fréquente.

Trois types de contraintes viennent s'ajouter à la difficulté de ce projet : les contraintes techniques, ergonomiques et juridiques.
Les principales contraintes techniques sont :

- Un temps d'exécution faible : de l'ordre de la seconde
- Un espace mémoire limité pour contenir le code de la (ou des) signature(s) de référence et le processus d'authentification lui-même : au maximum quelques kilos octets afin qu'ils puissent être stockés sur une carte à puce.
- La protection contre les attaques physiques de la (ou des) signature(s) de référence : carte à puce, …

S'ajoute à ces contraintes initiales une contrainte due au fait que la signature d'un individu évolue au cours du temps, il sera donc nécessaire de réactualiser régulièrement la (ou les) signature(s) de référence et donc d'établir un protocole très strict pour qu'un faux ne soit pas enregistré comme signature de référence. De plus, le fait de modifier la signature de référence pose le problème de l'apprentissage incrémental. En effet, cela implique des contraintes fortes sur la méthode de comparaison de signatures choisie car elle doit rester valable même si l'on modifie la (ou les) signature(s) de référence.

L'ergonomie est aussi un aspect important de ce projet. En effet, il est nécessaire de mettre en place des interfaces expliquant les conditions dans lesquelles sera utilisée la signature ainsi que les conséquences liées au fait que l'utilisateur signe. De plus, une attention particulière devra être portée à l'ergonomie de l'interface afin de fournir aux usagers des conditions stables de signature.

Aux contraintes techniques et ergonomiques s'ajoute une contrainte juridique. En effet, la Commission Nationale de l'Informatique et des Libertés (CNIL) impose que la (ou les) signature(s) de référence soi(en)t stockée(s) localement et non sur un serveur. Dans certains cas, une solution possible pourra être de stocker cette référence sur une carte à puce afin qu'elle soit mieux protégée contre les attaques physiques. Cela implique que le dispositif informatique utilisé soit équipé de lecteur de carte à puce. Dans le cas du téléphone portable, cette référence pourra être contenue dans la carte SIM.
Le fait d'utiliser la signature manuscrite comme moyen d'authentification induit que l'utilisateur soit informé du cadre juridique dans lequel sa signature va être utilisée. Et, il faudra vérifier la valeur juridique de ce mécanisme. Pour la plupart des documents papier signés actuellement, l'authenticité de la signature n'est pas vérifiée. Le problème apparaît s'il y a contestation de la signature. Il faut donc pouvoir prouver que la personne a bien signé elle-même. Pour cela, on peut envisager de rajouter de l'information au contenu du document à signer : nom, heure, date, lieu, numéro de série du hardware… pour pouvoir prouver que la personne titulaire du certificat a bien signé le document et mettre ainsi en place un mécanisme de non répudiation.

La problématique de la reconnaissance de signatures manuscrites n'est pas nouvelle - de nombreuses études en témoignent - mais elles prouvent aussi que le problème n'est pas totalement résolu. Les objectifs que nous nous sommes fixés sont de proposer de nouvelles

approches prenant en compte certaines caractéristiques propres à l'écriture mais aussi de mieux définir l'architecture du système idéal de comparaison de signatures. Dans le cadre de ce travail, nous nous intéressons au tracé en ligne de la signature représentée dans au moins trois dimensions : deux coordonnées spatiales (x, y), et une coordonnée temporelle t.

Le premier chapitre de ce mémoire présente les caractéristiques principales d'un système biométrique ainsi que les méthodes existant dans le domaine de l'authentification par signature manuscrite. Ce domaine faisant l'objet de nombreux travaux, nous présentons les principales méthodes utilisées successivement pour chacune des étapes de l'architecture classique d'un système d'authentification par signature manuscrite.

Le second chapitre présente nos contributions concernant l'extraction de caractéristiques sur des signatures manuscrites en ligne, ainsi que les différentes méthodes envisageables pour effectuer la sélection des caractéristiques les plus discriminantes parmi le grand nombre de possibilités existantes.

Le chapitre 3 décrit différentes possibilités d'amélioration de l'algorithme Dynamic Time Warping (DTW) que nous préconisons pour mieux mesurer la similarité entre les signatures et ainsi améliorer leur comparaison. Nous montrons également que l'étape de prétraitement des signatures a une forte influence sur les résultats de la comparaison fournis par DTW. Nous présentons donc une nouvelle approche que nous préconisons pour cette partie de la chaîne de traitement.

Le dernier chapitre démontre la faisabilité et l'efficacité de nos travaux grâce à la description détaillée des fonctionnalités et des caractéristiques principales du prototype développé pour la société ATOS dans le cadre de cette convention CIFRE.

CHAPITRE 1 - AUTHENTIFICATION ET SIGNATURE EN LIGNE

Dans ce chapitre, nous dressons un panorama des différentes méthodes d'authentification biométrique. Puis nous présentons plus en détails les concepts de l'authentification par signature manuscrite et les travaux réalisés dans ce domaine dans le monde universitaire et industriel.

1. BIOMETRIE

Après avoir présenté le contexte d'utilisation qui nous intéresse particulièrement, nous définirons plus précisément ce terme si couramment employé à l'heure actuelle et verrons les enjeux économiques en découlant. Nous aborderons les problèmes juridiques soulevés par ces nouvelles technologies. Enfin, nous décrirons les caractéristiques d'un système biométrique.

1.1. Contexte

La multiplication des outils informatiques nous oblige à saisir plusieurs fois par jour des combinaisons de caractères constituant un identifiant/mot de passe. Cette contrainte, qui vise à augmenter la sécurité, amène notamment les internautes à se simplifier la vie en privilégiant l'utilisation de mots de passe faciles à mémoriser, voire à les inscrire sur un petit papier scotché sous le clavier... Le constat est inquiétant : 91% des mots de passe utilisés par des internautes sont "connus", c'est-à-dire issus de l'environnement familier de la personne et jugés non viables par des spécialistes du cryptage :

- 21 % utilisent leur prénom, celui d'un membre de la famille ou de leur animal
- 15 % ont un lien avec une date clé (date de naissance, moment historique)
- 30 % des personnes partagent leur mot de passe avec leur partenaire

Plus généralement, les techniques d'authentification basées sur ce que l'on possède et sur ce que l'on sait présentent de nombreux inconvénients. Les objets permettant l'authentification sont souvent perdus ou volés et les mots permettant de s'identifier sont facilement oubliés. De plus, ce type de données est souvent partagé par plusieurs personnes. Par ailleurs, d'un point de vue sécurité, l'utilisation d'un mot de passe valide sur un réseau n'assure pas que la

personne qui s'est connectée est bien celle qu'elle prétend être. On sait seulement qu'elle possédait la bonne clé d'accès. L'identité et la protection des données privées ne peuvent pas être garanties et l'utilisation frauduleuse d'un de ces mécanismes ne peut pas être prouvée. Ces limitations des systèmes classiques d'authentification entraînent une perte de confiance et une augmentation des possibilités de fraude. La biométrie apporte une solution à ces différents problèmes. D'une part, le partage des données permettant l'authentification devient impossible. D'autre part, la confiance dans l'authentification est accrue puisque la personne doit être physiquement présente.

1.2. Définition

La biométrie est définie en France comme la science permettant l'identification d'individus à partir de leurs caractéristiques physiologiques ou comportementales [CAB03]. Ces caractéristiques doivent être universelles (exister chez tous les individus), uniques (permettre de différencier un individu par rapport à tout autre), permanentes (présentes tout au long de la vie), collectables (possibilité d'enregistrer les caractéristiques d'un individu avec l'accord de celui-ci) et mesurables (autoriser une comparaison future).
La biométrie permettrait donc l'identification d'une personne sur la base de caractères physiologiques ou de traits comportementaux automatiquement reconnaissables et vérifiables. L'avantage d'une telle identification est que chaque individu a ses propres caractéristiques physiques qui ne peuvent être ni changées, ni perdues, ni volées.
Les principaux moyens biométriques utilisés commercialement à l'heure actuelle sont :

- Les empreintes digitales
- La forme de la main
- La forme du visage
- La voix
- L'iris
- La signature manuscrite

D'autres approches biométriques sont l'objet de recherches et sont envisagées dans un futur proche :

- L'ADN
- La démarche
- La rétine

- La forme de l'oreille
- La dynamique de frappe au clavier

1.3. Le marché

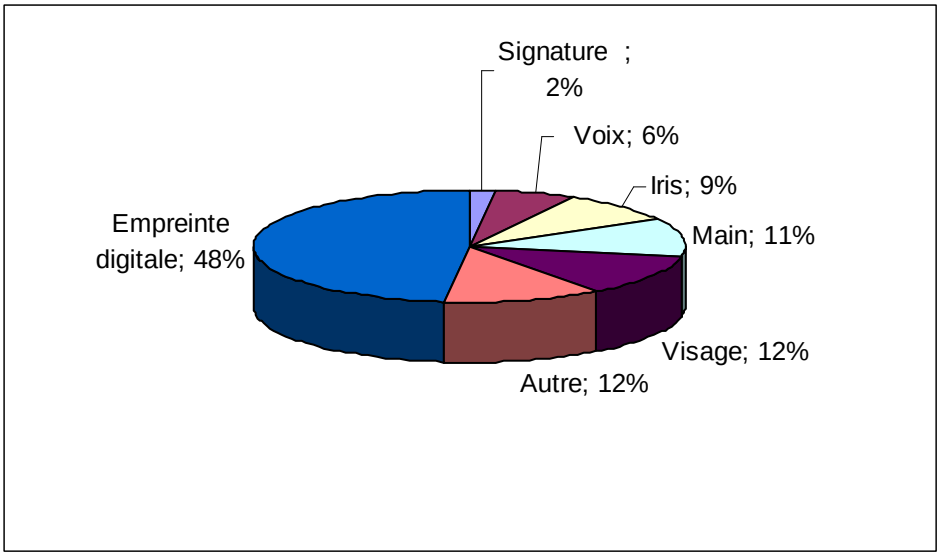


Figure 1. Marché de la biométrie en 2004 [IBG04].

L'utilisation, dans les logiciels, des empreintes digitales pour authentifier une personne a pris un essor considérable au cours de ces dernières années. Comme on peut le constater sur la Figure 1, qui représente une évaluation réalisée par l'International Biometric Group de la répartition du marché de la biométrie en 2004 [IBG04], l'authentification par empreinte digitale représente près de la moitié du marché de la biométrie. Les deux autres méthodes biométriques les plus développées sont basées sur la reconnaissance du visage et sur l'examen de la forme de la main. Tous les autres systèmes tels que la reconnaissance vocale, la reconnaissance basée sur les lignes dans la paume de la main, sur la signature manuscrite ou la façon de taper sur le clavier, restent très marginaux et ne représentent qu'un faible pourcentage des systèmes vendus.

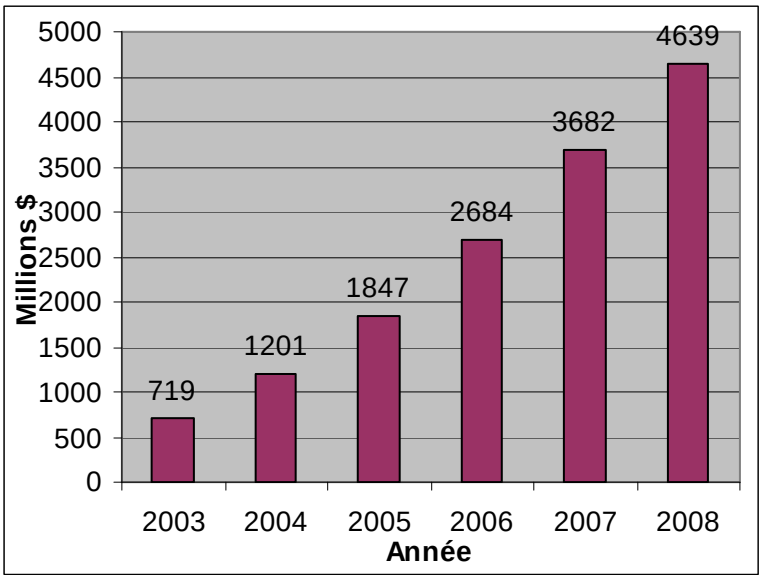


Figure 2. Evolution prévisionnelle des revenus générés par le marché de la biométrie.

La Figure 2 montre que la biométrie est aujourd'hui une technique d'authentification en plein essor. Son succès est lié à sa simplicité d'utilisation et à la non nécessité d'un mot de passe, mais du coté des utilisateurs ou clients potentiels, on rencontre plus ou moins de réticence. La réticence de certains utilisateurs à l'égard de la biométrie est principalement due à une crainte d'un usage détourné des données biométriques. Malgré cela, la demande est en forte croissance. Les demandes les plus fréquentes concernent le remplacement du mot de passe par une donnée biométrique à l'ouverture d'un logiciel et lors du contrôle d'accès aux locaux dont l'entrée est conditionnée par une authentification.

1.4. Comparaison des techniques

L'étude d'un produit biométrique se base principalement sur quatre points : la technologie et le coût associé, la simplicité d'usage, l'efficacité quantitative et juridique.

Plutôt que de comparer uniquement les performances de ces systèmes, il est nécessaire de tenir compte de l'environnement, de l'usage, de la facilité aussi bien de saisie que d'analyse, de stockage ou de vérification. En effet, chaque technologie possède des avantages et des inconvénients, acceptables ou inacceptables suivant les applications. Toutes les solutions ne sont donc pas concurrentes, elles n'offrent ni les mêmes niveaux de sécurité ni les mêmes facilités d'emploi.

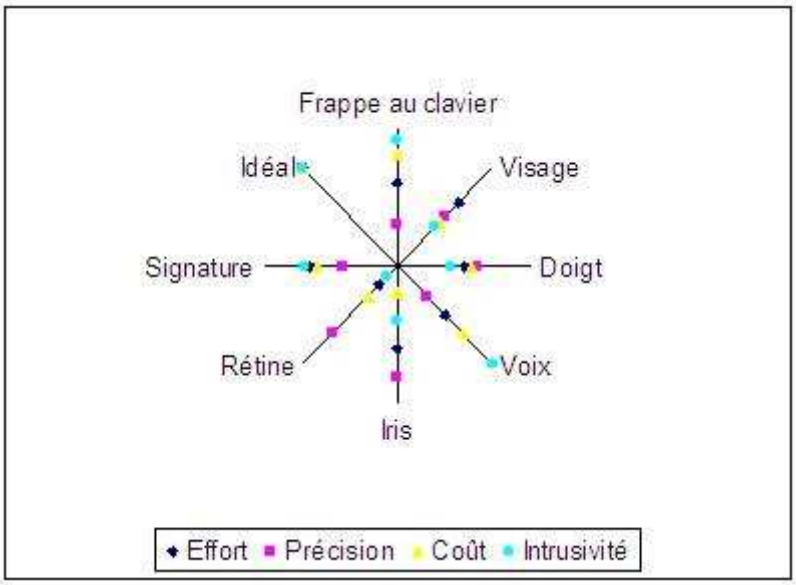


Figure 3. Comparatif des différentes technologies biométriques.

Et donc, comme on peut le constater sur la Figure 3, qui est le résultat d'une étude menée en 2003 par un groupe de travail sur la biométrie (International Biometric Group), aucune technologie biométrique existante n'est idéale [IBG04].

En comparaison des systèmes d'authentification utilisant un objet ou un mot de passe, qui offrent une réponse stable (oui ou non, 0% ou 100%) ; les informations biométriques sont plus fluctuantes et donnent des réponses en terme de pourcentage de taux de ressemblance entre un élément biométrique testé et un modèle enregistré (entre 0% et 100%, le 100% n'étant jamais atteint). Cette variation des résultats d'identification d'un individu est plus liée à la qualité de la capture de l'information biométrique (on n'a jamais deux images ou deux sons identiques), qu'à la modification de la caractéristique biométrique de l'individu qui est généralement stable dans le temps.

Il faut donc définir au sein de l'application un seuil de décision (acceptation ou refus) compris entre 0% et 100% de similarité entre le modèle et l'élément à tester. Ce seuil peut être différent pour chaque personne. Il peut également dépendre des contraintes imposées au système par l'application.

Les procédures et mesures de protection à mettre en place pour un système biométrique sont :

- Le codage des données biométriques
- La restriction des données biométriques encodées à la détermination de l'admissibilité d'une personne, on s'assure ainsi que ces données ne servent pas d'instrument de contrôle ou de surveillance non préalablement prévu

- L'assurance que les données recueillies et encodées ne permettent pas de reconstruire l'élément biométrique d'origine
- L'assurance que ce système n'est pas l'unique procédé d'identification mais qu'il est couplé à d'autres systèmes d'authentification classiques
- L'assurance que les données biométriques encodées ne permettent pas de retrouver un individu
- L'assurance de la mise en place de mesures visant à contrôler l'accès aux renseignements biométriques (qui et pourquoi)
- L'exigence de la présentation d'une ordonnance de tribunal avant d'autoriser les organismes externes, tels que services policiers et ministères gouvernementaux, à avoir accès aux renseignements biométriques.

1.5. Les inconvénients

La biométrie présente malheureusement à l'heure actuelle un certain nombre d'inconvénients. Notons son inconvénient majeur : aucune des mesures utilisées ne se révèle être totalement fiable. La raison réside bien dans l'une des caractéristiques majeures de tout organisme vivant : il s'adapte à l'environnement, il vieillit, il subit des traumatismes plus ou moins importants, bref il évolue et les mesures biométriques changent.
Prenons le cas le plus simple, celui des empreintes digitales mais la même chose s'applique à toute donnée physique. Suivant les cas, nous présentons plus ou moins de transpiration; la température des doigts n'est pas régulière (en moyenne, de 8 à 10° Celsius au-dessus de la température ambiante). Il suffit de se couper pour présenter une anomalie dans le dessin de ses empreintes. Bref, dans la majorité des cas, les mesures du capteur et du logiciel associé retourneront un résultat différent de la mesure initiale de référence. Or, il faut pourtant bien réussir à se faire reconnaître. En pratique, cela sera réalisé dans la plupart des cas car le système est amené à autoriser une marge d'erreur entre la mesure et la référence.

De manière générale, les faiblesses de ces systèmes ne se situent pas au niveau de la particularité physique sur laquelle ils reposent, mais bien sur la façon avec laquelle ils la mesurent, et la marge d'erreur qu'ils autorisent. Là encore, il convient de ne pas se laisser impressionner par une image illusoire de haute technologie - produit miracle.

De plus, les experts techniques mettent au passif de cette technologie, d'une part, son coût, d'autre part, la question de sa révocation. En effet, confronté à une personne qui a subtilisé un mot de passe ou une signature électronique, le titulaire du mot de passe ou de la signature peut facilement les remplacer ou les révoquer. La chose semble plus complexe pour une empreinte digitale ou rétinienne. Si un tiers s'approprie une identité biométrique du type empreintes digitales ou identité visuelle, il peut au moyen de ces identités biométriques passer tout type d'actes au nom de la victime. Comment la victime pourrait-elle alors révoquer sa propre empreinte digitale ou identité visuelle ? Les experts en sécurité sont partagés sur la question, même si, en majorité, ils semblent considérer que cette révocation est possible. Tous reconnaissent cependant la difficulté à mettre au passif cette protection technique.

1.6. Le cadre juridique

L'aspect légal est un point important de la biométrie. Cette technologie mettant en jeu un individu, personne physique, constitue une donnée à caractère personnel c'est-à-dire, selon la définition posée par la Loi n°2004-801 du 6 août 2004, une "information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres" [CAR97]. En cela, tout traitement portant sur la reconnaissance biométrique entre dans le champ d'investigation de la Commission Nationale de l'Informatique et des Libertés (CNIL) [CNI05]. Dans deux délibérations rendues le même jour, le 8 avril 2004 (délibérations n°04-017 et 04-018), la CNIL a fixé quelques points de repère qui montrent la vigilance dont elle fait preuve face à cette technologie.

Dans la première délibération, le centre hospitalier de Hyères envisageait de mettre en œuvre un traitement consistant à horodater les entrées et sorties de son personnel en s'appuyant sur un dispositif de reconnaissance de l'empreinte digitale. La CNIL a émis un avis défavorable à la mise en œuvre de ce traitement ayant pour objectif la gestion du temps de travail.

Pour motiver cet avis négatif, la CNIL s'appuie sur deux types d'arguments. D'une part, elle critique la centralisation des données biométriques sur un serveur central, voyant là une solution qui "n'est pas de nature à garantir la personne concernée de toute utilisation détournée de ses données biométriques", d'autre part, elle se fonde sur une disposition insérée dans le Code du Travail selon laquelle "nul ne peut apporter aux droits des personnes et des libertés individuelles ou collectives des restrictions qui ne seraient pas justifiées par la nature de la tâche à accomplir ni proportionnées au but recherché" (article L 120-2 du code du travail).

La CNIL considère dès lors que "seul un impératif de sécurité est susceptible de justifier la centralisation de données biométriques". Elle y voit au contraire dans le cas du centre hospitalier de Hyères un traitement disproportionné par rapport à la finalité recherchée, soit la gestion du temps de travail.

Dans la seconde délibération du même jour, la CNIL va en revanche donner un avis favorable à l'établissement public Aéroports de Paris pour un système de contrôle d'accès aux zones réservées, dans un objectif de sûreté, des aéroports d'Orly et de Roissy. Logiquement et compte tenu de la première délibération évoquée ci-dessus, la Commission retient ici que "seuls sont enregistrés sur le badge, le gabarit biométrique, le numéro du badge et le code PIN associé au badge" notant par là que les données biométriques résident avec la personne et que, au regard de l'application concernée, "ces données sont adéquates, pertinentes et non excessives".

"Ces deux délibérations ont été rendues sous l'empire de la loi ancienne de 1978 mais, quant aux règles de fond qu'elles posent, il n'y a aucune raison qu'elles ne soient pas prises en compte aujourd'hui. La biométrie est une technologie sans doute utile notamment à l'authentification, mais son usage doit sans doute respecter un certain nombre de règles légales à ne pas oublier" [ITE05].

1.6.1. L'absence de reconnaissance juridique de la biométrie

A l'heure actuelle, l'utilisation des systèmes biométriques ne fait pas l'objet d'un régime juridique particulier. La CNIL, dans le cadre de son rapport annuel d'activité présenté au mois de juillet 2004, a fait part de sa position sur les différentes techniques biométriques et les dangers découlant de leur utilisation. En effet, la CNIL considère qu'un élément d'identification biométrique ou sa traduction informatique sous forme de gabarit constitue une donnée à caractère personnel entrant dans le champ d'application de la loi "informatique et liberté" au même titre que d'autres données personnelles (nom, adresse, etc.). Dans ces conditions, la conservation ou le stockage des éléments biométriques d'identification s'apparente à la conservation d'une base de données et relève, en conséquence, de l'ensemble de la législation sur la protection des données. Plus particulièrement, la constitution de bases de données doit respecter les principes clés de finalité et de proportionnalité. En effet, la CNIL considère que "seul un impératif particulier de sécurité est susceptible de justifier la centralisation de données biométriques" [CNI04].

1.6.2. Les perspectives d'utilisation des techniques biométriques

Au regard de ce qui précède, la CNIL formule plusieurs propositions destinées à encadrer l'utilisation des techniques biométriques :

a) la non conservation de données biométriques dans des bases de données

Les technologies biométriques de reconnaissance qui ne reposent pas sur le stockage d'éléments d'identification biométrique dans une base de données ne soulèvent pas de difficultés particulières au regard de la loi "informatique et liberté", dès lors que l'élément d'identification biométrique est conservé par l'utilisateur et uniquement par lui (sur une carte à puce, etc.) ou sur un appareil dont il a l'usage exclusif (téléphone portable, appareils nomades, etc.).

b) le stockage de données biométriques lié à des impératifs de sécurité

Lorsqu'une base de données est constituée, le contrôle de finalité et de proportionnalité peut conduire à accepter la mise en œuvre de telles bases de données si un impératif particulier de sécurité le justifie (contrôle d'accès à des bâtiments hautement sécurisés de la Banque de France ; bâtiments de stockage de plutonium ; etc.).

1.6.3. La préférence de certaines données biométriques

A défaut d'une justification particulière, lorsqu'une base de données est constituée, il conviendra de préférer certains éléments biométriques tels que la forme de la main, la rétine etc. ne laissant pas de traces physiques plutôt que la constitution de fichiers ADN ou d'empreintes digitales, afin d'éviter les risques de détournement de tels fichiers.

C'est ainsi que la CNIL a rendu le 15 octobre 2002 un avis favorable à la mise en place d'un système de contrôle d'accès à une cantine scolaire utilisant des bases de données de gabarits de contour de la main.

Par contre, et dans le même cadre, la CNIL avait émis en 2000 un avis défavorable à un système de contrôle d'accès d'une cantine scolaire reposant sur une base de données d'empreintes digitales. En effet, cette base de données d'empreintes digitales pouvait être détournée de sa finalité première, et être exploitée à d'autres fins, ce que ne permet pas une simple exploitation du contour de la main.

Nous retrouvons encore une fois les principes premiers de finalité et de proportionnalité.

Les techniques biométriques aux fins d'identification et d'authentification sont promises à un avenir certain. Pour autant, au regard de l'importance des risques associés à l'utilisation des systèmes biométriques, et de leurs éventuelles conséquences juridiques (aux fins d'authentification, notamment) il est important que la législation encadre strictement leurs conditions d'utilisation.

1.7. But de l'authentification : vérification ou identification?

Le terme "authentification" couvre en fait deux sous problèmes : l'identification et la vérification.

L'identification consiste à déterminer, à partir d'une base de référence, la personne dont la donnée biométrique est la plus proche de celle testée. Dans ce cas, la réponse du système sera le nom de la personne ou le rejet de la donnée biométrique si aucune des références stockées dans la base n'est assez proche de la donnée testée.

La vérification correspond à une notion différente en ce sens qu'elle n'est pas reliée à une base de données. Cela consiste à vérifier si l'élément biométrique testé correspond bien à la personne qui prétend le posséder. Par conséquent, la réponse ne peut prendre que deux valeurs, l'acceptation ou le rejet de la donnée biométrique suivant le degré de similarité entre l'élément biométrique testé et une référence en tenant compte du niveau de sécurité souhaité.

Dans cette thèse, nous nous sommes plus intéressé au problème de la vérification qu'à celui de l'identification. L'identification est un problème de recherche du plus proche voisin parmi un ensemble de possibilités alors que la vérification est un problème de discrimination à deux classes, acceptation ou rejet. Par conséquent, les approches utilisées ne sont pas les mêmes pour ces deux problèmes. Alors que tous les modèles sont disponibles pour un problème d'identification, la difficulté de la vérification est accrue car on ne dispose que du modèle d'une personne à chaque fois pour prendre la bonne décision. A aucun moment du processus nous n'avons la possibilité de stocker et de comparer les données biométriques des différentes personnes impliquées. On ne peut donc pas effectuer de classification supervisée, en associant une classe à chaque individu, afin de rechercher et d'adapter des critères qui séparent au maximum les classes, qui augmentent la variance interclasses. Par conséquent, il est plus difficile de connaître les caractéristiques représentatives et discriminantes des données biométriques et qui permettraient une vérification facile de la personne. Dans le cadre de

l'identification, il faut maximiser la distance inter personnes alors qu'en vérification il faut minimiser la distance intra personne.

1.8. Architecture d'un système d'authentification biométrique

L'architecture générale d'un système d'authentification basée sur la biométrie est décrite dans la Figure 4.

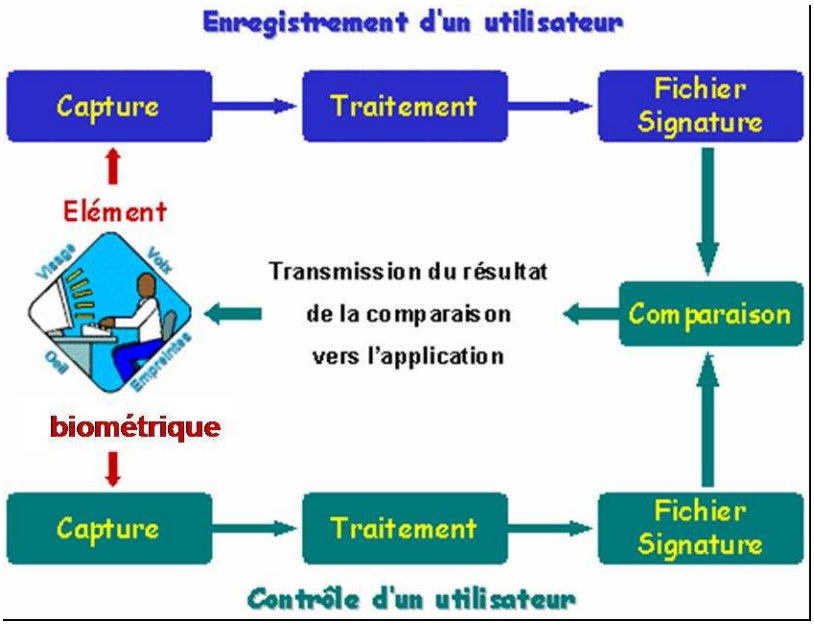


Figure 4. Architecture d'un système d'authentification biométrique.

Dans un premier temps, l'utilisateur doit s'enregistrer. Ce processus, appelé aussi enrôlement, est constitué d'une phase de capture de l'élément biométrique qui est répétée un certain nombre de fois pour avoir un aperçu de la variabilité de cet élément, d'une phase de traitement permettant l'extraction des caractéristiques de celui-ci et, à partir des données extraites, d'une phase de création d'un modèle représentatif de l'utilisateur.

Une fois enregistré, l'utilisateur peut alors s'authentifier. Les trois premières phases sont identiques à celles effectuées lors de l'enregistrement. La comparaison se fait alors entre les données extraites de l'élément biométrique testé et le ou les modèles enregistrés correspondants.

1.9. Evaluation

Lors de l'opération d'authentification, un capteur, de même nature que celui utilisé pour l'enrôlement, est utilisé pour générer un nouveau fichier brut qui va subir les mêmes

opérations d'analyse que le fichier modèle (référence). Un nouveau fichier signature sera produit. Une analyse de corrélation plutôt qu'une vérification d'identité est réalisée entre les deux fichiers signature car, bien évidemment, en pratique les deux fichiers signature ne sont jamais identiques. Il en résulte un coefficient de ressemblance qui peut varier de 0 à 100 %. Selon les critères de la sévérité souhaités par l'application en question, il suffit simplement de vérifier si la ressemblance est supérieure ou inférieure au seuil fixé à l'avance. Il va de soi que l'acceptation d'une transaction pour un distributeur de billets nécessite un seuil beaucoup plus sévère que celui autorisant l'accès à une salle de concert.

L'évaluation d'un système d'authentification ne peut pas être réalisée en utilisant uniquement le taux d'erreur classique parce que toutes les erreurs n'ont pas le même impact et les contraintes des applications peuvent changer. Nous devons donc différencier le taux de faux acceptés (FAR) indiquant les faux non détectés par le système et le taux de vrais rejetés (FRR) indiquant les signatures authentiques rejetées par le système. Le EER (Equal Error Rate) correspond au taux d'erreur pour lequel FAR est égal à FRR.

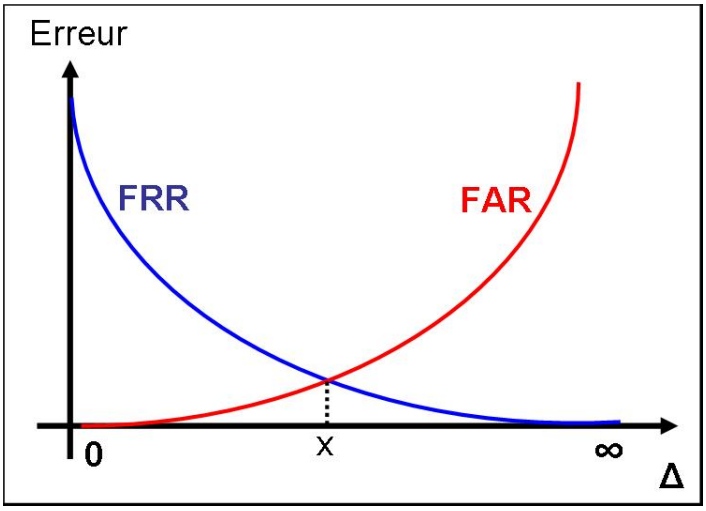


Figure 5. Evolution de FRR et de FAR en fonction du seuil.

Le graphe de la Figure 5 est purement démonstratif; Δ représente un paramètre du système, variant de 0 à l'infini. Très succinctement, on voit que plus la marge d'erreur Δ autorisée est importante, plus le taux de fausses acceptations augmente, c'est-à-dire que l'on va accepter de plus en plus de personnes qui ne sont pas autorisées (et donc la sécurité du système diminue). Par contre on voit que le taux de rejet des personnes autorisées diminue également, ce qui rend le système plus fonctionnel et répond mieux aux attentes des utilisateurs. A l'autre extrémité, si l'on diminue la marge d'erreur acceptée par le procédé de mesure biométrique, les tendances des 2 taux sont inversées : on va de moins en moins accepter des individus essayant de frauder mais on va, par la même occasion, avoir un taux de rejet sur des

personnes autorisées qui sera trop important pour être toléré dans la plupart des cas. Le compromis habituel est de régler le système selon la jonction des courbes, c'est à dire de fixer le paramètre Δ à la valeur x où le couple (FAR, FRR) est "minimal". Les valeurs de ces paramètres, FAR et FRR, doivent être fixées en fonction de l'application. En général, on ne peut pas fixer les deux simultanément. Pour un produit industriel, la valeur de FRR doit être faible afin d'éviter la répétition de la phase d'authentification quand une personne autorisée a été rejetée et doit faire un nouvel essai. Pour notre projet, l'objectif est de garder la valeur de FRR en dessous de 2% car les utilisateurs de systèmes d'authentification ne tolèrent pas de taux plus élevé.

Après cette présentation générale de ce qu'est la biométrie et des différents moyens d'authentification associés, nous allons décrire de manière plus détaillée l'authentification par signature manuscrite et dresser un panorama des travaux effectués dans ce domaine.

2. L'AUTHENTIFICATION PAR SIGNATURE MANUSCRITE

"La signature identifie celui qui l'appose, manifeste le consentement des parties aux obligations et confère l'authenticité à l'acte". (C. civ., art. 1316, al. 1 nouveau).

La signature manuscrite est depuis plusieurs siècles le moyen le plus répandu pour manifester sa propre volonté. Elle est aujourd'hui, et le demeurera sans doute dans le futur, le moyen biométrique d'authentification le plus utilisé.

L'utilisation de la signature manuscrite repose sur l'hypothèse que ce sont plus des mouvements instinctifs que des actes conscients qui sont impliqués dans la réalisation de la signature. Ce postulat implique que certaines caractéristiques de la signature sont stables donc constantes pour un signataire. Ainsi, la signature en ligne ou hors ligne peut être considérée comme une méthode biométrique comportementale. La principale difficulté concernant l'authentification est que la signature en entrée et la (ou les) signature(s) servant de référence ne sont pas exactement les mêmes. Pour que cette reconnaissance soit exacte, il faut que la variation existant entre les signatures d'une même personne soit inférieure à la distance entre les signatures de deux personnes différentes. Il faut donc essayer d'isoler les parties ou caractéristiques de la signature qui sont pratiquement constantes, de celles qui ne le sont pas.

Outre la variabilité habituelle, différentes raisons peuvent expliquer la variation de la signature :

c) le support et le stylet utilisés
d) l'importance du document sur lequel on appose la signature
e) le lieu et les conditions d'écriture

Les fonctions assurées par la signature manuscrite sur papier sont l'identification, l'adhésion au contenu, la garantie de l'intégrité, la constitution d'un original.

Un système d'authentification biométrique basé sur la signature manuscrite doit assurer les mêmes fonctions :

1. Fonction d'identification. Le système d'authentification doit être suffisamment fiable pour être reconnu comme moyen de non répudiation.
2. Fonction d'adhésion au contenu. Culturellement le fait d'apposer sa signature manuscrite signifie que l'on adhère au contenu indépendamment du support.
3. Fonction de garantie de l'intégrité. La garantie de l'intégrité du document peut être assurée par une fonction de hachage.
4. Fonction de constitution d'un original. La signature manuscrite sur papier ou sur interface graphique reste toujours unique : on ne refait jamais exactement la même signature.
5. Fonction psychologique. Le fait d'utiliser la signature manuscrite en amont de la signature électronique offre l'avantage de capter l'attention de l'individu sur l'importance de l'acte contrairement aux méthodes actuelles où l'on entre un code PIN.

2.1. Principes de fonctionnement

Classiquement, la conception d'un système d'authentification nécessite d'apporter des solutions à cinq problèmes :

1. Acquisition des données
2. Prétraitement
3. Extraction des caractéristiques et/ou parties stables
4. Comparaison (et donc décision)
5. Evaluation des performances

Dans le cas de la signature en ligne, ces problèmes se déclinent de la manière suivante :

1. L'acquisition des données se fait au moyen d'un stylet électronique. L'utilisateur signe à plusieurs reprises (au moins cinq fois) afin d'établir une référence. La signature de référence peut être, par exemple, une moyenne des signatures servant d'exemple.
2. Le prétraitement consiste à réduire le bruit, lisser le signal dans le temps et l'espace, normaliser les données (mise à l'échelle, centrage…) et à les coder.
3. Sur l'ensemble des supports, des données de deux types peuvent être extraites localement ou globalement. Elles peuvent concerner la forme ou la dynamique.
4. Les données relatives aux variations locales du signal en entrée concernent, par exemple, la position, la vitesse, l'accélération, … Ces dernières peuvent être couplées à des données plus générales comme la longueur, le temps total, des moyennes…
5. La comparaison entre la signature testée et la ou les signatures servant de référence correspond à un calcul de distance. Si la distance entre ces signatures est inférieure à un certain seuil, la signature est reconnue comme authentique sinon elle est reconnue comme un faux.
6. Afin d'évaluer le système, la constitution d'une base de test contenant un nombre important de signatures authentiques est indispensable.

2.2. Fausses signatures

Lorsqu'on évalue un système d'authentification par signature manuscrite, on doit prendre en compte trois types de faux : les faux aléatoires, les faux simples et les faux expérimentés [SAB95]. Les faux aléatoires sont réalisés par une personne ne connaissant pas la forme de la signature à imiter. Les faux simples sont des signatures pour lesquelles le libellé est identique mais la graphie différente. Les faux expérimentés sont réalisés par des personnes ayant accès à la fois à la forme et à la dynamique, voire à des informations sur la méthode d'authentification.

2.2.1. Types de Faux

Le faux aléatoire est obtenu en employant sa propre signature à la place de celle à imiter. A l'opposé du faux simple, le libellé d'un faux aléatoire est évidemment différent du libellé de l'authentique. Sa détection est donc a priori assez facile.

Le faux simple est rédigé sans tentative de copier la forme de la signature mais en connaissant le libellé c'est à dire le nom. C'est le faux le plus fréquemment rencontré en pratique. On peut identifier le scripteur du faux simple car ce dernier présente souvent un bon nombre de caractéristiques intrinsèques propres à son auteur.

Le faux par calque est obtenu en reproduisant fidèlement une signature authentique à l'aide d'un moyen quelconque de transfert de l'image de l'authentique sur un document. Il a toutes les caractéristiques d'un dessin. Cette technique est généralement bien adaptée aux systèmes d'authentification hors ligne. Le faux par calque manque de spontanéité, donne l'apparence de mouvements lents et d'une pression uniforme et les retouches sont souvent détectables.

On distingue deux types de faux par imitation, le faux par imitation servile et le faux par imitation libre. Lors d'une imitation servile d'une signature, le faussaire copie directement le modèle et s'y réfère aussi souvent que nécessaire. L'imitation servile de la signature authentique présente un dessin assez ressemblant à l'authentique. Parmi les divergences entre différents échantillons de ce type, on trouve les espacements, les alignements et l'inclinaison relative des lettres. On remarque également la présence d'une mauvaise inclinaison moyenne de l'écriture, un tracé lent et hésitant et la présence fréquente de retouches ou reprises.

Dans le cas d'une imitation libre, le faussaire procède par l'étude soignée de la signature authentique. Il mémorise l'image générale de la signature et le dessin des lettres, leurs espacements et autres détails picturaux. Le faussaire s'entraîne pour imiter la signature authentique, il compare le faux et l'authentique entre les essais et répète la même procédure jusqu'à entière satisfaction. A la différence du faux par imitation servile, celui-ci est caractérisé par une allure spontanée. Les principales divergences de ce faux par rapport à l'authentique résident dans les proportions relatives des lettres, des espacements, le's types des alignements et notamment un manque d'alternance des pleins et des déliés.

2.2.2. Remarques

On parle de faux par déguisement si le scripteur réalise une signature différente de sa signature habituelle dans le but de la renier ultérieurement.

On remarque que le faux par imitation libre est le faux le plus difficile à détecter et également le plus difficile à produire.

Enfin, notons que le pourcentage de faux acceptés est très difficile à évaluer car il est impossible d'obtenir des faux réalisés par des faussaires professionnels.

2.2.3. Création de fausses signatures : mode d'emploi

Afin de mieux détecter les fausses signatures, il peut être intéressant d'avoir une réflexion sur la manière dont les faussaires pourraient procéder.

La première difficulté concernant la création de fausses signatures est de collecter les données nécessaires c'est à dire :

- plusieurs exemples de signatures originales pour évaluer la forme de la signature, ses variations et la gestuelle du scripteur
- plusieurs enregistrements de vidéos de la personne en train de signer pour évaluer la dynamique de la signature et ses variations

Réaliser une imitation d'une signature requiert beaucoup d'entraînement car il est très difficile, voire impossible, d'imiter à la fois la forme et la dynamique. La durée de l'entraînement est proportionnelle à la difficulté de reproduction de la signature. En effet, une signature stable avec beaucoup de changements de direction et de rythme sera beaucoup plus difficile à reproduire qu'une signature relativement variable constituée d'un tracé simple et sans changement de rythme.

Une autre difficulté réside dans le fait qu'il n'est pas possible pour le faussaire de s'entraîner sur un système existant puisque celui-ci se bloque au bout d'un certain nombre de tentatives d'authentification. C'est le même principe qui est utilisé pour le code PIN d'une carte bancaire ou d'une puce de téléphone portable. Après, en général, trois échecs lors de la phase d'authentification, la carte ou le téléphone est verrouillé.

Le faussaire devra éventuellement essayer de trouver le fichier référence de l'utilisateur dont il souhaite imiter la signature.

Le faussaire devra donner une réponse aux questions :

- quelles sont les données collectées lors de la signature?
- quelles sont les caractéristiques extraites de la signature?
- comment sont-elles extraites de la signature?

La signature ne doit pas être identique aux signatures précédentes car un contrôle de rejeu est effectué. Cela signifie que ces fausses signatures ne pourront pas être l'œuvre d'un robot auquel on aurait fourni une vidéo de la personne en train de signer.

2.3. Avantages de l'utilisation de la signature manuscrite

En dépit des difficultés que nous venons de relever, les avantages de l'utilisation de la signature manuscrite comme un moyen d'authentification forte sont nombreux.

Concernant la pertinence de son utilisation, en apposant sa signature manuscrite, chaque signataire exprime – dans le sens propre du terme – l’empreinte de sa personnalité. Les juristes sont unanimes sur le fait que la signature électronique (i.e. le mot de passe) ne peut remplacer entièrement la signature manuscrite. L’authentification certaine des utilisateurs de

signatures électroniques ne peut être garantie qu'en y associant des caractéristiques biométriques. En effet, les cartes à puce, les codes confidentiels ainsi que les mots de passe ne représentent pas des références purement individuelles et en conséquence peuvent être sujets de manipulations ou de vols. De plus, contrairement aux mots de passe ou aux codes confidentiels, on n'oublie jamais sa signature. Par rapport aux autres technologies basées sur la biométrie physiologique, son utilisation ne nécessite pas généralement un coût supplémentaire élevé pour le capteur.

Concernant la fiabilité, chaque signature est unique, car elle reflète les propres habitudes, de nature autant physiologique que biomécanique, ainsi que le rodage individuel quotidien. Deux signatures ne peuvent jamais être exactement identiques sauf s'il s'agit d'une copie. Mais cela est automatiquement détectable.

Apposer sa signature sur un document est un acte bien accepté. En général, la signature a déjà fait l'objet de stockage au niveau, non seulement d'institutions financières, mais également au sein de diverses autres institutions. De plus, à ce jour, un des signes les plus fréquemment acceptés pour permettre la non répudiation ou la preuve d'engagement de l'individu est sa signature manuscrite. Son utilisation pour l'authentification est autant habituelle qu'acceptée, aussi bien pour les clients que pour les prestataires. A contrario, les autres procédés, notamment la prise d'empreintes digitales et la forme et l'aspect de l'iris, sont jugés trop invasifs pour un usage grand public. En effet, les systèmes basés sur l'empreinte digitale ont une connotation d'investigation criminelle et ceux basés sur l'iris nécessitent un contact très proche de l'œil avec le système d'acquisition. Le dernier avantage que l'on peut citer est que l'authentification par signature manuscrite est très facile à expliquer par rapport aux autres techniques d'authentification biométrique.

Toutes les raisons citées ci-dessus expliquent pourquoi la signature manuscrite a été retenue pour cette étude. Suivant la méthode de capture de la signature, on distingue deux familles de signatures : hors ligne et en ligne. Le paragraphe suivant décrit les différences existant entre les deux familles.

2.4. Différences entre hors ligne et en ligne

Dans un système hors ligne, la signature est effectuée sur un support papier puis scannée. La signature est donc assimilée à une image en niveaux de gris. C'est le cas notamment pour les systèmes de vérification de chèques. En hors ligne, on ne dispose pas de la dynamique de façon directe mais d'autres informations sont disponibles comme l'épaisseur du trait ou la variation d'intensité du niveau de gris constituant la signature. Au contraire, lors d'une

acquisition en ligne, le trait n'a pas d'épaisseur et est représenté avec la même intensité sur les systèmes ne permettant pas l'acquisition de la pression. Hormis pour l'étude de la forme, les techniques appliquées en hors-ligne ne peuvent donc pas, en général, être adaptées aux techniques en ligne puisqu'elles sont basées la plupart du temps sur l'étude des niveaux de gris de l'image [SAB97][SAN04] [WIR03].

Dans le cas d'un système en ligne, la signature est effectuée sur une tablette graphique ou tout autre support muni d'un stylet électronique. La signature est donc représentée par une suite de points définis par au moins 3 valeurs : x, y, t. Nous avons remarqué, lors de nos expérimentations, que les dispositifs actuels d'acquisition de l'écriture manuscrite en ligne sont loin d'offrir une ergonomie suffisante pour que les usagers les utilisent sans stress. En effet, la gêne occasionnée entraîne des efforts supplémentaires. Beaucoup de personnes adaptent ou modifient leur manière d'écrire et de signer lors du passage sur un support numérique. Cela est critique lorsqu'il s'agit de signer car on ne signe pas de la même manière sur papier ou avec un stylet et un temps d'adaptation au support numérique est donc nécessaire avant d'obtenir une stabilité suffisante de la signature.
Les problèmes liés à l'acquisition sont différents dans le cadre du en ligne et dans celui du hors ligne. En effet, en hors ligne, le papier utilisé pour signer peut être de différentes textures, le stylo a aussi une grande influence et enfin l'acquisition via le scanner peut donner des résultats différents suivant la résolution choisie. C'est aussi le cas pour les systèmes d'acquisition en ligne pour lesquels la résolution ou la fréquence d'acquisition ne sont pas fixées.

2.5. Variabilité des signatures manuscrites

2.5.1. Variation intra individu

Les signatures successives d'un même individu varient globalement et localement et diffèrent en orientation et en échelle. En effet, suivant le contexte, les signatures sont de longueurs et de durées différentes même si elles sont faites par un même scripteur et des variations aléatoires existent comme des ajouts ou retraits de traits. Par conséquent, comme nous l'avons déjà indiqué, si deux signatures de la même personne sont parfaitement identiques alors l'une d'entre elles peut être considérée comme un faux.
La plupart des articles traitant de l'authentification par signature manuscrite font état de personnes pour lesquelles le système d'authentification ne fonctionne pas. En effet, certaines

personnes ont une signature trop instable pour pouvoir établir un modèle représentatif de leur signature. Cela peut être dû à l'utilisation d'un nouveau support nécessitant une période d'adaptation. Il est notamment perturbant d'écrire avec un stylet sur une surface particulière ou encore d'écrire dans une zone restreinte. Dans [TAN01], l'auteur montre qu'il existe une forte corrélation positive entre une grande instabilité de la signature et une grande variance du temps total mis pour réaliser la signature. La durée totale de la signature peut donc être utilisée pour mesurer la variabilité intra individu d'une signature.

2.5.2. Variation inter individus

Les signatures sont très variées même pour des personnes d'un même pays. En effet, au-delà des habitudes culturelles, certaines personnes ont des signatures très complexes alors que d'autres écrivent uniquement leur nom. Cependant, on peut distinguer deux grandes catégories de signatures : les signatures occidentales et les signatures asiatiques. Les signatures occidentales peuvent elles-mêmes être classées en deux sous catégories : les paraphes très éloignés de la forme du nom telles les signatures européennes et les signatures très proches du nom telles certaines signatures anglo-saxonnes. Les signatures asiatiques sont très différentes des signatures occidentales; elles sont constituées de traits très courts séparés par des levés de stylet et orientés suivant un axe vertical. Par conséquent, les systèmes d'authentification par signatures manuscrites basés directement sur le style des signatures anglo-saxonnes ou asiatiques ne seront pas aussi performants dans le cas de signatures européennes.

2.5.3. Gestion de la variabilité des signatures

Il est très difficile de comparer deux systèmes de vérification de signature étant donné qu'aucune base de données internationale n'est disponible [JAI02].

Chacune des méthodes proposées a sa spécificité et est donc adaptée à un problème (identification ou vérification) et à un type de signatures (anglo-saxonnes, asiatiques ou européennes) voire à une base de données spécifique (celle qui a servi à faire les tests). Par conséquent, toute l'élaboration d'un système d'authentification dépend du but recherché. Une solution possible pour concevoir un système d'authentification plus générique serait peut être que celui-ci classifie au préalable les signatures en familles – paraphes, écritures… - pour leur appliquer ensuite un traitement spécifique en fonction de leur type [BOU97].

Le problème lors de l'évaluation des performances des systèmes repose sur la difficulté à créer des bases de signatures authentiques conséquentes. Une solution possible pourrait être de

générer de nouvelles signatures d'apprentissage représentatives de la variabilité de la signature à partir de celles existant déjà.

2.6. Bases de signatures

Les données acquises lors de la réalisation d'une signature sont une succession de points ordonnés dans le temps pour lesquels on peut disposer de plusieurs informations. Dans le cas présent, nous avons souhaité ne collecter que des informations disponibles sur l'ensemble des supports munis d'une interface graphique c'est à dire la position, le temps et le contact afin de détecter les levés de stylet. De plus, les expériences de la littérature montrent que les autres informations, telles que la pression ou l'angle du stylet avec la surface, ne sont pas des caractéristiques ayant donné jusqu'à présent de très bons résultats [TAN01][FRA03].

Les travaux présentés dans la suite de cette étude ont été évalués tout d'abord, sur une base de signatures en ligne constituée pour une compétition internationale : "Signature Verification Competition" (SVC) ayant eu lieu en 2004 dans le cadre de la conférence "International Conference on Biometric Authentication" (ICBA) [SVC04]. Cette base contient les signatures de 40 personnes. Pour chaque personne on dispose de 20 signatures authentiques et de 20 faux expérimentés (Figure 6). Les faux expérimentés ont été réalisés par des personnes ayant accès à une vidéo de la personne en train de signer.

La particularité de cette base est que les signatures enregistrées ne sont pas de véritables signatures mais des signatures créées pour l'élaboration de la base. La principale conséquence de ce choix est une plus faible stabilité de la signature que dans les cas concrets.

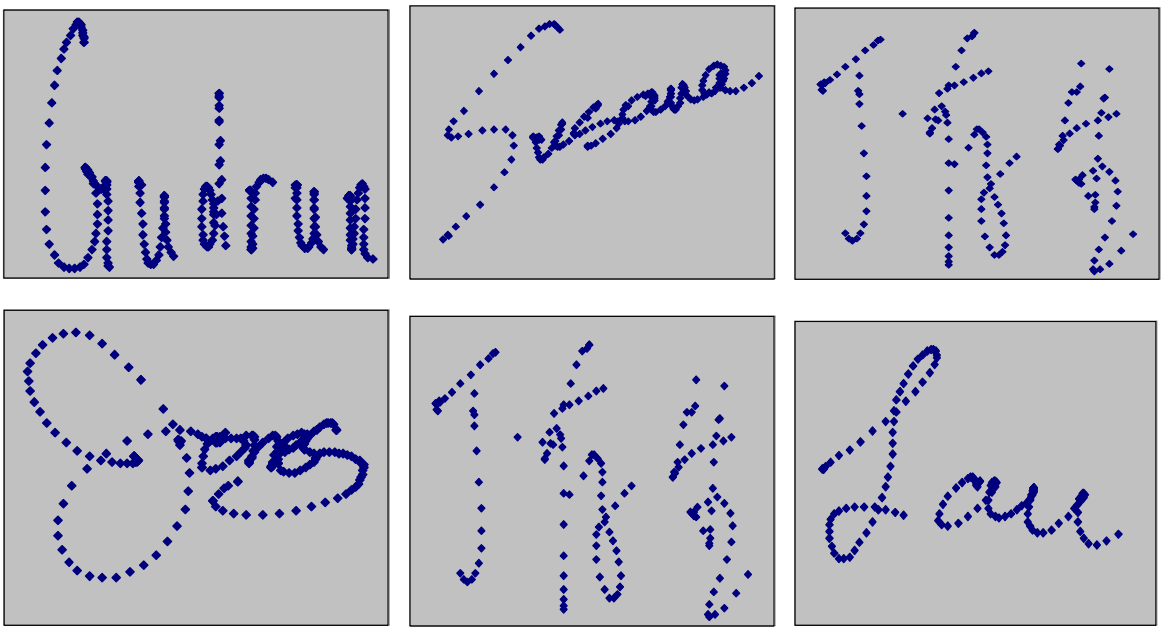


Figure 6. Exemples de signatures de la base SVC.

Nous allons maintenant étudier plus en détail le principe de fonctionnement d'un système d'authentification basé sur la signature manuscrite en ligne.

3. METHODES CLASSIQUES UTILISEES EN AUTHENTIFICATION DE SIGNATURES EN LIGNE

Depuis une cinquantaine d'années, l'authentification de signatures a été abordée par de nombreux chercheurs comme en témoignent le grand nombre de publications sur le sujet [HER77][HAS92][NAL97] et les états de l'art associés [GUP97b][GRI00][JAI02]. Les principales méthodes actuelles sont basées sur l'étude de la forme, de la cinématique et de la pression. Dans ce chapitre nous présentons, pour chaque étape du processus d'authentification, les différentes solutions exposées dans la littérature depuis 1994, date à laquelle est paru l'état de l'art de Plamondon et Leclerc [LEC94] qui est la suite de celui réalisé par Plamondon et Lorette [PLA89]. Etant donné que l'écriture a un comportement différent des signatures, nous ne nous attarderons pas sur les méthodes utilisées dans le domaine de l'authentification par l'écrit manuscrit [BEN03]. Après avoir présenté les travaux universitaires de ces dernières années dans le domaine de l'authentification par signature manuscrite, nous présenterons les produits industriels existant sur le marché ainsi que les améliorations à apporter aux systèmes existants.

3.1. Architecture et stratégie

La Figure 7 résume l'architecture classique d'un système d'authentification de signature en ligne. Beaucoup de travaux ont été réalisés pour tenter d'aboutir à des résultats optimaux pour chacune de ces étapes. Par contre peu de recherches ont encore été faites concernant l'architecture du système elle-même. Nous avons essayé de ne pas reproduire cette erreur lors de nos travaux et donc de travailler autant sur l'enchaînement voire la coopération entre traitements que sur les algorithmes eux-mêmes.

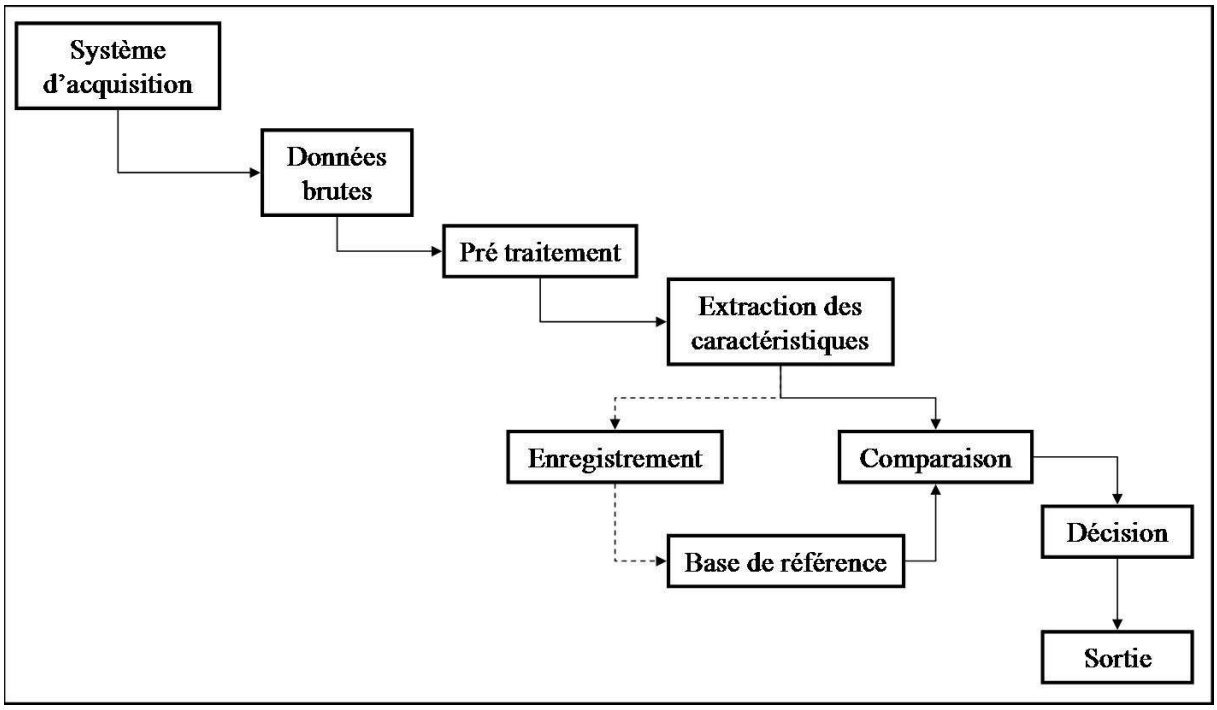


Figure 7. Schéma d'un système d'authentification.

Pour chacune des étapes constituant le schéma classique d'un système d'authentification par signature manuscrite en ligne, nous présentons les différentes voies explorées durant ces dix dernières années par l'ensemble des chercheurs. Puis nous ferons un récapitulatif des principaux systèmes existant avant d'étudier à quels niveaux des améliorations peuvent être apportées.

3.2. Acquisition et Prétraitement

3.2.1. Acquisition

L'acquisition de signatures manuscrites en ligne se fait au moyen d'un stylet électronique ou sur une tablette graphique.

Les supports classiques actuellement disponibles sur le marché sont au nombre de quatre (Figure 8).

Le **PDA et le stylet Anoto** sont les supports offrant le moins de possibilités. Leurs puissances de calcul sont relativement faibles et les données acquises sont uniquement la position du stylet et le temps. La **tablette graphique** est le support le moins cher et le plus répandu. De plus, c'est celui qui offre l'éventail le plus large au niveau des données d'acquisition. Aux informations de base, peuvent s'ajouter la pression, l'angle du stylet avec la tablette, etc. Le **Tablet PC** est un support intégré mais encore peu répandu. Son principal avantage est qu'il ne

nécessite pas de périphérique extérieur. De plus, on dispose de l'information de pression du stylet. Les caractéristiques des différents supports d'acquisition sont résumées dans le **Tableau** 1.

PDA ANOTO Tablet PC Tablette Graphique

Figure 8. Illustrations des supports classiques.

	SUPPORT			
CAPTEUR	PDA	Anoto	Tablet PC	Tablette Graphique
Coordonnées	●	●	●	●
Temps	●	●	●	●
Contact		●	●	●
Pression			●	●
Angle du stylet				●

Tableau 1. Caractéristiques des différents dispositifs d'acquisition.

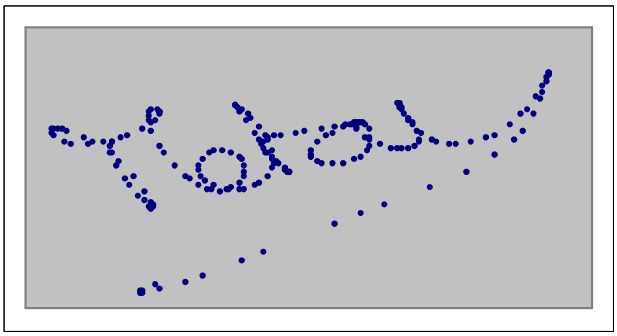


Figure 9. Exemple d'acquisition de signature en ligne.

Les données acquises lors de la réalisation d'une signature sont une succession de valeurs qui reflètent des informations sur les points de la signature. Au minimum, on dispose de la position des points et de la date mais on peut aussi connaître la pression ou encore l'angle du stylet avec la tablette (Figure 9).

En plus des supports classiques, d'autres supports d'acquisition ont été développés spécifiquement pour l'authentification par signature manuscrite. Certains chercheurs, souhaitant n'utiliser que la pression, ont développé leur propre système d'acquisition : un stylet ayant une plus grande sensibilité et une plus grande fréquence d'acquisition [KIK01].

Suivant le type de signatures à analyser, rapides comme les signatures anglo-saxonnes ou lentes comme les signatures japonaises, il est possible de régler la fréquence d'acquisition.
Deux autres systèmes ont été notamment proposés dans la littérature. L'un permet de capturer 5 signaux différents qui sont les forces exercées sur le stylet dans 3 directions et les angles entre le stylet et la surface d'écriture dans 2 directions [MAR98]. L'autre est basé sur l'utilisation d'une caméra [MAR97a]. Le principe consiste, après avoir repéré la position de départ du stylet, à effectuer un suivi du stylet dans chacune des images et ainsi à extraire l'évolution de la position du stylet au cours du temps. La plupart des études faites jusqu'à présent utilisaient des tablettes graphiques comme support d'acquisition et peu d'études traitent de l'influence du support d'acquisition sur les résultats obtenus. Aujourd'hui les types de support se multiplient et se démocratisent (mobiles, consoles, …), il pourrait donc être intéressant d'étudier l'influence du support choisi sur les résultats obtenus.

Etant donné que l'acquisition des données n'est pas toujours effectuée dans les mêmes conditions, il est nécessaire de réaliser un prétraitement. Cependant cette étape est moins importante que pour les systèmes hors ligne car, d'une part, seuls les points de la signature sont acquis et, d'autre part, dans le cas du hors ligne, le réglage de la numérisation peut être difficile.

3.2.2. Prétraitement

Le prétraitement consiste à préparer les données pour l'analyse : il peut s'avérer nécessaire de réduire le bruit, de lisser les tracés et de les coder sous une forme plus facilement exploitable.
Un des principaux prétraitements effectués consiste à rééchantillonner la signature, en procédant par interpolation linéaire, afin d'obtenir des points équidistants, soit pour avoir un nombre de points multiple de deux pour appliquer des méthodes basées, par exemple, sur les ondelettes comme dans [LEJ01], soit pour avoir le même nombre de points dans chaque signature et ainsi faciliter leur comparaison comme dans [MOH99] [RHE01] [ZHA96]. L'application d'un prétraitement est conditionnée par le choix des méthodes d'analyse utilisées. Pour faciliter la comparaison, le rééchantillonnage peut se faire en utilisant le même nombre de points pour toutes les signatures quelle que soit leur taille. Cependant, cela présente un inconvénient car on perd alors une grande partie de l'information sur la dynamique du tracé. De plus, cela nécessite une interpolation de la signature qui peut être très approximative surtout si la signature est rapide. Aussi certains chercheurs, afin de ne pas perdre trop d'informations, proposent de ne rééchantillonner que partiellement la signature

[JAI02]. Les zones de début et de fin de trait, ainsi que celles de changement de direction ne sont alors pas rééchantillonnées car ces points apportent une information plus importante que les autres points. Les informations liées à la dynamique sont calculées avant le rééchantillonnage puis recalculées par interpolation pour les points retenus. Une autre manière de normaliser la signature est d'utiliser la transformée inverse des coefficients de Fourier normalisés [KAS97].

Jusqu'à présent peu de méthodes d'authentification utilisent les mouvements du stylet pendant les levés donc cette information n'est pas prise en considération au moment de l'authentification [DIM02] [YAN95] car les anciennes tablettes ne permettaient pas d'obtenir cette information.

Le dernier aspect traité est la suppression des points redondants ou très proches, i.e. séparés par un ou deux pixels, pour minimiser les effets de résolution spatiale des tablettes graphiques [LEE96].

3.2.3. Normalisation

La normalisation a pour but de simplifier la comparaison entre les signatures. En effet, certaines caractéristiques extraites des signatures ne sont pas invariantes aux différentes transformations affines. Une étape de normalisation est alors indispensable si l'on ne veut pas tenir compte de ce type de variations lors de la comparaison. Ainsi, dans [HUA95], un filtre moyenneur est appliqué sur les signaux en entrée (x(t), y(t), v(t), a(t)) pour réduire le bruit dû à l'acquisition puis la signature est décalée de manière à ramener le centre de gravité de la signature à l'origine du repère mais aucune rotation n'est effectuée. Dans [KAS97], la normalisation porte sur les coefficients de Fourier issus du signal en entrée (x(t),y(t)). Dans [LEJ01], deux opérations sont effectuées. Tout d'abord, une rotation de la signature est effectuée de manière à ce que la pente du premier axe d'inertie de la signature soit nulle dans le repère de référence utilisé pour l'acquisition, puis une homothétie pour ramener l'ensemble des signatures dans un carré de même taille.

Plusieurs possibilités apparaissent lors de la mise en place du système. Dans [LEE96], étant donné que lors de l'enregistrement une ligne est présente sur la tablette, aucune normalisation n'est effectuée. La normalisation a comme principal inconvénient de supprimer des informations sur la manière avec laquelle la signature est réalisée comme la position, l'orientation et la taille de la signature par rapport au cadre. Il est également possible de sauvegarder certains indices (position, taille ou orientation) avant d'effectuer la normalisation

si l'on désire exploiter ces informations plus tard. C'est cette dernière solution qui nous paraît la plus intéressante et que nous avons donc retenue.

3.3. Caractérisation de la signature

Là encore plusieurs étapes peuvent être mises en évidence. Avant d'extraire certaines caractéristiques, différents auteurs font appel à une segmentation. Parmi les caractéristiques proposées, nous montrerons que certains choix doivent être opérés.

3.3.1. Segmentation

Le but de la segmentation est de pouvoir observer certaines parties de la signature à différents niveaux. Ainsi, au lieu de comparer les signatures dans leur ensemble, on divise les signatures en parties que l'on compare une à une. La somme des différences entre les parties stables observées dans deux signatures comparées nous donne une distance globale entre les signatures.

Une segmentation est nettement plus complexe dans le cadre de signatures européennes que dans le cas de signatures asiatiques ou anglo-saxonnes. En effet, ce type de signatures ne peut pas, en général, être segmenté au niveau des levés de stylet comme c'est le cas pour les signatures anglo-saxonnes ou asiatiques. Cela est dû au fait que les signatures anglo-saxonnes sont très proches de l'écriture du nom d'où un grand nombre de levés de stylet et les signatures asiatiques sont constituées d'une succession de traits délimités par des levés de stylet comme des lettres majuscules alors que les signatures européennes sont généralement plus proches du paraphe.

La difficulté de la segmentation est qu'elle doit être suffisamment stable pour un même individu et donc se baser sur des points "caractéristiques" de la signature. La principale méthode consiste à segmenter aux points où la vitesse est nulle ou faible comme les points de forte courbure. L'avantage de cette approche est qu'elle est stable par rapport à la taille de la signature. Dans [HUA95], après avoir calculé la vitesse instantanée en chaque point de la signature, une fenêtre glissante est utilisée pour parcourir la signature et détecter les points correspondant aux minimums locaux. Dans [DOL98], la segmentation de la signature se fait aux points où la vitesse verticale est nulle. D'autres approches existent notamment celle de [BRA93] utilisée dans [SCH97] et [RHE01] pour détecter les points importants de la signature et permettre la segmentation en ces points. L'importance d'un point dépend des variations de l'angle défini par ce point et ses points voisins. Fonctionnant sur le même

principe, la méthode de Brault [BRA93] est décrite de manière plus précise dans le chapitre 2 paragraphe 2.2.3.

Une fois segmentées, la comparaison des signatures est basée sur la différenciation entre les parties stables et non stables d'une signature, ces zones sont déterminées à partir des modèles de référence. Pour pouvoir isoler les parties de la signature qui sont constantes et celles qui varient, il est nécessaire d'utiliser une technique de comparaison telle que celles présentées au chapitre 3. Les segments des signatures, considérés comme trop variables (instables) seront ignorés lors de la phase de comparaison.

3.3.2. Extraction de caractéristiques

Cette étape est une des plus importantes car elle va conditionner la suite du traitement. En effet, après cette étape, la signature ne sera plus représentée par une suite de points mais par un vecteur constitué des valeurs de chacune des caractéristiques choisies.

Il est à noter que cette étape est souvent celle qui nécessite le plus de temps de calcul.

Etant donné qu'il est très difficile d'imiter à la fois la forme et la dynamique d'une signature manuscrite, l'étude des caractéristiques s'articule souvent suivant deux axes : forme et dynamique. Dans un premier temps, seront présentées les caractéristiques liées à la forme puis nous aborderons celles liées à la dynamique. Les caractéristiques retenues dans cette présentation des méthodes existantes sont celles le plus souvent citées dans la littérature.

On note $S = (Pt_1, Pt_2, \ldots, Pt_i, \ldots, Pt_n)$ une signature constituée de n points $Pt_i=(x_i,y_i,t_i)$.

3.3.2.1 Caractéristiques liées à la forme

Nalwa [NAL97] affirme que l'utilisation des caractéristiques dynamiques uniquement ne permet pas d'identifier l'auteur d'une signature. De plus, le fait d'utiliser la forme permet de mieux faire comprendre à l'utilisateur pourquoi sa signature a été rejetée. Le Tableau 2 fournit la liste des caractéristiques liées à la forme, ainsi que les méthodes de calcul et les commentaires associés. Nous nous sommes limité à celles le plus souvent utilisées :

Longueur totale du trait	$\sum_{i=1}^{n-1}(dist_{Eucl}(Pt_i, Pt_{i+1}))$
Distance entre le premier et le dernier point	$dist_{Eucl}(Pt_1, Pt_n)$
Rapport entre les déplacements vers la gauche et vers la droite	$\frac{-\sum_{i=1}^{n-1}\max(x_i - x_{i+1}, 0)}{\sum_{i=1}^{n-1}\min(x_i - x_{i+1}, 0)}$
Rapport entre les déplacements vers le haut et vers le bas	$\frac{-\sum_{i=1}^{n-1}\max(y_{i+1} - y_i, 0)}{\sum_{i=1}^{n-1}\min(y_{i+1} - y_i, 0)}$
Rapport entre les déplacements suivant X et suivant Y	$\frac{\sum_{i=1}^{n-1}(x_i - x_{i+1})}{\sum_{i=1}^{n-1}(y_i - y_{i+1})}$
Mesure de l'angle formé par la droite horizontale et celle joignant le premier et le dernier point	$Arc\tan\left(\frac{y_n - y_1}{x_n - x_1}\right)$
Somme des mesures d'angles formés par la droite horizontale et celle joignant le point considéré et le point suivant	$\sum_{i=1}^{n-1} Arc\tan(\frac{y_{i+1} - y_i}{x_{i+1} - x_i})$
Somme des valeurs absolues des mesures d'angles formés par la droite horizontale et celle joignant le point considéré et le point suivant	$\sum_{i=1}^{n-1}\left\lvert Arc\tan(\frac{y_{i+1} - y_i}{x_{i+1} - x_i})\right\rvert$
Déplacement horizontal moyen	$\frac{x_{max} - x_{min}}{nbPts}$
Histogramme horizontal défini sur N colonnes	
Nombre de traits dans la signature	Un trait est une composante connexe de la trace sur le support
Nombre d'intersections du tracé (peu stable)	

Tableau 2. Caractéristiques liées à la forme.

Les caractéristiques les plus souvent utilisées sont la longueur de la signature et le nombre de traits. Le nombre moyen de caractéristiques utilisées est de 2 ou 3. La caractéristique la plus stable est la longueur de la signature.

3.3.2.2 Caractéristiques liées à la dynamique

De même, pour la dynamique, le tableau 3 liste les caractéristiques qui sont classiquement utilisées :

Temps total	$t_n - t_1$
Vitesse moyenne	$\frac{\sum_{i=1}^{n-1}(dist_{Eucl}(Pt_i, Pt_{i+1}))}{t_n - t_1}$
Vitesse moyenne verticale	$\frac{\sum_{i=1}^{n-1}\lvert y_{i+1} - y_i\rvert}{t_n - t_1}$
Vitesse moyenne horizontale	$\frac{\sum_{i=1}^{n-1}\lvert x_{i+1} - x_i\rvert}{t_n - t_1}$
Vitesse maximale	$\max_{i=1,\ldots,n-1}\left(\frac{dist_{Eucl}(Pt_i, Pt_{i+1})}{t_{i+1} - t_i}\right)$
Accélération moyenne	$\frac{\sum_{i=1}^{n-1}(dist_{Eucl}(Pt_i, Pt_{i+1}))}{(t_n - t_1)^2}$
Accélération maximale	$\max_{i=1,\ldots,n-2}(\frac{dist_{Eucl}(Pt_i, Pt_{i+2})}{(t_{i+1} - t_i)^2})$

Tableau 3. Caractéristiques liées à la dynamique.

Les caractéristiques les plus souvent utilisées sont la durée totale du tracé de la signature, la vitesse moyenne suivant x et suivant y. Le nombre moyen de caractéristiques utilisées liées à la dynamique est de 4 ou 5. La caractéristique la plus stable est la durée totale de la signature. Le tableau 3 mentionne des caractéristiques globales. Certains auteurs proposent d'utiliser des caractéristiques plus locales telles que la vitesse instantanée ou l'accélération en chaque point. La vitesse instantanée est très sensible à la qualité de l'acquisition puisque le calcul se fait en utilisant une information très locale. Dans [HER77], l'accélération est utilisée pour étudier de façon qualitative l'activité musculaire lors de la réalisation de la signature, i.e. les forces

exercées sur le stylet. Cette caractéristique est notamment utilisée pour détecter les faux expérimentés.

3.3.2.3 Autres méthodes et bilan

Les études antérieures semblent prouver que les paramètres liés à la forme sont les plus fiables et les plus constants. En effet, les données relatives à la dynamique sont plus sensibles au système d'acquisition car la fréquence d'acquisition varie d'un système à l'autre.

L'utilisation de la pression comme caractéristique s'explique par le fait qu'elle possède l'avantage d'être d'une part une caractéristique cachée et difficile à reproduire [KIK01] et d'autre part elle est nouvelle par rapport au hors ligne. Cependant, aucune étude n'a prouvé que la pression soit propre à un scripteur et stable d'une signature à l'autre. De plus, les résultats obtenus avec l'utilisation de la pression uniquement ne sont pas concluants même lorsque l'on utilise un dispositif particulier pour l'acquisition de la pression [TAN01].

Certaines recherches ont essayé de modéliser la signature à l'aide de modèles connus. Ainsi, le but de la méthode proposée par [WES00] est de retrouver des informations sur la vitesse et l'accélération en utilisant les dérivées des B-Splines cubiques ajustées sur la trajectoire initiale. La méthode proposée par [LEJ01] est basée sur les ondelettes. Elle consiste en l'application d'une transformation en ondelettes sur les fonctions extraites de la signature (pression, vitesse suivant X et Y, angle du stylet et vitesse angulaire) pour compresser le signal et ne conserver que certains coefficients pour la suite du traitement. On peut aussi citer le travail de [MOH99] qui utilise des modèles auto régressifs pour modéliser la signature et utilise ensuite les coefficients déduits du modèle comme entrées d'un réseau de neurones.

Peu de travaux de recherche sont axés sur la recherche de nouvelles caractéristiques hormis [NAL97] qui ne s'est basé que sur la forme du tracé pour extraire des caractéristiques de la signature. Les efforts ont plutôt porté sur les méthodes de segmentation et de comparaison.

Comme le fait remarquer [GUP97a], l'étude de la forme est souvent négligée dans les systèmes d'authentification en ligne car la comparaison de forme n'est pas aisée et on obtient de bons résultats en utilisant uniquement la dynamique. Mais l'utilisation de la dynamique seule peut conduire à des aberrations [NAL97].

3.3.3. Sélection de caractéristiques

Parmi l'ensemble des caractéristiques définies précédemment, toutes ne sont pas discriminantes ou ne doivent pas être associées les unes avec les autres.

Plus le nombre de caractéristiques utilisées est élevé, plus les performances sont améliorées jusqu'à obtenir un maximum correspondant à une configuration optimale puis les performances décroissent à nouveau si on utilise trop de caractéristiques. De plus, l'utilisation d'un grand nombre de caractéristiques présente un certain nombre d'inconvénients. Parmi ceux-ci on peut citer tout d'abord la nécessité d'avoir un espace mémoire conséquent pour stocker les signatures ainsi qu'une augmentation exponentielle de la complexité de l'authentification en fonction du nombre de variables. De plus, étant donné que les signatures ne sont pas rigoureusement identiques, la distance entre la signature testée et la signature de référence va augmenter avec le nombre de caractéristiques et donc le seuil devra être augmenté d'où un risque plus important d'erreur. La mise en place d'une technique de sélection de caractéristiques peut donc s'avérer intéressante. Une des méthodes les plus adaptées à la sélection est l'utilisation d'algorithme génétique [XUH96]. Le principe est de coder l'utilisation ou non des différentes caractéristiques au niveau des chromosomes et d'utiliser, comme fonction de fitness à minimiser, le pourcentage d'erreur ou la distance entre les signatures du même individu. Le meilleur chromosome obtenu indique quelles sont les caractéristiques à retenir pour optimiser les performances du système.

Dans le cas de l'authentification par signature en ligne, le problème sera de disposer d'une base de signatures suffisamment importante pour effectuer la sélection, l'idéal étant de disposer de faux pour chaque individu de la base [RHE01].

3.4. Authentification ou reconnaissance

Après avoir choisi le principe sur lequel repose le système, c'est à dire la nature des caractéristiques retenues, il s'agit de le mettre en œuvre aussi bien au niveau de la phase d'enrôlement qu'à celui de la phase d'authentification elle-même. Nous abordons ici ces deux aspects.

3.4.1. Choix des modèles

La phase d'enregistrement consiste à créer un modèle, une signature de référence pour une personne à partir des signatures qu'elle a réalisées. Ces signatures sont qualifiées de signatures d'apprentissage. Par la suite, toute nouvelle signature sera comparée à ce modèle. L'idéal est de réaliser cette collecte de signatures sur plusieurs sessions pour avoir un meilleur aperçu de la variabilité des signatures. Un problème se pose si pour une personne donnée la variabilité des signatures est très importante. Une solution possible peut être d’augmenter le seuil mais cela rend le système vulnérable.

Pour répondre à cette problématique, certains auteurs prônent l'utilisation de plusieurs références. Pour choisir ces références, une méthode possible consiste à prendre M signatures parmi les N fournies. Différents critères de sélection sont possibles. Par exemple, ces références peuvent être choisies de façon à ce que toutes les signatures de l'ensemble des N signatures soient à une distance inférieure à un certain seuil d'au moins une signature des M signatures servant de référence collectées pendant l'enrôlement. Pour des raisons pratiques, on doit imposer un cardinal maximum à M pour des raisons algorithmique et à N pour des raisons conviviales.

[GUP97a] propose de combiner les deux possibilités pour la comparaison : créer une signature de référence à partir des signatures utilisées pour l'apprentissage, utiliser toutes les signatures de l'apprentissage pour la comparaison et calculer la distance moyenne ou minimale. Etant donné qu'il est très difficile de créer une signature de référence représentative des signatures d'apprentissage et de leur variabilité, la solution retenue consiste souvent à comparer la signature testée à l'ensemble des signatures d'apprentissage. D'ailleurs la méthode de [WIR97] consistant à construire une signature de référence comme moyenne des signatures effectuées lors de l'enregistrement ne donne pas des résultats très concluants. [HUA03] propose une méthode consistant à représenter l'ensemble d'apprentissage à l'aide d'un graphe regroupant l'ensemble des variations pour chaque segment de chacune des signatures effectuées lors de l'enrôlement autour d'une signature de référence. Ainsi, chaque segment de la signature testée est comparé à l'ensemble des variations du segment correspondant représentées dans le graphe modélisant l'ensemble des signatures d'apprentissage.

[VIE01] propose une méthode pour minimiser le rapport entre le nombre de signatures servant de référence et le nombre total de signatures demandées à l'utilisateur. L'idée de la méthode est de classer les signatures suivant leur distance par rapport au reste de l'ensemble d'apprentissage et de ne garder que les signatures les plus proches les unes des autres. L'avantage de cette méthode est d'éviter de conserver dans l'ensemble des signatures de référence des signatures non représentatives du signataire.

La méthode de [LEC99] consiste à rechercher parmi les signatures d'apprentissage le sous-ensemble le plus stable. Pour évaluer la stabilité entre deux signatures, une mise en correspondance, dont le but est de minimiser la somme des distances entre les points des deux signatures, est effectuée. La stabilité correspond alors au rapport entre le nombre de points n'ayant qu'un seul correspondant et le nombre total de correspondances. L'idée est d'évaluer

pour chaque sous-ensemble de l'ensemble des signatures d'apprentissage le niveau de stabilité et de choisir le sous-ensemble correspondant au maximum de stabilité.

Plutôt que d'effectuer des transformations sur les signatures pour passer dans un nouvel espace : l'espace des caractéristiques sélectionnées pour représenter les signatures, certains auteurs proposent de conserver toute l'information disponible en cherchant une méthode permettant de comparer directement deux tracés manuscrits en ligne. Ce type de technique est décrit dans la partie suivante.

3.4.2. Comparaison de signatures

Lorsqu'on cherche à comparer directement des signatures en ligne constituées de séquences de points ordonnés dans le temps, la difficulté réside dans le choix de la méthode de mise en relation des points constitutifs des deux tracés. Ainsi, la comparaison entre la signature testée et la signature servant de référence consiste à mesurer une distance ou une ressemblance. Si la distance entre ces deux signatures est inférieure à un certain seuil, la signature est reconnue comme authentique sinon elle est reconnue comme un faux.

On distingue les principales méthodes de comparaison suivantes :

a) Celles basées sur l'utilisation de méthodes nécessitant un apprentissage comme les chaînes de Markov cachées [FUE02][YAN95] ou les réseaux de neurones [FUE02] [LEJ01]. Dans le cadre de l'utilisation de ces méthodes, un individu est associé à une chaîne de Markov cachée ou à un réseau de neurones, construit à partir des signatures d'apprentissage, et qui sert de référence. L'inconvénient de ce genre de méthodes est qu'elles ne peuvent pas prendre en compte l'évolution de la signature au cours du temps. Elles présentent plus d'intérêt pour les systèmes d'identification que pour les systèmes de vérification.

b) Les autres méthodes sont basées sur un calcul de similarité entre signatures. Voici quelques unes des distances les plus utilisées :

- **Distance d'édition**

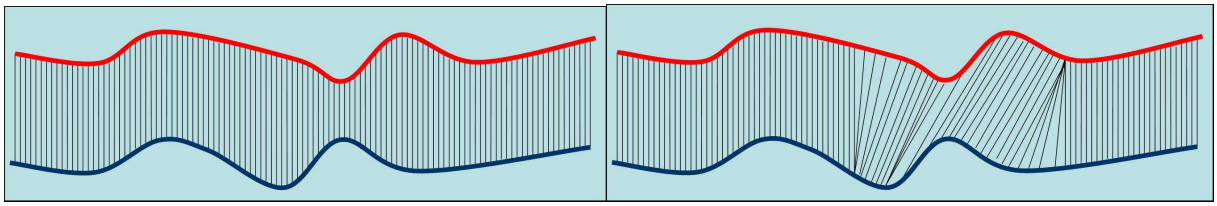
La *distance d'édition* entre deux chaînes x et y, d(x,y) est le coût correspondant à une suite minimum d'opérations d'édition élémentaires (insertion, suppression ou substitution d'un caractère) nécessaires pour transformer x en y [SCH04]. En fonction de la manière dont on code la signature (l'angle relatif entre les points consécutifs par exemple) la distance est invariante à une rotation, à une translation et à une homothétie.

Cette approche a été utilisée dans [GUP97a]. La signature est décomposée en deux ondes : x-profil (t,x) et y-profil (t,y), variations des coordonnées en fonction du temps. On code ensuite

les pics et les vallées dans ces deux ondes. Pour cela, on procède par divisions successives aux endroits de plus faibles amplitudes. On réitère l'opération jusqu'à ne plus obtenir de vallées. Le codage est ensuite amélioré en indiquant l'amplitude du pic (plus l'amplitude est grande, plus le symbole est répété). De plus, on ramène les informations des deux profils dans le même codage afin de représenter la signature par une suite de symboles (i.e. un mot). La comparaison de signatures se ramène donc à la comparaison de mots. L'algorithme utilisé est celui de Wagner et Fisher [WAG74]. L'utilisation basique de la méthode ne donne pas des résultats très concluants pour tous les scripteurs.

- **Distance élastique ou Dynamic Time Warping (DTW)**

Etant donné que deux signatures n'ont pas exactement le même rythme, et que les décalages temporels ne sont pas linéaires, il semble intéressant d'utiliser une distance qui tienne compte de cette variabilité. DTW est un outil qui réalise une mise en correspondance point à point entre deux signaux et qui présente l'avantage d'être insensible aux légères différences dans le rythme et aux différences de longueur de chaînes (Figure 10).



(a) **(b)**

Figure 10. Mise en correspondance point à point entre deux signatures sans prise en compte des décalages temporels (a) et en appliquant l'algorithme DTW (b).

Dynamic Time Warping est une application des techniques de programmation dynamique développées par Bellman dans les années 50 [BEL57]. Cet algorithme est particulièrement utilisé dans le domaine de la reconnaissance de la parole. Cette méthode d'une complexité polynomiale permet de trouver pour chaque élément d'une courbe, le meilleur élément correspondant dans l'autre courbe relativement à un certain voisinage et à une certaine métrique [PLA88]. Il reste donc à définir ces critères pour obtenir un appariement optimal. Dans son implémentation classique, DTW prend en compte le type de déplacement dans la séquence : synchronisation, étirement ou compression.

Les points initiaux de chacune des deux courbes C et C' sont mis en correspondance. Soient Pt_i le ième point de la courbe C et Pt'_j le jème point de la courbe C', les points Pt_i et Pt'_j ayant été mis en correspondance (Figure 11). Soient d_1, d_2 et d_3 les distances suivantes :

$$\begin{cases} d_1 = Dist(Pt_{i+1}, Pt'_j) \\ d_2 = Dist(Pt_i, Pt'_{j+1}) \\ d_3 = Dist(Pt_{i+1}, Pt'_{j+1}) \end{cases}$$

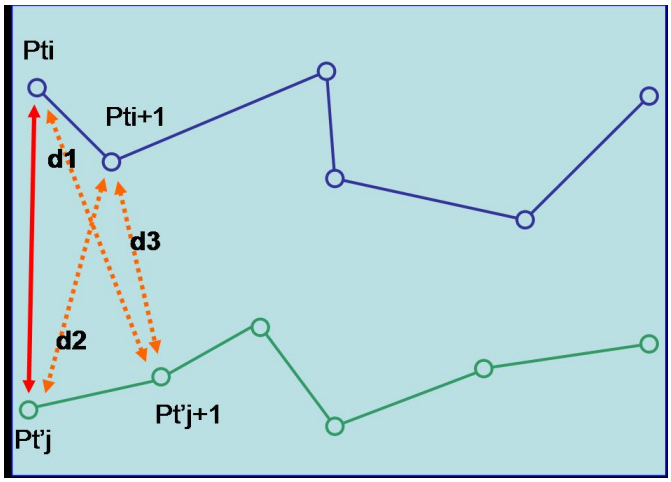


Figure 11. Principe de l'algorithme DTW.

Alors sont mis en correspondance les points Pt_k et Pt'_l réalisant :

$$Dist(Pt_k, Pt'_l) = \min(d_1, d_2, d_3)$$

Classiquement, la métrique utilisée est la distance spatiale euclidienne entre les points des signatures.

Une fois la mise en correspondance des points effectuée, on peut calculer la distance entre les deux courbes en calculant la somme des distances entre couples de points correspondants. Les transformations autorisées dans la recherche de la correspondance sont l'allongement et le rétrécissement suivant l'axe temporel d'un signal relativement à l'autre. Le but de ces ajustements est de minimiser la différence entre les deux signaux. Il existe deux approches pour rechercher ces ajustements : l'une asymétrique et l'autre symétrique. Dans le cas asymétrique, on cherche à établir une correspondance entre la signature testée et un modèle de signature alors que, dans le cas symétrique, on recherche aussi une correspondance entre le modèle et la signature testée. Cette deuxième alternative qui tient compte de la réunion des deux appariements donne de meilleurs résultats [HAS92].

Cette méthode de comparaison de courbes est très utilisée dans le domaine de l'authentification par signature manuscrite en ligne. Calculer des distances entre signatures avec la méthode du DTW permet de mettre en place un système de vérification plus flexible, plus efficace et plus adaptatif que les systèmes basés sur les réseaux de neurones ou les chaînes de Markov cachées car l'apprentissage peut être incrémental. De plus, cet aspect est

très important lorsque l'on envisage d'élaborer une méthode d'authentification qui prenne en compte l'évolution de la signature au cours du temps.

- **Regional correlation**

La technique *Regional correlation* a été introduite par Herbst et Morrissey en 1976 [HER76] puis affinée par Herbst et Liu [HER77]. Pour comparer la signature de référence et la signature testée, chacune d'elles est partitionnée en segments puis une recherche de correspondance entre les segments est effectuée au moyen d'une mesure de corrélation basée sur l'accélération. La segmentation de la signature est effectuée à partir des levés de stylet. Le calcul de la corrélation dépend de la longueur du segment ; plus le segment est long, plus il est difficile de trouver le meilleur alignement entre les segments en raison de l'accumulation de faibles erreurs. Cette méthode est relativement robuste par rapport aux variations pour une signature donnée car elle n'est pas perturbée par des pauses ou des éléments manquants.

- **Split and merge matching**

Split and merge matching a été introduit par Wu et al. en 1997 [WUL97]. Cette méthode recherche tout d'abord l'alignement optimal entre la signature testée et la signature de référence puis calcule la distance entre les signatures. L'algorithme présente des particularités propres aux signatures chinoises. Ainsi, chaque signature est décomposée en "mots", c'est à dire un ensemble de traits en contact. Chaque mot de la signature représente une séquence. La mesure de similarité utilisée entre deux séquences est obtenue par la formule (*). Soient V et W construits à partir de plusieurs exemples $V_0, V_1, \ldots, V_{s-1}$ et $W_0, W_1, \ldots, W_{s-1}$ respectivement avec $V_i = (v_i^0, v_i^1, ..., v_i^{T-1})$ et $W_i = (w_i^0, w_i^1, ..., w_i^{T-1})$. S est le nombre d'exemples et T la dimension de V_i et W_i avec v_i^j un attribut de l'exemple V_i. On procède par interpolation pour que V_i et W_i soient de même longueur.

$$Dist(V,W) = \sum_{i=0}^{S-1}\sum_{j=0}^{T-1}\left|(v_i^j - \overline{v_j}) - (w_i^j - \overline{w_j})\right| \qquad (*)$$

Pour effectuer la mise en correspondance de deux séquences, chaque séquence est découpée en deux sous-séquences et chaque sous-séquence est mise en correspondance avec la sous-séquence correspondante. Cet algorithme est effectué récursivement jusqu'à atteindre une profondeur préfixée ou jusqu'à ce que la distance entre les sous-séquences soit inférieure à un seuil prédéfini. Les sous-séquences correspondant à la signature testée sont modifiées par compression et/ou étirement de telle sorte que la fusion des sous-séquences modifiées donne un trait de même longueur que la séquence issue de la signature de référence. La distance totale entre les signatures est alors obtenue en calculant la somme des distances entre mots.

- **Distance entre les distributions**

Au lieu de calculer une distance à partir des points appariés, il est possible de calculer un indice de corrélation entre les coordonnées des points [PAL04]. Il est au préalable nécessaire de normaliser les coordonnées pour obtenir la correspondance des centres de gravité. On obtient alors des résultats du type de ceux présentés figures 12 et 13 lors du calcul des corrélations.

Graphes associés aux quantiles (50 points)
En abscisse, la signature testée et en ordonnée une signature d'apprentissage

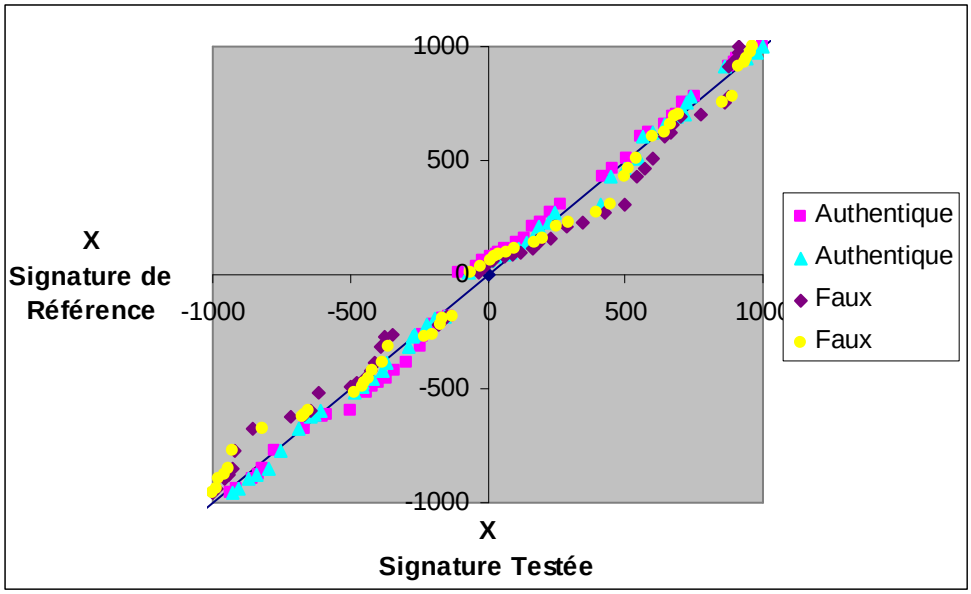


Figure 12. Quantiles suivant X pour 4 signatures relativement à une signature fixée.

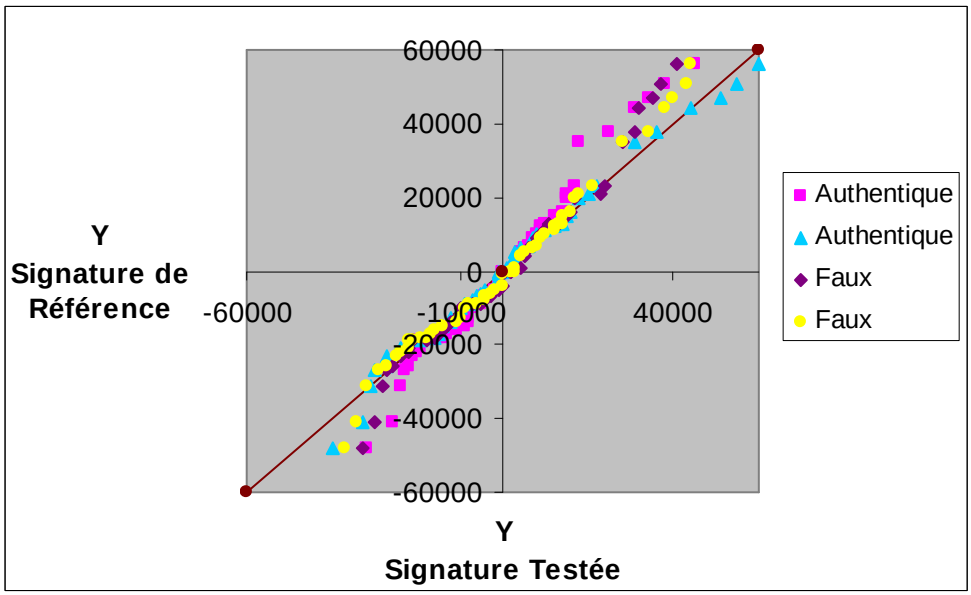


Figure 13. Quantiles suivant Y pour 4 signatures relativement à une signature fixée.

Sur les Figure 12 et 13, les faux expérimentés conduisent aux valeurs en jaune et violet foncé et les deux autres sont des signatures authentiques. Comme on peut le remarquer sur cet exemple, rien ne semble distinguer les faux des signatures authentiques; le faux jaune est même au contraire plus près de la signature d'apprentissage que certaines signatures authentiques.

- **Mesure de similarité ER2**

Après avoir effectué la mise en correspondance de la signature testée *St* et d'une signature d'apprentissage *Sa* à l'aide de l'algorithme DTW, on obtient deux ensembles de points de même cardinal *n* de la forme St={(x_{11},y_{11},t_{11}), …, (x_{1i},y_{1i},t_{1i}), …, (x_{1n},y_{1n},t_{1n})} et Sa={(x_{21},y_{21},t_{21}), …, (x_{2i},y_{2i},t_{2i}), …, (x_{2n},y_{2n},t_{2n})} où le point (x_{1i},y_{1i},t_{1i}) est mis en correspondance avec (x_{2i},y_{2i},t_{2i}).

La mesure de similarité spatiale est donnée par la formule [LEI04] :

$$SimS_{ER} = \frac{\left[\sum_{i=1}^{n}\left(x_{1i}-\overline{x_1}\right)\left(x_{2i}-\overline{x_2}\right)+\sum_{i=1}^{n}\left(y_{1i}-\overline{y_1}\right)\left(y_{2i}-\overline{y_2}\right)\right]^2}{\sum_{i=1}^{n}\left(x_{1i}-\overline{x_1}\right)^2\sum_{i=1}^{n}\left(y_{1i}-\overline{y_1}\right)^2+\sum_{i=1}^{n}\left(x_{2i}-\overline{x_2}\right)^2\sum_{i=1}^{n}\left(y_{2i}-\overline{y_2}\right)^2}$$

La distance spatiale devient :

$$DistS_{ER} = 1000 * (1 - SimS_{ER})$$

La mesure de similarité temporelle est donnée par la formule :

$$SimT_{ER} = \frac{\left[\sum_{i=1}^{n}\left(t_{1i}-\overline{t_1}\right)\left(t_{2i}-\overline{t_2}\right)\right]^2}{\sum_{i=1}^{n}\left(t_{1i}-\overline{t_1}\right)^2\sum_{i=1}^{n}\left(t_{2i}-\overline{t_2}\right)^2}$$

La distance temporelle devient :

$$DistT_{ER} = 1000 * (1 - SimT_{ER})$$

L'avantage de cette approche est que l'ordre de grandeur des deux distances est le même et donc que la fusion des deux distances est plus simple.

La distance spatio-temporelle devient donc par exemple :

$$DistST_{ER} = \frac{DistS_{ER} + DistT_{ER}}{2}$$

Après avoir effectué la mise en correspondance des deux signatures S_1 et S_2 et calculé la distance entre les deux courbes, on obtient une mesure de similarité par la formule :

$$Sim(S_1, S_2) = \exp(\frac{-dist(S_1, S_2)}{2 * M_d})$$

où $M_d = \max_{\substack{i \neq j \\ i,j=1 à n}} (Dist(Sa_i, Sa_j))$, *n* désigne le nombre de signatures d'apprentissage et *Sa* les signatures d'apprentissage.

3.4.3. Seuil de décision et évaluation des performances

- **Adaptation des seuils**

Comme nous l'avons vu auparavant, les performances d'un système d'authentification de signature sont généralement évaluées suivant deux critères : le pourcentage de faux acceptés (FAR), le pourcentage de vrais rejetés (FRR). Il est à noter que ces deux variables FAR et FRR sont très corrélées et si l'une d'elles augmente l'autre diminue.

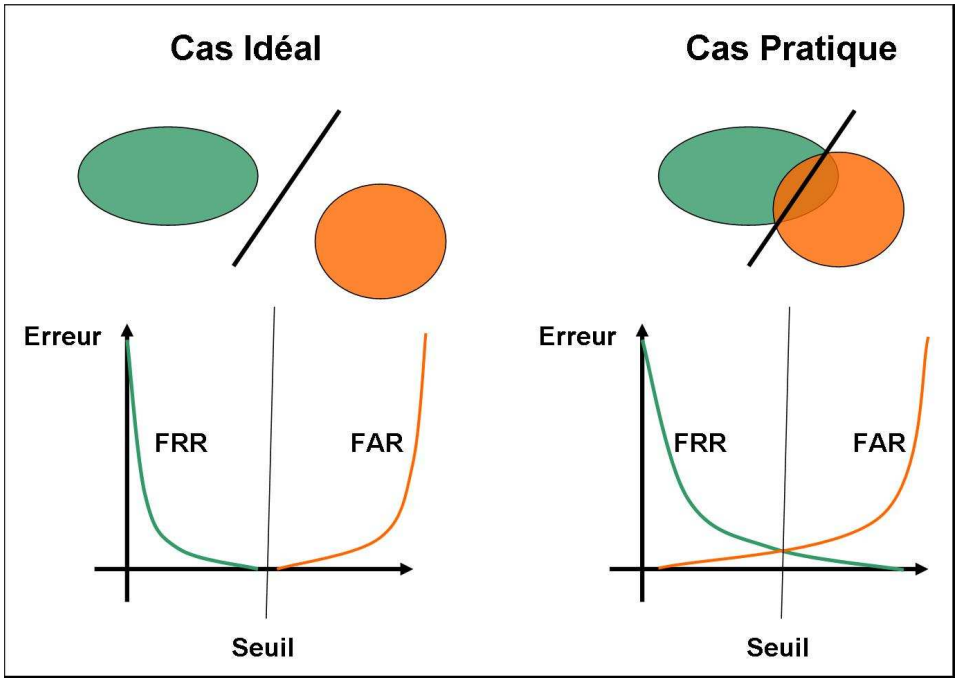


Figure 14. Illustration de la difficulté de la décision.

La corrélation entre les valeurs de FAR et de FRR est illustrée par la Figure 14. Elle est principalement due à la difficulté d'isoler la classe de signatures d'un individu par rapport aux autres signatures.

Les buts de l'authentification peuvent être différents suivant le type d'utilisation. Pour les systèmes de gestion des accès, le pourcentage de faux acceptés est plus élevé. Pour les systèmes sécurisés, le pourcentage de vrais rejetés doit être plus élevé.

Pour ce qui est du choix d'un seuil individuel, il dépend de deux facteurs : la variabilité de la signature et le taux de FRR ou FAR souhaité.

L'idéal serait de disposer de faux aléatoires et de faux expérimentés [DOL98] mais cela n'est pas réaliste surtout si l'on envisage la création d'un produit industriel.

La variabilité d'une signature a une très grande influence sur le choix du seuil. La difficulté réside dans son estimation. Cependant, une indication peut être donnée par la variation du temps mis pour signer, cette variation semble corrélée à la variabilité intra signataire [TAN01]. Une autre étude [KHO03] a abordé le problème de l'adaptation du seuil au scripteur. La solution préconisée pour que la distance tienne compte de la variabilité intra scripteur consiste à tronquer la distance en soustrayant une valeur comme par exemple la distance moyenne intra apprentissage. Dans ce cas aussi, un biais est introduit en utilisant des faux pour déterminer le seuil optimal pour chaque individu. Dans [JAI02], plusieurs approches ont été envisagées pour déterminer un seuil propre à chaque caractéristique et à chaque scripteur. Ainsi, pour chaque caractéristique utilisée pour décrire la signature, le seuil individuel est calculé en ajoutant une tolérance dépendant de la distance minimale, de la distance moyenne ou encore de la distance maximale entre la signature testée et l'ensemble des signatures d'apprentissage par rapport à cette caractéristique. Les auteurs montrent que le meilleur résultat est obtenu en considérant la distance minimale. Dans [WUL97], le seuil propre à chaque caractéristique est déterminé en se basant sur la moyenne et l'écart type calculés sur l'ensemble d'apprentissage.

- **Evaluation des performances**

Il est très difficile de comparer les performances de deux systèmes de vérification de signatures étant donné qu'aucune base de données internationale n'est disponible. De plus, le protocole expérimental diffère selon les laboratoires. En effet, les signatures ne sont pas acquises dans les mêmes conditions, les bases d’apprentissage et de test ne sont pas toutes de la même taille, le nombre de références par individu n’est pas fixe, les faux ne sont pas de "vrais" faux car ils ne sont pas réalisés par des faussaires professionnels. Par conséquent, le pourcentage de faux acceptés est très difficile à évaluer.

La compétition d'authentification par signature manuscrite ("Signature Verification Competition") organisée dans le cadre de la conférence intitulée "International Conference on Biometric Authentication" (ICBA) a permis de confronter différentes méthodes dans les mêmes conditions expérimentales. Parmi les différentes méthodes proposées, les meilleurs résultats (EER=2.79% pour des faux aléatoires et EER=2.84% pour des faux expérimentés) ont été obtenus par des universitaires turques de l'université de Sabanci [YEU04]. Ces résultats donnent un aperçu des performances à atteindre voire à dépasser. On remarque également que la détection des faux aléatoires qui, a priori, peut sembler triviale ne l'est pas.

En effet, le taux de faux aléatoires acceptés est relativement élevé et du même ordre de grandeur que celui pour les faux expérimentés. Une explication possible est que ce taux est fortement dépendant de la variabilité intra scripteur.
Notons également que les méthodes développées jusqu'alors ne sont, ni toujours réalistes, ni industrialisables. En effet, certaines méthodes supposent que le nombre de signatures d'apprentissage est supérieur à 5, elles tiennent rarement compte de la taille de la référence stockée ou encore du temps nécessaire pour l'authentification [HEN04]. Or, dès que l'on rajoute des contraintes sur les systèmes d'authentification, les performances chutent [MAI04]. Notre travail s'inscrit dans une démarche de signature de documents électroniques où la signature en ligne sera utilisée comme facteur d'authentification. Aussi, pour nous, ces dernières contraintes sont-elles importantes.

Nous ne pouvons dans cet état de l'art faire une description exhaustive de toutes les méthodes et systèmes d'authentification par signatures manuscrites en ligne. Néanmoins, nous avons synthétisé dans un tableau les principales caractéristiques des systèmes majeurs produits ces dernières années et ayant donné lieu à des publications depuis 1994.
Le Tableau 4 récapitule également les taux d'erreurs associés aux différents systèmes lorsque ceux-ci étaient précisés.

Auteurs	Signal en entrée et description des paramètres	Apprentissage et/ou base de test specimens(S)xwriters(W)	Méthode de comparaison	Taux d'erreurs	Commentaires
Dimauro, Impedovo, Pirlo [DIM94]	position et levés de stylo	apprentissage : 1000 signatures authentiques (50Sx20W) test : 800 signatures authentiques (40Sx20W) faux : 800 ((4x20S)x10W)	Comparaison des éléments séparés par les levés et baissés de stylo	Type I : <1,7% Type II : <12%	
Dimauro, Impedovo, Pirlo, Sarcinella [DIM02]	x, y, vx, vy	apprentissage : 45 signatures authentiques (3Sx15W) test : 45 signatures authentiques (3Sx15W) faux : 750 signatures authentiques (50Sx15W)	Vitesse pondérée par le taux de stabilité de la région	Type I : 3,2% Type II : 2,6%	Une zone très stable est plus difficile à imiter qu'une zone peu stable
Dolfing [DOL98]	x, y, pression, inclinaison du stylo	apprentissage : 1530 signatures authentiques (15Sx51W) validation : 255 signatures authentiques (5Sx51W) test : 510 signatures authentiques (10Sx51W) faux : 3000 amateurs (51W) 240 professionnels	HMM (1modèle par personne)	Type I : 1,9% Type II : 1,9%	Minimisation de EER
Fuentes, Garcia-Salicetti, Dorizzi [FUE02]	x, y, pression	apprentissage : 765 signatures authentiques (15Sx51W) test : 765 signatures authentiques (15Sx51W) faux : 1470(30S/individu-2)+1530(30S/individu)+ 200(10S pour 20 individus)	HMM (1modèle par personne) + MLP (1modèle par personne) + SVM (Support Vector Machine)	Type I : 4,62% Type II : 8,25%	HMM traduit mal la variabilité d'un scripteur MLP isole bien un scripteur
Gupta, Joyce [GUP97a]	x, y, t	apprentissage : 295 signatures authentiques (5Sx59W) test : 609 signatures authentiques faux : 325 (expérimentés)	Calcul du nombre d'altérations entre 2 séquences(ou mots) (Algorithme de Wagner et Fisher 1974)	Type I : 14% Type II : 15%	
Huang, Yan [HUA95]	x, y, t	apprentissage : 180 signatures authentiques (9Sx20W dont 1 gaucher) test :sign authentiques (6-11)Sx20W faux : (10-20)Sx20W	Multiple regional matching	Type I : 0,05% Type II : 0,02%	Combine le mouvement du stylet levé et baissé
Jain, Griess and Connel [JAI02]	local : ecart et angle entre 2 points consécutifs, vitesse relative entre 2 points global : nombre de traits	apprentissage : 1232 signatures authentiques (10-42Sx102W) faux : 60 (3Sx20W) imitateurs non professionnels	DTW	Type I : 3,3% Type II : 2,7% ---------- Type I : 2,8% Type II : 1,6%	Seuil commun pour la classification ---------- Seuil adapté à chaque scripteur pour la classification

Kashi, Hu, Nelson, Turin [KAS97]	23 paramètres globaux dont 2 relatifs au temps : durée totale de l'écriture de la signature durée pendant laquelle le stylo est en contact avec le support	542 signatures authentiques (6Sx59W) faux : 325	HMM	Type I : 2,5% Type II : 2,5%	
Kholmatov, Yanikoglu [KHO03]	x(t), y(t)	apprentissage : 752 signatures authentiques (8Sx94W) test : 182 signatures (94W) faux : 313 expérimentés	DTW	Type I : 1,4% Type II : 1,4%	Apprentissage du seuil pour différencier les signatures authentiques des faux en utilisant des faux
Lam, Kamins [LAM89]	x, y	signatures authentiques : 8=8Sx1W faux : 152=8Sx19W (expérimentés)	Analyse discriminante Jackknife	Type I : 0% Type II : 2,5%	
Lee, Berger, Aviczer [LEE96]	x, y, temps total, vitesse moy, vitesse/x positive moyenne, vitesse/y positive moyenne, durée pdt laquelle vitesse/x<0 / temps pdt lequel le stylo est appuyé,...	apprentissage : 5063 signatures authentiques = (13-1000)Sx105W + 248 Chinese(23W) + 26 Arabic (2W) + 13 Tamil (1W) + 13 Hebrew (1W) faux : 1148 simples 3466 expérimentés 1148 timed	Prehard majority classifier	Type I : 2,5% Type II : 2,5% ---------- Type I : 0% Type II : 7%	Minimisation de EER ---------- Type I mis à 0
Lejtman, George [LEJ01]	Pression, vx, vy, angle du stylo, vitesse angulaire	apprentissage : 9-11Sx41W chinese + 9-11Sx41W latin faux : 8-11Sx41W chinese + 8-11Sx41W latin	Réseau de neurones	Type I : 0% Type II : 0,1%	
Lin, Chen [LIN98]	Coordonnées des points et vitesses	apprentissage : 225 signatures authentiques (5Sx45W) test : 900 signatures authentiques (20Sx45W) faux : 20Sx45W (20 faux pour chaque personne)	Comparaison des points importants par une méthode de relaxation	Type I : 6,4% Type II : 16%	Signatures chinoises
Martens and Claesen [MAR97a]	forces exercées sur le stylo suivant 3 axes et angle d'inclinaison du stylo	apprentissage : 180 signatures authentiques(10Sx18W) test : 180 signatures authentiques(10Sx18W) faux : 14555 = 20Sx17W (signatures authentiques des autres personnes) + 15Sx41W	Kernel approach PDF(Probability Density Function)	Type I : 0,4% Type II : 0,4%	Minimisation de EER
Martens and Claesen [MAR97b]	xform, xmotion,forces exercées sur le stylo suivant 3 axes et angle d'inclinaison du stylo	apprentissage : 270 signatures authentiques(15Sx18W) test : 90 signatures authentiques(5Sx18W) faux : 340 = 20Sx17W (signatures authentiques des autres personnes)	DTW	Type I : 7-8% Type II : 7-8%	Minimisation de EER

Martens and Claesen [MAR98]	forces exercées sur le stylo suivant 3 axes et angle d'inclinaison du stylo	apprentissage : 180 signatures authentiques(10Sx18W) test : 180 signatures authentiques(10Sx18W) faux : 340 = 20Sx17W (signatures authentiques des autres personnes)+39W signatures aléatoires	DTW(forme, vitesse, consistence)	Type I : 0,6% Type II : 0,6%	Minimisation de EER
Mohankrishnan, Lee, Paulik [MOH99]	x, y, temps total de la signature, DTW distorsion (DTW signature décomposée en 8 segments)	apprentissage : 1600 signatures authentiques (100Sx16W)+faux expérimentés test : 800 sign authentiques (50Sx16W) +faux expérimentés faux : casual 1600=100Sx16W + expérimentés 1920=(8x30)Sx8W	Reseau de neurones (1 / scripteur)	Type I : 0,78% Type II : 1,6%	
Nalwa [NAL97]	x, y	1452 signatures authentiques (204W) dont 6 signatures pour l'apprentissage pour chaque scripteur 1150 faux	Cross relation (5 paramètres coordonnées du centre de masse, the torque, mesure de la courbure de l'ellipse…,)	Type I : 3,6% Type II : 3,6%	Minimisation de EER
Parker [PAR02]	x(t), y(t)	690 signatures authentiques (50Sx14W) leave one-out	Distance entre pixels noirs de la signature testée et un masque constitué des signatures de la base d'apprentissage	Type I : 0% Type II : 0%	Pas de faux
Rhee, Cho, Kim [RHE01]	x, y, vx, vy, angle entre segments, durée totale	apprentissage : 5250 signatures authentiques (105Sx50W) test : 500 signatures authentiques (10Sx50W) faux : 20Sx50W expérimentés 50Sx49W aléatoires	Dynamic matching(faux de type aléatoire) Distance euclidienne entre certaines caractéristiques des segments (faux de type expérimentés)	Type I : 3,4% Type II : 3,4% ---------- Type I : 0,49% Type II : 0,49%	Faux de type aléatoire ---------- Faux de type expérimentés
Rokbani, Alimi [ROK03]	x(t), y(t)	apprentissage : 525 signatures authentiques (20Sx15W) test : 450 signatures authentiques (30Sx15W) faux : 56S par personne	Distance entre deux graphes (matching par 2-ppv) + Comparaison de caractéristiques	Type I : 7,4% Type II : 6,5%	Prise en compte de l'évolution de la signature
Schmidt, Kraiss [SCH97]	x, y, pression	apprentissage : 81 signatures authentiques (3Sx27W) test :sign authentiques (8-17)Sx27W faux : 48 expérimentés	Multiple-to-one matching algorithm	Type I : 0% Type II : 12%	
Tanabe, Yoshihara, Kameya, Mori [TAN01]	Pression	signatures authentiques : 160=(20-30)Sx6W faux : 600=100Sx6W	DP Matching	Type I : 6% Type II : 6%	
Wessels and Omlin [WES00]	x(t), y(t) pression et inclinaison du stylo	apprentissage : 750 signatures authentiques (15Sx50W) test : 750 signatures authentiques (15Sx50W) faux : 50-80	HMM	Type I : 0% Type II : 13-15%	

Wirtz [WIR97]	position, vitesse, acceleration	6000 signatures authentiques (20W) 6000 faux (20W)	Dynamic Programming Matching + Stroke based DPM	Type I : 9,89% Type II : 9,89%	Minimisation de EER
Wu, Lee, Jou [WUL97]	positions(x,y) et vitesses en chaque point / à x et à y	apprentissage : 100 signatures authentiques (10Sx10W) test : 200 signatures authentiques (20Sx10W) faux : 4 imitateurs	Comparaison des suites de valeurs (position et vitesse) par "split-and-merge matching algorithm"	Type I : 13,5% Type II : 2,8%	Signatures chinoises (3 signatures pour créer une référence et 7 pour définir le seuil)
Wu, Lee, Jou [WUL98]	ux(t), uy(t) vitesse et accélération	apprentissage : 270 signatures authentiques (10Sx27W) test : 560 signatures authentiques (30Sx27W) faux : 650 4 imitateurs	Spectre logarithmique	Type I : 1,4% Type II : 2,8%	Signatures chinoises
Yang, Widjaja, Prasad [YAN95]	angles entre l'axe des x et le segment constitué de 2 points consécutifs	apprentissage : 248 signatures authentiques (8Sx31W) test : 248 signatures authentiques (8Sx31W) faux : autres signatures (16Sx30W pour chaque personne)	HMM (1modèle par personne)	Type I : 1,75% Type II : 4,44%	Stabilité des données par rapport à la rotation, la translation et le changement d'échelle

Tableau 4. Récapitulatif des différentes méthodes d'authentification par signature manuscrite en ligne.

Comme on peut le constater sur le Tableau 4, les protocoles d'expérimentation ne sont pas identiques. Ainsi le nombre de signatures d'apprentissage varie de 3 à 100, le nombre de scripteurs de 15 à 100, les signatures peuvent être asiatiques, européennes ou anglo-saxonnes les données en entrée peuvent être uniquement les positions x,y et le temps associé ou uniquement la pression, le critère optimisé peut être FAR, FRR ou EER… Il est donc très difficile de comparer les performances des différents systèmes et, par conséquent, de déterminer quelle est la meilleure approche. On distingue deux grandes classes de méthodes : celles basées sur un calcul de distance et celles basées sur des méthodes d'apprentissage telles HMM ou réseaux de neurones. Cette deuxième approche semble donner de bons résultats (EER inférieur à 1%) mais sur des bases de signatures de taille relativement faible (nombre de scripteurs inférieur à 40). Les performances de la méthode présentée dans [KHO03] - basée sur un calcul de distance, utilisant relativement peu de signatures d'apprentissage (8) et testée sur une base conséquente de signataires (proche de 100) - nous donnent un aperçu des performances à atteindre, i.e. une valeur de EER de l'ordre de 1%.
Aux systèmes présentés ci-dessus développés dans un cadre universitaire, il faut ajouter les systèmes industriels.

4. PRODUITS INDUSTRIELS

4.1. Présentation des logiciels

Il est difficile, voire impossible, de connaître le principe de fonctionnement des logiciels d'authentification développés par les industriels. Etant donnée l'importance de ce marché, les entreprises ont tendance à protéger leurs inventions en gardant leurs recherches confidentielles.
Voici néanmoins une liste non exhaustive des sociétés ayant développé des logiciels d'authentification en ligne par la signature:
Pour PDA :

- **CIC** (Communication Intelligence Corporation)
- **Cyber SIGN Incorporated**
- **MMI Group**
- **Romsey Associates Ltd.**
- **Valyd (ex-Adaptec)**

Pour PC :

- **Checkmate Electronics (NASDAQ - CMEL)**
- **CIC** (Communication Intelligence Corporation)
- **Cyber SIGN Incorporated**
- **Gateway File System Inc. Imaging (e-clips)**
- **MMI Group**
- **SoftPro (App Informatif Davos)**
- **Valyd (ex-Adaptec)**
- **Wondernet**

Nous nous sommes principalement intéressé aux offres disponibles sur PDA. Une des raisons de ce choix est que, généralement, les solutions pour PC utilisent des caractéristiques non disponibles sur PDA - du moins pour l'instant - telles que pression, inclinaison du stylet… et qu'à l'inverse les données en entrée disponibles sur un PDA (forme et dynamique) sont disponibles sur tous les supports munis d'une interface graphique.
De plus, les logiciels d'authentification pour PC nécessitent une puissance de calcul supérieure à celles disponibles sur les supports mobiles actuels. On peut souligner également que dans le cas de solution pour PC, les modèles sont généralement stockés sur un serveur.
Les données retenues pour décrire les différents produits sont :

- Les systèmes d'exploitation sur lesquels ils fonctionnent
- L'espace mémoire nécessaire pour stocker le logiciel
- La durée de l'identification
- Le type de stockage du modèle
- Le nombre de signatures pour l'apprentissage
- Les deux sociétés leaders sur le marché sont CIC et Cyber SIGN.

A priori, seules cinq entreprises proposent un logiciel permettant l'authentification en ligne par signature sur PDA. Cependant, étant donné la fragilité du marché dans le domaine de l'informatique, ces informations ne sont pas définitives : l'évolution est continue. En effet, de nombreuses sociétés qui proposaient des logiciels d'authentification ne sont plus présentes à l'heure actuelle sur le marché.

Société	CIC	Cyber SIGN Incorporated	Romsey Associates Ltd.	MmiGroup Corporation	Valyd
Produit	Sign-On	Log-on-Lock	PDALock	Sign Q	eSignLogon
Nombre de signatures pour l'apprentissage	3	3	3 (4KB)	4(20-500B)	
Stockage du modèle	local (codé avec l'algorithme triple DES)		local (crypté)		
Durée de l'identification (en ms)			1	1	
Système d'exploitation	Palm OS 3.3, Windows CE 3.0	Windows CE	Palm OS 3.3, iPack, Windows CE	Palm OS 3.3, Windows CE	Palm OS, Windows CE
Espace mémoire nécessaire	75KB (Palm OS) 200KB (Windows CE)		75KB	140KB	
Remarques		utilisation de la pression vente de kit de developpement pour BSA	Dérive de la signature prise en compte Utilise la technologie PenFlow	7 niveaux de sécurité	
Possibilité de télécharger une version demo	oui	oui	oui	oui	non

Tableau 5. Comparatif des différents logiciels d'authentification par signature manuscrite en ligne.

Comme le montre le **Tableau** 5, tous les logiciels fonctionnent sous Windows CE et Palm OS (à partir de la version 3.3).

L'espace mémoire nécessaire pour l'installation est compris entre 75KB et 200KB et il est en général plus important sur Palm OS que sur Windows CE.

Les temps annoncés pour l'identification sont les mêmes pour *PDALock* et pour *Sign Q*, ils sont proches de 1ms. Pour ce qui concerne les autres produits, aucun renseignement n'est donné sur la durée de l'identification. On peut remarquer qu'il faut faire un compromis entre le degré d'authentification et le temps de calcul maximum toléré.

Un aspect important des méthodes proposées est le choix des caractéristiques utilisées pour la création des modèles de signatures. On remarque que certains algorithmes – *Log-on-Lock* de Cyber SIGN et *PDALock* de Romsey Associates Ltd. - utilisent comme information la pression alors que ces dispositifs sont sensés fonctionner sur des PDA qui ne disposent pas de capteur de pression.

La principale différence entre les produits se situe au niveau du stockage des modèles. En effet, l'espace nécessaire varie de moins de 1KB à plus de 4KB. L'autre différence se situe au niveau du type de protection i.e. le type de codage utilisé. Le produit développé par CIC utilise l'algorithme triple DES pour chiffrer les données relatives au modèle. Les autres sociétés ne communiquent aucune information sur le type de chiffrement utilisé.

Le nombre de signatures nécessaires à l'apprentissage oscille entre 3 et 4 et certains logiciels comme *Sign On* permettent de réactualiser le modèle à tout instant.

Deux des produits industriels – *Sign Q* de MMI Group Corporation et *PDALock* de Romsey Associates Ltd.- présentent un lien avec le monde de la recherche. Par conséquent, on pourra accorder plus d'attention à ces produits puisque les méthodes utilisées doivent avoir un fondement scientifique.

On peut citer pour exemple le système d'authentification développé par Jin-Whan Kim, Professeur au Sungsim College of Foreign Languages, dont les domaines de recherches sont l'authentification de signature en ligne, la reconnaissance en ligne de caractères, le traitement de la voix et les systèmes utilisant la biométrie, pour la société MMI Group Corporation.

Les paramètres utilisés sont :

- La vitesse
- L'accélération
- La pression
- La forme (coordonnées, direction)
- Le nombre de levés et de baissés de stylet

- Le nombre de points de croisement
- Le temps total de la signature
- Le nombre de traits
- Le temps entre deux traits consécutifs
- Le nombre total de points

On peut paramétrer le niveau de sécurité. Il en existe 7, depuis basique jusqu'à haute sécurité. (Taux d'erreur annoncé : environ 0%)

4.2. Bilan

Aucune des solutions proposées n'a été développée en Europe et donc la question se pose sur l'efficacité de ces solutions sur des signatures européennes.

Il est très difficile de comparer ces produits entre eux car tous annoncent des taux d'identification proches de 100% en des temps très brefs, de l'ordre de la milliseconde.

Néanmoins, des logiciels sortent du lot en raison de l'ajout de certaines fonctionnalités. Ainsi les produits vendus par Romsey Associates Ltd. et par Cyber SIGN prennent en compte la dérive de la signature alors que ceux proposés par MMI Group et par Cyber SIGN permettent de paramétrer le niveau de sécurité.

Par contre, le problème juridique du litige lié à la répudiation n'est abordé par aucun des logiciels.

Ces systèmes fonctionnent comme des boîtes noires et il est très difficile voire impossible d'avoir des informations sur les méthodes mises en place. De plus, les performances annoncées n'offrent que peu d'informations puisque les bases de test ne sont pas décrites (nombre d'utilisateurs, …). Il n'est donc pas possible de donner un avis raisonnable sur ce genre de système.

Pour finir, notons que ces logiciels sont présentés comme des moyens de verrouiller l'accès à un Palm. Mais rien n'empêche un voleur de le formater…

5. CONCLUSION

5.1. Les principaux systèmes

L'ensemble des systèmes présentés dans ce chapitre suivent tous la même architecture globale : acquisition, prétraitement, caractérisation, comparaison et décision mais chacun a travaillé à l'amélioration d'une ou plusieurs étapes. La majeure partie des travaux porte sur les phases

d'acquisition et de comparaison. Ainsi plusieurs auteurs ont travaillé sur les données en entrée du système et notamment étudié l'intérêt de la pression ou de l'angle du stylet avec la surface d'écriture. Pour ce qui concerne la comparaison, on distingue deux grands axes d'études : les méthodes utilisant des méthodes d'apprentissage telles les réseaux de neurones ou les HMM et les méthodes basées sur le calcul de distances entre la signature testée et les signatures d'apprentissage sans changement d'espace de représentations telles que DTW.

Dans les méthodes classiques, les caractéristiques sélectionnées ont concerné autant la forme que la dynamique des signatures. Peu de caractéristiques originales ont été testées jusqu'à présent.

Il est également important de garder à l'esprit que la vérification de l'authenticité par signature ne permet pas de disposer d'information sur la variabilité des caractéristiques inter scripteurs mais que la seule information disponible est la distance intra scripteur calculée sur les signatures d'apprentissage.

Etant donné que les systèmes n'ont pas été testés sur la même base de signatures, il est très difficile de déterminer la pertinence d'une méthode par rapport à une autre mais il semble que de nombreux progrès restent à faire en authentification de signatures manuscrites surtout concernant les faux expérimentés.

5.2. Que reste-t-il à faire?

On constate que les différentes recherches effectuées dans le cadre de l'authentification par signature portent sur l'amélioration des différentes étapes mais pas sur l'architecture. Il semble donc opportun de mener de nouvelles expérimentations afin de juger de la pertinence des différents traitements effectués et de leur influence sur les performances. Une réflexion sur leur ordonnancement dans la chaîne des traitements pourrait également être intéressante.

La signature évolue au cours du temps comme l'écriture. Or cet aspect de la signature n'est encore que peu intégré dans les systèmes d'authentification proposés dans la littérature. Cependant ce constat entraîne de nombreuses contraintes sur le système d'authentification. En effet, cela impose de réactualiser la base des signatures servant de référence comme le souligne [JAI02] dans ses perspectives. Pour tenir compte de l'évolution de la signature, on pourrait affecter un poids plus important à la dernière signature reconnue comme authentique.

Outre la caractérisation des signatures, les méthodes de classification et de comparaison des signatures doivent elles aussi être améliorées afin d'être notamment adaptables et adaptées à chaque scripteur mais aussi beaucoup moins sensibles aux caractéristiques des dispositifs d'acquisition.

Une autre voie possible de recherche est de déterminer des caractéristiques stables dans le temps et de les utiliser en priorité par rapport aux caractéristiques jugées moins stables.

L'influence de la phase de prétraitement des données sur les résultats fournis par les différentes méthodes de comparaison a été très peu étudiée jusqu'à présent. Il nous paraît pourtant opportun de mener une telle étude. Les techniques de fusion des classificateurs pourraient également s'avérer un atout supplémentaire pour détecter les faux expérimentés en faisant par exemple coopérer des méthodes utilisant l'approche "caractéristiques" avec des méthodes de comparaison directe des tracés manuscrits. De même, d'autres combinaisons de classificateurs ou de mesures de similarité entre tracés pourraient être testées.

Nous avons exploré plusieurs de ces voies dans nos travaux de recherche. Les méthodes développées et résultats obtenus sont présentés dans les chapitres suivants.

CHAPITRE 2 – NOUVELLES METHODES DE CARACTERISATION

L'objectif de ce chapitre est de présenter de nouvelles méthodes de caractérisation de la signature basées sur une analyse multi résolution des tracés manuscrits. L'apport de ces caractéristiques est par la suite évalué à l'aide de plusieurs méthodes de sélection de caractéristiques.

Les deux premières parties de ce chapitre sont consacrées à l'acquisition et au prétraitement. La troisième partie présente les différentes alternatives pour la création et la gestion des modèles de signatures. La dernière partie décrit la phase de construction des différentes caractéristiques ainsi que les méthodes de sélection utilisées pour évaluer leur pertinence.

1. ACQUISITION

La première difficulté que nous avons rencontrée a concerné l'acquisition des coordonnées des points à intervalles de temps réguliers. En effet, le système d'exploitation Windows n'étant pas un système temps réel, l'acquisition se fait régulièrement uniquement si d'autres programmes n'utilisent pas les ressources systèmes à ce moment là! Nous avons envisagé deux grandes approches afin de résoudre ce problème :

- La première consistait à capturer la position et le temps uniquement lorsque le stylet se déplace sur le support. Mais l'acquisition du temps n'était alors pas suffisamment précise pour pouvoir ensuite faire des mesures de vitesse instantanée.
- La seconde consistait à capturer la position du stylet à intervalles de temps réguliers. Plusieurs timers ont été testés. Le choix qui a été fait consiste à utiliser un timer avec une fréquence élevée et à ne retenir les points que s'il y a déplacement.

La fréquence d'acquisition variant d'une tablette à l'autre notre premier travail a consisté à mettre en place un dispositif d'acquisition résolvant ce problème. Cette phase nous a également permis de vérifier que, pour avoir des données utilisables pour l'authentification, la fréquence d'acquisition des points doit être au moins supérieure à 100Hz.

Il est à noter que bien peu d'articles traitent de ces difficultés aussi bien en analyse de signatures qu'en reconnaissance de l'écriture manuscrite en ligne.

Comme nous utilisons les coordonnées des points de la signature, avant de comparer les signatures entre elles, il est nécessaire d'effectuer des prétraitements afin de normaliser les tracés des signatures. En effet, une mesure de distance n'est pas invariante dans le cas de transformations telles que rotation, translation ou homothétie, appliquées, avec différents paramètres, sur chacun des éléments à comparer.

2. PRETRAITEMENT

Nous limitons la première étape à une normalisation des données acquises puis nous abordons l'étude de la réduction du nombre de points décrivant les signatures acquises; phase que nous pouvons voir comme une phase de préparation des données avant l'application de nos méthodes de comparaison de signatures.

2.1. Normalisation

Dans notre cas, la normalisation est réalisée en trois étapes. Tout d'abord, on détermine la direction de l'axe principal d'inertie de la signature, c'est-à-dire la pente de la droite des moindres carrés du nuage de points formant la signature. Sur l'exemple de la Figure 15, l'axe d'inertie est en rouge.

Pour cela, nous calculons les variances suivant l'axe X, puis suivant l'axe Y. L'axe d'inertie ou la droite des moindres carrés a un coefficient directeur de $\frac{\beta}{\alpha}$.

Où $\alpha = \sqrt{\left(\frac{1}{2}+\frac{Var\_x-Var\_y}{2D}\right)}$ et $\beta = \text{Sign}(\text{Cov\_xy})\sqrt{\left(\frac{1}{2}-\frac{Var\_x-Var\_y}{2D}\right)}$

et $D = \sqrt{\left((Var\_x-Var\_y)^2+4Cov\_xy\right)}$

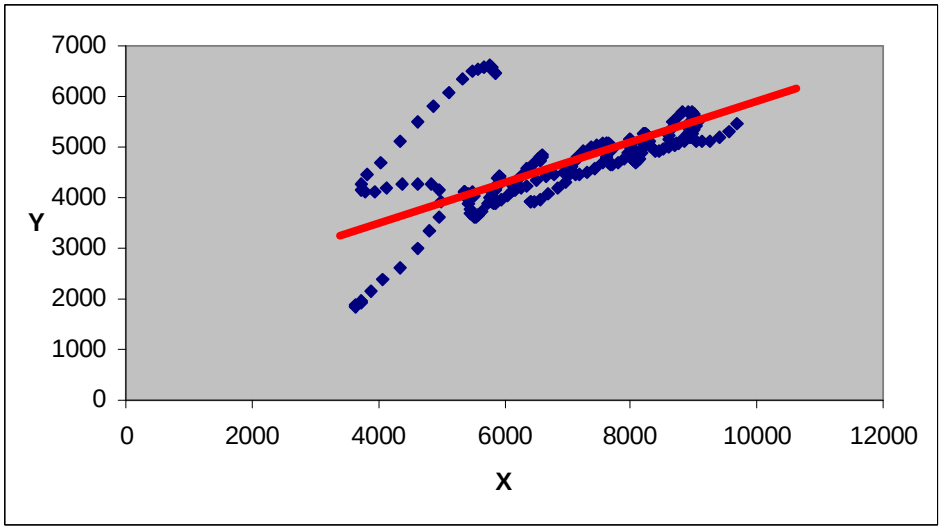


Figure 15. Axe d'inertie de la signature.

On effectue ensuite une rotation de façon à ce que l'axe d'inertie soit horizontal. La Figure 16 illustre l'effet de la rotation sur l'exemple.

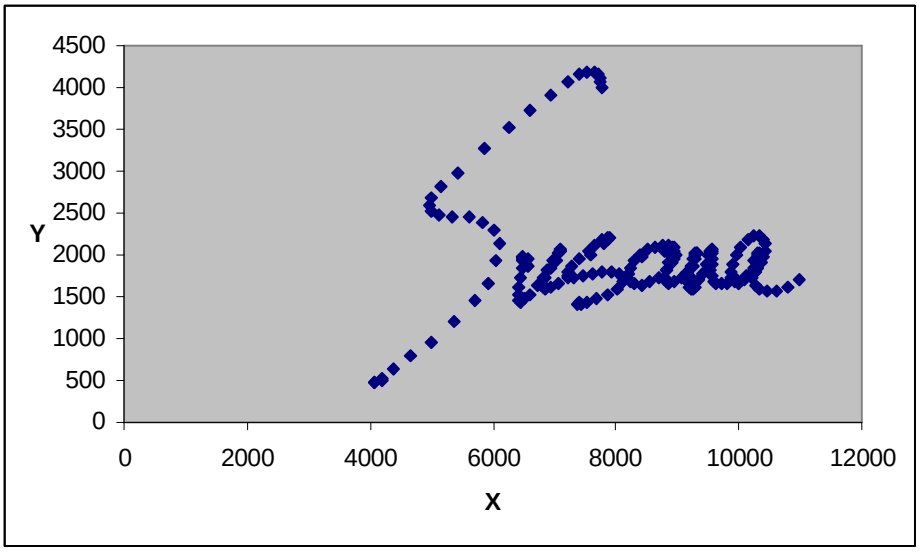


Figure 16. Signature redressée.

Puis, une homothétie est effectuée de manière à ce que toutes les signatures soient contenues dans un rectangle de même largeur, fixée à 300 pixels. L'homothétie consiste tout simplement à appliquer un coefficient multiplicateur ou réducteur, aux dimensions réelles, suivant les besoins. Le centre de l'homothétie est le centre de gravité de la signature. Le fait de ne pas imposer la hauteur du rectangle englobant permet de conserver les proportions de la signature. Une translation est ensuite effectuée sur la signature de manière à ce que le centre de gravité soit confondu avec l'origine du repère durant toute la suite du traitement (Figure 17).

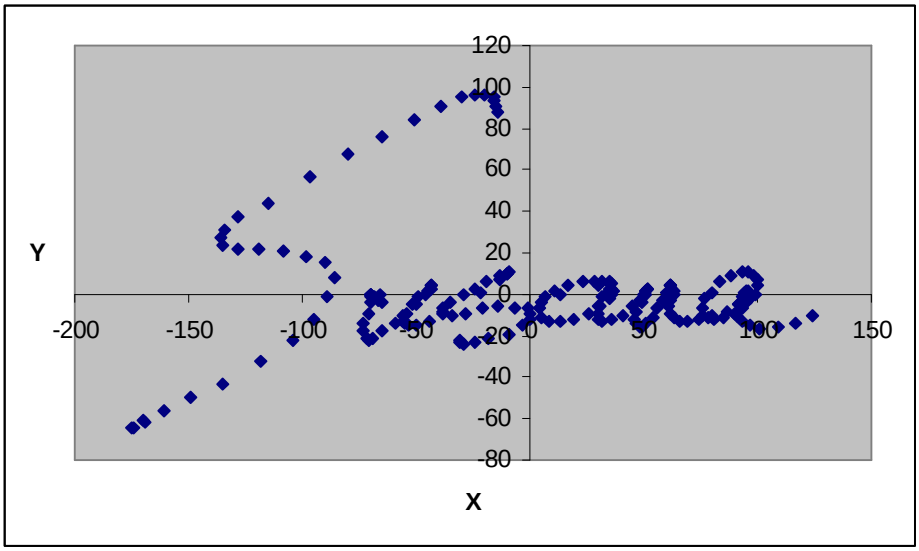


Figure 17. Translation pour positionner le centre de gravité à l'origine du repère.

Cette méthode d'alignement est celle qui donne les meilleurs résultats. En effet, les alignements prenant comme référence le point initial ou en ajustant le point le plus à gauche sur l'axe des ordonnées et le point le plus bas sur l'axe des abscisses engendrent une plus grande dispersion lorsque l'on superpose les signatures [IGA04]. La Figure 18 illustre l'influence du choix de la translation et l'impact sur la superposition des points des signatures. Ce choix est très important car il a une influence sur la comparaison à venir des points des signatures.

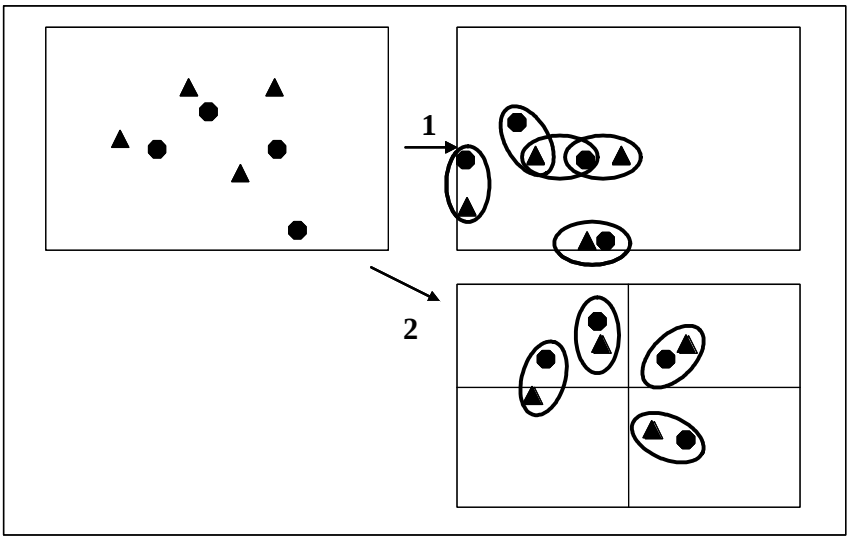


Figure 18. Influence de la translation.

(1 : Translations parallèlement aux axes, 2 : Translation par rapport au centre de gravité)

La Figure 19 montre alors la superposition des signatures d'un individu après normalisation, ces dernières devant être classées dans une même classe par le système d'authentification malgré leurs différences.

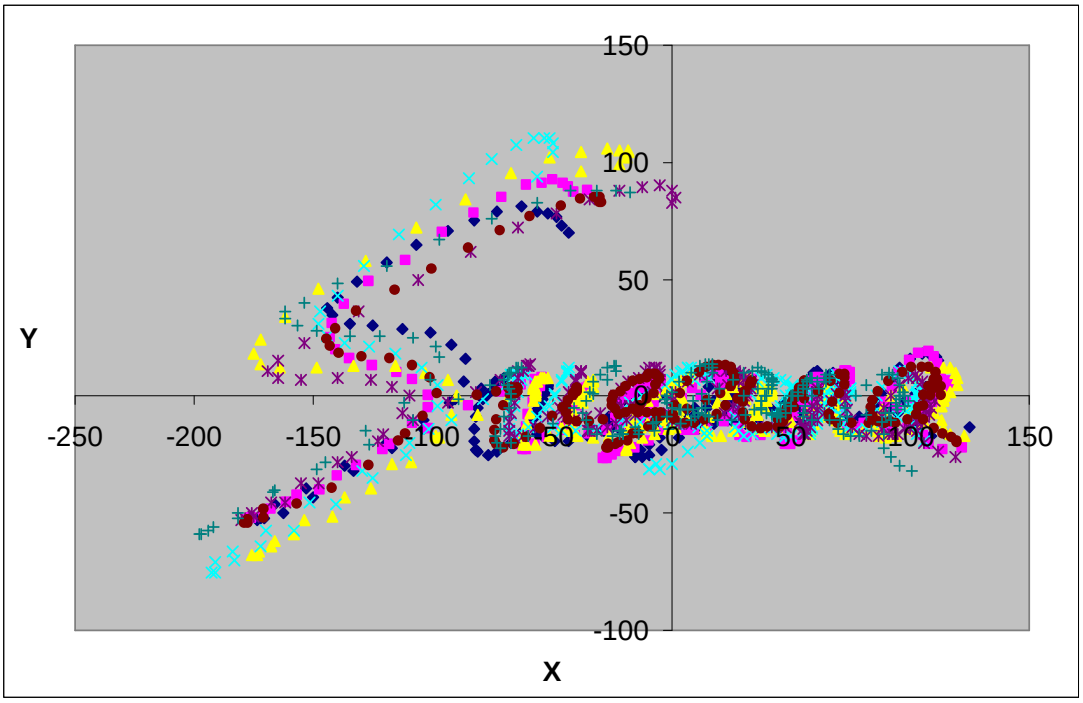


Figure 19. Superposition des signatures d'un individu.

2.2. Réduction du nombre de points de la signature

Peu d'articles traitent du problème de la sélection de points représentatifs de la signature dans le but de réduire la taille de données à stocker ou d'améliorer l'analyse durant la phase d'authentification. Habituellement, les points les plus intéressants sont supposés être ceux qui permettent d'effectuer une segmentation pour étudier ensuite les différentes parties indépendamment [LEE04]. Au contraire, pour nous, la réduction du nombre de points a plusieurs objectifs primordiaux :

- éviter de conserver l'intégralité des points de la signature afin de minimiser la quantité de données à stocker et d'augmenter le niveau de sécurité en conservant moins de données relatives aux signatures ayant servi à l'apprentissage et ainsi éviter les attaques de type rejeu,
- minimiser la durée du processus d'authentification car moins de données sont traitées
- améliorer les performances des méthodes de comparaison. En effet, les données brutes peuvent être considérées comme bruitées car trop précises : de légères variations peuvent perturber la phase de comparaison. Par conséquent, une réduction du nombre de points constitue un lissage qui peut permettre d'accroître la robustesse du système et donner au système une certaine invariance au matériel d'acquisition.

La conservation uniquement des points stables de la signature peut permettre d'augmenter l'efficacité de la méthode d'authentification [DIM02]. Nous insistons particulièrement sur

l'optimisation du choix des meilleurs points utilisés pour la comparaison [WIR04a], cela afin de rendre le processus d'authentification le plus rapide et le plus efficace possible.

La principale difficulté de la sélection de points vient de la difficulté de définir un critère de choix qui soit pertinent. Les trois critères que nous avons utilisés sont la conservation de la représentation spatiale de la signature, la stabilité spatiale et temporelle des points et l'importance ou représentativité des points [BRA93]. La réduction du nombre de points peut être comparée à une approximation polygonale. En effet, le but de l'approximation polygonale est de rechercher une forme globale qui coïncide au mieux avec la forme d'origine tout en limitant la perte d'information. L'approximation polygonale ne s'effectue habituellement pas sur des données en ligne. Néanmoins, moyennant quelques adaptations, nous considérons que l'approximation polygonale appliquée à un tracé en ligne devrait donner un bon compromis entre précision et réduction du volume de données à conserver.
Notre première idée a été de construire une méthode de réduction du nombre de points d'un tracé en se basant sur l'approximation polygonale proposée par Wall [WAL84]. Pour montrer la pertinence de cette méthode, des comparaisons avec d'autres méthodes de sélection de points ont été réalisées. La première consiste à choisir les points de manière aléatoire, la seconde à utiliser un algorithme génétique, la troisième est basée sur la méthode de Brault et la dernière repose sur l'analyse locale de la vitesse.

L'autre difficulté de la phase de sélection de points concerne l'évaluation de la pertinence des points sélectionnés. En effet, pour pouvoir comparer les méthodes de sélection entre elles, il faut pouvoir déterminer quelles sont celles qui permettent la meilleure authentification. La connaissance, uniquement, de la position des points sélectionnés sur la signature ne permet pas de comparer les méthodes entre elles. Par conséquent, l'évaluation ne pourra avoir lieu qu'au travers de l'utilisation d'un processus d'authentification complet et en comparant les pourcentages de FAR et de FRR. Les résultats présentés dans la suite de ce chapitre ont été obtenus sur la base SVC décrite précédemment.

2.2.1. Réduction aléatoire

De manière à ne pas fixer arbitrairement le nombre de points N à conserver, nous avons choisi de fixer ce nombre à un quart de la moyenne du nombre de points des signatures définies comme signatures d'apprentissage avec un maximum de 50 points. Etant donné que les signatures n'ont pas le même nombre de points, ce sont les indices relatifs des points qui sont considérés et non les coordonnées spatiales, les tracés étant paramétrés par le temps. Par

conséquent, nous choisissons N indices relatifs de points de la signature. Dans la Figure 20, nous donnons deux exemples d'ensembles de points retenus, ils sont représentés par de larges carrés.

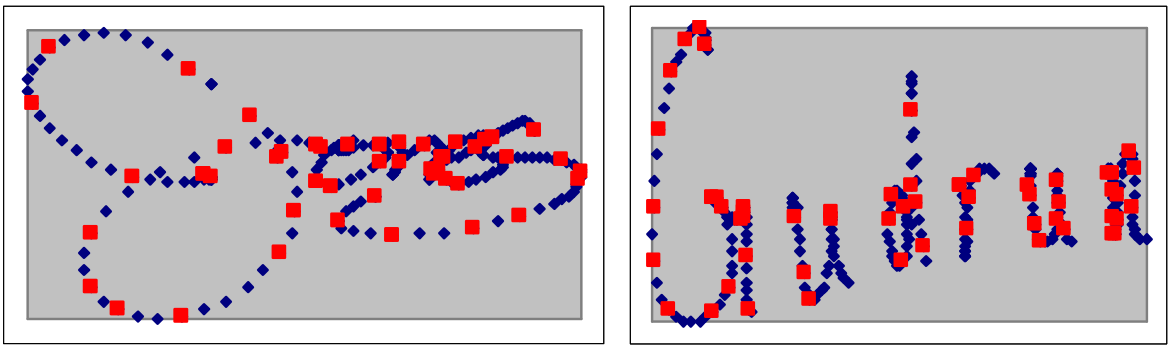


Figure 20. Illustration de la réduction aléatoire du nombre de points.

Evidemment, cette méthode de sélection ne correspond pas à une optimisation dans la résolution du problème, même si cela permet de réduire le temps de calcul de la phase d'authentification. Nous l'utilisons ici uniquement à titre de comparaison avec les autres méthodes de sélection non aléatoire des points représentatifs des signatures.

2.2.2. Réduction par algorithme génétique

Dans cette méthode, le critère utilisé pour effectuer la sélection des points à conserver est la stabilité des points. Pour évaluer cette stabilité, on calcule une distance entre les signatures de la base d'apprentissage. L'intérêt de cette méthode vient du fait qu'elle utilise l'ensemble des signatures d'apprentissage pour effectuer la réduction du nombre de points.

Rappelons qu'un algorithme génétique est un processus itératif de recherche de l'optimum pour une fonction appelée *fitness* [GOL89]. Il manipule une population formée de chromosomes. Chaque chromosome représente le codage d'une solution potentielle du problème à résoudre, et est constitué d'un ensemble d'éléments appelés gènes pouvant prendre plusieurs valeurs. A chaque itération, une nouvelle population de chromosomes est générée par modification des gènes. Au fur et à mesure de l'évolution, les chromosomes vont tendre vers l'optimum de la fonction objectif. La création d'une nouvelle population à partir de la précédente se fait par application des opérateurs génétiques que sont la sélection, le croisement et la mutation. La sélection permet de sélectionner les chromosomes qui optimisent le mieux la fonction. Le croisement permet de générer deux chromosomes nouveaux à partir de deux chromosomes sélectionnés tandis que la mutation modifie la valeur d'un ou plusieurs gènes du chromosome.

Dans notre cas, le but est de réduire le nombre de points de la signature en les sélectionnant et en conservant les points les plus adaptés au problème de l'authentification. Notre objectif plus particulier est de déterminer les points stables, en un certain sens, pour un signataire. C'est pour cela que nous utilisons l'ensemble des signatures d'apprentissage. La fonction fitness à minimiser est la moyenne des distances, au sens de l'algorithme DTW, entre les couples de signatures de l'ensemble d'apprentissage associé à un signataire. Comme précédemment, étant donné que les signatures n'ont pas le même nombre de points, nous utilisons les indices relatifs des points à la place des coordonnées comme gènes des chromosomes. Par conséquent, nous recherchons les indices des points qui minimisent la fonction objectif. Un chromosome représente le codage d'une solution partielle du problème. Dans le cas présent, un chromosome correspond aux indices de l'ensemble des points de la signature et un gène correspond à l'indice d'un point de la signature (Figure 21). Un gène prend la valeur 1 si le point est retenu comme stable et significatif et 0 dans le cas contraire.

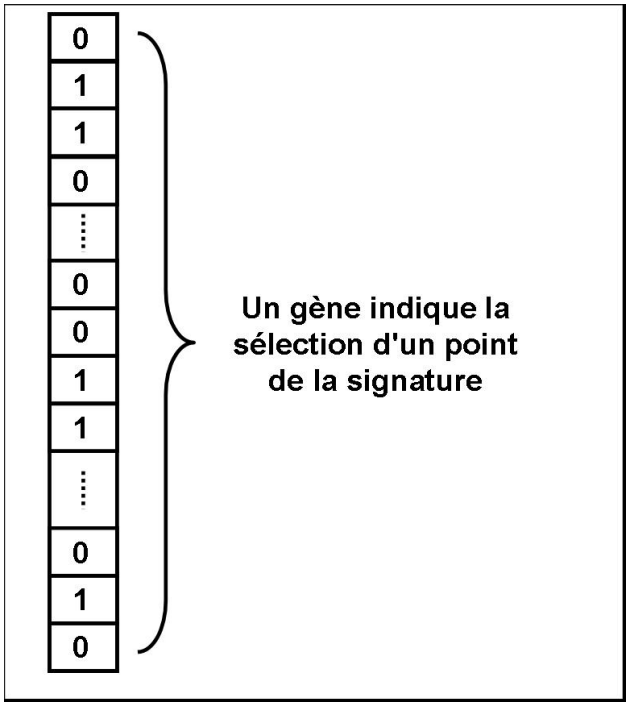


Figure 21. Illustration d'un chromosome associé à l'ensemble des points de la signature.

Le nombre de points utilisés dans la comparaison est trouvé automatiquement : c'est le nombre de 1 dans le chromosome. La taille de la population a été fixée à 100 chromosomes. Comme on peut le voir sur la Figure 22, la moyenne des distances au sens de DTW entre signatures d'apprentissage n'évolue quasiment plus beaucoup, bien avant 100 générations. Afin d'éviter le phénomène d'apprentissage "par cœur", nous avons décidé de générer seulement 20 générations.

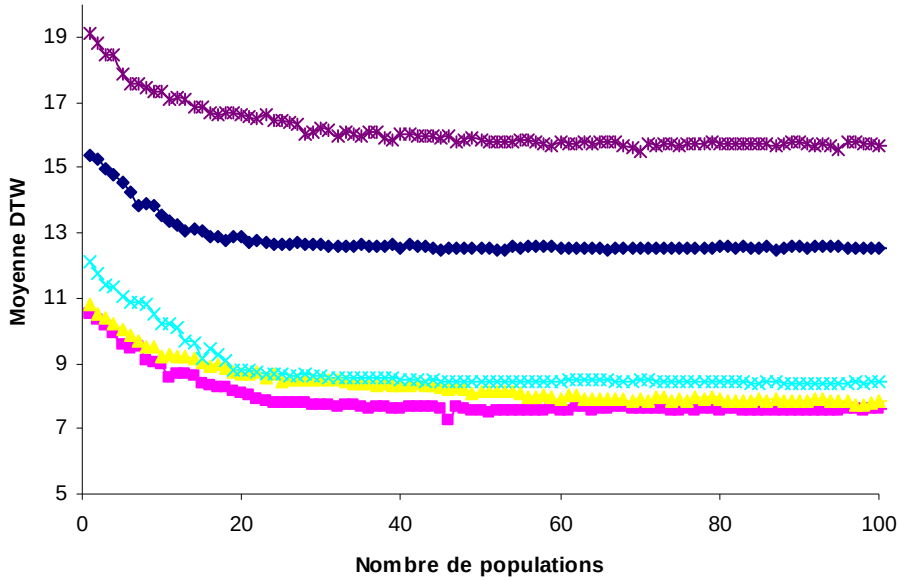


Figure 22. Evolution de la moyenne des distances intra signatures d'apprentissage pour cinq scripteurs.

Avec cette méthode, le choix des points n'est pas réalisé directement en fonction des caractéristiques de la signature mais de manière à minimiser la distance entre les signatures d'apprentissage.

On peut noter que les points retenus par l'algorithme génétique ne sont pas situés exclusivement aux points de forte courbure contrairement aux méthodes d'approximation polygonale. On constate également que le nombre de points retenus est proche de 50. Deux exemples sont présentés dans la Figure 23.

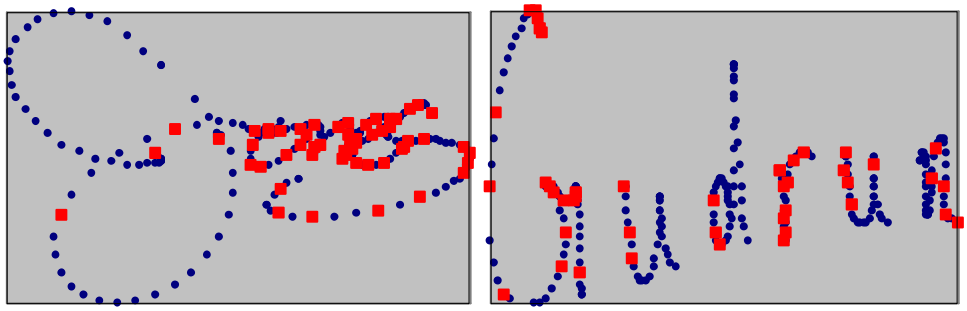


Figure 23. Illustration des points retenus par algorithme génétique.

2.2.3. Réduction par la méthode de Brault

Pour cette méthode, le critère de conservation d'un point est sa représentativité [BRA93]. Cette méthode est souvent utilisée pour segmenter les signatures [RHE01]. L'évaluation de la représentativité d'un point est basée sur la courbure locale du tracé de la signature.

Soit $S=(Pt_1, Pt_2, \ldots, Pt_N)$ une signature constituée de N points. La représentativité R d'un point d'indice i dépend de la variation d'angle entre le point sélectionné et les points voisins. Elle est illustrée sur la Figure 24.

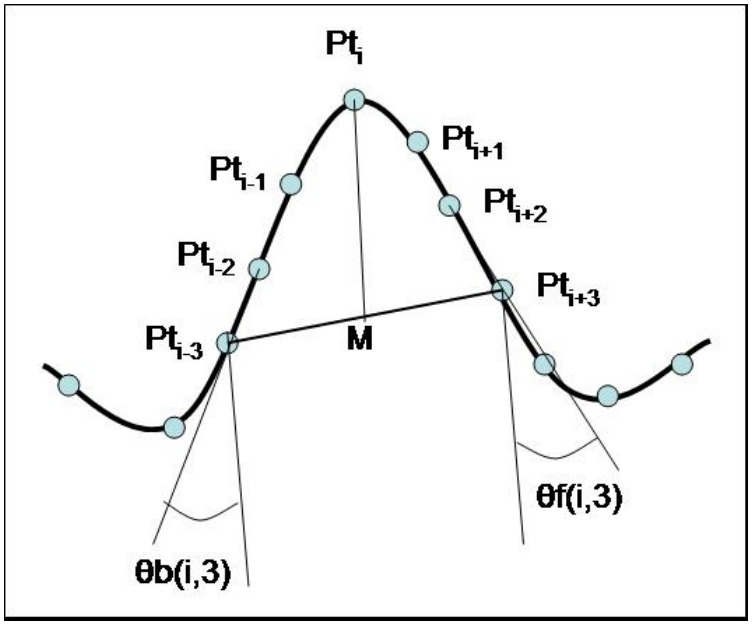


Figure 24. Illustration du calcul de la représentativité du point Pt_i avec $n=3$.

Nous précisons maintenant la méthode de calcul. Soient $\theta_b(i,n)$ l'angle entre la tangente au point Pt_{i-n} et la droite passant par le point Pt_i et le milieu du segment $[Pt_{i-n}, Pt_{i+n}]$ et $\theta_f(i,n)$ l'angle entre la tangente au point Pt_{i+n} et la droite passant par le point Pt_i et le milieu du segment $[Pt_{i-n}, Pt_{i+n}]$.

Pour pouvoir définir un point de forte courbure, on définit un angle θ_{max}. Cet angle est lié à la courbure maximum recherchée. Les points V_i pris en considération pour le calcul de la courbure du point Pt_i doivent vérifier la condition C :

$$\left|\theta_b(i,n)\right| \le \theta_{\max} \text{ et } \left|\theta_f(i,n)\right| \le \theta_{\max} \qquad \text{(C)}$$

Pour chaque point Pt_i, on calcule la contribution de chaque point se situant dans le voisinage V_i du point Pt_i à l'aide de la formule suivante :

$$IMP(i,n)=\cos(\theta_b(i,n))\times\cos(\theta_f(i,n))$$

Puis on calcule la représentativité du point d'indice i :

$$R(i) = \sum_{n=N_0(i)}^{N_d(i)} IMP(i,n)$$

avec $N_0(i)=K\times M(i)$, où $K>2$ et $M(i)$ le nombre de premières paires de points ne respectant pas la condition *(C)*. $N_d(i)$ correspond au nombre de points maximum tels que la condition *(C)* soit toujours respectée. Les points se situent là où la fonction *R(i)* est maximum.

Cette méthode utilise donc deux paramètres θmax et n. Pour le choix de ces paramètres, nous avons utilisé les valeurs préconisées dans [SCH97] i.e. $\frac{3\pi}{8}$ pour θmax et 3 puis 5 pour n afin de détecter les courbes serrées et larges. Les points retenus sont ceux qui correspondent à des maximums locaux de la représentativité R. Aux points retenus par cette méthode sont ajoutés les points correspondant aux extrémités des traits car ils sont aussi considérés comme des points représentatifs de la signature.

Pour l'exemple suivant θ_{max} est fixé à $\frac{3\pi}{8}$ et n est fixé à 3 puis à 5. Les points retenus correspondent à l'union des deux ensembles de points obtenus avec les deux paramètres.

La Figure 25 montre deux exemples des points retenus par cette méthode.

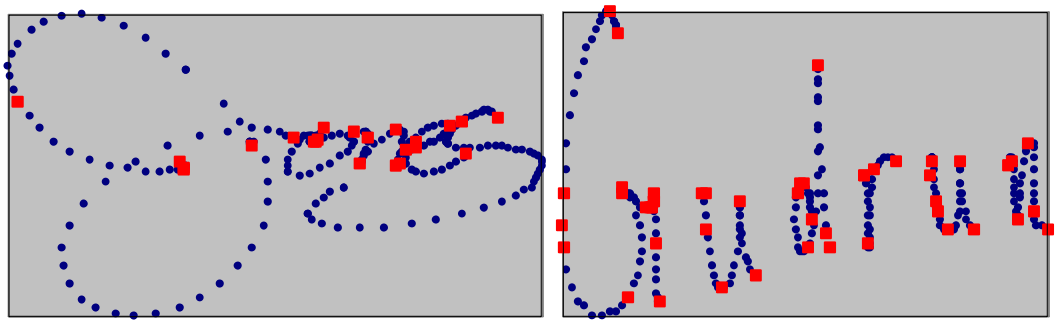


Figure 25. Illustration des points retenus par la méthode de Brault.

2.2.4. Sélection des points de vitesse minimum

Encore une fois, nous conservons dans cette méthode les points en fonction de leur représentativité. Dans ce cas, un point sera considéré comme représentatif de la signature s'il correspond à un minimum local de la vitesse. Le calcul de la vitesse instantanée v_i se fait selon la formule suivante :

$$v_i = \frac{dist_{Eucl}(Pt_i, Pt_{i+1})}{t_{i+1} - t_i}$$

Ce principe est souvent retenu pour effectuer la segmentation de signatures [HUA95]. Le seul paramètre intervenant dans cette méthode est le nombre de voisins pris en compte pour définir si un point est véritablement un minimum local. Comme précédemment et pour les mêmes raisons, nous avons ajouté les points situés aux extrémités des traits à l'ensemble des points retenus. La Figure 26 montre un exemple des points retenus.

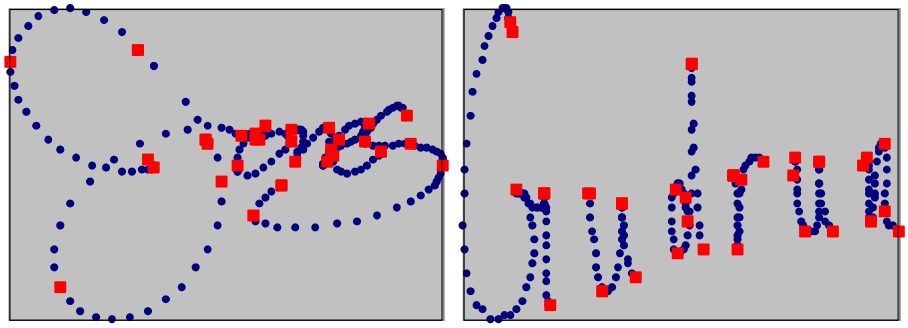


Figure 26. Illustration des points correspondant à des minimums locaux de la vitesse.

2.2.5. Réduction par approximation polygonale

Le critère utilisé pour réduire le nombre de points de la signature est la conservation de la représentation spatiale. L'idée est de voir la signature comme une suite de tracés rectilignes et de chercher à conserver suffisamment de points pour avoir une bonne approximation de ce tracé. Ici, les points sont considérés essentiels si la ligne polygonale qu'ils définissent approxime assez bien la forme de la signature. Pour déterminer ces points, nous avons choisi d'utiliser la méthode d'approximation polygonale proposée par Wall [WAL84]. Cette méthode d'approximation procède de proche en proche et introduit un nouveau sommet lorsque l'erreur commise en remplaçant un morceau du tracé par une ligne droite devient trop importante. Cette erreur est fonction d'un calcul d'aire. Un seuil est défini : il représente l'erreur cumulée c'est à dire l'aire entre la courbe et le segment l'approximant à chaque étape du processus.

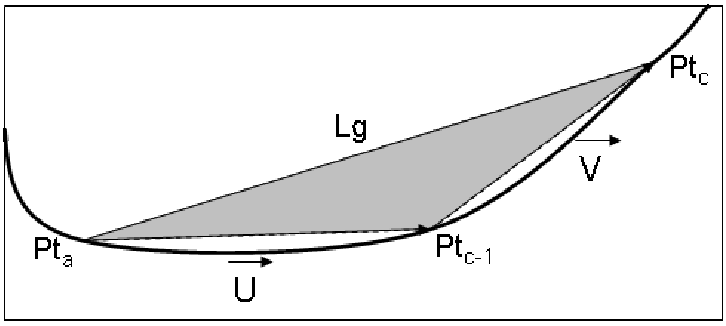


Figure 27. Illustration du calcul d'aire entre la courbe et le segment l'approximant.

La formule utilisée pour calculer une valeur représentative de l'erreur commise est :

$$\boldsymbol{Erreur = Erreur + \left\| \vec{U} \wedge \vec{V} \right\|}$$

Un nouveau sommet est ajouté quand : $\boldsymbol{Erreur > Lg \times Epsilon}$ où *Lg* désigne la longueur du segment approximant et *Epsilon* est une constante. La Figure 27 illustre le principe de l'algorithme sur un exemple. Si le cumul de l'erreur déjà commise auparavant, c'est-à-dire l'aire délimitée par les sommets Pt_a, Pt_c et Pt_{c-1}, est inférieure à *LgxEpsilon* alors on peut supprimer le sommet Pt_{c-1} sinon on le conserve comme point représentatif du tracé.

Ainsi, nous déterminons les sommets d'une ligne polygonale par un processus itératif. Cela présente deux avantages. D'une part, il n'est pas nécessaire de stocker tous les points de la courbe et d'autre part, les points de l'approximation polygonale sont extraits dans l'ordre de leur apparition lors de la réalisation de la signature.

Néanmoins, cet algorithme requiert de choisir le paramètre *Epsilon* lié à la qualité de l'approximation que l'on souhaite obtenir. Nous avons considéré deux méthodes pour choisir la valeur de ce paramètre. Dans un premier temps, nous avons cherché à estimer un seuil fixe, valable pour toutes les signatures indépendamment du signataire. Dans un deuxième temps, nous avons pensé qu'un seuil individualisé déterminé de manière automatique en fonction du signataire, c'est-à-dire de l'ensemble des signatures dont on dispose, permettrait de s'adapter à la forme de chaque type de signature.

2.2.5.1 Seuil fixe

La valeur du seuil correspond au degré de précision désiré pour la représentation de la signature. Nous avons fait un compromis entre la qualité de la représentation et le volume de données. Le seuil doit permettre de lisser la signature, de réduire le bruit tout en préservant suffisamment d'information pour caractériser la signature de manière efficace. De plus, un seul seuil pour tous les signataires semble raisonnable étant donné que les signatures ont été normalisées. La Figure 28 montre des exemples de résultats obtenus sur deux signatures en utilisant un seuil fixe déterminé expérimentalement (seuil = 100).

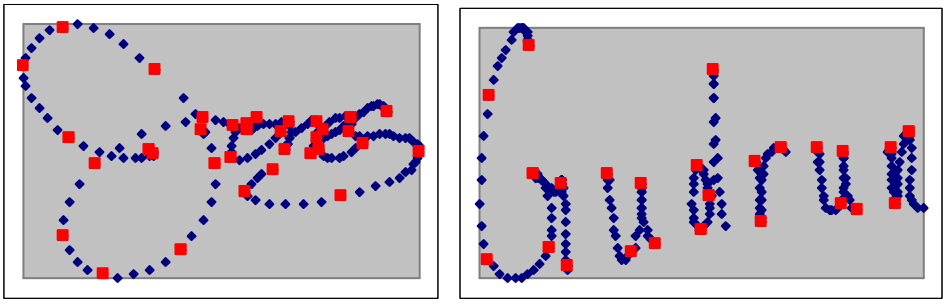


Figure 28. Illustration des points retenus par approximation polygonale.

2.2.5.2 Seuil individualisé

La polygonalisation prend en compte comme seul paramètre le degré de précision souhaité. Pour évaluer l'influence de ce paramètre, nous avons cherché le meilleur seuil d'approximation pour chaque personne, suivant deux critères. C'est la meilleure correspondance au sens de la stabilité pour chaque signataire que nous cherchons à atteindre grâce à ce seuil. Deux définitions de la stabilité ont été utilisées. D'une part, en considérant seulement le nombre de segments de l'approximation polygonale, d'autre part en considérant

les distances entre signatures où ne sont conservés que les points essentiels. Dans un premier cas, nous avons considéré le nombre de vecteurs nécessaires pour représenter les signatures d'apprentissage. Le premier critère utilisé consiste donc à minimiser la variance du nombre de segments. Le second critère que nous avons aussi étudié a pour but de minimiser la moyenne des distances au sens de DTW entre les signatures d'apprentissage. Nous avons fait varier le seuil de l'approximation polygonale afin de déterminer sa valeur optimale en fonction des deux critères précédents. Nous avons déterminé ainsi, pour chaque critère, le meilleur seuil pour chaque signataire. Cependant, les résultats obtenus en utilisant le premier critère n'ayant pas été concluants, seuls les tests utilisant le deuxième critère seront présentés par la suite.

La moyenne des distances DTW entre les signatures d'apprentissage évolue de manière irrégulière en fonction du seuil sélectionné (Figure 29). Cependant, le meilleur seuil donne généralement de meilleurs résultats qu'un seuil correspondant à l'utilisation de tous les points. La Figure 30 montre un exemple des points retenus en utilisant un seuil individualisé.

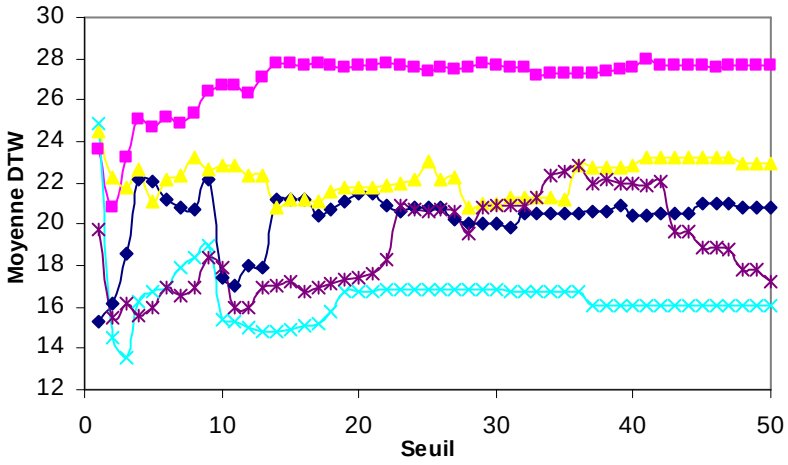


Figure 29. Illustration de l'évolution de la moyenne des distances DTW pour 5 signataires.

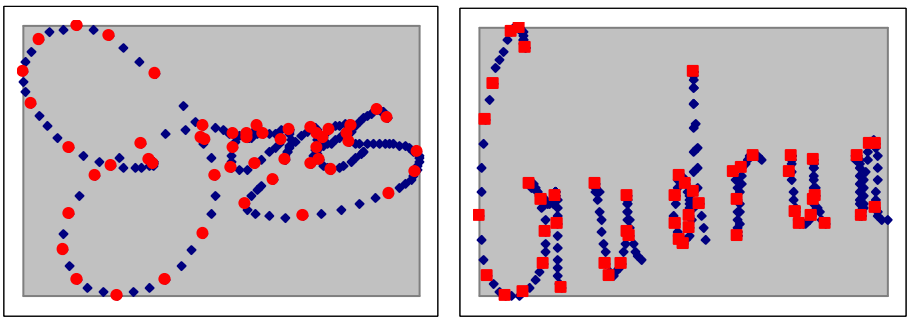


Figure 30. Illustration des points retenus par approximation polygonale avec un seuil individualisé obtenu par minimisation de la moyenne des distances DTW.

2.2.6. Bilan préliminaire

Nous avons déjà dit que l'évaluation ne peut se faire que sur un système complet d'authentification. L'utilisation de la version de base de DTW nous a néanmoins permis d'obtenir des résultats préliminaires concernant ce prétraitement que nous préconisons sur des signatures manuscrites en ligne. Ces résultats nous permettent d'ores et déjà d'affirmer que le fait de sélectionner des points pertinents améliore notablement les performances des systèmes d'authentification comme le montre la Figure 31.

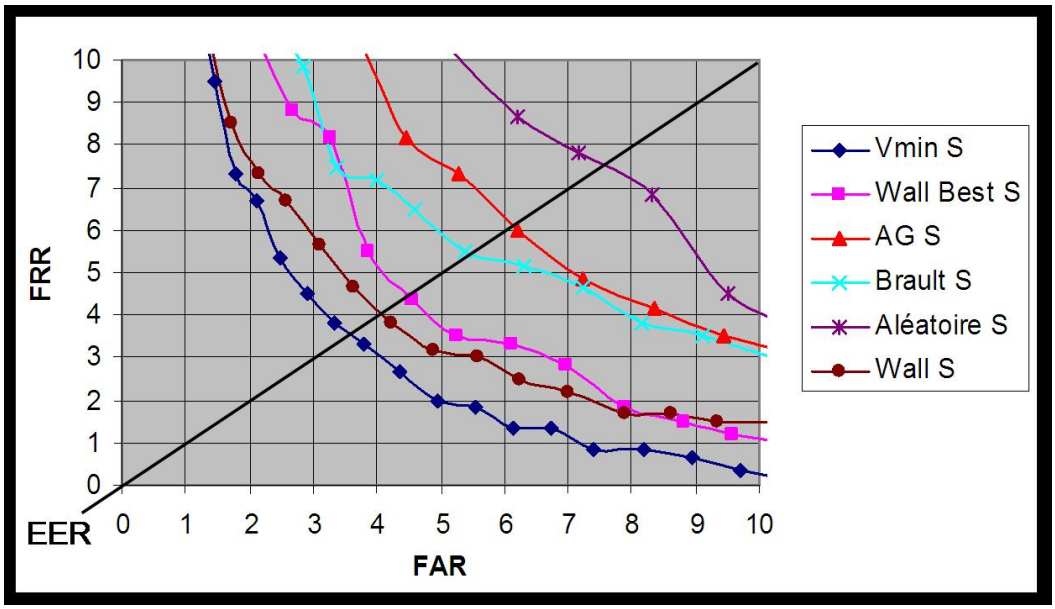


Figure 31. Résultats obtenus à partir de chacune des méthodes présentées (S=Spatiale) basées sur la distance calculée avec DTW.

Concernant les diverses méthodes proposées pour réduire le nombre de points, comme on pouvait s'y attendre, les taux de reconnaissance les moins bons sont obtenus avec un choix aléatoire de points. Cependant, la dégradation des performances n'est pas très importante ce qui nous montre la robustesse de l'algorithme de comparaison. Les points retenus par l'utilisation de la méthode de Brault donnent de moins bons résultats que les autres méthodes. Concernant la méthode de Wall, étant donné que le premier critère de recherche d'un seuil individualisé (stabilité du nombre de points retenus) donne de moins bons résultats que le deuxième critère, minimisation de la moyenne des distances intra signataire, seuls les résultats pour le deuxième critère sont donnés. Les meilleurs résultats de sélection de points sont obtenus en considérant les points correspondant aux minimums de la vitesse. La valeur de EER pour cette méthode est de 3,7% et la valeur de FAR de 5% pour une valeur de FRR égale à 2%. A titre de comparaison, si on utilise l'ensemble des points de la signature, on obtient une valeur de EER similaire et une valeur de FAR de 4,8% pour un taux de FRR égal à 2%.

Les résultats sont très proches de ceux obtenus à partir des points correspondant aux minimums de vitesse instantanée alors que le nombre de points utilisés pour la comparaison a été réduit au minimum d'un facteur 5.

Rappelons que ces résultats ont été obtenus avec l'algorithme DTW classique et qu'il est envisageable d'obtenir des résultats différents lors du couplage de ce prétraitement avec d'autres méthodes de comparaison de courbes. Ainsi, il pourra être intéressant de tester à nouveau l'ensemble des techniques de sélection de points avec la méthode de comparaison de courbes qui sera considérée comme la meilleure à la fin de notre travail.

Une fois la phase de prétraitement étudiée, il est nécessaire de se pencher sur la phase d'apprentissage qui correspond dans notre cas à la sélection, la création et la gestion du ou des modèles de la signature.

3. CREATION ET GESTION DES MODELES

La création du modèle représentatif de la signature de chaque utilisateur doit répondre à deux contraintes contradictoires. Il faut à la fois limiter le nombre de représentants et optimiser leur représentativité. Une autre contrainte, imposée par le contexte industriel de notre projet, est l'irréversibilité du modèle i.e. il ne doit pas être possible de reconstituer complètement la ou les signatures ayant servi à sa création.

3.1. Modèle basé sur la moyenne

Dans ce cas, le modèle est en fait une moyenne calculée sur l'ensemble des signatures produites par l'utilisateur lors de l'enregistrement. Les informations stockées correspondent la plupart du temps à la moyenne M et à l'écart type V de chacune des caractéristiques des signatures A1, A2, …, A5 de l'ensemble d'apprentissage. Comme on peut le constater sur la Figure 32, ce choix présente des inconvénients. En effet, la signature S1 est considérée comme authentique alors que la signature S2 ne l'est pas alors qu'elles sont à la même distance de la signature d'apprentissage A5. Le nombre de signatures d'apprentissage est trop limité pour considérer l'ensemble comme représentatif de la distribution globale des signatures.

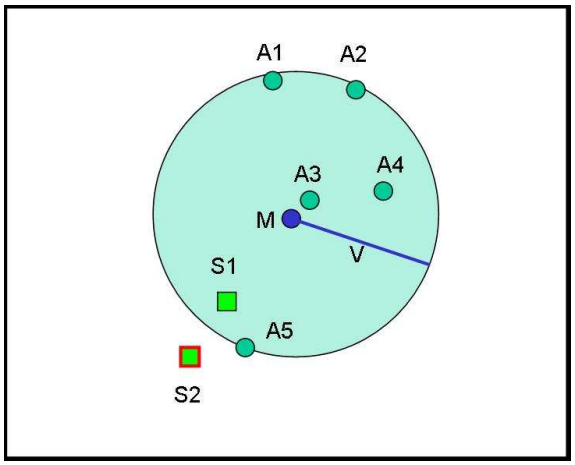


Figure 32. Modèle basé sur la moyenne (● : signatures d'apprentissage, ■ : signatures à tester).

Une alternative à ce choix est de considérer un modèle plus représentatif de la variabilité de la signature.

3.2. Modèle à plusieurs références

En considérant toutes les signatures d'apprentissage, on conserve une information sur toutes les signatures d'apprentissage, c'est-à-dire le vecteur de caractéristiques de chacune des signatures de l'ensemble d'apprentissage. Cette solution permet d'être moins sensible à la variabilité des signatures même lorsque l'on choisit un seuil de distance plus strict que pour la première solution (cf. rayon des cercles). Ce modèle est donc une solution possible au problème mentionné précédemment (Figure 33).

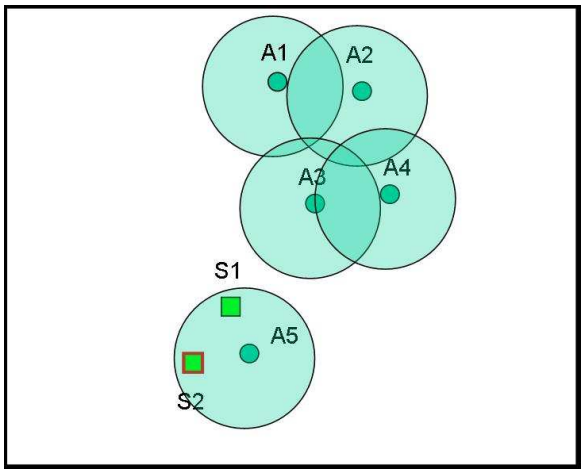


Figure 33. Modèle constitué de chacune des signatures d'apprentissage (● : signatures d'apprentissage, ■ : signatures à tester).

Ce modèle permet d'avoir un modèle plus représentatif de la variabilité de la signature d'un individu notamment lorsqu'un individu signe de façons différentes selon le contexte. Pour des raisons de continuité, une alternative à cette approche serait de considérer l'enveloppe

convexe de l'ensemble d'apprentissage ou plutôt du domaine constitué par la réunion des sphères associées aux signatures d'apprentissage.

3.3. Nombre de signatures utilisées

A notre avis, la base d'apprentissage doit contenir au moins 5 signatures de référence pour avoir un aperçu de la variabilité de la signature d'un individu. Il est difficile d'en demander plus sans que l'utilisateur abandonne l'idée d'utiliser sa signature pour s'authentifier. Cependant, plus le nombre de signatures d'apprentissage est élevé plus le modèle sera représentatif de la signature car il prendra en compte plus précisément les spécificités et les variations de la signature.

Le plus dur reste à faire puisqu'il nous faut maintenant choisir la manière dont nous allons caractériser les signatures pour pouvoir les comparer les unes aux autres.

Le paragraphe suivant présente de nouvelles approches permettant de caractériser une signature à la fois sur le plan spatial et sur le plan temporel de manière originale.

4. EXTRACTION DE CARACTERISTIQUES

Ici on considère l'ensemble des points de la signature. L'extraction de caractéristiques est une étape importante car elle correspond à un changement d'espace de représentation. En effet, durant cette étape, nous décidons quelles informations sont conservées et lesquelles sont supprimées car non utilisées pour la prise de décision.

Nous avons vu dans l'état de l'art les caractéristiques qui sont couramment utilisées par les chercheurs travaillant sur le sujet, voici maintenant les caractéristiques que nous avons mises en place.

4.1. Caractéristiques choisies : classiques ou novatrices

Nous proposons d'utiliser une approche fractale pour extraire des caractéristiques pertinentes des signatures. Trois représentations des signatures sont considérées. Une signature est tout d'abord vue comme un tracé continu, puis comme une suite de vecteurs et enfin comme un ensemble de points. Chacune de ces approches apporte une part d'information, soit spatiale, soit temporelle sur l'aspect global de la signature.

Des caractéristiques liées à la densité de la signature et au degré de superposition de la signature sont également calculées.

Une fois la liste des caractéristiques définie, la dernière partie de cette section est consacrée à l'étude de différentes méthodes de sélection de caractéristiques.

4.1.1. Dimension fractale ou Complexité de la signature

La notion de dimension fractale se situe dans la famille des dimensions non entières définies à la suite des travaux de Hausdorff ou Minkowski. Un ensemble dont on ne peut mesurer la longueur, ou l'aire, et qui admet une mesure finie lorsque l'on considère comme mesure une puissance particulière de pas est un ensemble fractal. Cette puissance particulière est appelée dimension de Hausdorff ou dimension fractale de l'élément étudié [MAN84]. Le calcul de la dimension fractale repose sur l'existence d'une relation en loi puissance entre une mesure et un paramètre variable en fonction de l'échelle d'observation à laquelle s'exerce la mesure. On cherche alors à repérer cette relation en faisant varier l'échelle d'observation et en effectuant différentes mesures du phénomène. Un graphe d'évolution peut être associé à ces mesures pour être ensuite analysé. Selon Mandelbrot, la dimension fractale caractérise le degré d'irrégularité ou de fragmentation d'un ensemble. Elle peut donc être utilisée pour quantifier la complexité d'une courbe. L'intérêt principal de la dimension fractale est d'extraire de la signature des éléments propres à l'individu et non discernables à l'œil nu. Un grand nombre de méthodes existent pour estimer la dimension suivant le type de données en entrée. Nous présenterons ici trois méthodes originales adaptées aux signatures en ligne. Après l'acquisition, si l'on considère que la signature n'est qu'un ensemble de points, on peut reproduire le tracé effectué en joignant les différents points. C'est alors uniquement la forme de la signature qui est étudiée (approche hors ligne). Par contre, si l'on utilise directement l'ensemble des points fournis par la tablette (approche basée sur la vectorisation et méthode locale), dans ce cas, les indications sur la chronologie du tracé et sur la dynamique sont prises en compte. Ces différentes approches pour étudier la signature sont décrites dans les prochains paragraphes.

4.1.1.1 Fractalité de l'écriture manuscrite

Dans [BOU97] qui traite de la classification de l'écrit par les méthodes fractales et prouve le comportement fractal de l'écriture, une partie de la thèse traite de l'application de ces paramètres fractals à l'authentification de signatures manuscrites hors ligne. Ces paramètres ont permis de définir une classification en quatre familles : les paraphes, les signatures simples, les signatures légèrement enjolivées et les signatures complexes. Ces premiers

résultats encourageants nous ont incité à poursuivre dans cette voie et étudier ces paramètres sur des signatures en ligne [WIR04c].

4.1.1.2 Approche hors ligne

Cette première méthode a pour entrée l'image de la signature reconstituée par application de l'algorithme de Bresenham entre les couples de points temporellement consécutifs. Une illustration du résultat de l'application de l'algorithme de Bresenham réalisant une interpolation entre les points fournis par le dispositif d'acquisition est donnée à la Figure 34. Ensuite, comme en hors ligne, seule la forme est prise en compte. La valeur de la dimension fractale de X est donnée par l'expression suivante où ε désigne l'échelle d'observation, n la dimension de l'espace dans lequel on fait l'étude et A la mesure associée dans cet espace :

$$D(X) = \lim_{\varepsilon \to 0} \left(n - \log(A(X_\varepsilon)) / \log \varepsilon \right)$$

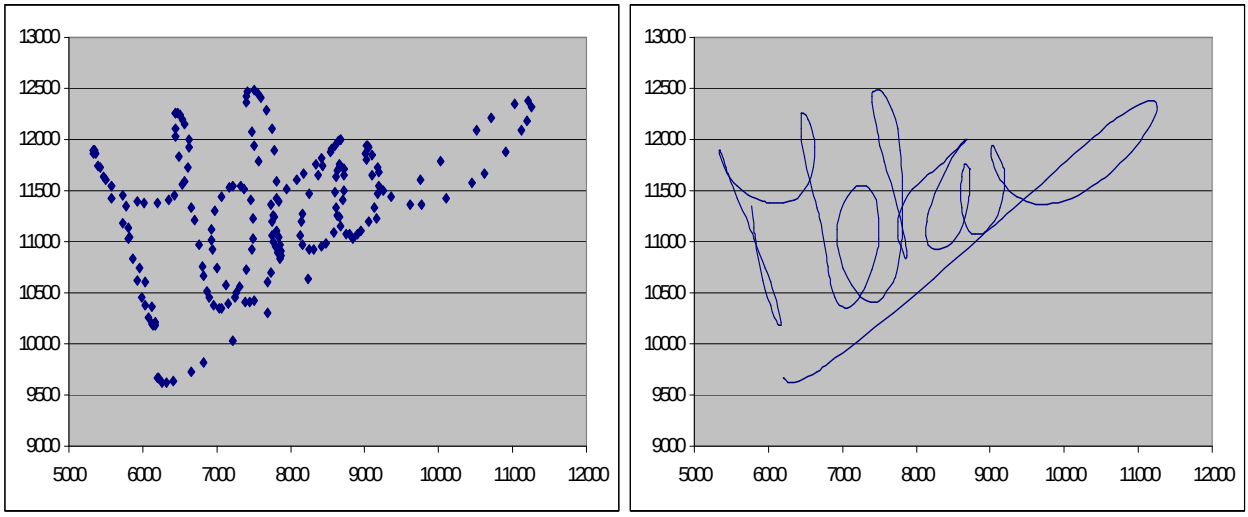


Figure 34. Signature avant et après Bresenham.

Dans notre cas, X représente la signature contenue dans un plan (n=2), A représente l'aire. X_ε est le dilaté de X dans une dilatation morphologique dont l'élément structurant est paramétré par une longueur ε. Dans la pratique, ε est un entier dont les valeurs varient de 1 à k. L'aire de la signature ainsi que les aires de ses différents dilatés peuvent être calculées. A chaque dilatation, on perd des détails sur le tracé ainsi on simule une étude du tracé à plus grande échelle. La Figure 35 représente les dilatés d'une signature à différentes étapes.

Figure 35. Représentation des dilatés d'une signature : approche hors ligne.

Dans la pratique on peut remarquer, comme on le voit sur la Figure 36, que la courbe représentative de $\log(A(X_\varepsilon)/\varepsilon)$ en fonction de $log(\varepsilon)$ peut être approximée par deux segments de droite consécutifs. L'évolution de l'aire peut donc être estimée par les pentes de ces deux droites obtenues par régressions linéaires. Le point de changement de modèle est obtenu par recherche du point le plus éloigné du segment joignant le premier et le dernier point de la courbe. On prendra comme première caractéristique fractale la pente la plus forte soit entre le premier point et la cassure soit entre la cassure et le dernier point.

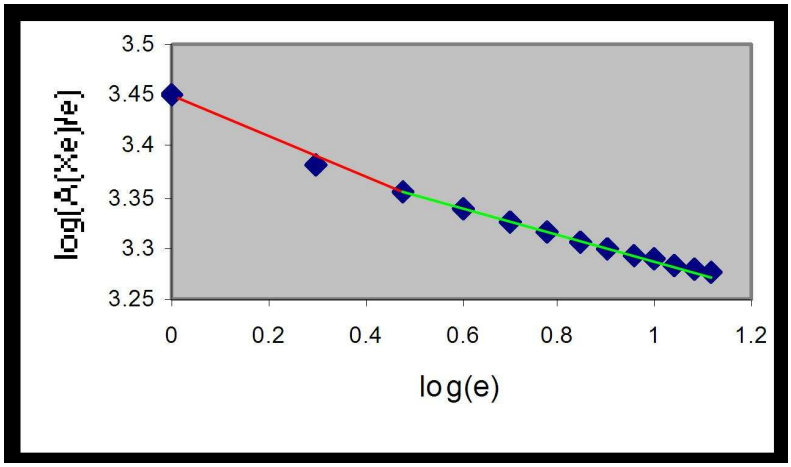


Figure 36. Evolution de l'aire des dilatés de la signature en fonction de la taille de l'élément structurant avec des échelles logarithmiques.

La dimension fractale d'un tracé dans un espace de dimension deux, et donc d'une signature, est comprise entre 1 et 2. La dimension fractale d'un trait droit est 1. Plus le tracé de la signature sera complexe, plus la dimension fractale sera proche de 2. La dimension fractale de la signature, dont les dilatés sont représentés à la Figure 35, est égale à 1,3. Pour cette signature, on peut déterminer la dimension fractale à partir de 5 itérations.

4.1.1.3 Approche basée sur la vectorisation

Cette deuxième méthode plus proche des données en ligne se base sur la liste des points constituant la signature obtenue par le système d'acquisition. L'intérêt de cette méthode est de prendre en compte la temporalité du tracé, plus explicitement les intervalles de temps séparant les points successifs du tracé, contrairement à la méthode précédente. Pour calculer cette dimension, nous avons réalisé une approximation polygonale de la signature en utilisant l'algorithme de Wall [WAL84]. Le paramètre étudié en fonction d'une échelle d'observation est le nombre de vecteurs nécessaires pour décrire l'ensemble de la signature pour un seuil donné. Le changement d'échelle est modélisé par la variation de l'erreur d'approximation considérée dans l'algorithme de Wall. L'erreur d'approximation augmentant au fur et à mesure des itérations, progressivement le tracé de la signature deviendra très rectiligne. On perd de plus en plus de détails. Ce principe est donc lié à une représentation multi échelle du tracé de la signature. La formule devient alors :

$$D(X) = \lim_{\varepsilon \to 0} \left(1 - \frac{\log\left(N(X_\varepsilon)/\varepsilon\right)}{\log \varepsilon} \right)$$

Où N(Xε) représente le nombre de vecteurs nécessaires pour approximer la signature à l'échelle ε. X_ε est une approximation de la signature obtenue par une étape de vectorisation caractérisée par un seuil d'erreur ε. Pour calculer cette dimension fractale, on approxime successivement le tracé de la signature par un ensemble de segments en faisant varier la précision. Une illustration est donnée dans la Figure 37.

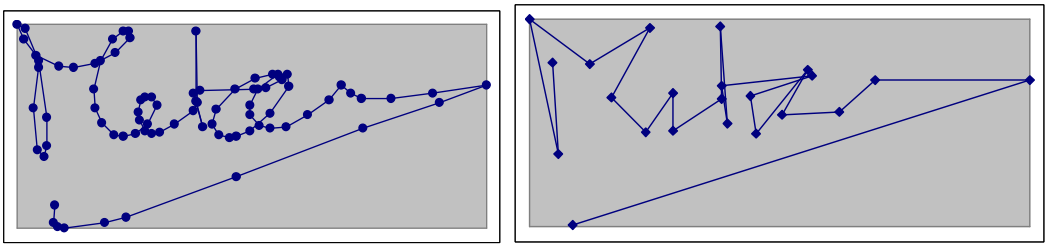


Figure 37. Evolution des représentations en fonction de l'erreur d'approximation dans la construction de la ligne polygonale.

Pour chaque échelle d'observation de la signature, on calcule le nombre de vecteurs obtenus pour approximer la signature. Puis, on représente l'évolution du logarithme du nombre de vecteurs en fonction du logarithme du seuil d'approximation. A chaque itération, on diminue la précision de l'approximation de la signature et donc on prend en considération de moins en moins les détails du tracé.

Dans la pratique, on peut remarquer que la courbe représentative de $\log\left(N(X_\varepsilon)/\varepsilon\right)$ en fonction de log(ε) peut être approximée par deux segments de droite consécutifs. Cette seconde

dimension fractale est ensuite calculée de la même façon que précédemment à partir du graphe d'évolution. On approxime la courbe par deux segments de droite obtenus par régression linéaire puis on considère la pente la plus forte en valeur absolue comme la dimension fractale.

4.1.1.4 Calcul local

Cette méthode consiste à procéder comme pour la première méthode mais, au lieu de considérer la signature reconstituée avec l'algorithme de Bresenham, on utilise la liste des points constituant la signature. Le masque utilisé pour la dilatation est le même que pour la première méthode. La Figure 38 représente les dilatés d'une signature considérée comme une suite de points. L'intérêt de cette méthode est que l'on opère directement sur les données de la signature et donc on n'assimile pas la signature à une suite de segments comme c'est le cas pour le calcul d'un grand nombre de caractéristiques classiques comme la longueur ou la somme des angles.

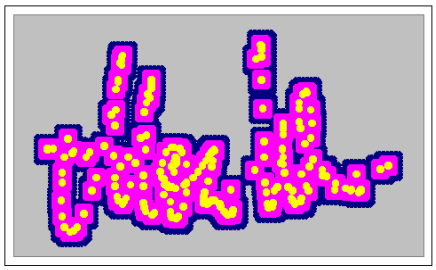


Figure 38. Représentation des dilatés d'une signature : approche en ligne.

Il est à noter que la courbe obtenue en considérant l'aire des dilatés de la signature en fonction du nombre d'itérations (Figure 39) présente des formes plus marquées que dans les études précédentes mais on peut toujours l'approximer par deux demi-droites.

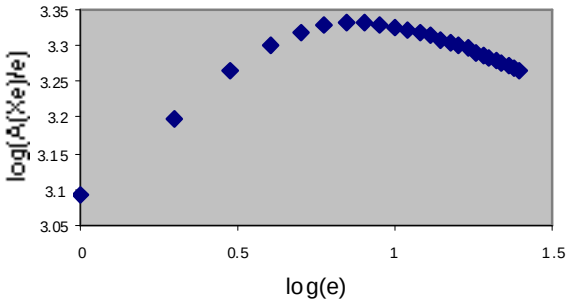


Figure 39. Evolution de l'aire des dilatés de la signature en fonction de la taille de l'élément structurant.

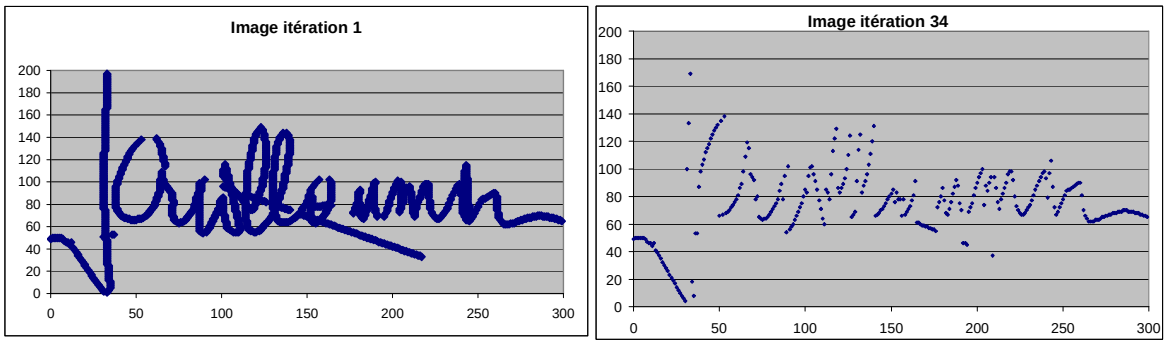
Chaque segment peut être interprété : la première demi-droite à gauche correspond aux dilatations effectuées avant le contact entre les différents points dilatés. La dimension d'un ensemble discret de points est inférieure à 1. Elle est donc liée à la vitesse ou plus exactement

à la régularité de la vitesse d'exécution de la signature. Plus le tracé sera saccadé, plus la pente de la première demi-droite sera importante. La pente de la seconde demi-droite correspond aux dilatations effectuées après le contact entre les différents éléments dilatés. Elle est donc liée à la forme du tracé de la signature. La dimension d'une courbe est comprise entre 1 et 2. Nous considérons donc les deux pentes comme des caractéristiques différentes et complémentaires.

4.1.2. Dimension de profondeur

La dimension de profondeur caractérise le degré de superposition verticale d'une courbe reconstituée à l'aide de l'algorithme de Bresenham. Cette approche permet de prendre en compte les boucles présentes dans la signature, même si elles sont imparfaitement fermées.

Pour calculer la dimension de profondeur, on détermine les profils supérieur et inférieur de la signature. A chaque étape, on supprime ces profils si ils ne provoquent pas l'apparition de colonnes ne contenant aucun pixel noir. On répète cette opération jusqu'à obtenir un tracé invariant. La Figure 40 représente la signature initiale après reconstruction du tracé et le résultat obtenu après la dernière itération.



(a) (b)

Figure 40. Tracé de la signature après une itération appliquée sur la signature (a) et le tracé invariant obtenu au final (b).

Puis, de façon analogue au cas de la dimension fractale, on peut représenter l'évolution du logarithme du nombre de pixels constituant la signature en fonction du logarithme du nombre d'itérations effectuées. Sur la courbe obtenue, on repère la cassure i.e. le point le plus éloigné du segment joignant le premier et le dernier point. La dimension de profondeur est alors estimée à partir de la pente la plus forte entre le premier point et la cassure ou entre la cassure et le dernier point.

Plus la dimension de profondeur est grande, plus la courbe est constituée de "tracés" superposés. Comme la dimension fractale, cette dimension est invariante par translation.

4.1.3. Dimension de masse ou Densité de la signature

L'évolution de la densité de la signature est étudiée en fonction de la taille d'une zone d'influence autour d'un point de focalisation. A partir du centre du rectangle englobant la signature, on définit un rectangle centré sur ce point. A chaque étape, on compte le nombre de points de la signature contenus dans le rectangle puis on augmente la taille du rectangle. Une illustration de la méthode de calcul de la dimension de masse est donnée à la Figure 41.

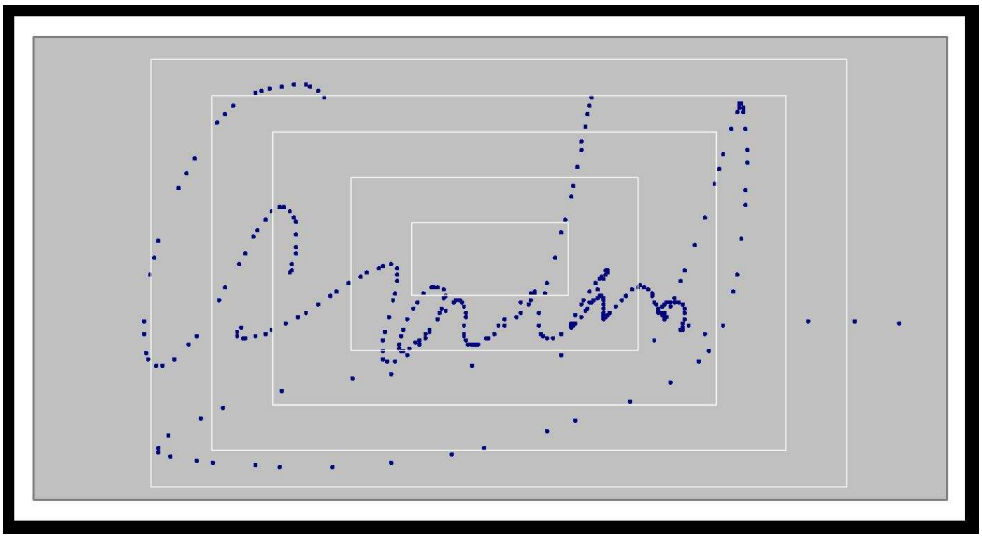


Figure 41. Illustration de la méthode utilisée pour calculer la dimension de masse.

On modélise ensuite l'évolution du logarithme du nombre de points contenus dans chacun des rectangles en fonction du logarithme de la taille des rectangles. La dimension de masse est estimée à partir de la demi-droite de plus forte pente. Plus la dimension sera élevée, plus la signature sera dense.

4.2. Nouvelles caractéristiques vs. caractéristiques classiques

Pour tester nos caractéristiques, nous avons élaboré un logiciel de vérification dans lequel le modèle d'une signature est constitué de la moyenne et de l'écart type pour l'ensemble des caractéristiques calculées à partir des signatures d'apprentissage. Soient m_i et σ_i, respectivement la moyenne et l'écart type de la caractéristique i pour l'ensemble des signatures de référence, soient n le nombre de caractéristiques utilisées dans le processus de décision, α un seuil fixé et t_i la valeur de la caractéristique i de la signature testée. La règle de décision est la suivante :

si $\max_{i=1..n}(|t_i - m_i|/\sigma_i) < \alpha$, alors la signature est acceptée. Sinon la signature est considérée comme fausse.

Ceci peut être réalisé quel que soit l'ensemble de caractéristiques considérées et quelle que soit la valeur du seuil.

Nous avons donc conçu un système d'authentification fonctionnant suivant le modèle précédent et utilisant les quatre dimensions fractales, la dimension de profondeur, la densité plus un certain nombre de caractéristiques traditionnelles suivantes : angle entre l'horizontale et la droite passant par le premier et par le dernier point, somme des angles entre les points de la signature, somme absolue des angles entre les points de la signature, longueur de la signature, distance entre le premier et le dernier point, rapport entre les déplacements vers la gauche et vers la droite, rapport entre les déplacements vers le haut et vers le bas, rapport entre les déplacements suivant X et suivant Y. Nous disposons donc d'un ensemble de 14 primitives. Dans la suite de ce chapitre, nous utiliserons les notations suivantes :

Classiques

- RHB : Rapport entre les déplacements vers le haut et vers le bas;
- APD : Angle entre l'horizontale et la droite passant par le premier et par le dernier point;
- SA : Somme des angles entre les points de la signature;
- SAA : Somme des valeurs absolues des angles entre les points de la signature;
- L : Longueur de la signature;
- DPD : Distance entre le premier et le dernier point;
- RGD : Rapport entre les déplacements vers la gauche et vers la droite;
- RXY : Rapport entre les déplacements suivant x et suivant y;

Fractales

- DF : Dimension Fractale;
- DV : Dimension Vecteur;
- DFL1 : Dimension Fractale Locale 1;
- DFL2 : Dimension Fractale Locale 2;
- DP : Dimension de Profondeur;
- DM : Dimension de Masse.

La Figure 42 montre les résultats de la vérification obtenus à l'aide des caractéristiques classiques seules, fractales seules et classiques plus fractales. Nous avons fait varier le seuil afin de définir différents systèmes caractérisés par différentes valeurs de FAR et de FRR. Les valeurs de FAR et de FRR sont les plus représentatives d'un système de vérification.

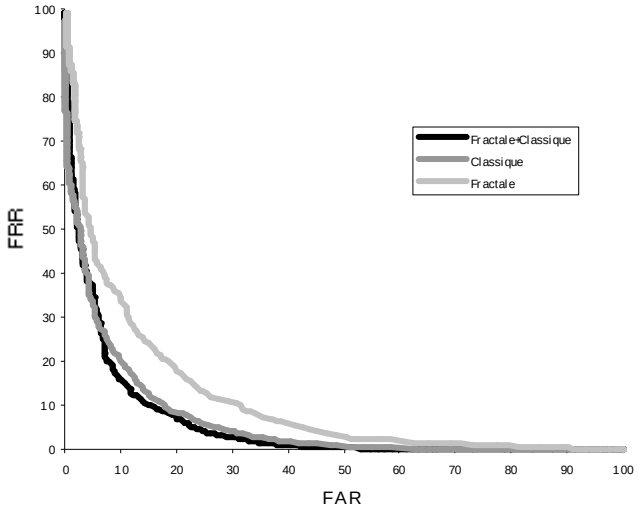


Figure 42. FAR vs FRR pour différentes valeurs de α dans un système de vérification basé sur une représentation vectorielle des signatures.

Même si les caractéristiques liées à la dimension fractale seule sont moins pertinentes pour authentifier une signature que les caractéristiques classiques seules, les résultats montrent qu'elles améliorent la vérification par signature si elles sont utilisées en complément des caractéristiques classiques.

Afin d'évaluer le pouvoir discriminant des nouvelles caractéristiques proposées, nous avons aussi choisi d'effectuer une sélection parmi l'ensemble des caractéristiques. Pour cela, plusieurs méthodes de sélection de caractéristiques ont été appliquées sur l'ensemble des caractéristiques à la fois classiques et nouvelles.

4.3. Sélection de caractéristiques

Parmi l'ensemble des caractéristiques utilisables, toutes ne sont pas assez discriminantes. Il est important de supprimer les caractéristiques qui pourraient perturber les résultats afin d'éviter ensuite des erreurs lors de la prise de décision durant la phase de comparaison. Plus le nombre de caractéristiques utilisées est élevé (dans la limite du raisonnable), plus les performances sont améliorées. Mais l'utilisation d'un grand nombre de caractéristiques présente un certain nombre d'inconvénients. Tout d'abord, il est nécessaire d'avoir un espace mémoire conséquent. De plus, étant donné que les signatures ne sont pas rigoureusement identiques, la distance entre la signature testée et la signature de référence va augmenter et donc le seuil devra être augmenté. Il peut donc être utile d'effectuer une sélection parmi ces caractéristiques [XUH96][KET05]. Pour sélectionner les variables les plus pertinentes, plusieurs méthodes ont été envisagées :

- calcul de la variance intra classe et interclasses après avoir normalisé les valeurs des caractéristiques en divisant chaque valeur par le maximum pour chaque caractéristique
- une ACP
- le test de Fisher
- un algorithme génétique pour trouver la meilleure combinaison des caractéristiques

Il est à noter que ces méthodes sont très dépendantes de la taille de la base de signatures utilisée. Par conséquent, plus la base sera importante, plus les résultats pourront être considérés comme exacts car vérifiés sur une grande variété de signatures. Les résultats présentés pour chacune des méthodes ont été obtenus sur la base SVC décrite précédemment.

4.3.1. Rapport variance intra classe sur variance interclasses

Le but recherché est de trouver les variables qui minimisent la distance intra individu et qui maximisent la distance inter individus. Une solution possible est d'évaluer le rapport entre la variance intra classe et la variance interclasses pour chaque caractéristique. Ces variances doivent être normalisées au préalable pour pouvoir comparer les valeurs des différentes caractéristiques. Plus ce rapport sera faible, plus la variable sera discriminante. Cette méthode nous permet de classer les variables suivant leur pouvoir discriminant. Il faut ensuite fixer un seuil à partir duquel on ne tient pas compte des autres caractéristiques ou fixer un nombre maximum de caractéristiques.

Ainsi, pour chacune des caractéristiques citées précédemment, nous avons calculé le rapport entre la moyenne de la variance intra classe et la variance inter classes. Le Tableau 6 présente les résultats obtenus. Les caractéristiques ont été classées suivant leur pouvoir discriminant, du plus élevé au plus faible.

DPD	0.062
DF	0.108
L	0.168
RHB	0.172
RGD	0.202
DFL1	0.207
DP	0.261
DFL2	0.273
DM	0.315
RXY	0.335
DV	0.361
SAA	0.457
APD	1.260
SA	2.055

Tableau 6. Rapport entre la variance intra classe et la variance inter classes pour chacune des caractéristiques.

Au vu des résultats obtenus, les caractéristiques novatrices présentées dans ce chapitre ont un fort pouvoir discriminant, notamment la dimension fractale calculée à partir du tracé reconstitué de la signature. Le tableau semble également indiquer, que suivant ce critère, les caractéristiques se basant sur des calculs d'angles sont peu discriminantes.

4.3.2. Analyse en composantes principales

L'analyse en composantes principales (en abrégé ACP) est une méthode de réduction du nombre de variables permettant des représentations géométriques des individus. Cette réduction ne sera possible que si les **p** caractères initiaux ne sont pas indépendants et ont des coefficients de corrélation non nuls.

	L	RXY	RHB	RGD	DF	DP	DV	SA	SAA	APD	DPD	DFL1	DFL2	DM
L	1	0.23	-0.13	0.68	0.73	0.78	-0.01	0.42	0.72	-0.5	0	-0.43	-0.79	-0.12
RXY	0.23	1	0.15	-0.41	0.22	-0.1	-0.19	-0.03	0.09	0.01	0.48	-0.05	-0.2	0.01
RHB	-0.13	0.15	1	-0.23	0.13	-0.27	-0.16	-0.34	-0.04	0.87	-0.07	-0.13	-0.04	0.04
RGD	0.68	-0.41	-0.23	1	0.45	0.78	0.04	0.42	0.51	-0.47	-0.6	-0.25	-0.54	-0.1
DF	0.73	0.22	0.13	0.45	1	0.54	-0.34	0.43	0.49	-0.19	0.04	-0.76	-0.77	0.18
DP	0.78	-0.1	-0.27	0.78	0.54	1	0.01	0.41	0.55	-0.51	-0.14	-0.29	-0.67	-0.17
DV	-0.01	-0.19	-0.16	0.04	-0.34	0.01	1	-0.19	0.03	-0.13	0.11	0.36	0.11	-0.44
SA	0.42	-0.03	-0.34	0.42	0.43	0.41	-0.19	1	0.57	-0.45	-0.14	-0.28	-0.33	-0.02
SAA	0.72	0.09	-0.04	0.51	0.49	0.55	0.03	0.57	1	-0.35	-0.13	-0.27	-0.58	-0.24
APD	-0.5	0.01	0.87	-0.47	-0.19	-0.51	-0.13	-0.45	-0.35	1	-0.05	0.11	0.25	0.08
DPD	0	0.48	-0.07	-0.6	0.04	-0.14	0.11	-0.14	-0.13	-0.05	1	-0.03	0.01	-0.02
DFL1	-0.43	-0.05	-0.13	-0.25	-0.76	-0.29	0.36	-0.28	-0.27	0.11	-0.03	1	0.32	-0.11
DFL2	-0.79	-0.2	-0.04	-0.54	-0.77	-0.67	0.11	-0.33	-0.58	0.25	0.01	0.32	1	-0.05
DM	-0.12	0.01	0.04	-0.1	0.18	-0.17	-0.44	-0.02	-0.24	0.08	-0.02	-0.11	-0.05	1

Tableau 7. Matrice de variance covariance.

La matrice de variance covariance apporte de nombreuses informations sur les caractéristiques novatrices. Tout d'abord, la dimension de profondeur est liée à la fois à la forme et à la complexité de la signature. La dimension fractale, quelle que soit la méthode de calcul, apporte une information différente des caractéristiques classiques. Pour finir, la dimension de masse apporte une information complémentaire des caractéristiques classiques et des caractéristiques fractales.

L'ACP est une méthode factorielle car la réduction du nombre des caractères ne se fait pas par une simple sélection de certains d'entre eux, mais par la construction de nouveaux caractères synthétiques obtenus en combinant les caractères initiaux au moyen des "facteurs". C'est une méthode linéaire car les facteurs sont des combinaisons linéaires des caractéristiques initiales. L'inconvénient de cette méthode est qu'elle ne prend pas en compte la variance intra classe. Si une caractéristique a une très grande variance inter individus bien qu'elle possède également une très grande variance intra individu, elle aura une forte influence sur la détermination des axes principaux.

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14
L	-0.4	0.06	0.13	-0.16	0.14	-0.06	0.08	-0.17	0.11	-0.17	0.56	0.4	-0.1	0.46
RXY	-0.01	0.35	0.49	-0.07	-0.1	-0.38	0.35	-0.35	-0.27	-0.25	-0.18	0.02	-0.01	-0.25
RHB	0.11	0.43	-0.29	-0.45	-0.06	-0.04	-0.1	0.01	0.04	-0.24	0.19	-0.36	0.51	0.11
RGD	-0.35	-0.22	-0.33	-0.05	0.17	-0.03	0.12	-0.06	-0.07	-0.15	0.35	-0.05	0.01	-0.72
DF	-0.35	0.35	-0.02	0.03	0.08	0.19	-0.11	-0.03	-0.29	0.18	0.05	-0.55	-0.52	0.11
DP	-0.38	-0.12	-0.02	-0.09	0.24	0	0.19	0.44	0.09	-0.52	-0.48	-0.08	-0.02	0.15
DV	0.05	-0.38	0.19	-0.44	0.22	0.18	-0.51	-0.36	-0.31	-0.16	-0.17	0.01	-0.01	0.01
SA	-0.28	-0.07	0.01	0.25	-0.62	-0.09	-0.36	0.24	-0.43	-0.21	0.08	0.09	0.16	0.04
SAA	-0.33	-0.02	0.04	-0.23	-0.4	-0.2	-0.27	-0.21	0.64	0.11	-0.21	-0.08	-0.16	-0.13
APD	0.26	0.34	-0.3	-0.33	-0.07	-0.04	-0.14	0.25	-0.09	-0.12	-0.08	0.48	-0.49	-0.17
DPD	0.07	0.19	0.62	0.01	0.13	0.26	-0.26	0.43	0.23	-0.09	0.27	-0.04	0.04	-0.31
DFL1	0.23	-0.32	0.09	-0.13	0.09	-0.71	-0.08	0.28	-0.04	0.03	0.24	-0.31	-0.22	0.1
DFL2	0.35	-0.18	-0.04	0.17	-0.3	0.26	0.13	-0.22	0.19	-0.59	0.19	-0.24	-0.33	0.06
DM	0.03	0.26	-0.16	0.54	0.4	-0.29	-0.47	-0.22	0.14	-0.26	-0.08	0.05	0.02	-0.01

Tableau 8. Définition des facteurs.

Le Tableau 8 confirme les résultats obtenus par l'étude des variances puisque les caractéristiques novatrices représentent des éléments importants pour la construction des nouveaux axes. En effet, parmi les quatre caractéristiques (RGD, DF, DP et DFL2) ayant un poids important dans la composition du premier facteur, trois sont des caractéristiques novatrices. Concernant le deuxième axe, deux parmi les quatre plus importantes sont également des caractéristiques novatrices. La proportion de caractéristiques novatrices parmi les caractéristiques de poids importants est le même pour les deux axes suivants, V3 et V4.

Qualité cumulée
36.66
53.59
66.23
76.71
82.77
88.28
92.15
94.97
96.98
98.45
99.26
99.62
99.89
100

Tableau 9. Qualité cumulée des axes principaux.

Par contre le Tableau 9 indique que les caractéristiques sont assez peu corrélées et qu'il semble donc difficile de construire par combinaison linéaire des variables synthétiques discriminantes en nombre réduit. L'analyse montre en effet qu'il est nécessaire de retenir 6 facteurs pour parvenir à expliquer 85% de la variance, les deux premiers d'entre eux ne permettant d'expliquer que la moitié de la variance.

4.3.3. Test de Fisher

Le but de ce test est également de trouver les caractéristiques qui minimisent la distance intra individu et qui maximisent la distance inter individus. Cette méthode permet d'ordonner les variables en fonction de leur capacité de discrimination. Ce test permet de tester l'homogénéité d'une population mais ne permet pas de savoir quel élément de la population est différent. Il s'utilise dans le cas de deux classes pour évaluer le pouvoir discriminant des caractéristiques. Afin d'adapter ce test à un problème à plusieurs classes, une classe par individu, on effectue ce test en considérant les individus deux par deux. Pour chaque caractéristique, nous évaluons l'homogénéité de la population suivant le test de Fisher. La moyenne de ces valeurs est calculée pour chaque caractéristique pour l'ensemble des couples de classes. La moyenne est ensuite calculée sur chacune des caractéristiques.

Origines de la variance	Sommes des Carrés des Ecarts (SCE)	Degrés de liberté	Carrés Moyens (CM)	F
Inter-classes	$SCE_{inter} = \sum_j n_j (\overline{X}_{.j} - \overline{X}..)^2$	p-1	$CM_{inter} = \frac{SCE_{inter}}{p-1}$	$\boldsymbol{F_{obs} = \frac{CM_{inter}}{CM_{intra}}}$
Intra-classe	$SCE_{intra} = \sum_j \sum_i (X_{ij} - \overline{X}.j)^2$	N-p	$CM_{intra} = \frac{SCE_{intra}}{N-p}$	

p est le nombre de signataires et N le nombre total de signatures (nombre de signataires x nombre de signatures)

L'objectif de ce test est de vérifier si les variances de deux populations sont égales. Deux hypothèses sont envisageables :

- H0 : $CM_{inter} = CM_{intra}$
- H1 : CM_{inter} différent de CM_{intra}

On peut montrer que le rapport $F = \frac{VarInter / p-1}{VarIntra / N-p}$, si H0 est vérifiée, suit la loi de Fisher-Snedecor à (p-1, N-p) degrés de liberté.

La statistique $\boldsymbol{F_{obs} = \frac{CM_{inter}}{CM_{intra}}}$ est utilisée pour évaluer l'hypothèse H0.

Si F_{obs} est supérieur à F_{seuil} lu dans la table de la loi de Fisher-Snedecor pour un risque d'erreur α fixé alors l'hypothèse H0 est rejetée. En d'autres termes, si $F_{obs} > F_{seuil}$, alors la caractéristique étudiée a un effet significatif en moyenne sur la discrimination des individus, sinon la caractéristique étudiée n'a pas d'effet significatif en moyenne sur la discrimination des individus.

Le Tableau 10 présente les résultats obtenus pour chaque caractéristique.

L	59.67
DF	46.57
DV	45.00
DPD	36.94
DP	27.08
RHB	26.89
RGD	25.79
RXY	23.70
DFL1	18.95
DM	16.80
DFL2	15.47
SAA	11.99
APD	5.59
SA	1.92

Tableau 10. Résultat du test de Fisher : F2,∞ = 4.61 pour une probabilité de P=99.9%.

Une nouvelle fois, la classification obtenue par le test de Fisher montre clairement que les caractéristiques relatives à la dimension fractale ont des propriétés intéressantes. Elles ont des taux élevés et donc possèdent un grand pouvoir de discrimination. Mais toutes les caractéristiques sont discriminantes avec une grande confiance étant donné que la valeur de F obtenue pour chacune d'elles est supérieure au seuil $F_{2,\infty}$, excepté la somme des angles.

4.3.4. Algorithme génétique (AG)

Pour construire le classificateur optimal à l'aide d'un AG, chaque caractéristique est associée à un gène, donc un chromosome correspond à un ensemble de caractéristiques. Si une des caractéristiques est sélectionnée, le gène correspondant est fixé à 1 et à 0 dans le cas contraire. De plus, la valeur du seuil de notre classificateur, déjà utilisée au paragraphe 4.2., est codée sur les 8 derniers gènes de chaque chromosome de manière à ce que lui aussi soit déterminé au cours du processus d'optimisation. Un chromosome est donc constitué de 22 gènes (14 caractéristiques et 8 bits pour le seuil). La structure d'un chromosome est schématisée à la Figure 43.

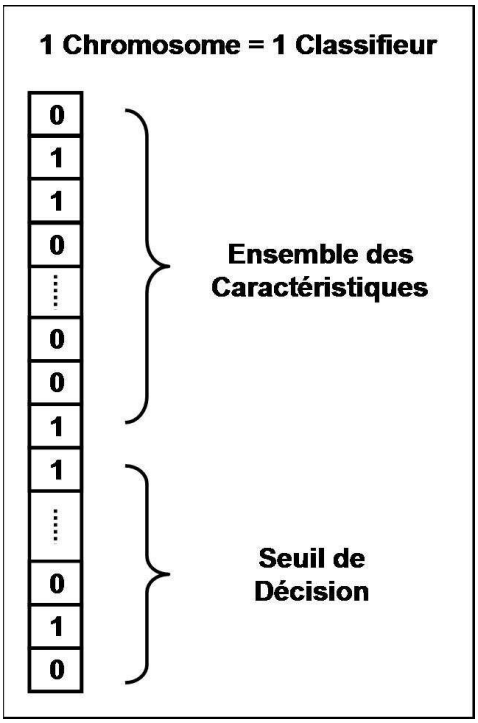


Figure 43. Illustration de la structure d'un chromosome i.e. un classificateur.

Les paramètres utilisés dans notre AG sont les suivants :

- taux de mutation : 0,01%
- croisement : 3 points de coupure
- sélection : 4-Tournoi

On applique un algorithme génétique pour trouver la meilleure combinaison des caractéristiques et le meilleur seuil au sens de la minimisation du taux d'erreur.

Pour mettre en place une minimisation des taux de FAR et de FRR, il faut donc réaliser un algorithme génétique multi objectifs [OLI03]. Le principe consiste à évaluer, pour chaque chromosome, les taux de FAR et de FRR (Figure 44) puis à lui affecter un score correspondant à la position du point, dont les coordonnées sont FAR et FRR, par rapport aux autres chromosomes. Plus un point (chromosome) est près des axes, plus on lui affectera un score élevé. La fonction de fitness utilisée est alors le rang du chromosome. On obtient ainsi, par association de chromosomes de même rang, plusieurs courbes parallèles correspondant à plusieurs niveaux de score. Une représentation schématique des rangs obtenus pour différents chromosomes est donnée à la Figure 44.

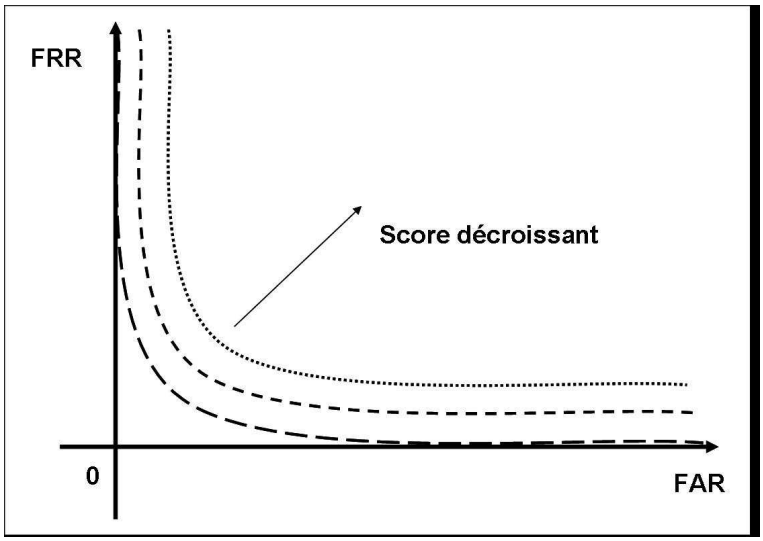


Figure 44. Illustration des courbes obtenues à partir de l'ensemble des chromosomes issus de la population.

Le résultat final obtenu est un ensemble de chromosomes minimisant à la fois le taux de FAR et celui de FRR. Dans nos tests, nous avons utilisé une population de 2000 chromosomes et généré 100 populations.

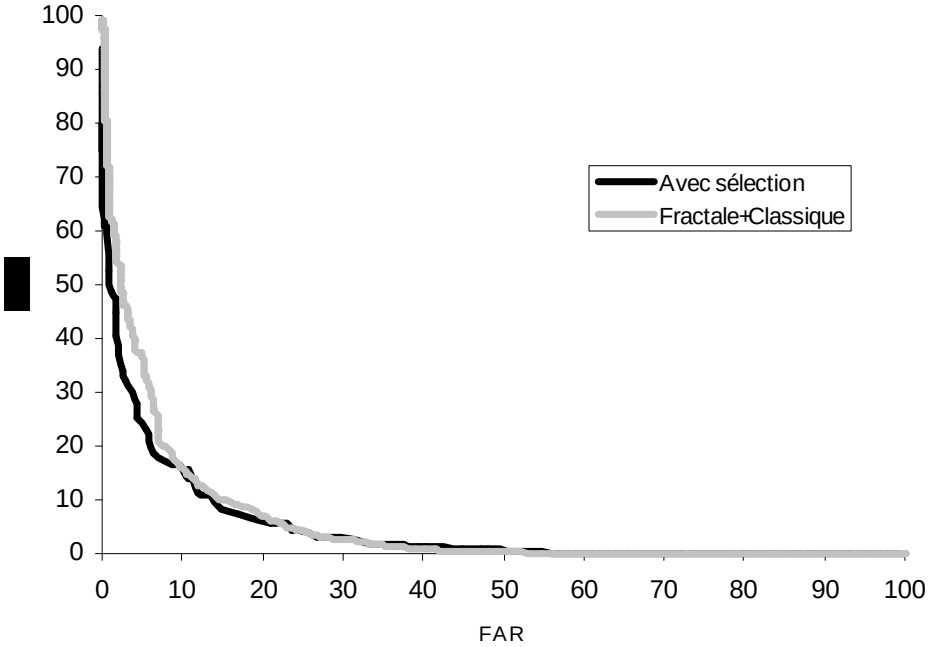


Figure 45. Comparaison des performances obtenues avec et sans sélection des caractéristiques par algorithme génétique multiobjectifs.

La Figure 45 montre l'amélioration apportée aux résultats de vérification par signature par l'utilisation d'un algorithme génétique. Nous avons aussi constaté que l'augmentation du nombre de caractéristiques n'est pas nécessairement une bonne chose d'où l'intérêt de la phase de sélection. Ainsi, le meilleur résultat, FAR=9,5% et FRR=7%, est obtenu avec les 5

caractéristiques suivantes : dimension fractale, dimension vecteur, somme des angles, distance entre le premier et le dernier point, et la dimension de masse. Parmi les caractéristiques retenues, plus de la moitié sont des caractéristiques novatrices.
Notre sélection par algorithme génétique indique également (Figure 45) que les résultats sont meilleurs avec 5 caractéristiques qu'avec 14 (fractale + classique).

4.3.5. Bilan

Concernant les nouvelles caractéristiques que nous proposons, on peut noter que les 3 dernières dimensions sont spécifiques aux signatures en ligne car elles considèrent la signature comme une séquence de points alors que la première nécessite une reconstruction de la signature. Le principal désavantage de la dimension fractale est qu'elle requiert un temps de calcul plus long que la majorité des caractéristiques traditionnelles. Cela est dû à l'aspect itératif de la méthode fractale. Cependant le nombre d'itérations peut être réduit car les droites de régression peuvent être obtenues à partir de 4 itérations.
Concernant la sélection de caractéristiques avec l'aide de l'algorithme génétique multi objectifs, l'inconvénient est qu'il faut disposer d'une base conséquente pour pouvoir la mettre en œuvre. Plus la base sera importante, plus les résultats pourront être considérés comme fiables, proches de la réalité. Le principal intérêt de ces méthodes est de diminuer le nombre de caractéristiques utilisées sans pour autant réduire les performances en terme d'authentification. Cette étape de sélection ne peut se faire que pendant la phase de conception et pas de façon personnalisée. En effet, on ne dispose pas d'une base suffisamment conséquente de signatures au moment de l'enregistrement. Cependant on peut envisager de supprimer les caractéristiques dont la variance est très grande.
Un autre point à souligner est que les caractéristiques utilisées sont, pour la plupart, liées à la forme de la signature et par conséquent permettent de détecter principalement les faux aléatoires. Les deux dernières dimensions fractales peuvent permettre de détecter les faux expérimentés mais il est nécessaire d'ajouter des caractéristiques liées à la dynamique aux caractéristiques actuelles. Notre principal objectif de validation de nouvelles caractéristiques extraites à partir de la forme de la signature et liées à la dimension fractale a été réalisé avec succès. Nous avons aussi démontré qu'il n'est pas toujours intéressant d'augmenter le nombre de caractéristiques descriptives. Certes, de nombreuses améliorations sont encore possibles, les résultats obtenus étant encourageants.

CHAPITRE 3 – NOUVELLES METHODES DE COMPARAISON

De nombreux articles sur l'authentification des signatures en ligne mentionnent les bonnes performances de l'algorithme DTW, nous avons pensé pouvoir compléter l'approche précédente en utilisant ce type d'approche. Notre étude est basée sur un calcul de dissimilarités entre la signature testée et la (ou les) signature(s) d'apprentissage.

La première partie de ce chapitre est consacrée aux améliorations apportées à cette méthode souvent utilisée dans le cadre de l'authentification par signature manuscrite. La seconde partie présente différentes approches pour prendre en compte les spécificités propres à chaque signataire tant au niveau du modèle de référence que du seuil de similarité.

1. AMÉLIORATIONS DE DYNAMIC TIME WARPING

L'algorithme DTW comporte plusieurs étapes, d'abord la mise en correspondance de deux suites, ensuite le calcul de la distance entre ces séquences. Nous aborderons successivement ces deux aspects.

1.1. Mise en correspondance

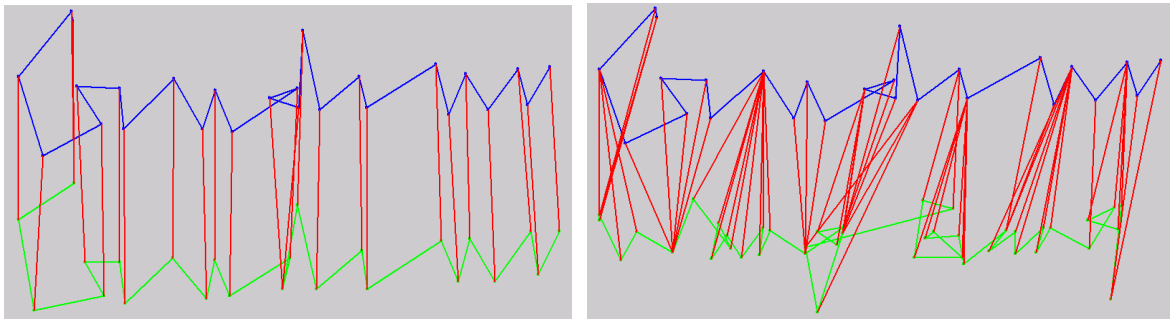
Le but de l'algorithme Dynamic Time Warping est de rechercher la meilleure mise en correspondance entre deux ensembles de points ordonnés. Nous étudions, dans les deux paragraphes suivants, différentes possibilités de codage des courbes correspondant aux signatures afin de voir l'impact de ces codages sur la mise en correspondance et ainsi déterminer celui permettant d'avoir la mise en correspondance la plus proche de celle que l'on pourrait faire visuellement [WIR04b].

Rappelons qu'avant de commencer la mise en correspondance, les signatures sont normalisées et lissées grâce à la méthode de suppression des points "inutiles" présentée au paragraphe 2.2. La réduction du nombre de points influe, à notre avis, grandement sur les résultats pouvant être fournis par l'algorithme DTW puisque celui-ci cherche à mettre en correspondance les points de chaque courbe. Plus le nombre de points est faible, plus il semble aisé de les mettre

en correspondance. Dans cette partie, nous utilisons donc toujours les représentations des signatures obtenues après réduction (lissage) du nombre de points représentant les signatures.

1.1.1. Approche points

Il est tout d'abord possible d'utiliser simplement les coordonnées des points retenus comme représentatifs pour effectuer la mise en correspondance des signatures grâce au DTW. La qualité de la mise en correspondance sera évaluée en comparant la mise en correspondance réalisée de manière automatique à une correspondance réalisée visuellement. La Figure 46 montre le résultat; d'une part, de la mise en correspondance automatiquement obtenue de deux signatures authentiques et, d'autre part, d'une signature authentique et d'un faux aléatoire. Le premier résultat est vraiment bon car il est très proche du résultat attendu. Quant au second résultat, il est très différent du précédent puisque de nombreux points de la signature authentique ont plus d'une correspondance avec les points de la fausse signature. Cet exemple montre que les mises en correspondance multiples d'un point peuvent permettre de définir un nouveau critère pour la détection de faux.

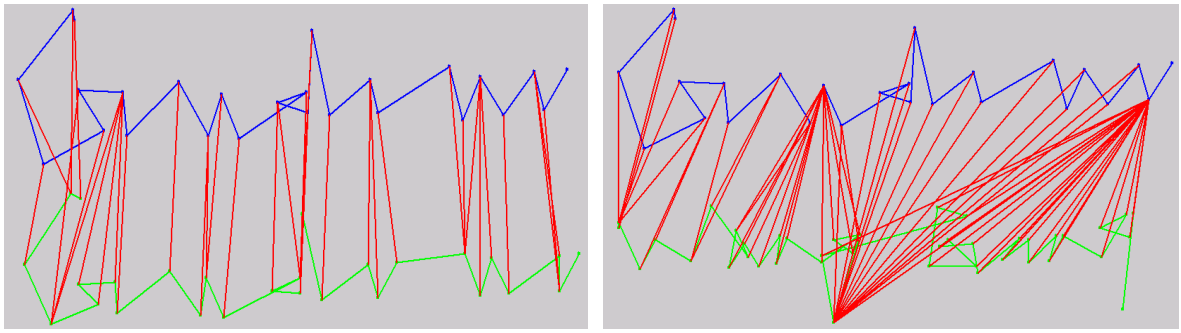


(a) (b)

Figure 46. Illustrations de la mise en correspondance en utilisant les coordonnées spatiales (a) de deux signatures authentiques et (b) d'une signature authentique et d'un faux.

1.1.2. Approche segments : Longueur, durée ou angle

Lorsque les signatures sont considérées comme une suite de segments (et non plus comme une succession de points), d'autres paramètres peuvent être utilisés dans le même but de mise en correspondance de signatures. Comme on peut l'observer sur la Figure 47, le résultat de la mise en correspondance entre deux signatures authentiques en utilisant DTW uniquement sur la longueur des vecteurs n'est pas celui attendu. Le résultat de la mise en correspondance est très différent de celui réalisé par un humain.



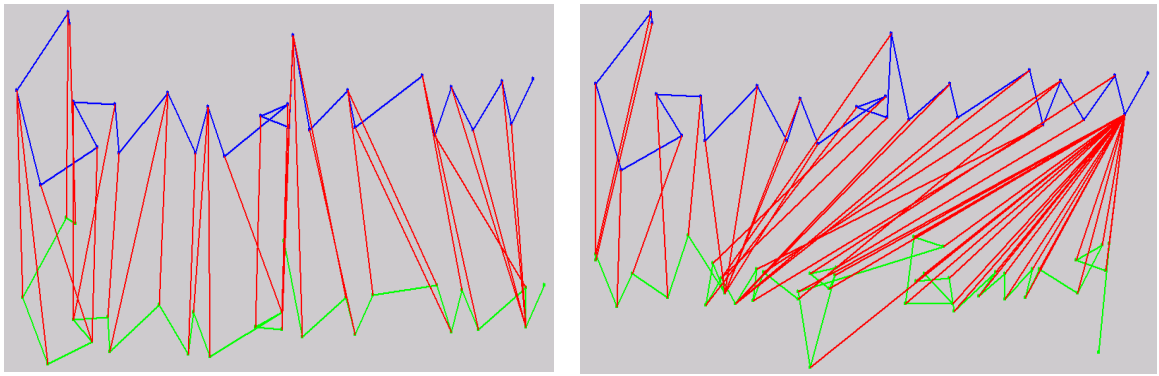
(a) (b)

Figure 47. Illustrations de la mise en correspondance en utilisant la longueur des vecteurs (a) de deux signatures authentiques et (b) d'une signature authentique et d'un faux.

Ce médiocre résultat nous montre que l'information de longueur n'est pas suffisante pour comparer deux vecteurs. En représentant un vecteur seulement par sa longueur, nous avons perdu trop d'information.

La seconde approche consiste à utiliser la durée du tracé entre les deux extrémités du segment comme seule information disponible. Le résultat de cette approche est similaire au précédent. La durée n'apporte pas assez d'information pour effectuer la mise en correspondance entre les deux signatures. Comme précédemment, les erreurs de la mise en correspondance sont dues au fait qu'en considérant uniquement un critère, beaucoup de vecteurs ont plusieurs correspondants possibles.

Dans une troisième approche, nous utilisons la mesure absolue de l'angle entre le vecteur et l'horizontale comme seule information. Dans ce cas, le résultat est meilleur que dans les deux précédentes approches car ce critère semble plus représentatif ou plus variable que les autres. Comme on peut le voir sur la Figure 48, la direction du tracé est plus significative que la longueur des segments au sein des signatures en ligne.



(a) (b)

Figure 48. Illustrations de la mise en correspondance en utilisant l'angle absolu des vecteurs (a) de deux signatures authentiques et (b) d'une signature authentique et d'un faux.

Des erreurs restant présentes par rapport au résultat souhaité, on peut en conclure que la meilleure mise en correspondance est obtenue en utilisant uniquement les coordonnées des sommets de l'approximation polygonale de la signature. En fait, utiliser les coordonnées de ces points est très similaire à la réalisation de la fusion de différentes données utilisées dans chacune des approches étudiées ensuite. Ainsi dans la suite, nous utilisons la mise en correspondance à partir des coordonnées des points sélectionnés durant la phase de prétraitement.

Les tests réalisés en considérant la signature comme une suite de segments (angle, longueur, durée) pour calculer la distance entre les signatures n'ont pas donné de bons résultats. Le meilleur résultat est obtenu en considérant le rapport entre la durée de deux segments consécutifs. Nous obtenons une valeur de EER égale à 20% et une valeur de FAR de 37,5% pour un taux de FRR égal à 2%. Ce résultat souligne l'importance de considérer les coordonnées des points dans le calcul de la distance car il existe un grand nombre de vecteurs proches si on utilise uniquement les caractéristiques correspondant aux vecteurs.

1.2. Mesure de dissimilarité entre signatures

Habituellement, la dissimilarité entre les signatures est la somme des distances euclidiennes entre les points en correspondance dans les deux signatures.

Soient S1 et S2 deux signatures, et nbCorresp le nombre de points en correspondance dans S1. On note Pt_i un point quelconque de S1 et P'_i l'ensemble des points correspondants dans S2. La distance entre les deux signatures S1 et S2 est donnée par la formule suivante :

$$dist(S1,S2) = \sum_{i=1}^{nbCorresp} \sum_{P' \in P'_i} DistPt(Pt_i, P')$$

1.2.1. Normalisation

Dans un premier temps, nous avons choisi de normaliser la distance en divisant par le nombre de correspondances afin d'être indépendant du nombre de points de la signature. La formule de la distance devient donc :

$$Dist(S1,S2) = \frac{1}{\sum_{i=1}^{nbCorresp} |P'_i|} dist(S1,S2)$$

Le résultat peut être vu comme la moyenne des distances entre points mis en correspondance dans les deux signatures.

1.2.2. Utilisation d'informations locales

Lors de la mise en correspondance des points de deux signatures, dans le but d'améliorer encore la signification de la distance entre deux points ou vecteurs au sens du DTW, nous avons choisi d'utiliser des informations différentes de l'approche classique qui tient compte uniquement des coordonnées.

Nous avons exploré d'autres voies pour calculer la distance entre points ou vecteurs en ajoutant une information locale complémentaire au lieu d'utiliser à nouveau les coordonnées. Là aussi, trois types d'information sont considérés : la longueur, la durée et l'angle des vecteurs constitués par les couples de points, pour comparer les signatures. Premièrement, l'information locale utilisée est le rapport r_{ik} des longueurs de deux vecteurs consécutifs $s_{i-1,k}$ et $s_{i,k}$ pour une signature Sk. Nous définissons la distance entre les vecteurs $s_{i,1}$ et $s_{j,2}$ mis en correspondance comme la différence en valeur absolue des deux rapports r_{i1} et r_{j2}. Deuxièmement, l'information locale utilisée est le rapport des durées de deux segments consécutifs. La distance est calculée comme précédemment. Troisièmement, l'information locale ajoutée est l'angle entre le segment précédent et le segment suivant. La distance entre deux segments est la différence entre les mesures des deux angles. Tous les angles ont des mesures comprises entre 0 et 360 degrés.

D'autres tests ont été effectués en utilisant des informations liées à la dynamique comme la vitesse instantanée aux points sélectionnés mais aucun résultat concluant n'a été obtenu.

L'utilisation de la distance de Mahalanobis a été envisagée dans le but d'améliorer la signification de la distance entre points en prenant en compte la variabilité en chaque point de la signature d'apprentissage mais les résultats n'ont pas été plus concluants. Pour calculer cette distance de Mahalanobis, une étape d'initialisation est nécessaire. Pour cela, on effectue tout d'abord la mise en correspondance classique des deux courbes. Puis on mesure la variation suivant x et suivant y pour chaque point.

$$d_{Mahalanobis}(Pt_i, Pt'_j) = \sqrt{\left(\frac{x_i - x'_j}{\overline{dx}}\right)^2 + \left(\frac{y_i - y'_j}{\overline{dy}}\right)^2}$$

La principale raison du mauvais résultat est que l'inverse de la matrice de variance covariance est très instable car le déterminant de cette matrice est proche de 0.

Les conclusions de cet ensemble d'expérimentations sont encore une fois que la méthode la plus simple, c'est à dire l'utilisation des coordonnées des points après réduction/lissage, donne les meilleurs résultats car conserve le maximum d'information.

1.2.3. Prise en compte de l'appariement

Afin de minimiser le rejet de signatures authentiques, nous proposons d'apporter une autre modification au calcul de la dissimilarité entre signatures. Habituellement, dans le cas de plusieurs correspondances pour un point d'une des signatures, toutes les distances entre ce point et ces correspondants sont ajoutées. Nous pensons qu'il est préférable de calculer uniquement la distance entre les premiers points se correspondant afin de ne prendre en considération qu'une seule fois chaque appariement et d'éviter ainsi un cumul d'erreurs. Cette modification de l'algorithme de DTW est illustrée dans la Figure 49.

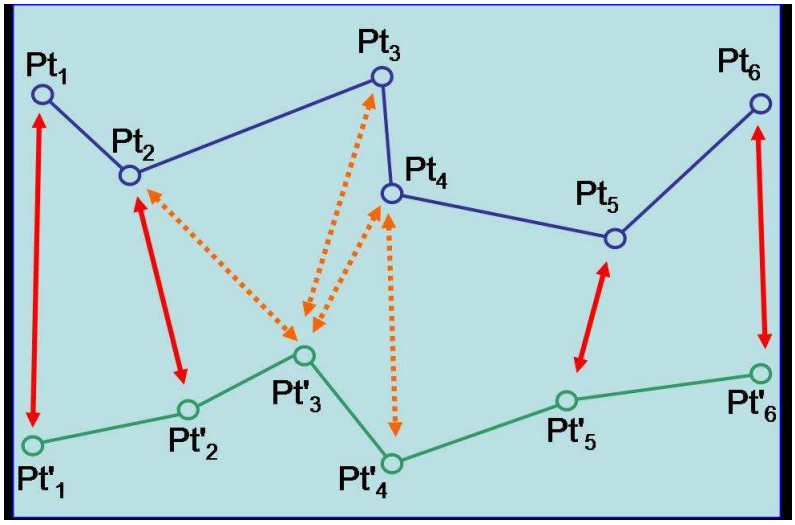


Figure 49. Illustrations des distances prises en considération : en rouge, les distances comptabilisées dans le calcul de la dissimilarité entre les signatures.

Nous précisons P'i={Pt'$_{i1}$, Pt'$_{i2}$, …} où les points sont indicés suivant les temps croissants. Ainsi la formule de distance est modifiée et devient :

$$dist(S1,S2) = \sum_{i=1}^{nbCorresp} DistPt(Pt_i, Pt'_{i,1})$$

Et la formule de la distance normalisée devient :

$$Dist(S1,S2) = \frac{1}{nbCorresp} dist(S1,S2)$$

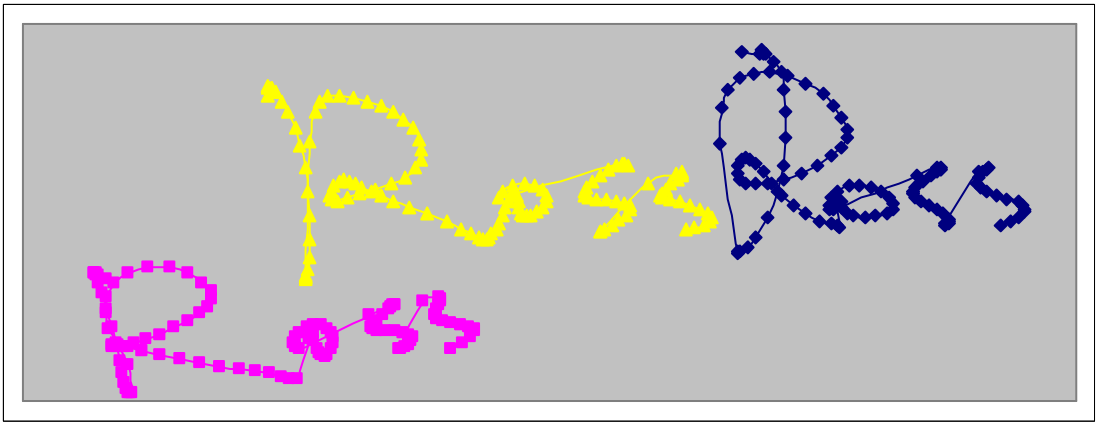


Figure 50. Illustration des erreurs de classification à partir de la distance spatiale.

Sur la Figure 50, trois signatures authentiques d'un même utilisateur sont représentées. On constate que la signature en bleu est différente des deux autres sur le plan spatial notamment au niveau du premier trait. Pour cette signature, le rejet de la signature d'un point de vue spatial semble donc logique. Pour pallier ce type de problème, nous avons essayé de calculer de différentes manières la mesure de dissimilarité entre deux signatures.

1.2.4. Prise en compte de la variabilité intra scripteur

Afin de prendre en compte la variabilité propre à chaque signataire pour améliorer les performances de l'authentification, l'approche consiste à utiliser la seule information disponible dans un système réel, à savoir l'ensemble des signatures d'apprentissage. Le but est de réduire le rejet de signatures authentiques lorsque la variabilité intra scripteur est grande et d'augmenter le rejet de faux lorsque la variabilité intra scripteur est faible.

Pour évaluer cette variabilité intra scripteur, on calcule la distance entre les signatures d'apprentissage Sa_i et on détermine la plus petite distance intra apprentissage *minDistApp*.

$$\boldsymbol{minDistApp = \min_{i,j,i\neq j} Dist(Sa_i, Sa_j)}$$

La formule de la distance entre deux signatures prenant en compte la variabilité intra scripteur devient donc :

$$DistVar(S1, S2) = Dist(S1, S2) - minDistApp$$

D'autres solutions ont été envisagée pour évaluer l'importance de la variabilité : la moyenne des distances intra apprentissage et le maximum des distances intra apprentissage. Mais ces valeurs risquaient d'être trop importantes dans le cas où la variabilité serait importante et le risque était par conséquent d'accepter trop de faux.

L'avantage de cette méthode est que l'on peut ensuite comparer les distances obtenues pour différents scripteurs à un même seuil puisque la prise en compte de la variabilité se fait au niveau du calcul de la distance. De plus, pour estimer cette variabilité, nous n'avons utilisé que des informations dont on dispose dans un cas réel de mise en place d'un processus de vérification, c'est-à-dire les signatures d'apprentissage

1.2.5. Distance Temporelle

Parmi les méthodes de sélection des points représentatifs présentées au paragraphe 2.2, aucune n'utilise d'informations temporelles globales. Seule la méthode basée sur les minimums de la vitesse utilise une information temporelle locale. De plus, les coordonnées des points retenus sont déjà utilisées pour évaluer la qualité de la mise en correspondance. Par conséquent, afin de prendre en considération l'aspect temporel global de la signature, au lieu de calculer une distance euclidienne entre les coordonnées des points, nous avons utilisé le temps écoulé entre le premier point et chaque point considéré. Les avantages de prendre en compte une information temporelle globale plutôt qu'une information locale, comme la vitesse instantanée, sont une moins grande sensibilité, d'une part, aux variations intra scripteur et, d'autre part, au bruit dû à l'acquisition.

Avant de pouvoir comparer ces paramètres entre différentes signatures, une normalisation linéaire du temps sur l'intervalle [0,1] est effectuée. Afin d'éviter des erreurs dues à la normalisation, cette distance sera calculée uniquement si la différence entre la durée totale de la signature testée et des signatures d'apprentissage n'est pas trop importante. La formule pour le calcul de la distance entre deux points Pt et Pt' est donnée ci-dessous :

$$DistPt_{temporelle}(Pt, Pt') = \left| t(Pt) - t(Pt') \right|$$

La Figure 51 illustre les résultats obtenus lorsque l'on utilise une distance temporelle plutôt que spatiale selon la méthode de sélection des points représentatifs utilisée. Cette figure est à comparer avec la Figure 31 récapitulant les résultats obtenus en utilisant l'algorithme DTW classique.

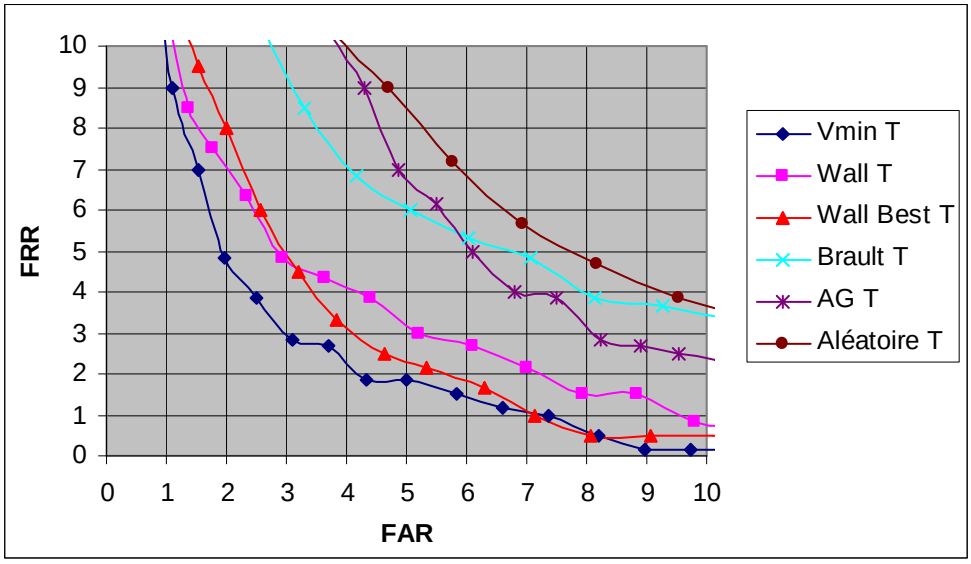


Figure 51. Résultats obtenus avec la distance temporelle pour chacune des méthodes présentées (T=Temporelle).

Pour l'ensemble des méthodes de sélection de points, nous observons une amélioration significative des performances présentées Figure 31. Toutefois, l'ordre des méthodes est conservé. La méthode basée sur les minimums de la vitesse donne les meilleurs résultats. La valeur de EER pour cette méthode est de 3% et la valeur de FAR de 4,3% pour une valeur de FRR égale à 2%. Si l'on considérait l'ensemble des points de la signature, on obtiendrait une valeur de EER égale à 4,2% et une valeur de FAR égale à 5,8% pour une valeur de FRR égale à 2%. La référence aux seuls points retenus permet donc d'améliorer les performances d'authentification lorsqu'on considère la dimension temporelle.

Si l'on compare ces résultats avec ceux obtenus lors de l'utilisation des caractéristiques globales (dimensions fractales + classiques), après une sélection des plus pertinentes, une valeur de EER beaucoup plus importante (proche de 12%) était obtenue (Figure 45). L'approche par DTW en considérant une distance temporelle permet d'améliorer les résultats très nettement, quelle que soit la méthode de sélection de points utilisée, comme le montre la Figure 51.

1.2.6. Distance Curviligne

Afin de compléter l'information apportée par la distance temporelle, une mesure de dissimilarité a été mise en place permettant de prendre en compte l'efficacité du tracé. En effet, pour une même durée, la longueur du tracé peut être très variable suivant la vitesse d'exécution. La distance curviligne nous permet donc de comparer le rythme du tracé entre deux signatures. Le principe de calcul consiste à comparer la distance parcourue depuis le premier point d'acquisition jusqu'au point considéré.

Une normalisation linéaire de la longueur sur l'intervalle [0,1] est effectuée pour pouvoir comparer des signatures n'ayant pas des longueurs identiques. Afin d'éviter des erreurs dues à la normalisation, cette distance sera calculée uniquement si la différence entre la longueur totale de la signature testée et des signatures d'apprentissage n'est pas trop importante.

Après avoir effectué la mise en correspondance des points sélectionnés en se basant toujours sur les coordonnées des points, une distance entre les signatures est calculée suivant la formule suivante :

$$Dist(S1,S2) = \sum_{i=1}^{n} \left| \frac{Longueur(Pt_i)}{Longueur(S1)} - \frac{Longueur(Pt'_i)}{Longueur(S2)} \right|$$

où $Longueur(Pt_i) = \sum_{j=2}^{i} Dist(Pt_{j-1}, Pt_j)$.

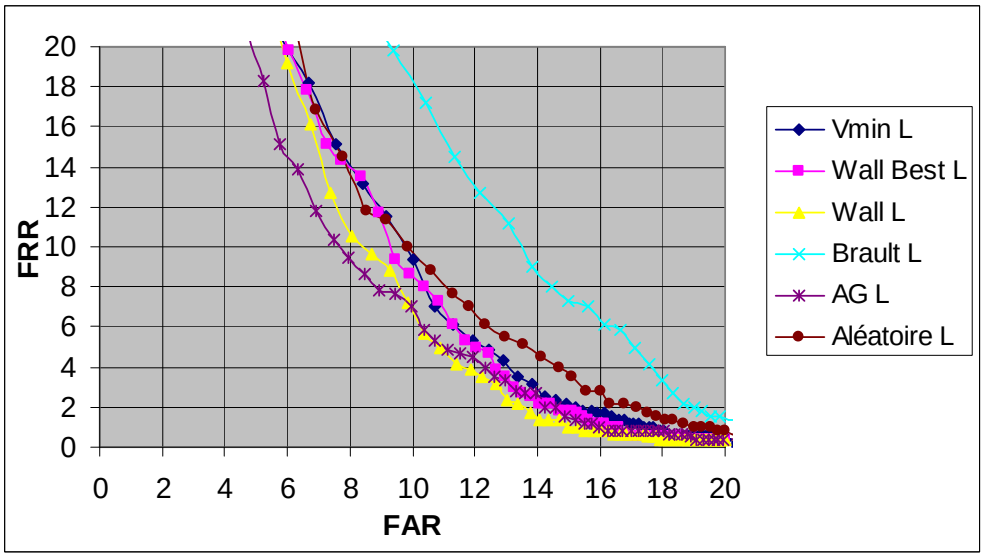


Figure 52. Résultats obtenus avec la distance curviligne pour chacune des méthodes présentées (L=Curviligne).

Les résultats obtenus en considérant la distance curviligne (Figure 52) sont nettement moins bons que ceux obtenus en considérant la distance spatiale ou temporelle. Avec cette distance, la valeur de EER est proche de 10% en considérant les points de vitesse minimum et de 8% en considérant les points sélectionnés par algorithme génétique.

1.2.7. Combinaisons des distances

Les trois distances spatiale, temporelle et curviligne utilisant des informations différentes, il nous a paru intéressant de voir si elles étaient complémentaires. Nous avons donc modifié le calcul de la dissimilarité entre deux signatures de sorte qu'elle prenne en compte ces trois distances en les combinant.

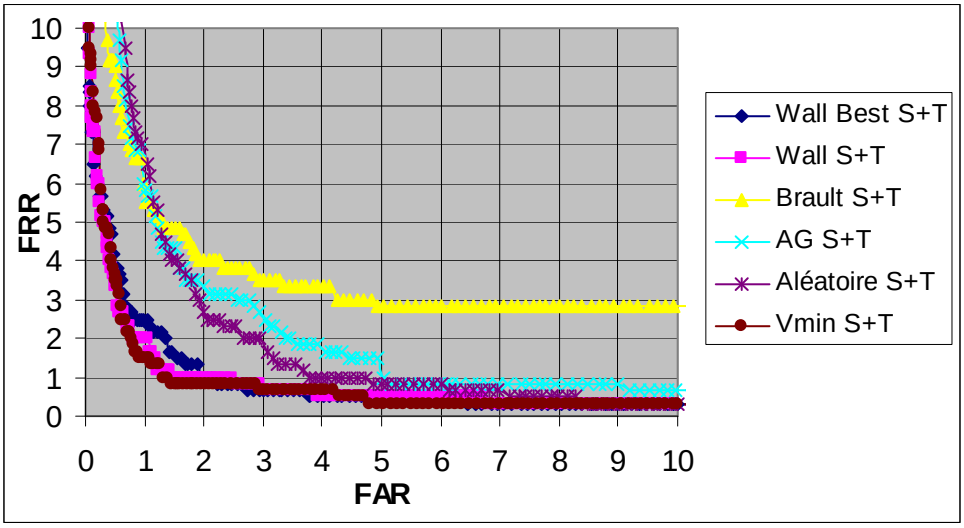


Figure 53. Résultats obtenus avec la combinaison des distances spatiale et temporelle pour chacune des méthodes présentées.

Comme on peut le constater sur la Figure 53, la combinaison des deux distances spatiale et temporelle permet d'améliorer les résultats de chacune des méthodes sauf pour la méthode utilisant les algorithmes génétiques. Ceci peut s'expliquer par le fait que les points retenus par l'algorithme génétique l'ont été suivant un critère uniquement spatial ce qui n'implique pas une stabilité temporelle. Comme précédemment, la méthode de réduction de points basée sur les minimums locaux de la vitesse permet d'obtenir les meilleurs résultats. La valeur de EER pour cette méthode est de 1,3%.

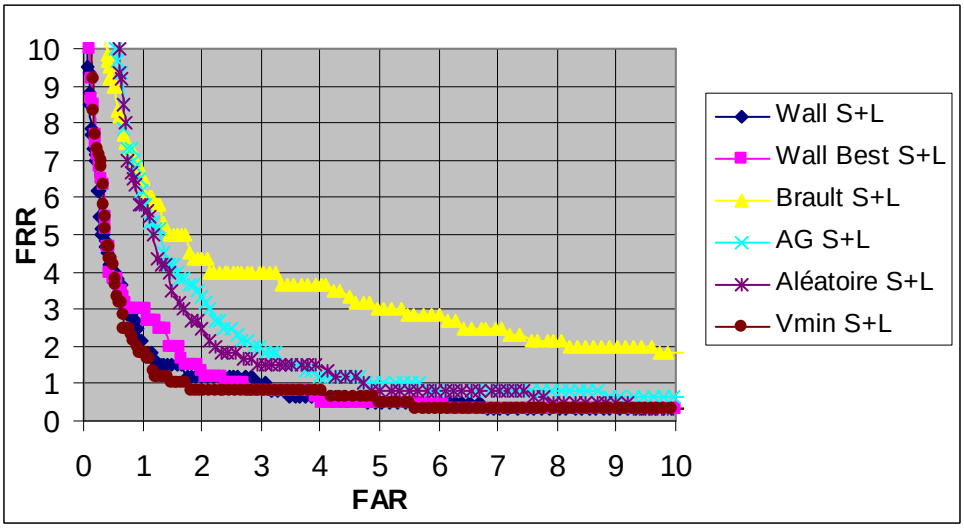


Figure 54. Résultats obtenus avec la combinaison des distances spatiale et curviligne pour chacune des méthodes présentées.

Alors que les performances obtenues avec la distance curviligne étaient nettement inférieures à celles obtenues avec la distance spatiale, la combinaison des deux distances permet

d'améliorer les résultats obtenus en considérant uniquement la distance spatiale (Figure 54). La valeur de EER est réduite de 1,6% à 1,3%, soit une réduction de 18%.

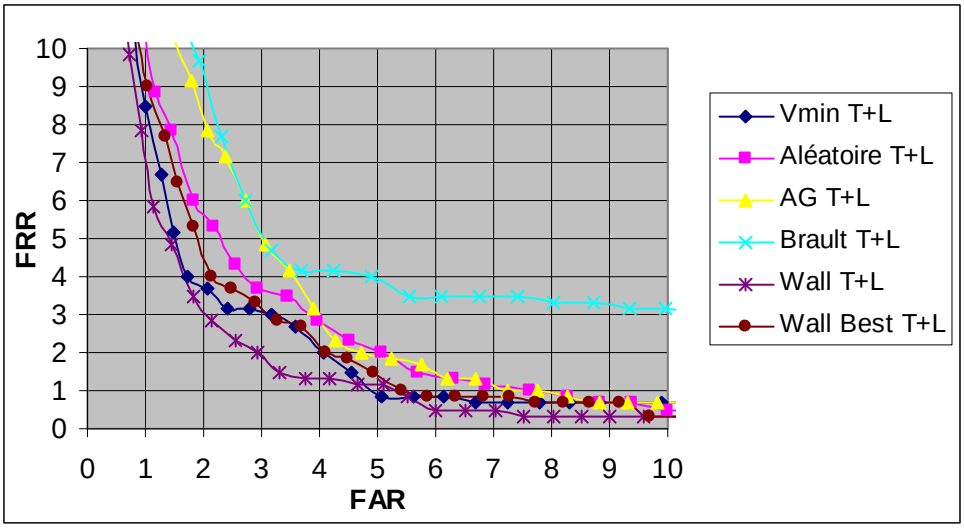


Figure 55. Résultats obtenus avec la combinaison des distances temporelle et curviligne pour chacune des méthodes présentées.

Comme précédemment, la combinaison des distances temporelle et curviligne permet d'améliorer les résultats obtenus en utilisant les distances séparément (Figure 55) même si les performances sont moins bonnes que celles obtenues lorsqu'on utilise la distance spatiale et la distance temporelle. La valeur de EER obtenue est de 2,5%.

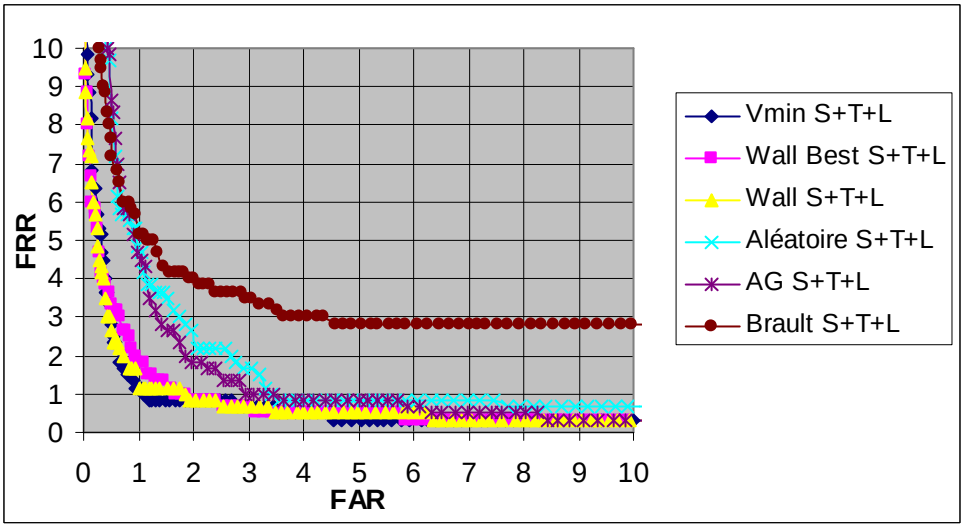


Figure 56. Résultats obtenus avec la combinaison des distances spatiale, temporelle et curviligne pour chacune des méthodes présentées.

La combinaison des trois distances est la combinaison qui permet d'obtenir le meilleur résultat, soit une valeur de EER égale à 1,1% (Figure 56). Cela confirme que les trois distances sont complémentaires et apportent une information différente. Par rapport aux

résultats obtenus en utilisant uniquement la distance spatiale, on constate une amélioration de plus de 30% des performances.

Voici les résultats obtenus en considérant différentes combinaisons des différentes distances :

	Distance Spatiale	Distance Temporelle	Distance Curviligne	Distance S+T	Distance L+T	Distance L+S	Distance S+T+L	Distance S+min(T,L)
EER	1,6%	3,2%	9,5%	1,3%	3,2%	1,3%	**1,1%**	**1,3%**

Tableau 11. Résultats obtenus en testant différentes combinaisons des différentes distances.

Dans un premier temps, les tests étaient effectués avec des poids similaires pour les différentes distances.

Le but de ce dernier test est de rechercher la meilleure combinaison linéaire des différentes distances.

La nouvelle distance est obtenue avec la formule suivante :

$$d = \alpha \times distS + \beta \times distT + \gamma \times distL$$

où $\begin{cases} \alpha + \beta + \gamma = 1 \\ \alpha, \beta, \gamma \in [0,1] \end{cases}$

Le cube de variation des paramètres est exploré systématiquement avec un pas de 0,1.

On observe une légère amélioration des résultats lorsque la combinaison des distances est la suivante :

$$d = 0{,}2 \times distS + 0{,}6 \times distT + 0{,}2 \times distL$$

La valeur de EER passe de 1,1% à 0,9%, soit un gain de l'ordre de 20%.

On constate également que :

- aucune variable utilisée seule n'est meilleure qu'une combinaison des autres variables
- toutes les variables sont nécessaires dans le calcul de la nouvelle variable

La distance temporelle semble être plus discriminante que les variables liées à la forme de la signature.

Les résultats présentés jusqu'à présent ont été obtenus sur la base SVC. Afin de confirmer ces conclusions, des tests dans les mêmes conditions ont été effectués sur deux autres bases (ATOS, MCYT). Ces résultats sont présentés chapitre 4, paragraphe 4.3.

2. CHOIX DU SEUIL ET ADAPTATIVITE AUX SCRIPTEURS

Nous n'avons présenté jusqu'à présent que des courbes liant FAR à FRR mais lors de la mise en place d'un système réel, un seuil devra être choisi sur la mesure de dissimilarité soit pour chaque scripteur soit globalement.

Ainsi le choix de ce seuil de décision dépend de deux facteurs : la variabilité de la signature et le taux de FRR ou de FAR souhaité. L'évaluation du processus d'authentification en identification plutôt qu'en vérification nous a permis d'estimer les performances que l'on pouvait atteindre si on utilisait un seuil adapté à l'utilisateur.

Pour avoir une meilleure estimation du gain apporté par la mise en place d'un seuil adaptatif, nous avons calculé a posteriori le seuil optimal, par minimisation de la somme de FAR et de FRR, pour chaque individu puis la valeur de EER globale obtenue en prenant en compte ce seuil. Etant donné l'importance du gain obtenu en considérant un seuil individualisé, nous avons cherché à déterminer ce seuil adapté à chaque scripteur de manière automatique en considérant comme seule information disponible les signatures d'apprentissage d'un individu.

On cherche à expliquer une variable *s*, ici le seuil optimal, à l'aide d'une combinaison linéaire de *NbVar* variables supposées indépendantes $x^{(1)}$, ..., $x^{(NbVar)}$ à partir de *NbObsv* observations :

$$s = \sum_{i=1}^{NbVar} \alpha_i x^{(i)} + \beta + e$$

où e désigne l'erreur résiduelle (moyenne nulle).

Afin de savoir s'il existe une relation entre les variables $x^{(i)}$ et *s*, une solution possible est de calculer le coefficient de corrélation multiple R2 :

$$R^2 = \frac{R_{YX} R_{XX}^{-1} R_{XY}}{R_{YY}} \text{ avec } R_{XY} = {}^t\tilde{X}\tilde{Y}$$

Si la valeur de cette corrélation est importante, les valeurs des α_i seront estimées à partir des observations de la variable *s* pour des valeurs fixes de $x^{(i)}$.

Le test de significativité globale de la régression mesure la valeur explicative de l'ensemble des variables $x^{(i)}$ sans préciser celle (ou celles) qui apporte(nt) une réelle explication de la variable *s*. L'hypothèse testée est la suivante :

- H : aucune variable $x^{(i)}$ n'explique *s*.

$$F = \frac{NbObsv - NbVar - 1}{NbVar} \times \frac{R^2}{1 - R^2}$$

Si H est vérifiée, alors F suit une loi de Fisher-Snedecor à n1 = NbVar et n2 = NbObsv-NbVar-1 degrés de liberté.

Si F est supérieur à la valeur lue dans la table de Fisher-Snedecor, alors l'hypothèse précédente est rejetée et on en conclut qu'il existe au moins une variable $x^{(i)}$ explicative de *s*.

2.1. Recherche d'une corrélation linéaire entre le seuil optimal et les caractéristiques de la signature

Une recherche de corrélation entre les caractéristiques de la signature et le seuil optimal de chaque individu a été effectuée. Les caractéristiques retenues pour caractériser la signature sont la longueur, la durée et la vitesse moyenne car ces caractéristiques ont une forte stabilité intra scripteur ainsi qu'un pouvoir discriminant élevé.

Les variables utilisées sont :

- la longueur moyenne de la signature sur l'ensemble d'apprentissage divisée par la variance de la longueur au carré
- la vitesse moyenne de la signature sur l'ensemble d'apprentissage divisée par la variance de la vitesse au carré
- la durée moyenne de la signature sur l'ensemble d'apprentissage divisée par la variance de la durée au carré

Les résultats obtenus étant moins bons que ceux obtenus avec un seuil fixe, d'autres corrélations ont été recherchées, comme par exemple entre la vitesse moyenne et le seuil optimal.

Aucune corrélation (<0,2) n'existe entre ces variables et le seuil optimal déterminé pour chaque utilisateur (le test de significativité est inférieur à 3 alors que $F_{3,36,99\%}$=4,38).

2.2. Recherche d'une corrélation entre le seuil optimal et les distances intra apprentissage

Afin de caractériser l'ensemble d'apprentissage de chaque utilisateur, nous avons calculé 3 valeurs : la moyenne, le minimum et le max des distances intra signatures d'apprentissage. Ensuite, une corrélation entre chacune de ces valeurs et le seuil optimal a été recherchée.

Les variables utilisées sont :

- la distance moyenne entre les signatures d'apprentissage
- la variance de la distance entre les signatures d'apprentissage

Pour évaluer la qualité de la corrélation, un seuil estimé est calculé pour chaque utilisateur et la méthode d'authentification est évaluée en considérant ce seuil.
La valeur de la corrélation reste là aussi relativement faible (<0,5) mais la significativité du test est plus importante. Dans le cas où l'on utilise les points correspondant aux minimums de vitesse on obtient une corrélation de l'ordre de 0,3 et une valeur de significativité égale à 8,14 supérieure à $F_{2,37,99\%}$=7,37. Ce résultat indique donc que le seuil optimal peut être expliqué par au moins une des deux variables utilisées.
Le seuil optimal ne semble donc pas lié à la forme ou à ses variations ni à la dynamique ou ses variations mais plutôt à la distance entre les signatures d'apprentissage.

2.3. Signatures problématiques : instabilité forte

La plupart des auteurs font état de personnes pour lesquelles le système d'authentification ne fonctionne pas. Certaines signatures de personnes sont en effet trop inconstantes pour établir un modèle. Les personnes concernées sont essentiellement des personnes qui ne signent pas régulièrement, ainsi les personnes de moins de 20 ans et les catégories socioprofessionnelles qui n'utilisent pas la signature dans le cadre de leur travail.
Il semble qu'une grande instabilité de la signature soit liée à une grande variance dans le temps total nécessaire pour réaliser une signature. Pour [HER77], la variation de temps entre 2 signatures successives d'un même signataire est inférieure à 10ms. Un traitement particulier devra être étudié pour gérer ce type de problème. Nous n'avons pas eu suffisamment de temps pour l'aborder dans le cadre de cette thèse.

3. BILAN

Nous avons montré au travers de différentes expérimentations des approches proposées qu'il est possible d'améliorer la mise en correspondance déduite de DTW. Nous observons que la meilleure mise en correspondance est obtenue en utilisant les coordonnées des points sélectionnés par analyse de la vitesse instantanée. Mais la mise en correspondance pourrait être améliorée dans le cas d'une succession de plusieurs petits segments. En effet, si les coordonnées des points sont proches, des erreurs peuvent plus facilement apparaître. La polygonalisation de la signature permet de réduire cet effet de bord en réduisant le nombre de points et en les espaçant. Donc, prendre en compte l'information locale peut permettre de réduire les erreurs de mise en correspondance. Lorsque nous considérons la qualité de la mise en correspondance, avant de calculer la distance, il est possible de détecter des faux que l'on ne pourrait pas distinguer en considérant la distance globale. La mesure de la qualité de la

mise en correspondance peut être plus précise et d'autres mesures peuvent être utilisées. Pour le calcul de la dissimilarité, prendre en compte la différence de longueur, de durée ou d'angle entre deux vecteurs au lieu des coordonnées des points ne permet pas d'améliorer la qualité de l'authentification.

Notre étude montre que les méthodes utilisées traditionnellement pour segmenter la signature peuvent être utilisées pour réduire le nombre de points représentatifs des signatures. En effet, la réduction du nombre de points représentatifs permet de lisser les données initiales et d'éliminer le bruit contenu dans les signatures avant de les comparer, ce qui permet d'améliorer les résultats de la mise en correspondance et donc du système d'authentification. Le gain le plus important de la réduction du nombre de points apparaît lorsque l'on considère la distance temporelle. Les points retenus correspondent donc à des points significatifs et stables de la signature. De plus, la réduction du nombre de points d'une signature permet, d'une part, de limiter l'espace mémoire pour stocker la référence et, d'autre part, de réduire le temps de calcul car les traitements sont effectués sur un nombre plus réduit de points.

L'amélioration des résultats obtenus en combinant les différentes distances – spatiale, temporelle et curviligne – semble confirmer la complémentarité de ces trois approches. Néanmoins, le résultat de la recherche de la combinaison optimale des différentes distances semble indiquer que la distance temporelle est la distance qui permet de discriminer au mieux les signataires.

L'adaptation au scripteur par la mise en place d'un seuil individualisé se révèle relativement complexe lorsque l'on utilise uniquement pour information les signatures d'apprentissage. La recherche de corrélation entre le seuil optimal et les caractéristiques de la signature ou les distances intra apprentissage n'ont pas permis d'en déduire une méthode pour fixer les seuils individualisés.

CHAPITRE 4 - MISE EN ŒUVRE ET PERFORMANCES

Dans ce chapitre, nous présentons tout d'abord l'environnement dans lequel s'intégrera le module d'authentification par signature manuscrite en ligne que nous proposons, notamment par une brève description du fonctionnement des certificats électroniques. Puis nous présentons les résultats obtenus en appliquant les méthodes décrites dans les précédents chapitres sur deux nouvelles bases de test : une base non publique constituée au sein de la société Atos Worldline et une base publique constituée en Espagne dans le cadre d'un projet de recherche sur la biométrie multimodale. La dernière partie de ce chapitre décrit les différents éléments qui constituent le prototype réalisé soit sur le PC soit sur le PDA.

1. CONTRAINTES INDUSTRIELLES

1.1. Sécurité

Etant donné que l'on souhaite réaliser un prototype commercialisable, certains choix ont été faits en fonction des contraintes à prendre en compte : temps de traitement, taille de la mémoire nécessaire au stockage de la référence, prise en compte de l'évolution de la signature au cours du temps, ergonomie… Il faut donc faire un compromis entre le degré d'authentification et le temps de calcul maximum (puissance de calcul du support).

Concernant l'acquisition, les données fournies par le dispositif devront être protégées (par chiffrement et horodatage des données par exemple) lors de leur transfert entre le support d'acquisition et le lieu de stockage du modèle de la signature. Le but est d'éviter les interceptions des informations circulant entre le dispositif d'acquisition et le système effectuant l'authentification ainsi que leur utilisation dans des attaques de type rejeu.

Concernant les informations relatives à la signature stockée, la principale contrainte est la suivante : le pattern enregistré ne devra pas être réversible, c'est à dire que les données stockées pour constituer la référence ne devront pas permettre de reconstituer la signature complète. Le but est encore d'éviter les actes de type rejeu.

De manière générale, il faut pouvoir se protéger des attaques de type rejeu. Etant donné qu'une personne ne peut pas reproduire deux fois de façon identique sa signature, une signature identique ou trop proche de la référence ou des signatures précédentes sera rejetée. Pour cela, une solution possible consiste à conserver les temps d'exécution des 10 dernières signatures reconnues comme authentiques et à les comparer au temps de la signature testée. La probabilité qu'une personne effectue deux fois une signature avec le même temps à la milliseconde près est quasiment nulle.

Afin de mieux appréhender le contexte dans lequel s'insère notre travail (réalisation du prototype), nous allons rappeler quelques généralités sur les infrastructures à clés publiques ou PKI qui gèrent l'utilisation des certificats.

1.2. Infrastructures à clés publiques

La cryptographie existe depuis toujours mais elle reposait, jusqu'à un passé récent, sur l'idée qu'un échange protégé d'informations ne pouvait se baser que sur le partage d'une clé secrète commune. En 1976, Whitfield Diffie et Martin Hellman ont décrit la cryptographie à clé publique et notamment comment partager une information secrète sur la base de clés non confidentielles, bouleversant ainsi le paradigme de la cryptographie à tout jamais [DIF76]. Ainsi, on a considéré qu'un utilisateur connaissait la clé publique d'une personne simplement en consultant un annuaire ou un serveur Web et ainsi il la considérait comme vraie.
Malheureusement le problème de l'authentification n'est pas réglé. En effet, comment s'assurer qu'une clé publique appartient bel et bien à la personne que vous allez joindre? Un faussaire peut tenter de vous faire chiffrer de l'information à l'aide de sa clé publique alors que vous croyez qu'elle appartient à votre interlocuteur. Si le subterfuge fonctionne, son auteur pourra lire des informations qui ne lui étaient pas destinées. Aussi le cryptosystème doit permettre l'authentification des clés publiques. Il faut pouvoir être sûr qu'une clé publique appartient bien à celui qui prétend en être le propriétaire, et qui possède par conséquent la clé privée associée. Pour cela et pour permettre d'établir la confiance lors des échanges électroniques, interviennent des tierces parties de confiance ("Trusted Third Party"). Cette opération s'effectue à l'aide d'un certificat émis par un tiers certificateur. Ce certificat est en fait une attestation de l'identité du titulaire de la clé publique à laquelle il se rattache.

1.3. Le certificat

Un certificat est l'équivalent d'une carte d'identité ou d'un passeport. Un passeport contient des informations concernant son propriétaire (nom, prénom, adresse, …), la signature manuscrite, la date de validité, ainsi qu'un tampon et une présentation (forme, couleur, papier) qui permettent de reconnaître que ce passeport n'est pas un faux, qu'il a été délivré par une autorité bien connue. Un certificat électronique contient des informations équivalentes. Le certificat s'appuie sur un protocole normalisé X509 V3 qui permet d'associer à la clé des informations spécifiques à l'entité (physique ou morale) à laquelle elle se rapporte.

La signature électronique contenue dans le certificat est calculée sur les informations contenues dans le certificat. La signature est l'empreinte de ces informations chiffrées avec la clé privée de l'autorité de certification qui a délivré ce certificat. Le mécanisme de vérification du certificat est identique à celui utilisé pour vérifier une signature, c'est la clé publique de l'autorité de certification qui sera ensuite utilisée par le destinataire. Cependant, cette vérification de certificat dépend de la connaissance que l'on a de l'autorité de certification qui l'a émis : la confiance est absolue si on a accès directement à la clé publique de l'autorité, sinon il faut accepter de passer par une autre autorité ayant elle-même certifié la clé publique de l'autorité émettrice du certificat. Le problème réside en la confiance que l'on peut apporter à une autorité de certification disjointe ; en effet la confiance est matérialisée par le certificat or s'il est construit sur des informations subjectives, la certification croisée peut être compromise. Ce format de certificat permet néanmoins une distribution sûre des clés publiques.

La publication des certificats (donc a fortiori des clés publiques) est faite en utilisant des structures d'annuaires de type LDAP (Lightweight Directory Access Protocol). Les certificats révoqués sont regroupés dans des listes : CRL (Certificate Revocation List) qui sont des structures de données signées et dont le format est défini par le protocole X509 V2 CRL ; ce format peut permettre une diffusion des CRL via les annuaires LDAP.

Les listes de révocation doivent d'une part être protégées pour éviter toute corruption, d'autre part être accessibles en permanence et le plus à jour possible (notion de temps réel).

Un certificat numérique doit en général, après émission par une autorité de certification, être stocké et conservé par son propriétaire. Ce stockage pose un certain nombre de problèmes. En effet, il doit offrir une protection contre une utilisation par un tiers, la perte de données liée à l'instabilité du support. Par ailleurs se pose le problème de la standardisation des supports et de leur moyen de lecture.

Nous avons vu le rôle important du certificat dans la cryptographie à clé publique. Nous allons à présent définir plus précisément les infrastructures qui permettent de les gérer.

1.4. L'infrastructure de Gestion de Clés

Une PKI, Public Key Infrastructure, (en français : I.G.C. i.e. Infrastructure de Gestion de Clés) est une structure à la fois technique et administrative permettant une mise en place, lors de l'échange de clés, de relations de confiance entre des entités morales et/ou physiques et/ou logiques. Une Infrastructure de Gestion de Clés offre un environnement de confiance, ainsi qu'un ensemble de garanties et services relatifs aux certificats de clés publiques

La PKI permet la gestion de clés publiques à grande échelle. Quelques exemples de mise en œuvre des PKI pour des applications courantes : S/MIME (Secure/Multipurpose Internet Mail Extensions) pour la messagerie électronique, HTTPS et SSL (Secure Socket Layer) pour les communications Web, SSH (Secure Shell) et SSF (Secure Shell Français) pour des applications interactives, IPSec (IP Security) pour les communications entre équipements (de réseaux ou station).

La PKI est composée de plusieurs modules distincts :

Au niveau des entités matérielles et logicielles, on trouve trois entités (Figure 57). Une autorité de certification (AC), chargée de créer les certificats et les listes de révocation ainsi que d'assurer la publication de ceux-ci. En plus de ce rôle, elle signe et enregistre (dans la base de données) toutes les transactions et erreurs. Une autorité d'enregistrement (AE) qui divise la PKI en domaines d'opérations. Elle est chargée de valider les demandes de création, de renouvellement et de révocation de certificats. Elle route les communications de et vers l'AC. Elle maintient également une base de données signée des log ainsi qu'un annuaire de publications des certificats, voire des listes de révocation. Cet annuaire est le plus souvent au format LDAP. Un opérateur de certification (OC) qui assure la partie technique de la fourniture et de la gestion du cycle de vie des certificats. Son rôle consiste à mettre en œuvre une plate-forme opérationnelle, fonctionnelle, sécurisée, dans le respect des exigences énoncées dans la Politique de Certification (PC) et dont les modalités sont détaillées dans la Déclaration des Pratiques de Certification (DPC). Il est d'ailleurs dépositaire de la clé privée de l'Autorité de Certification

Avant de délivrer un certificat, les autorités de certification sont chargées de faire les vérifications nécessaires sur l'identité des demandeurs de certificats.

L'autorité de certification s'appuie généralement sur deux autres entités qui travaillent par délégations : l'Autorité d'Enregistrement et l'Opérateur de Certification. Cependant elle garde

la responsabilité des procédures et des principes de certification ; c'est elle qui fait appliquer la politique de certification et elle est responsable pour ses utilisateurs du niveau de confiance fourni par l'IGC.

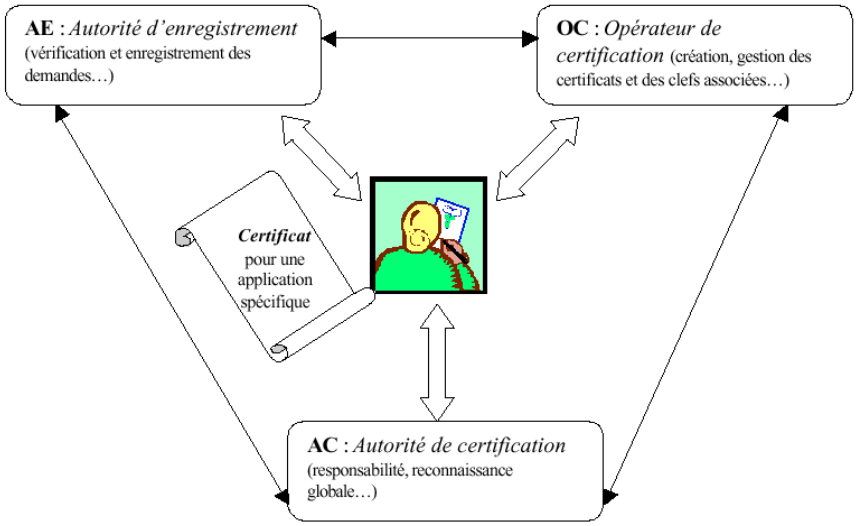


Figure 57. Architecture d'une Infrastructure de Gestion de Clés.

Après cette présentation du contexte du projet, nous présentons la solution mise en place pour renforcer l'authentification du propriétaire du certificat électronique en remplaçant le code PIN utilisé actuellement par une authentification par signature manuscrite en ligne.

2. ACQUISITION DE SIGNATURES EN LIGNE

L'utilisation d'un nouveau support peut nécessiter une période d'adaptation notamment concernant le fait d'écrire avec un stylet sur une surface particulière ou encore le fait d'écrire dans une zone restreinte. Par conséquent, d'une part, il faudrait adapter le nombre de tentatives en fonction de l'évolution de la maîtrise du support et, d'autre part, réactualiser la base de données car avec l'habitude certaines caractéristiques de l'individu peuvent varier comme par exemple la durée totale de la signature.

Pour évaluer l'influence de ce type de biais sur notre méthode d'authentification, il faut pouvoir effectuer des tests en continu sur plusieurs mois et donc disposer d'une base conséquente de signatures.

Ainsi, une base de signatures a été constituée avec l'aide de la société ATOS et de son personnel afin d'évaluer les différentes propositions d'améliorations des systèmes

d'authentification mises en œuvre. De plus, une nouvelle base de signatures (MCYT) a été rendue publique début 2005 et nous a permis de compléter notre expérimentation.

2.1. Présentation de la base de signatures ATOS

Cette base nous a permis de mieux contrôler les protocoles, les dispositifs et l'acquisition. Le protocole utilisé pour l'acquisition des signatures est présenté dans l'Annexe. L'acquisition des signatures s'est faite sur un Tablet PC et sur un PDA suivant le protocole suivant (Figure 58 et Annexe) :

- Familiarisation avec le dispositif de capture (écrire son nom, prénom : mots que l'on écrit de façon automatique)
- Entraînement à la signature sur le dispositif de capture jusqu'à obtenir une signature que la personne reconnaît comme conforme
- Enregistrement du modèle de la signature (5 signatures)
- Enregistrement de signatures de test (5 signatures)

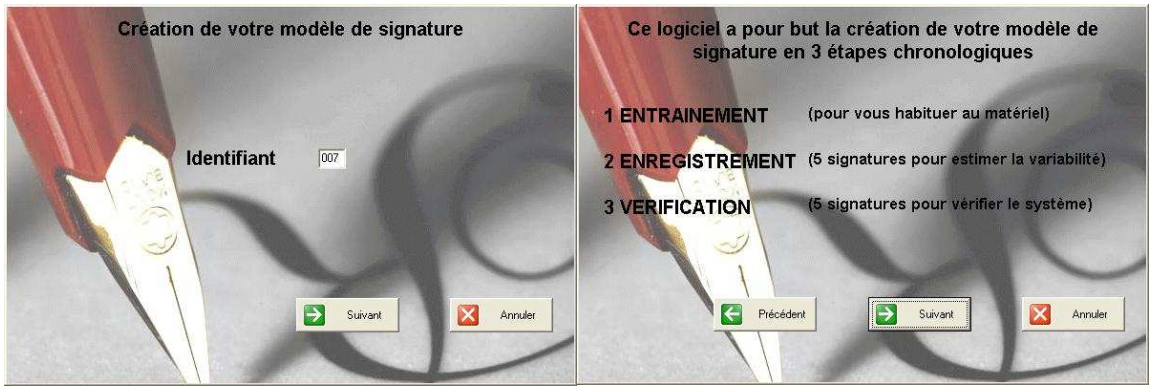


Figure 58. Différents écrans proposés durant la phase d'acquisition des signatures manuscrites sur le Tablet PC.

Nous avons privilégié le PDA car il constitue le support envisagé à terme pour l'application commerciale. En complément, un Tablet PC a également été utilisé afin d'étudier la stabilité des résultats inter supports. La constitution de cette base de signatures nous a permis d'en déduire plusieurs informations importantes.

Tout d'abord, le nombre de signatures pour constituer le modèle de la signature doit être inférieur ou égal à 5. En effet, à partir de 5 signatures, l'attention diminue et l'enregistrement est considéré comme long.

La deuxième information importante est la difficulté de signer sur un support comme le PDA. Trois raisons expliquent cette difficulté. Premièrement, la taille du stylet est relativement petite et fine. Deuxièmement, la taille de l'écran est petite ce qui contraint fortement les personnes ayant une signature de grande taille ou avec des accélérations importantes. Et,

troisièmement, il n'est pas possible d'appuyer le poignet sur le support pour signer. La dernière information que l'on peut retenir est que l'acquisition sur Tablet PC est plus facile que sur PDA. Ceci est dû principalement au fait que cette acquisition est très proche au niveau de l'ergonomie de celle sur papier : le stylet utilisé a la forme d'un stylo et le cadre de saisie de la signature est relativement grand. Enfin, on constate que le contact du stylet avec la surface des deux supports utilisés est plus "glissant" que celui que l'on a l'habitude de connaître lorsque l'on utilise un stylo sur du papier.

Un autre aspect qui plaide plus en faveur du Tablet PC est que l'acquisition effectuée sur un PDA est de moins bonne qualité que sur un Tablet PC. En effet, la résolution étant moins importante sur un PDA, les coordonnées des points de la signature sont moins précises. On peut comparer ce phénomène au problème de la résolution des photos utilisées pour la reconnaissance du visage par exemple. Ainsi, la qualité de l'acquisition a des répercussions importantes sur l'ensemble du processus.

Pour effectuer l'évaluation des différentes méthodes développées, nous avons donc constitué une base de signatures contenant 1000 signatures authentiques réalisées par 100 personnes. Parmi les 10 signatures authentiques de chaque signataire, 5 ont été utilisées pour construire le modèle de la signature et les autres pour les tests. Chaque signataire de la base est représenté par 5 modèles : un pour chaque signature d'apprentissage.

La Figure 59 présente le format du fichier de sortie fourni par le dispositif d'acquisition.

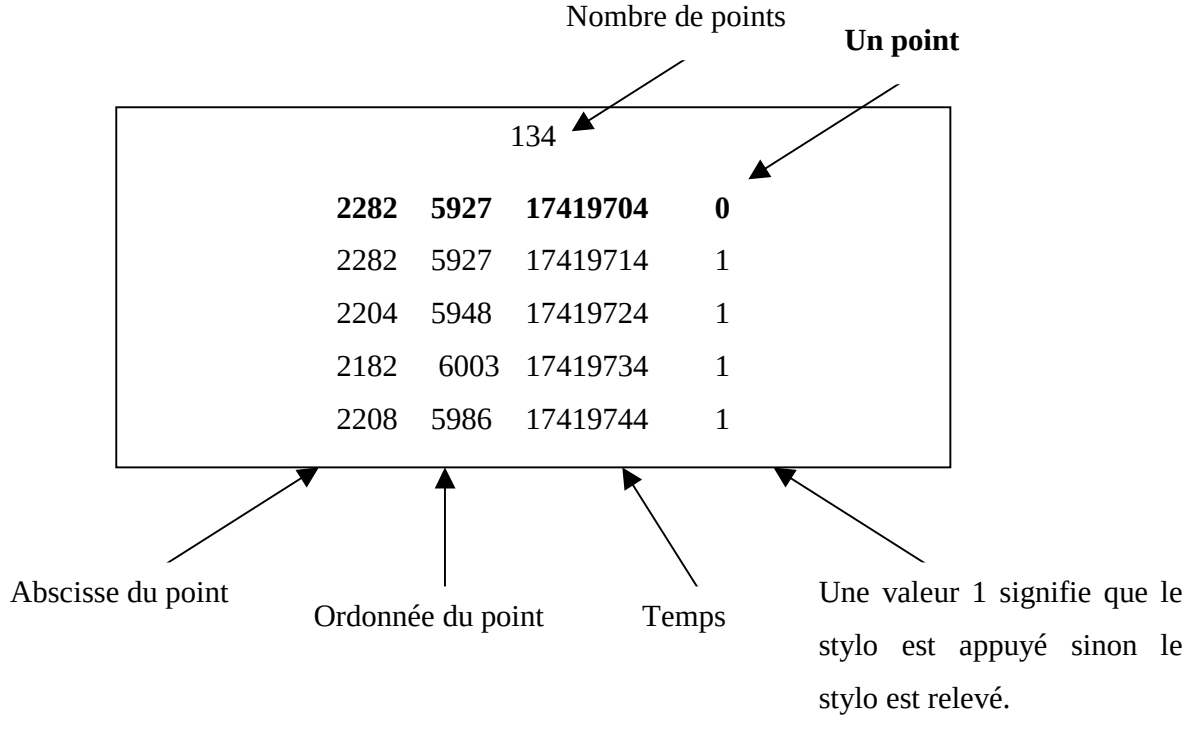


Figure 59. Format des fichiers obtenus au terme de l'acquisition.

2.2. Présentation de la base de signatures MCYT

En complément des bases SVC et ATOS, une autre base de signatures a été utilisée pour valider les méthodes proposées. Cette base, référencée dans la suite par base MCYT (Ministerio de Cienca y Technologia), est une partie de la base biométrique multimodale (empreinte digitale et signature) créée en Espagne [ORT03]. La création de cette base a été motivée par le manque de base de données biométriques publiques et donc par la difficulté de comparer les différents systèmes biométriques entre eux chacun étant testé sur des bases différentes.

L'acquisition de la base de signatures a été faite sur une tablette Wacom. Cinq informations sont récupérées avec une fréquence de 100Hz : x, y, pression, angle du stylet par rapport au plan horizontal et angle du stylet par rapport au plan vertical. Etant donné que l'on ne souhaite utiliser que les coordonnées, le temps et le contact, les fichiers de la base ont été reformatés pour réduire l'information à celle disponible sur tous les supports.

La base contient les signatures de 100 personnes. Pour chaque personne, on dispose de 25 signatures authentiques et de 25 faux expérimentés. Il est à noter qu'au moment de la rédaction de la thèse, cette base n'étant pas disponible depuis longtemps, aucun résultat n'a encore été publié sur cette base.

Nous allons maintenant présenter les résultats obtenus sur les bases ATOS et MCYT afin de confirmer les précédentes conclusions obtenues sur la base SVC.

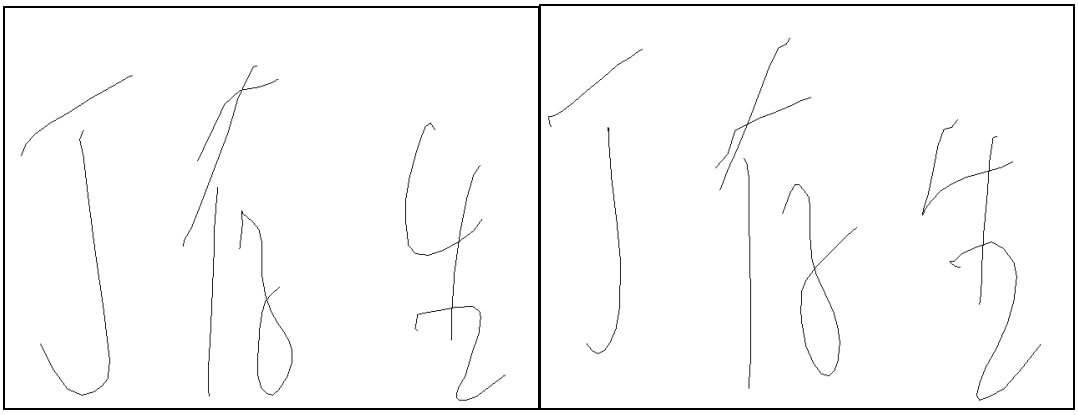
3. ARCHITECTURE CHOISIE

3.1. Hypothèse de travail

Le système proposé est basé sur un constat qui fait l'unanimité parmi les chercheurs : il est très difficile, voire impossible, pour un faussaire d'imiter à la fois la forme de la signature et de réaliser la signature de la même façon que l'auteur (même dynamique). Le temps d’écriture d’une signature est globalement plus long pour un faussaire que pour le signataire original.

La méthode de comparaison a donc été découpée en deux étapes complémentaires. Une première étape consiste à comparer la forme de la signature testée avec la forme des références pour éliminer les faux aléatoires et une partie des faux semi aléatoires. En effet, pour les faux de type expérimenté, la forme est très proche de la signature authentique ; il faut

donc utiliser d'autres caractéristiques (plus liées à la dynamique) pour éliminer ce type de faux. Pour illustrer la difficulté de différencier des faux expérimentés des signatures authentiques, nous avons choisi 3 signatures proposées pour un même signataire – 2 signatures authentiques (Figure 60) et 1 faux (Figure 61) - sur lesquelles nous avons calculé des caractéristiques simples : la longueur, la durée et le nombre de traits.



(1) (2)

Figure 60. Deux exemples de signatures authentiques.

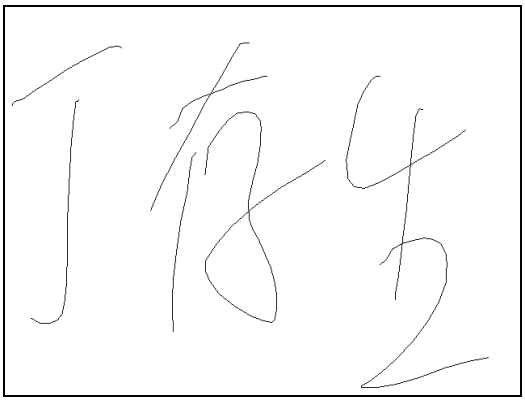


Figure 61. Un exemple de faux expérimenté.

Paramètre	**Signature authentique (1)**	**Signature authentique (2)**	**Signature imitée Figure** 61
Longueur totale	30226.96	29602.12	32259.78
Temps total	2513	1973	2694
Nombre de traits	9	9	9

Tableau 12. Comparaison des caractéristiques globales des signatures authentiques et des faux expérimentés.

Comme on peut le constater sur cet exemple, distinguer une signature authentique d'une signature imitée peut s'avérer très difficile.

La deuxième étape a donc pour but de détecter les faux semi aléatoires et les faux expérimentés en comparant à la fois la forme et la dynamique des signatures.

3.2. Description et performances de l'architecture

Lorsqu'il est évident qu'une signature diffère de celle à laquelle elle est comparée, il n'est pas nécessaire d'appliquer des traitements coûteux en temps de calcul, comme le calcul de DTW. C'est pourquoi nous avons organisé l'ensemble du processus sous la forme d'une approche "Coarse to fine" (Figure 62). Ainsi la première étape a pour but de détecter uniquement les faux évidents à partir de caractéristiques globales et rapides à calculer. Après avoir éliminé ces faux grossiers, seuls certains faux et les signatures authentiques doivent être analysés par des méthodes plus élaborées. La principale contrainte de cette première étape est de ne pas rejeter les signatures authentiques. Etant donné que les caractéristiques utilisées doivent être relativement stables, nous avons opté pour des caractéristiques globales : longueur et durée de la signature.

Soient *Lt* et *Dt* respectivement la longueur et la durée de la signature testée et *Li* et *Di* respectivement la longueur et la durée de la ième signature d'apprentissage. La règle de décision est la suivante :

Si $Lt > 1{,}4 \times \max_{i=1 à n}(Li)$ *ou* $Lt < 0{,}6 \times \min_{i=1 à n}(Li)$, alors la signature est considérée comme un faux

sinon nous passons à l'étape suivante.

Nous appliquons le même principe avec la durée de la signature. Cette première étape permet d'obtenir les taux suivants :

	FAR	FRR
Base ATOS	52%	0%
Base SVC	42%	0,2%
Base MCYT	33%	1,3%

Tableau 13. Résultats au terme de la première étape sur chacune des bases.

La variation des résultats obtenus au terme de la première étape illustre le fait qu'il est très difficile de comparer des résultats obtenus sur des bases de signatures différentes. En effet, suivant le type de personnes sollicitées pour l'acquisition, les résultats peuvent être très variables. Le milieu socio-professionnel a une grande influence sur la variabilité de la signature. Ainsi une personne qui n'a pas l'occasion de signer ou qui signe très rarement aura une signature moins stable qu'une personne amenée à signer régulièrement dans le cadre professionnel. Cela est principalement dû au fait que l'on est de moins en moins amené à signer dans la vie de tous les jours hors du monde professionnel.

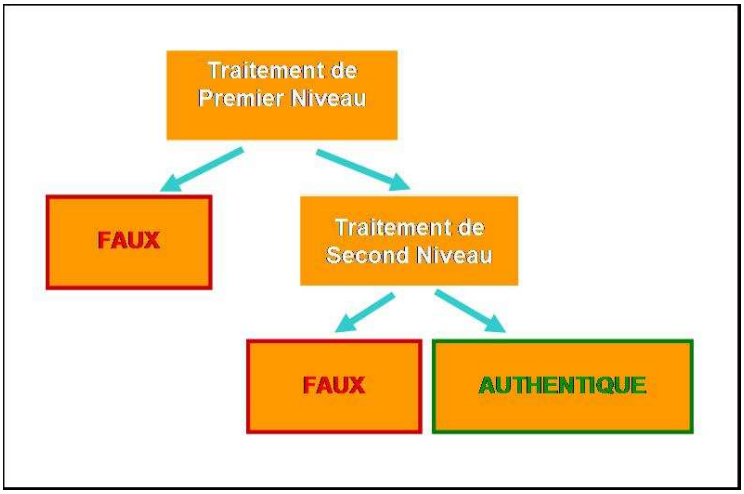
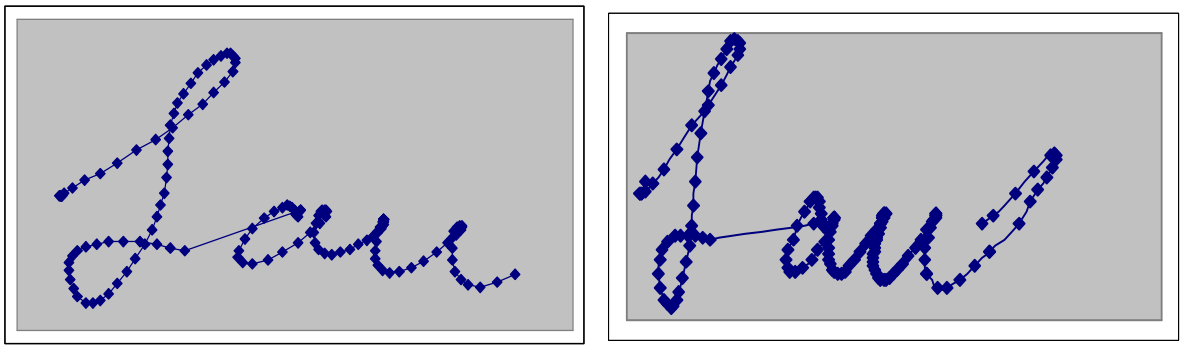


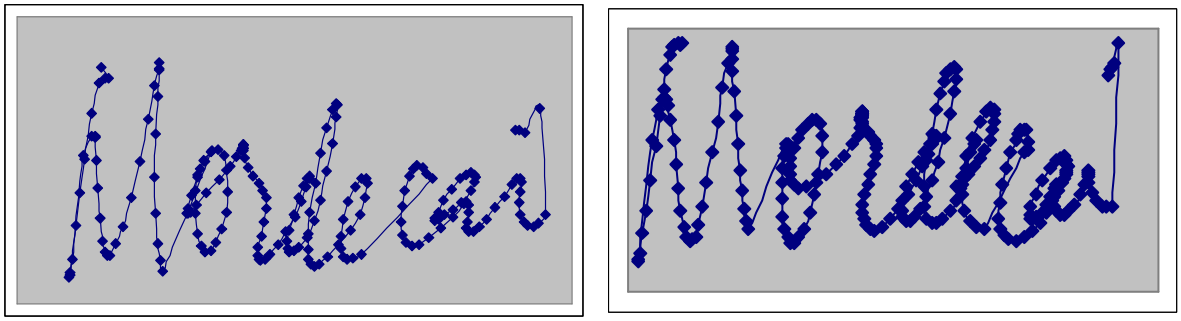
Figure 62. Traitement par niveau, architecture Coarse to Fine.

Sur la base SVC, le traitement de premier niveau permet d'obtenir un taux de FRR de 0,2%. Parmi les 600 signatures authentiques testées, deux signatures authentiques sont rejetées dès cette première étape. Les signatures rejetées sont représentées ci-dessous.



(a) (b)

Figure 63. Exemple de signature rejetée au terme de la comparaison de longueur : (a) signature d'apprentissage, (b) signature authentique rejetée.



(a) **(b)**

Figure 64. Exemple de signature rejetée au terme de la comparaison de durée : (a) signature d'apprentissage, (b) signature authentique rejetée.

Dans ces deux cas, la durée totale de la signature est nettement supérieure à la durée moyenne. Cela semble dû à un problème au niveau de l'acquisition. Il semblerait que le stylet était toujours détecté alors que le signataire avait fini sa signature. Par conséquent, il semble tout à fait logique que ces deux signatures soient rejetées. Cela souligne le fait que l'acquisition doit se faire sur un support permettant de visualiser les données enregistrées par le stylet comme un Tablet PC ou un PDA, pour permettre de voir ce qui est effectivement acquis comme tracé par le logiciel.

Les tests effectués lors de l'enrôlement pour la constitution de la base Atos, plus précisément durant la phase de vérification, permettent de réduire les erreurs dues à des problèmes d'acquisition comme c'est le cas sur les exemples de la Figure 63 et de la Figure 64. De plus, la possibilité de voir le résultat de l'acquisition en temps réel permet à l'utilisateur de faire un contrôle visuel du résultat ce qui n'est pas le cas lors de l'utilisation de la plupart des tablettes graphiques. Le protocole d'acquisition proposé permet d'avoir une base de meilleure qualité.

Le but de l'étape suivante est de détecter les faux qui n'ont pas été détectés lors de la première phase. Comme l'approche consistant à ne pas considérer les doublons dans le calcul de la dissimilarité pourrait réduire aussi la distance entre signatures authentique et fausse, avant de calculer la distance entre les signatures, nous effectuons une estimation globale de la qualité de la mise en correspondance. Pour cela, nous comparons le nombre de fois où un point d'une signature a plus d'un point correspondant dans l'autre signature. Cela permet de définir un nouveau critère de vérification des signatures. Si le nombre de mises en correspondance multiples est trop important, cela signifie que la signature testée est un faux.

Ainsi, avant de calculer la dissimilarité entre les signatures, nous comparons la mise en correspondance entre signatures d'apprentissage avec les mises en correspondance entre les signatures S_t et S_i. Soit $Matching(S_1,S_2)$ le nombre de points qui ont plus d'un point correspondant lorsque l'on effectue la mise en correspondance entre la signature S_1 et la signature S_2. La règle de décision est :

Si $\min_{i=1 \grave{a} n}(Matching(St,Si)) < (2 \times \max_{\substack{i=1 \grave{a} n \\ j=1 \grave{a} n \\ i \neq j}}(Matching(Si,Sj))$, alors la signature est considérée comme potentiellement authentique sinon elle est rejetée. Au terme de cette étape, on obtient les résultats présentés dans le Tableau 14.

	Aléatoire	Algorithme Génétique	Brault	Wall	Wall Best	Vmin
FAR	54,2%	48,4%	47,9%	32,7%	36,4%	30,2%
FRR	0,2%	0,2%	0,2%	0,9%	0,2%	0,2%

Tableau 14. Résultats obtenus sur la base SVC au terme des deux premières étapes suivant la méthode de sélection des points.

On observe déjà des différences suivant le type de critère utilisé pour conserver les points. Les méthodes basées sur l'algorithme de Wall et sur l'étude des minimums de la vitesse donnent des résultats nettement meilleurs, au minimum 10% de faux rejetés en plus, qu'avec les autres méthodes. Une des principales raisons est que les points retenus sont relativement espacés et donc le risque de confusion entre deux points lors de la mise en correspondance diminue.

Etant donné que l'on souhaite réaliser au final un prototype ayant la possibilité de prendre en compte l'évolution de la signature au cours du temps, cela impose des contraintes sur le choix de la méthode de comparaison entre signatures testées et de référence. En effet, cette méthode ne doit pas nécessiter un réapprentissage complet comme c'est le cas lorsque l'on utilise des systèmes basés sur les réseaux de neurones ou les chaînes de Markov. Cela explique le choix de DTW pour mesurer la dissimilarité entre signatures.

On calcule ensuite la dissimilarité, au sens de DTW telle que définie au paragraphe 1.2.7 du chapitre 3, entre les signatures. Soient *St* la signature testée et *Si* la ième signature d'apprentissage. La règle de décision utilisée est la suivante :

Si $\min_{i=1 à n}(DTW(St,Si)) < \alpha$, alors la signature est acceptée sinon elle est considérée comme fausse.

Nous avons choisi de faire varier α de façon à définir différents systèmes. En effet, l'objectif d'un système d'authentification peut être, soit de minimiser le taux de FRR pour éviter d'avoir un rejet trop important de vrais utilisateurs, soit de minimiser le taux de FAR pour limiter l'accès à des personnes non autorisées. La qualité de ces systèmes est donc représentée dans un espace à deux dimensions : FAR et FRR. Pour l'instant et à la vue des résultats présentés au paragraphe 2 du chapitre 4, la valeur de α est fixe pour l'ensemble des signataires.

3.3. Comparaison et validation des résultats obtenus

Les résultats énoncés jusqu'à présent utilisaient la base de signatures SVC. Afin de juger de la pertinence de nos méthodes, quelle que soit la base de test utilisée, nous avons refait une campagne de test en utilisant notre base de signatures (base ATOS). Voici les résultats obtenus.

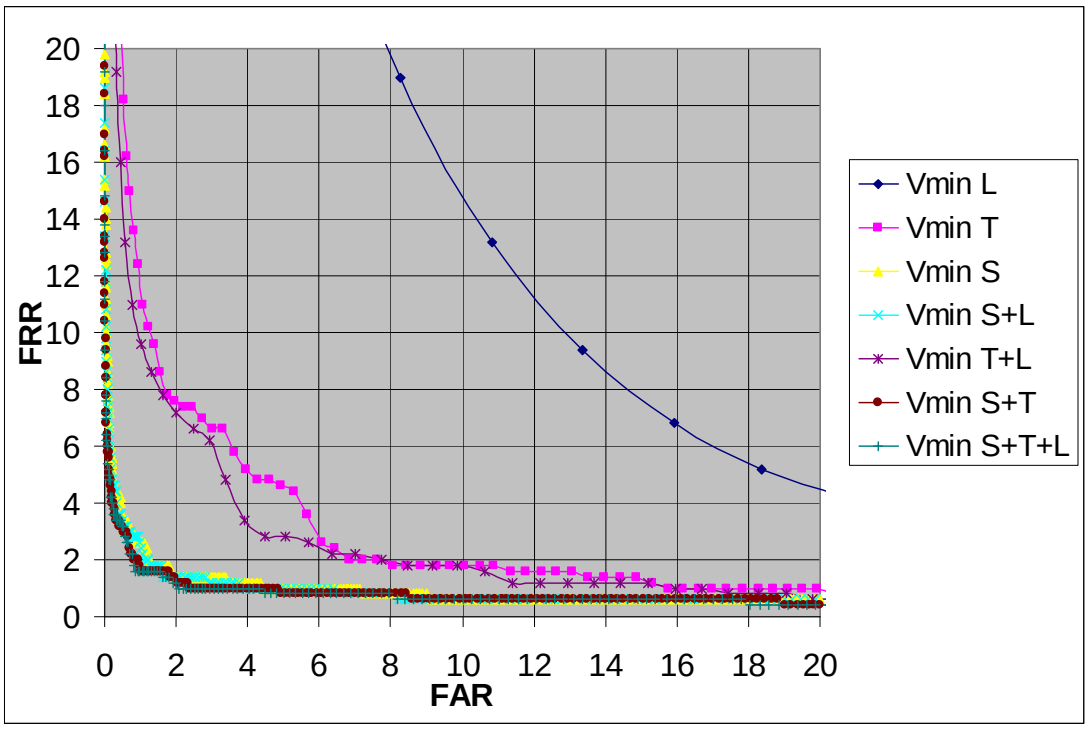


Figure 65. Comparatif des résultats obtenus sur la base ATOS (Tablet) avec les différentes distances : spatiale (S), temporelle (T) et curviligne (L).

Comme sur la base SVC, les meilleurs résultats sont obtenus avec la combinaison des trois distances : spatiale, temporelle et curviligne (Figure 65). Cependant la meilleure valeur de EER obtenue est 1,6% ce qui est légèrement supérieur à la valeur obtenue sur la base SVC : 1,1%. Deux raisons peuvent expliquer ce phénomène. Premièrement, les signatures de la base ATOS sont réalisées uniquement par des personnes dont la langue maternelle est le français d'où une variabilité moins importante des formes des signatures que pour la base SVC qui contenait des signatures anglo-saxonnes et asiatiques. La deuxième raison est que le nombre de signatures authentiques par individu est moins important dans la base ATOS (10 signatures par personne) que dans la base SVC (20 signatures par personne) et donc le taux de FRR est calculé sur 5 signatures ce qui rend la comparaison très sensible au bruit alors que le taux de FAR est calculé sur 900 signatures.

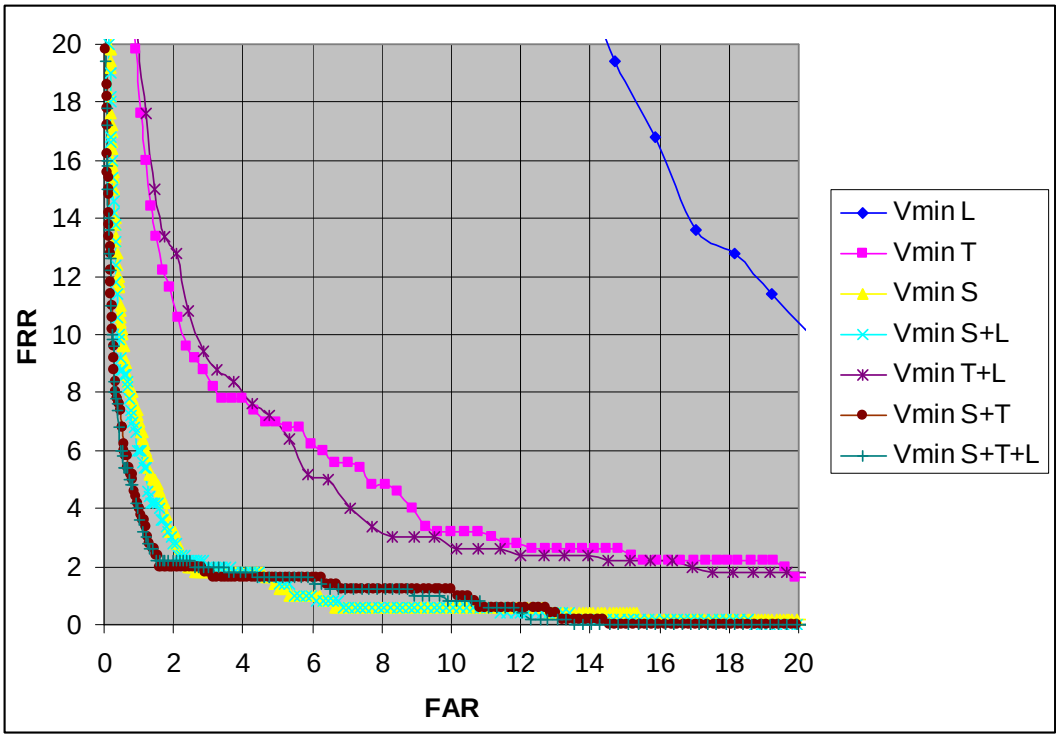


Figure 66. Comparatif des résultats obtenus sur la base ATOS (Palm) avec les différentes distances : spatiale (S), temporelle (T) et curviligne (L).

Les résultats obtenus sur la base ATOS sont moins bons lorsque l'acquisition se fait sur Palm plutôt que sur Tablet PC (Figure 66). Ceci confirme les impressions des personnes ayant participé à la construction de cette base, à savoir la difficulté de signer sur un PDA. Le meilleur résultat, EER égal à 2%, est obtenu en combinant la distance spatiale et la distance temporelle. Contrairement aux résultats précédents, la distance curviligne n'apporte pas plus d'informations mais dégrade les performances obtenues en considérant les distances spatiales et temporelles.

Nous présentons maintenant les résultats obtenus sur la base MCYT à la Figure 67.

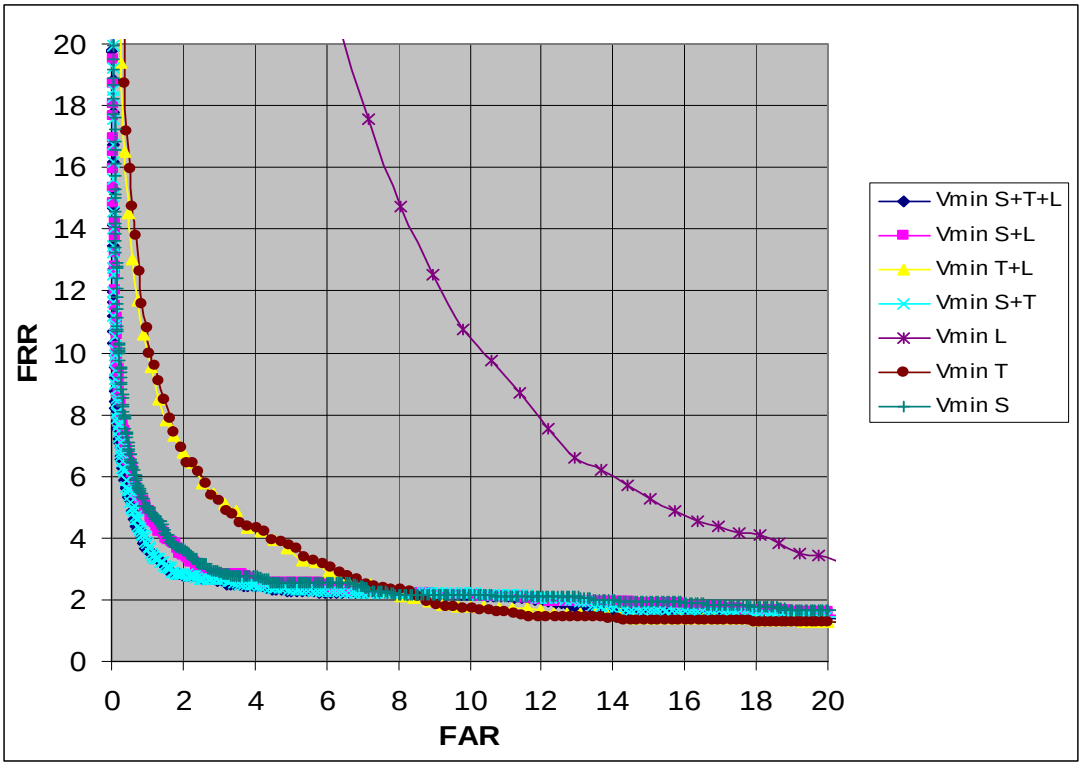


Figure 67. Comparatif des résultats obtenus sur la base MCYT avec les différentes distances : spatiale (S), temporelle (T) et curviligne (L).

Les résultats obtenus sur la base MCYT confirment les précédents résultats. La combinaison des différentes distances (spatiale, temporelle et curviligne) permet d'obtenir les meilleurs résultats, soit un taux de EER égal à 2,7%. La méthode d'authentification que nous proposons semble donc relativement robuste. Les performances obtenues sur la base MCYT sont cependant moins bonnes que celles obtenues sur la base SVC. Une des raisons à cette baisse de performance est que la première étape de la phase de comparaison ne semble pas adaptée à la base. En effet, le taux de FAR obtenu est de 32,3% et le taux de FRR de 1,3%. L'étude des cas d'erreur concernant les signatures authentiques rejetées montre que les rejets sont justifiés. En effet, on constate, notamment pour 5 individus, une grande variabilité de la signature. Certaines signatures authentiques se révèlent être singulièrement différentes des signatures d'apprentissage et l'acceptation de ces signatures même par un expert humain semble peu probable.

La variation intra individu semble donc plus importante pour cette base que pour les bases utilisées précédemment.

Les tests ont donc été refaits en utilisant un nombre de signatures d'apprentissage plus important, 10 signatures d'apprentissage au lieu de 5. Au terme de la première phase, le taux de FAR est de 43,6% et celui de FRR de 0,3%. Ces valeurs sont plus proches de celles obtenues sur les bases SVC et ATOS. Le processus complet d'authentification permet

d'obtenir un taux de EER de 1,5%, soit une amélioration des performances de près de 45%. Ce résultat confirme le fait que plus le nombre de signatures réalisées lors de l'enrôlement est important, plus le modèle créé est représentatif de la variabilité de la signature de l'utilisateur et plus la méthode d'authentification est fiable.

3.4. Evaluation en situation "réelle"

Afin d'être cohérent avec l'application finale souhaitée, nous avons réalisé des tests dans une situation plus proche de la réalité. Ainsi, au lieu de tester les signatures une par une, nous les avons testées trois par trois pour simuler ainsi le fait que l'utilisateur a le droit à trois tentatives pour s'authentifier, comme c'est le cas actuellement pour le code PIN.
La règle est donc la suivante : si aucune des 3 signatures consécutives n'est considérée comme authentique alors la personne qui a signé est considérée comme un faussaire. Sur la base SVC, avec ce protocole de test, nous avons obtenu un taux de FAR de 0,6% pour un taux de FRR égal à 0%. Dans la méthode d'évaluation utilisée précédemment, le taux de FAR est proche de 20% pour un taux de FRR de 0%. La prise en compte de trois signatures consécutives permet donc d'améliorer de manière significative les performances notamment pour les personnes ayant une signature très variable mais aussi de limiter l'influence des erreurs dues à des problèmes d'acquisition comme nous l'avons vu au paragraphe précédent.

Après avoir présenté de manière détaillée le module d'authentification par signature manuscrite en ligne, nous allons décrire dans le prochain chapitre comment il s'intègre dans l'architecture globale du prototype.

4. DEVELOPPEMENT D'UN PROTOTYPE POUR PDA

Dans le cadre de la dématérialisation des documents, la signature manuscrite de documents numériques entraînant une signature numérique du document est un plus pour conserver les habitudes des utilisateurs. De plus, l'ajout de l'image de la signature sur le document numérique permet de conserver l'aspect visuel des documents papier.

Le développement d'un prototype de système complet d'authentification forte implique une étude de risque, une identification des différentes attaques possibles et les parades associées.

Un prototype d'authentification par signature manuscrite utilisant comme périphérique d'acquisition un PDA a été développé. L'objectif de ce prototype est de signer numériquement des documents à l'aide d'un certificat contenu sur une carte à puce. Au lieu d'utiliser un code PIN pour accéder au certificat, on veut le remplacer par une signature manuscrite. Par conséquent, il faut remplacer le code PIN par la signature de référence de l'utilisateur. Le protocole d'identification est le suivant : l'utilisateur a le droit à 3 tentatives pour s'authentifier. Si les 3 signatures sont reconnues comme des faux, le système se bloque. Sinon, l'authentification de la signature manuscrite donne accès à la clé privée afin de signer ou d'authentifier des documents. Il faudrait pouvoir régler le seuil d'acceptation en fonction de l'importance du document. Plus le document est sensible, plus le FAR doit tendre vers 0 mais l'inconvénient est une augmentation du taux de FRR.

De plus, il faudra gérer et protéger les transmissions entre le PC et les différents supports. En particulier pour le PDA, la synchronisation avec le PC devra éventuellement être précédée d'un cryptage des données dans le cas de transmission de données confidentielles.

La signature comme moyen de sécurisation ne peut pas être utilisée trop souvent (par exemple à chaque démarrage) car sinon risque de banalisation de la signature entraînant une modification de celle-ci. Elle doit être réservée pour les événements importants : validation de contrat, … ou alors définir plusieurs signatures suivant le rôle de chacune.

4.1. Etapes menant à la signature électronique d'un contrat

Tout d'abord, on choisit au niveau du PC le document à signer au format ".doc" ou ".pdf" (Figure 68). Différentes informations sont présentées à l'utilisateur afin d'identifier de manière précise le document à signer : le nom du fichier, la date de dernière modification, la taille du

fichier. Une fois ces données validées, elles sont transmises au PDA par une synchronisation. Au niveau du PDA, les données relatives au document sont affichées pour que l'utilisateur sache quel document est signé. L'utilisateur est ensuite invité à signer sur le PDA. Il faut alors synchroniser le PDA pour transmettre la signature au PC. La signature est ensuite analysée et comparée au modèle contenu dans la carte à puce reliée au PC par un lecteur de carte à puce. Si la signature est reconnue comme authentique, la signature du document est générée.

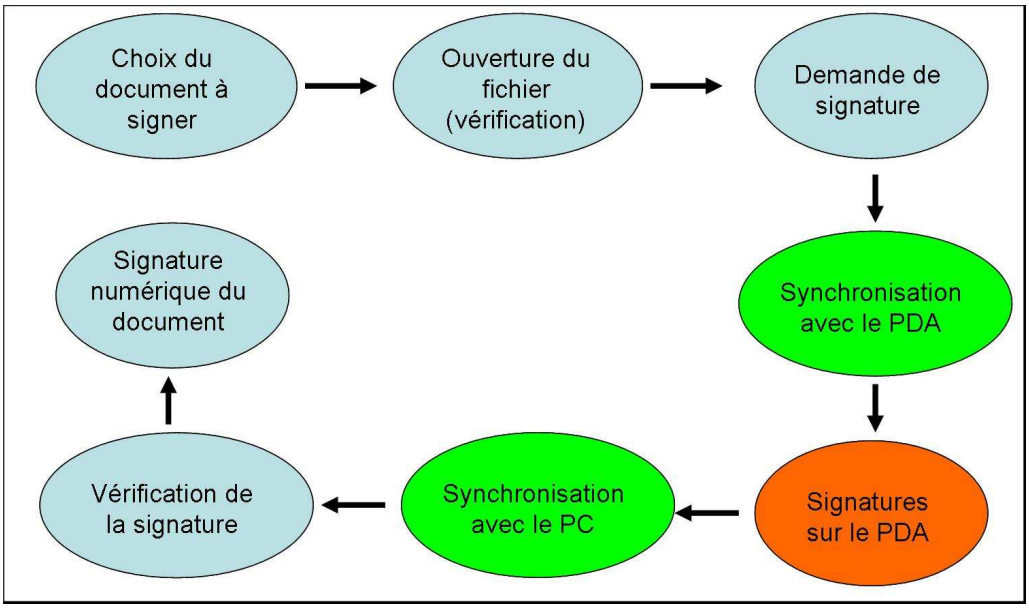


Figure 68. Enchaînement des différentes étapes menant à la signature électronique du document.

4.2. Logiciels associés

L'interface au niveau du PC a été développée en Java afin de pouvoir gérer avec des outils "open source" l'accès à la carte à puce, mais le module d'authentification est une dll développée en C++ afin d'avoir un temps d'exécution relativement faible. De plus, cette architecture permet de séparer la partie interface de la partie authentification et donc permet une évolution de l'authentification de façon simple en remplaçant la dll. La librairie utilisée pour l'accès à la carte à puce est IAIK. Les traitements d'authentification qui sont relativement lourds en terme de puissance de calcul et de consommation mémoire sont effectués sur le PC car le PDA ne dispose pas à l'heure actuelle d'une grande puissance de calcul.

Actuellement plusieurs systèmes d'exploitation existent pour les PDA, les deux principaux étant Windows CE et Palm OS. Le PDA utilisé pour réaliser le prototype est un Palm. Ainsi les développements réalisés dans le cadre de ce travail sont spécifiques au système d'exploitation Palm OS. Concernant les développements relatifs au PDA, deux programmes

ont été réalisés. Le premier permet de gérer l'interface au niveau du Palm et le deuxième, appelé conduit, permet d'effectuer la synchronisation du Palm et du PC, i.e. le transfert des informations relatives au contrat du PC au PDA ainsi que le transfert de la signature du PDA au PC (Figure 69).

A ces différents programmes, il faut ajouter aussi un module d'enregistrement qui permet de créer le modèle d'une signature d'un individu à partir des signatures acquises sur le Palm. Comme pour l'authentification, la création du modèle se fait à l'aide d'un programme en C++.

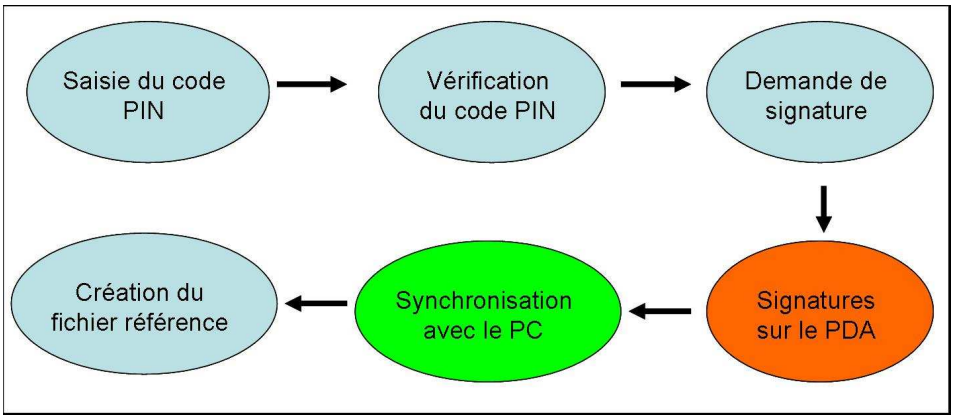


Figure 69. Enchaînement des différentes étapes menant à la création du modèle de signature à partir d'acquisitions effectuées sur le PDA.

4.3. Caractéristiques du module de signature numérique

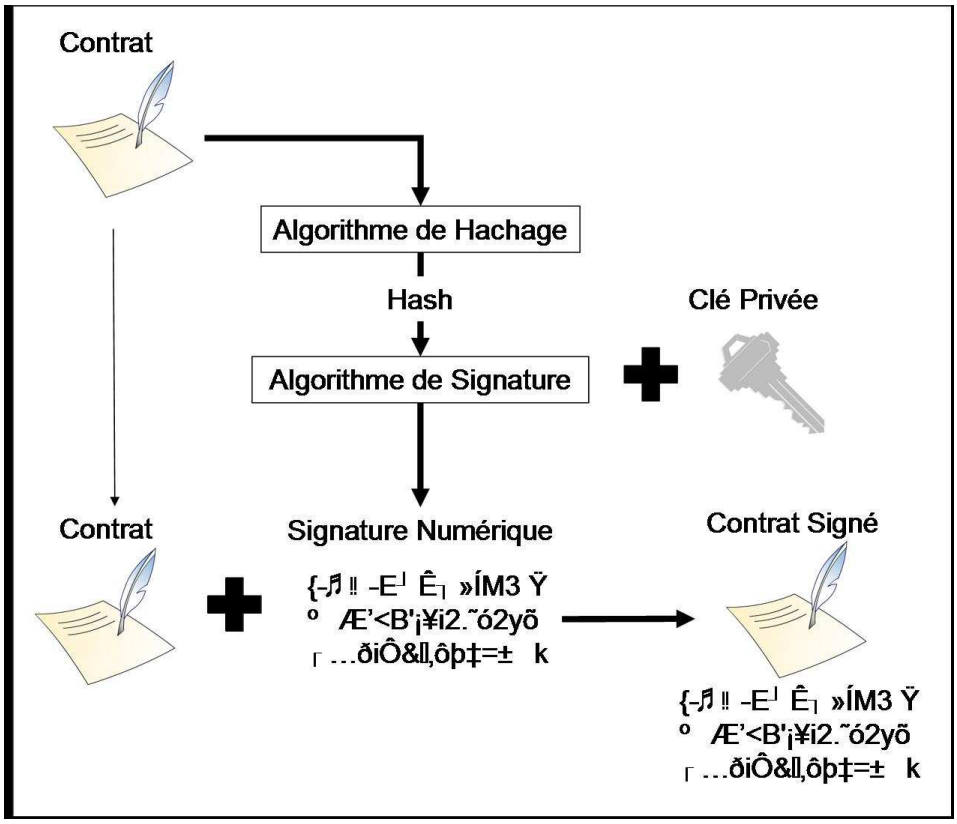


Figure 70. Etapes menant à la signature numérique d'un contrat.

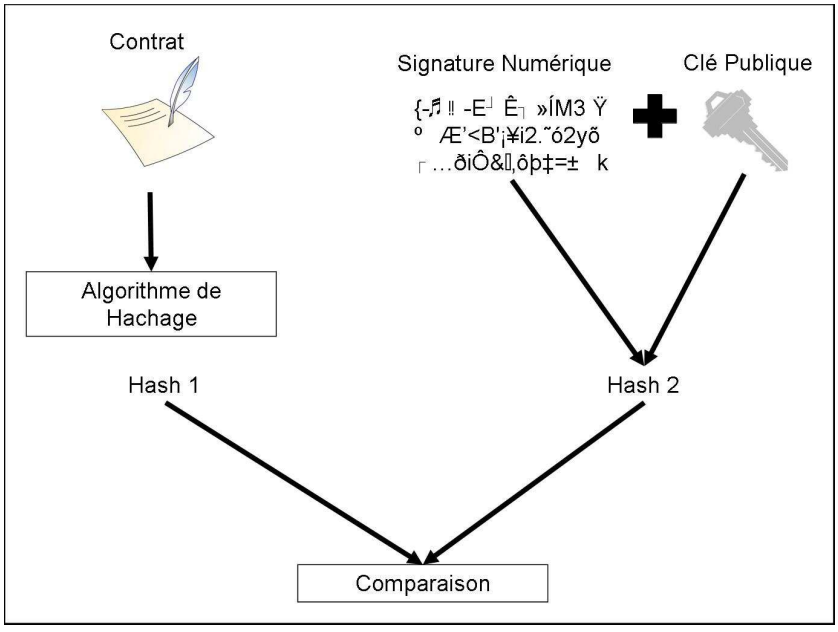


Figure 71. Etapes menant à la vérification de la signature numérique d'un contrat.

4.4. Caractéristiques du module d'authentification

Comme nous l'avons vu précédemment, les valeurs de FAR et de FRR doivent être définies en fonction du type d'application. Pour un produit industriel, la valeur de FRR doit être faible pour éviter la répétition de la phase d'authentification quand la personne autorisée est rejetée.
La sélection des caractéristiques ne peut se faire que pendant la phase de conception et donc pas de façon individuelle car cela nécessiterait de disposer d'une base conséquente de signatures pour pouvoir définir les caractéristiques qui minimisent la distance intra classe et qui maximisent la distance inter classes.
Le réglage du seuil peut être effectué par l'utilisateur mais dépend aussi de la variabilité de la signature de celui-ci.

4.5. Ergonomie

L'ergonomie est un élément très important dans le développement d'un logiciel sur PDA étant donné que l'utilisateur ne dispose que de moyens d'interactions réduits le plus souvent à un stylet, un écran relativement petit…

Cet aspect du logiciel a donc été étudié en collaboration avec un laboratoire d'ergonomie informatique. Les conseils fournis par Christian Bastien nous ont permis de développer une interface à la fois fonctionnelle et conviviale.
Le fait de n'utiliser que les coordonnées (x,y), le temps et le contact, permet d'utiliser un stylet inerte pour l'acquisition sur un écran tactile. Cela présente un avantage certain si l'on envisage un déploiement sur des bornes accessibles à un large public. En effet, le coût de déploiement ainsi que les risques de vol ou de détérioration de matériel sont moins importants.

CONCLUSION ET PERSPECTIVES

1. CONCLUSION

Dans cette thèse, nous avons abordé le problème de l'authentification par signature manuscrite en ligne.

Après une étude des travaux antérieurs, nous avons tout d'abord cherché à extraire des caractéristiques significatives de la signature notamment en utilisant différentes approches basées sur la multi résolution. Cette approche repose sur l'aspect fractal de l'écriture. Ces caractéristiques mises en place ont un pouvoir discriminant relativement supérieur à celui des caractéristiques classiques. Cependant, ces caractéristiques étant globales et pas suffisamment discriminantes lorsqu'on les utilise seules, nous avons ensuite étudié et proposé différentes façons de mesurer la similarité entre signatures.

Tout d'abord, afin d'améliorer ce calcul de dissimilarité entre signatures et de répondre aux contraintes posées notamment sur la taille en mémoire du modèle, nous nous sommes également focalisé sur la réduction du nombre de points représentatifs de la signature et avons démontré les nombreux avantages obtenus.

Ensuite, dans le but d'améliorer les résultats obtenus en utilisant une méthode efficace appelée Dynamic Time Warping, nous avons cherché à améliorer la mise en correspondance des points des signatures.

Après avoir effectué cette sélection des points les plus représentatifs de la signature, nous avons présenté différentes approches pour augmenter la signification de la distance en utilisant l'ensemble des informations disponibles et pas seulement les coordonnées des points.

Nous avons finalement validé notre travail, c'est à dire construit une architecture utilisant les méthodes proposées pour l'authentification de signatures manuscrites. Les tests que nous avons effectués sur des bases réelles de signatures (SVC, ATOS, MCYT) montrent que les choix effectués étaient judicieux.

2. PERSPECTIVES

La recherche d'un seuil individualisé pour chaque scripteur est à compléter pour améliorer les performances en tenant compte des spécificités intra signatures stables pour chaque utilisateur et se rapprocher des performances optimales obtenues en identification.

Au niveau de l'architecture de la méthode de comparaison, une alternative à l'approche séquentielle de type "coarse to fine" pourrait consister en la fusion entre les différents classificateurs, c'est à dire ceux définis à partir des caractéristiques extraites de la signature et ceux définis à partir des différentes distances.
Une approche plus locale du problème, consistant en une segmentation de la signature pour détecter les parties stables, peut aussi être une voie de recherche. Pour l'instant, toutes les approches que nous avons mises en place traitent les signatures dans leur globalité.
Tout au long de cette thèse, nous avons cherché des méthodes permettant de prendre en compte l'évolution de la signature au cours du temps. Cependant, nous n'avons pas pu réaliser une étude sur le long terme mettant en avant cette évolution malgré la mise en place de la base Atos. Une méthode possible pour la prise en compte de l'évolution de la signature au cours du temps consisterait à ajouter à la base de références les signatures acceptées au premier essai.
Le travail sur les faux expérimentés reste encore à améliorer.
Une autre solution à envisager pour accroître les performances consisterait à utiliser, couplée avec l'authentification par la signature, l'authentification par mot de passe manuscrit [PAR89][SRI01].
Jusqu'à présent les signatures utilisées lors des phases d'enregistrement et d'acquisition sont celles réalisées habituellement : aucune indication n'est donnée au signataire. Ce faisant, on n'utilise pas les propriétés importantes que possède l'écriture manuscrite à savoir la conservation des rapports de taille et de temps. Par conséquent, une piste à explorer serait de faire varier la taille du cadre d'acquisition de la signature. Une autre voie pourrait être de demander à l'utilisateur de signer à plusieurs vitesses différentes [LEE96].

Modifier le calcul de la dissimilarité entre deux signatures permet de réduire la distance entre deux signatures authentiques en ne tenant pas compte des petites différences grâce à la réduction du nombre de points. Même si l'utilisation des coordonnées des points permet d'obtenir de très bons résultats, nous pensons que la combinaison avec des informations locales pourrait permettre d'améliorer les performances dans certains cas. Par conséquent, une de nos principales perspectives est de fusionner l'information locale avec les coordonnées pour calculer la distance entre signatures.

PUBLICATIONS

Communications dans des congrès internationaux avec comité de lecture

Matthieu WIROTIUS, Audrey SEROPIAN, Nicole VINCENT, "Writer Identification from Gray Level Distribution", International Conference on Document Analysis and Recognition (ICDAR), Edinburgh (Ecosse). pp. 1168-1172. Août 2003.

Matthieu WIROTIUS, Nicole VINCENT, "Stroke Inner Structure Invariance in Handwriting", International Graphonomics Society (IGS), Scottsdale (USA), pp.308-311, Novembre 2003.

Matthieu WIROTIUS, Jean-Yves RAMEL, Nicole VINCENT, "New Features for Authentication by On-Line Handwritten Signatures", International Conference in Biometric Authentication (ICBA), Hong-Kong, pp. 577-583, Juillet 2004.

Matthieu WIROTIUS, Jean-Yves RAMEL, Nicole VINCENT, "Improving DTW for Online Handwritten Signature Verification", International Conference in Image Analysis and Recognition (ICIAR), Porto (Portugal), pp. 786-793, Septembre 2004.

Matthieu WIROTIUS, Jean-Yves RAMEL, Nicole VINCENT, "Selection of Points for On-Line Signature Comparison", International Workshop On Frontiers in Handwriting Recognition (IWFHR), Tokyo (Japon), pp. 503-508, Octobre 2004.

Matthieu WIROTIUS, Jean-Yves RAMEL, Nicole VINCENT, "Comparison of point selection for characterizing on-line signature", Biometric Technology for Human Identification, Orlando (USA), pp. 307-313, Mars 2005.

Matthieu WIROTIUS, Jean-Yves RAMEL, Nicole VINCENT, "Contribution of Global Temporal Information for Authentication by On-Line Handwritten Signatures", International Graphonomics Society (IGS), Salerno (Italie), pp. 266-270, Juin 2005.

Communications dans des groupes nationaux, séminaires, forums

Matthieu WIROTIUS, "On-line signature verification based on shape", Journées Doctorales, Tours, France, Mai 2003, pp. 105-106.

Matthieu WIROTIUS, "Authentification de scripteurs", Journée Jeunes Chercheurs S.A. 5.2. et GRCE, Tours, France, Juin 2003.

Matthieu WIROTIUS, "Authentification par signature manuscrite sur PDA", Journée PDA'03, Tours, France, Décembre 2003.

Matthieu WIROTIUS, "Sélection de points pour la comparaison de signatures en ligne", Journée Jeunes Chercheurs, La Rochelle, France, Décembre 2004.

Matthieu WIROTIUS, "Comparaison des méthodes de sélection de points pour caractériser les signatures on-line", Journée Thématique S.A. 5.2. du GDRI3 et GRCE, Paris, France, Mai 2005.

Matthieu WIROTIUS, "Premiers résultats des travaux de recherches menés dans le domaine de la reconnaissance dynamique de signature", Identech 2005, Marseille, France, Juin 2005.

BIBLIOGRAPHIE

[BEL57] R. Bellman, "Dynamic Programming", Princeton University Press, 1957.

[BEN03] A. Bensefia, T. Paquet et L. Heutte, "Information Retrieval based writer identification", Proceedings of the International Conference on Document Analysis and Recognition (ICDAR'03), Edimburg, Scotland, pp. 946-950, 2003.

[BOU97] V. Bouletreau, "Vers un classement de l'écrit par des méthodes fractales", Thèse de doctorat, INSA Lyon, 195 p., 1997.

[BRA93] J. Brault et R. Plamondon, "Segmenting handwritten signatures at their perceptually important points", IEEE Pattern Analysis and Machine Intelligence (PAMI), vol. 15, no. 9, pp. 953-957, 1993.

[CAB03] C. Cabal, "Méthodes scientifiques d'identification des personnes à partir des données biométriques et techniques de mise en oeuvre", rapport ministériel, no. 938 Assemblée Nationale, 226 p., 2003.

[CAR97] J. Carbonnier, "Droit Civil, Introduction", PUF, collection Thémis, 25ème édition, 198 p., 1997.

[CNI04] CNIL, délibération 04-018 du 8 avril 2004.

[CNI05] Commission Nationale de l'Informatique et des libertés (CNIL), http://www.cnil.fr

[DIF76] W. Diffie et M.E. Hellman, "New directions in cryptography", IEEE Transactions on Information Theory, 22, pp. 644-654, 1976.

[DIM94] G. Dimauro, S. Impedovo and G. Pirlo, "Component-oriented algorithms for signature verification", International Journal of Pattern Recognition and Artificial Intelligence, vol. 8, no 3, pp. 771-794, 1994.

[DIM02] G. Dimauro, S. Impedovo, R. Modugno, G. Pirlo, L. Sarcinella, "Analysis of Stability in Hand-Written Dynamic Signatures", Proceedings of the 8th International Workshop on Frontiers in Handwriting recognition (IWFHR'02), Ontario (Canada), pp. 259-263, 2002.

[DOL98] J. G. A. Dolfing, E. H. L. Aarts et J. J. G. M. Van Oosterhout, "On line Signature verification with Hidden Markov Models", Proceedings of the 14th International Conference on Pattern Recognition (ICPR'98), Australie, vol. 2, pp. 1309-1312, 1998.

[FRA03] K. Franke et L.R.B. Schomaker, "Pen orientation characteristics of on-line handwritten signatures", International Graphonomics Society (IGS), Scottsdale (USA), pp.224-227, 2003.

[FUE02] M. Fuentes, S. Garcia-Salicetti et B. Dorizzi, "On line signature verification: fusion of a Hidden Markov Model and a Neural Network via a Support Vector Machine", Proceeding of the 8th International Workshop on Frontiers in Handwriting Recognition (IWFHR'02), Ontario (Canada), pp. 253-258, 2002.

[GOL89] David Goldberg, "Genetic Algorithms in Search, Optimization, and Machine Learning", Addison-Wesley Professional, 432 p., 1989.

[GRI00] D. Griess et K. Jain, "On-line Signature Verification", Michigan State University Technical Report TR00-15, 2000.

[GUP97a] G. Gupta et R. C. Joyce, "A Study of Shape in Dynamic Handwritten Signature Verification", Rapport technique, Computer Science Dept, James Cook University of North Queensland, 1997.

[GUP97b] G. Gupta et A. Mc Cabe, "A Review of Dynamic Handwritten Signature Verification", Rapport technique, Computer Science Dept, James Cook University of North Queensland, 1997.

[HAS92] T. Hastie, E. Kishon, M. Clark et J. Fan, "A Model for Signature Verification", Rapport technique, AT&T Bell Laboratories, 1992.

[HEN04] O. Henniger et K. Franke, "Biometric User Authentication on Smart Cards by Means of Handwritten Signatures", First International Conference on Biometric Authentication (ICBA'04), Hong-Kong (Chine), pp. 547-554, 2004.

[HER76] N. M. Herbst et J. F. Morrissey, "Signature Verification Method and Apparatus", U.S. Patent, 3,983,535, 1976.

[HER77] N. M. Herbst et C. N. Liu, "Automatic Signature Verification based on accelerometry", Research Rep. RC-5810, IBM, J. Res. Dev., vol. 2, pp. 245-253, 1977.

[HUA95] K. Huang et H. Yan, "On-line signature verification based on dynamic segmentation and global and local matching", Optical Engineering, vol.34, no.12, pp. 3480-3487, 1995.

[HUA03] K. Huang et H. Yan, "Stability and style-variation modelling for on-line signature verification", Pattern Recognition, vol. 36, pp. 2253-2270, 2003.

[IBG04] Biometrics Market and Industry Report 2004-2008, http://www.biometricgroup.com.

[IGA04] J.J. Igarza, L. Gomez, I. Hernaez, I. Goirizelaia, "Searching for an optimal reference system for on-line signature verification based on (x,y) alignment", First International Conference on Biometric Authentication (ICBA'04), Hong-Kong (Chine), pp. 519-525, 2004.

[ITE05] O. Iteanu, "Biométrie, une technologie sous surveillance?", http://solutions.journaldunet.com/0502/050209_juridique.shtml, 2005.

[JAI02] A. K. Jain, F. D. Griess et S. D. Cornell, "On-line Signature Verification", Pattern Recognition, vol. 35, no. 12, pp. 2963-2972, 2002.

[KAS97] R. S. Kashi, J. Hu, W. L. Nelson et W. Turin, "On-line Handwritten Signature Verification using Hidden Markov Model Features", Proceedings of the International

Conference on Document Analysis and Recognition (ICDAR'97), Ulm (Allemagne), pp. 253-257, 1997.

[KET05] H. Ketabdar, J. Richiardi et A. Drygajlo, "Global Feature Selection for On-Line Signature Verification", International Graphonomics Society (IGS), Salerno (Italie), pp. 59-63, 2005.

[KHO03] A. Kholmatov et B. Yanikoglu, "An Improved Decision Criterion for Genuine/Forgery Classification in On-Line Signature Verification", Proceedings of ICANN/ICONIP, 2003.

[KIK01] M. Kikuchi et N. Akamatsu, "Development of Speedy and High Sensitive Pen System for Writing Pressure and Writer Identification", Proceedings of the International Conference on Document Analysis and Recognition (ICDAR'01), pp. 1040-1044, 2001.

[LEC99] V. Di Lecce, G. Dimauro, A. Guerriero, S. Impedovo, G. Pirlo, A. Salzo et L. Sarcinella, "Selection of reference signatures for automatic signature verification", Proceedings of the International Conference on Document Analysis and Recognition (ICDAR'99), pp. 597-600, 1999.

[LEC94] F. Leclerc et R.Plamondon, "Automatic Signature Verification: the State of The Art 1989-1993", International Journal of Pattern Recognition and Artificial Intelligence, vol. 8, no. 3, pp. 643-660, 1994.

[LEE96] L. L. Lee, T. Berger et E. Aviczer, "Reliable on-line human signature verification systems", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 18, no. 6, pp. 643-647, Juin 1996.

[LEE04] J. Lee, H.S. Yoon, J. Soh, B. Tae Chun, Y. Koo Chung, "Using geometric extrema for segment-to-segment characteristics comparison in on-line signature verification", Pattern Recognition, vol. 37, pp. 93-103, 2004.

[LEI04] H.Lei, S. Palla, V. Govindaraju, "ER^2 : an Intuitive Similarity Measure for On-line Signature Verification", Proceeding of the International Workshop on Frontiers in Handwriting Recognition (IWFHR'04), pp. 191-195, Tokyo, Japon, 2004.

[LEJ01] D. Z. Lejtman et S. E. George, "On-line handwritten signature verification using wavelets and back-propagation neural networks", Proceedings of the International Conference on Document Analysis and Recognition (ICDAR'01), pp. 992-996, 2001.

[LIN98] C.-F. Lin et C.-W. Chen, "A new approach to the verification of chinese signatures with variant orientations and scales using relaxation and state-space search methods", Pattern Recognition, vol. 31, no.6, pp. 665-674, 1998.

[MAI04] D. Maio, D. Maltoni et R. Cappelli, "FVC2004: Third Fingerprint Verification Competition", Proceedings of the First International Conference on Biometric Authentication (ICBA'04), Hong-Kong (Chine), pp. 1-7, 2004.

[MAN84] B. Mandelbrot, "Les objets fractals", Flammarion, Paris, 203 p., 1984.

[MAR97a] R. Martens et L. Claesen, "Dynamic Programming Optimisation for On-line Signature Verification", Proceedings of the International Conference on Document Analysis and Recognition (ICDAR'97), pp. 653-656, 1997.

[MAR97b] R. Martens et L. Claesen, "On-line Signature Verification : Discrimination emphasised", Proceedings of the International Conference on Document Analysis and Recognition (ICDAR'97), pp. 657-660, 1997.

[MAR98] R. Martens et L. Claesen, "Incorporating local consistency information into the online signature verification process", International Journal on Document Analysis and Recognition, vol. 1, no 2, pp. 110-115, 1998.

[MOH99] N. Mohankrishnan, W. Lee et M. Paulik, "A performance evaluation of a new signature verification algorithm using realistic forgeries," Proceedings of the International Conference on Image Processing (ICIP'99), pp. 575-579, 1999.

[NAL97] V. S. Nalwa, "Automatic On-Line Signature Verification," Proceedings of the IEEE, vol. 85, no.2, pp. 215-239, 1997.

[OLI03] L.S. Oliveira, R. Sabourin, F. Bortolozzi et C. Y. Suen, "Feature Selection for Ensembles: A Hierarchical Multi-Objective Genetic Algorithm Approach", Proceedings of the International Conference on Document Analysis and Recognition (ICDAR'03), Edimburg, Scotland, pp. 676-680, 2003.

[ORT03] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J.-J. Igarza, C. Vivaracho, D. Escudero et Q. –I. Moro, "MCYT baseline corpus: a bimodal biometric database", IEEE Proceeding Vision, Image and Signal Processing, vol. 150, no. 6, 2003.

[PAL04] S. Palla, H. Lei et V. Govindaraju, "Signature and Lexicon Pruning Techniques", Proceeding of the International Workshop on Frontiers in Handwriting Recognition (IWFHR'04), Tokyo (Japon), pp. 474-478, 2004.

[PAR89] M. Parizeau et R. Plamondon, "What types of scripts can be used for personal identity verification", Computer Recognition and Human Production of Handwriting, pp. 77-90, 1989.

[PAR02] J. R. Parker "Simple Distances Between Handwritten Signatures", Rapport technique, Computer Vision Laboratory, Computer Science Dept, Calgary University, 2002.

[PLA88] R. Plamondon et M. Parizeau, "Signature verification from position, velocity and acceleration signals : A comparative Study", Proceedings of the 9th International Conference on Pattern Recognition (ICPR'88), Rome (Italie), vol. I., pp. 260-265, 1988.

[PLA89] R. Plamondon et G. Lorette, "Automatic Signature Verification and Writer Identification – The State of the Art", Pattern Recognition, vol. 22, no.2, pp. 107-131, 1989.

[RHE01] T. H. Rhee, S. J. Cho et J. H. Kim, "On-line signature verification using model-guided segmentation and discriminative feature selection for skilled forgeries", Proceedings of the International Conference on Document Analysis and Recognition (ICDAR'01), pp. 645-649, 2001.

[ROK04] Rokbani N., et Alimi M.A. (2004), "Un système de vérification de signatures manuscrites en ligne pour PDA", Proceeding Conférence Internationale Francophone sur l'Ecrit et le Document (CIFED'2004), La Rochelle, France, 2004.

[SAB95] R. Sabourin et G. Genest, "Définition et évaluation d'une famille de représentations pour la vérification hors-ligne des signatures", Traitement du Signal, vol. 12, n. 6, pp. 585-596, 1995.

[SAB97] R. Sabourin, G. Genest et F. J. Prêteux, "Off-Line Signature Verification by Local Granulometric Size Distribution", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 9, pp. 976-988, 1997.

[SAN04] C. Santos, E.J.R. Justino, F. Bortolozzi et R. Sabourin, "An Off-Line Signature Verification Method based on the Questioned Document Expert's Approach and a Neural Network Classifier", International Workshop On Frontiers in Handwriting Recognition (IWFHR), Tokyo (Japon), pp. 498-502., 2004.

[SCH04] S. Schimke, C. Vielhauer et J. Dittman, "Using Adapted Levenshtein Distance for On-Line Signature Authentication", Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04), Cambridge (UK), vol. 2, pp. 931-934, 2004.

[SCH97] C. Schmidt et K. –F. Kraiss, "Establishment of personalized templates for automatic signature verification," Proceedings of the International Conference on Document Analysis and Recognition (ICDAR'97), Ulm (Allemagne), pp. 263-267, 1997.

[SRI01] S.N. Srihari, "Handwriting Identification", Rapport technique, CEDAR TR01-1, 2001.

[SVC04] http://www.cs.ust.hk/svc2004/download.html

[TAN01] K. Tanabe, M. Yoshihara, H. Kameya et S. Mori, "Automatic signature verification based on the dynamic of pressure", Proceedings of the International Conference on Document Analysis and Recognition (ICDAR'01), pp. 1045-1049, 2001.

[VIE01] C. Vielhauer, R. Steinmetz et A. Mayerhöfer, "Transitivity based enrollment strategy for signature verification systems", Proceedings of the International Conference on Document Analysis and Recognition (ICDAR'01), pp. 1263-1266, 2001.

[WAG74] R.A. Wagner et M.J. Fisher, "The String-to-String Correction Problem", Journal of the ACM, vol. 21, no. 1, pp. 260-265, 1974.

[WAL84] K. Wall et P. Danielsson, "A fast sequential method for polygonal approximation of digitized curves", Computer Vision, Graphics and Image Processing, vol.28, pp.220-227, 1984.

[WES00] T. Wessels et C. W. Omlin, "A hybrid system for signature verification", International Joint Conference on Neural Networks, vol. 05, no. 5, pp. 5509, 2000.

[WIR03] M. Wirotius, A. Seropian et N. Vincent, "Writer Identification from Gray Level Distribution", Proceedings of the International Conference on Document Analysis and Recognition (ICDAR'03), Edinburgh (Ecosse). pp. 1168-1172, 2003.

[WIR04a] M. Wirotius, J.-Y. Ramel, N. Vincent, "Selection of Points for On-Line Signature Comparison", International Workshop On Frontiers in Handwriting Recognition (IWFHR), Tokyo (Japon), pp. 503-508, 2004.

[WIR04b] M. Wirotius, J.-Y. Ramel, N. Vincent, "Improving DTW for Online Handwritten Signature Verification", International Conference in Image Analysis and Recognition (ICIAR), Porto (Portugal). pp. 786-793, 2004.

[WIR04c] M. Wirotius, J.-Y. Ramel, N. Vincent, "New Features for Authentication by On-Line Handwritten Signatures", International Conference in Biometric Authentication (ICBA), Hong-Kong, pp. 577-583, 2004.

[WIR97] B. Wirtz, "Average prototypes for stroke-based signature verification", Proceedings of the International Conference on Document Analysis and Recognition (ICDAR'97), vol.1, pp. 268-272, 1997.

[WUL97] Q.-Z. Wu, S.-Y. Lee et I.-C. Jou, "On-line signature verification based on split-and-merge matching mechanism", Pattern Recognition Letters, vol. 18, pp.665-673, 1997.

[WUL98] Q.-Z. Wu, S.-Y. Lee et I.-C. Jou, "On-line signature verification based on logarithmic spectrum", Pattern Recognition, vol. 31, no.12, pp.1865-1871, 1998.

[XUH96] Y. Xuhua, T. Furuhashi, K. Obata et Y. Uchikawa, "Selection of features for signature verification using the genetic algorithm", Computers and Industrial Engineering, vol. 30, no. 4, pp.1037-1045, 1996.

[YAN95] L. Yang, B. K. Widjaja et R. Prasad, "Application of Hidden Markov Models for signature verification", Pattern Recognition, vol. 28, no.2, pp.161-170, 1995.

[YEU04] D.-Y. Yeung, H. Chang, Y. Xiong, S. George, R. Kashi, T. Matsumoto et G. Rigoll, "SVC2004: First International Verification Competition", International Conference in Biometric Authentication (ICBA), Hong-Kong, pp. 16-22, 2004.

[ZHA96] P. Zhao, A. Higashi, et Y. Sato, "On-line signature verification by adaptively weighted DP matching", IEICE Transactions on Information and Systems, vol. E79-D, no. 5, pp. 535–541, 1996.

ANNEXE

Protocole d'Acquisition de signatures manuscrites en ligne

Nous indiquons ici les différentes étapes de la constitution de la base ATOS. Les participants doivent d'abord être convaincus de l'intérêt pour la société et de la sécurité assurée par le protocole d'expérimentation. Puis viennent l'enrôlement et l'acquisition de signatures de test qui doivent être le plus convivial possible.

Constitution d'une base de signatures

1. CADRE DU PROJET

Le but du projet est l'évaluation et la mise au point d'une technique d'authentification par signature manuscrite en ligne permettant d'activer des opérations de signatures numériques par certificats x509.

1.1. La motivation

Il n’est pas possible de valider un système d’authentification de signature sans une expérimentation à grande échelle réalisée sur des signatures naturelles, variées et représentatives. Il est donc nécessaire de créer une base de signatures répondant à un certain nombre de critères.

Rappelons quel est le cadre d’utilisation d’un logiciel d’authentification par signature. Il se déroule en deux phases distinctes. D’abord l’enrôlement a pour but d’apprendre les caractéristiques du signataire, il repose sur quelques signatures servant de modèle. Ensuite des authentifications proprement dites quand l’utilisateur en éprouve le besoin. L’authentification doit être faite à partir d’une signature qui est comparée aux modèles établis.

1.2. Les garanties

- Les personnes participant à la formation de la base de signatures sont informées sur le but de sa création.
- Les signatures sont transformées en fichier de données brutes (suite de données spatio-temporelles) puis cryptées. En aucun cas, les signatures ne sont conservées sous forme d'images.
- L'utilisation des signatures sera réalisée sans jamais faire mention du nom du signataire et sans reproduction d'une image de la signature.
- Les informations concernant le signataire ne seront pas liées à l'identité de la personne, elles seront utilisées pour des statistiques et l'établissement de corrélations. L'anonymat est totalement assuré.
- L'acquisition se fera sur 100 personnes et sera basée sur le volontariat.
- La durée totale de l'enregistrement est inférieure à 10 minutes par volontaire.
- Les signatures réalisées doivent être les signatures habituelles de l'individu et réalisées de façon naturelle.
- La base de fichiers constituée ainsi que les données personnelles seront détruites à la fin de l'expérimentation.

1.3. Enregistrement de l'individu

La personne doit remplir un formulaire contenant les informations suivantes : âge, sexe, main utilisée pour l'écriture, langue maternelle et fréquence usuelle d'utilisation de la signature (Paragraphe 3).

1.4. Protocole d'enregistrement de signatures

L'acquisition est faite sur 2 supports différents : PDA (Palm OS) et Tablet PC.

Par conséquent les actions décrites ci-dessous sont à répéter 2 fois, une fois pour chaque support.

- Entraînement à la signature sur tablette graphique jusqu'à obtenir une signature que la personne reconnaît comme conforme

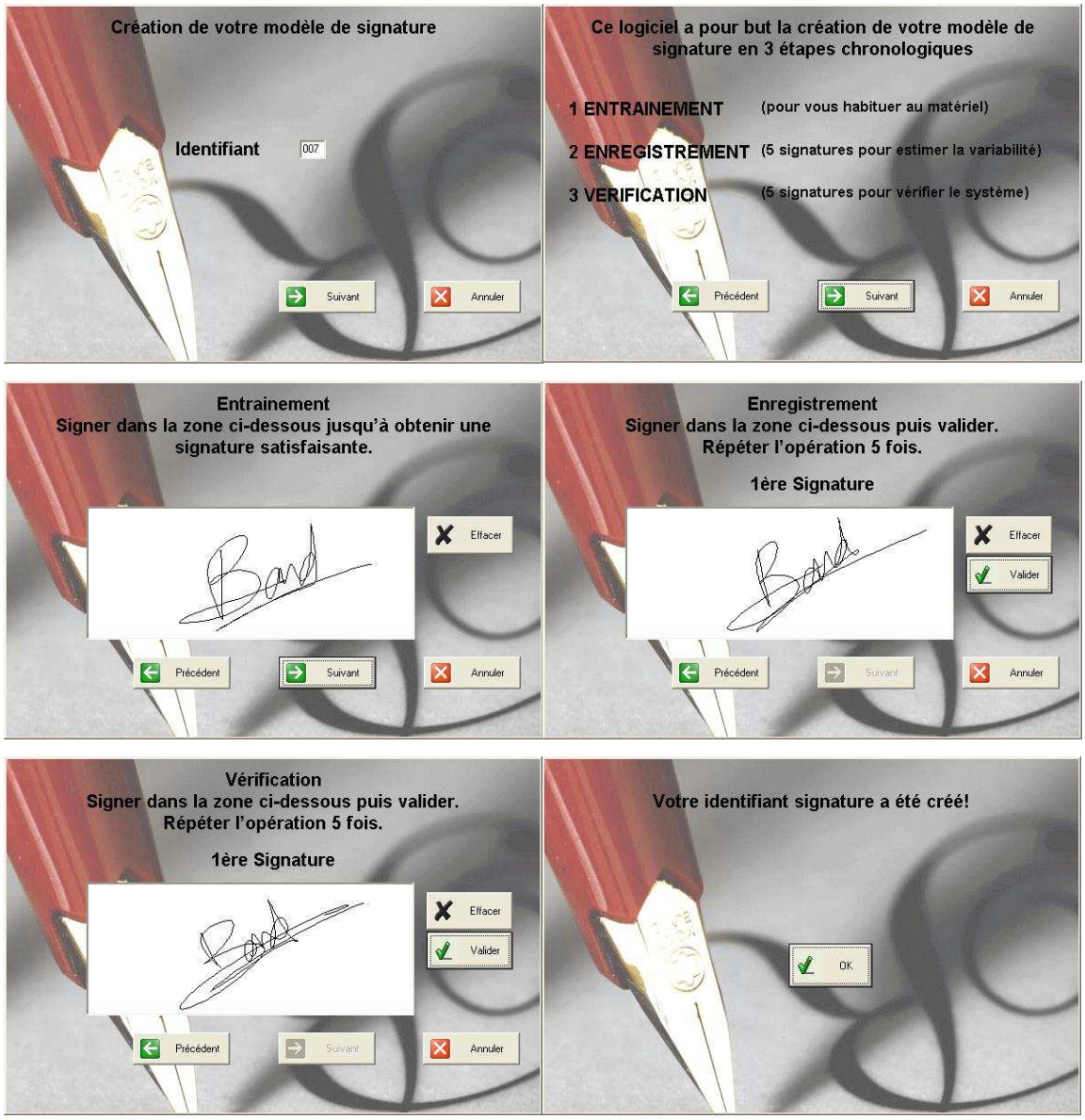
- Enregistrement du modèle de la signature simulant un enrôlement (5 signatures)
- Enregistrement de signatures de test simulant une authentification (5 signatures)

On obtient ainsi un minimum de 20 signatures pour chaque individu : 10 sur chaque support.

1.5. Lieu d'acquisition

Bureau i 116

2. INTERFACE DU LOGICIEL D'ACQUISITION



3. FORMULAIRE

Numéro	Age	Sexe (M/F)	Langue Maternelle	Ecriture (Gaucher/Droitier)	Fréquence d'utilisation de la signature (Plusieurs fois par : Jour/Semaine/Mois)	Accord pour utiliser la signature
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						

4. RESULTATS

Une invitation pour la présentation de l'avancement de l'ensemble des travaux sera envoyée aux volontaires.