République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la recherche scientifique
Université 8Mai 1945 – Guelma
Faculté des sciences et de la Technologie
Département d'Electronique et Télécommunications

Mémoire de fin d'étude
pour l'obtention du diplôme de Master Académique

Domaine : **Sciences et Technologique**
Filière : **Electronique**
Spécialité : **Systèmes Electroniques**

Vérification des signatures manuscrite : cas de faux par imitation

---

Présenté par :
Naili Yaaqoub

Atrous Mounia

Sous la direction de :

Bourouba Houcin

Mai 2014

Table de matière

Table de Figure

Table d'abréviation

LED	Light Emitting Diode
PIN	Personal Identification Number
BD	Base de Données
FRR	False Rejection Rate
FAR	False Acceptancc Ratc
EER	Equal Error Rate
DET	Detection Error Tradeoff
TVA	Taux de Vraie Acceptation
TFA	Taux de Fausse Acceptation
TEE	Taux d'Erreur Egale
ROC	Receiver Operating Characteristic
T.B.C	Taux de personne Bien Classée
CMC	Cumulative Match Characteristics
DPI	dot per inch
LBP	Local binary pattern
LBPH	Local binary pattern histogramme
NDG	Niveau de gris
SVM	Séparateurs à vastes marges « support vector machines »
LS-SVM	Least squares support vector machines
RLSC	Regularized Least-Squares Classification

KKT	Karush–Kuhn–Tucker
RBF	radial basis function
OCR	Optical Character Recognition
KNN	K de plus proche voisin « K Nearest Neighbors »
GPDS	Groupe de Traitement numérique du signal

Remerciements

Nous tenons tout d'abord à remercier à tout instant

DIEU tout puissant qui nous a éclairé la vie par

le savoir et nous a aidé à réaliser ce travail de fin

d'études.

Nous remercions, après, toute personne qui nous

a conseillé, guidé, encouragé et soutenu tout au

long de cette année et qui a contribué de prés ou de

loin à l'aboutissement de ce travail.

En particulier notre encadreur :

Mr :BOUROUBA Hocine

Pour sa disponibilité, son aide précieuse ainsi que pour

l'attention qu'il nous a porté tout au long

de la réalisation de ce projet...

Dédicace

Pour la plus chère de la présence ... de la tour affectif doux et publié

...à La seule motivation pour poursuivre ma carrière à l'école

Pour la m'a appris que la vie est une lutte et le succès le plus récent, et à la ... volonté du cœur pour voir

... Pour le pur esprit de ma mère

pour m'on chère Papa

... De faire le deuil de leur peine et se réjouir de joie

Sœurs et frères long de ma carrière et l'école Cindy

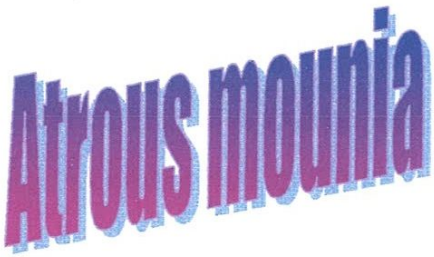
.Plus cher pour les gens, mon mari, "Hocine"

Pour l'un d'eux le sens le plus élevé de la fraternité et de l'amitié de mes amis Loyal

Pour ma carrière que je n'ai jamais connue dans toute l'école

et

Pour touts les étudient de ma promo mastère



Dédicaces

Je dédie humblement ce mémoire à :

A celui qui m'a toujours ouvert ses bras et soutenue dans tout ce que j'ai entrepris ; ce lui qui a su être bon, gentille et compréhensive avec moi ; celui dont je regrette l'absence à cette étape importante de ma vie ; celle qui me manque terriblement aujourd'hui mon très chère et adorée papa.

A celle qui s'est toujours de vouée et sacrifiée pour moi ; celle qui m'a aidée du mieux qu'elle pouvait pour réussir ; celle qui m'a accompagnée tout au long de ce parcours périlleux ; celle qui a toujours été là dans mes moments de détresse , ma très chère mère.

A ceux qui m'on toujours encouragées et soutenues moralement, mes très chers amies.

A mes très chère sœurs Saoussen et à m'on très cher frère Abd Elghafour qui ma énormément aidée et à qui je témoigne mon affection et ma profonde reconnaissance.

A mon très cher professeur bourouba houssine qui m'a toujours encouragée et soutenue depuis le début de ma thèse ; celui qui a toujours su trouver les mots pour me redonner la force de continuer et d'aller au bout de cette aventure qu'est la thèse ! !

Yakoub Naili

Introduction Générale

Depuis les attentats du 11 septembre 2001 aux Etats-Unis, et la médiatisation qui en a été faite, la sécurité est devenue une préoccupation internationale. Une des conséquences a été un meilleur contrôle du taux migratoire par les pays, en s'assurant mieux de l'identité des voyageurs. Cette préoccupation de sécurité est d'autant plus aigue dans notre temps moderne que les moyens de communication se multiplient. La nécessité de la protection civile d'une part et la lutte contre les fraudes et les crimes d'autre part, placent au centre un dispositif sécuritaire pour de nombreux domaines comme par exemple le transport, le secteur bancaire, les services publics, etc. Le dénominateur commun, est d'offrir des moyens simples, pratiques, fiables, pour vérifier ou identifier une personne, sans l'assistance d'une autre personne.

Afin de répondre à ces besoins liés à la sécurité, la biométrie se présente comme une technologie potentiellement puissante. En effet, les différents moyens biométriques visent à utiliser des caractéristiques comportementales et/ou physiologiques spécifiques à chaque personne.

La vérification de la signature est parmi les axes les plus importants de la vérification biométrique de l'identité. En effet, la signature a toujours été le moyen le plus accepté socialement et légalement pour l'identification et l'authentification de tout document officiel. Elle est facile à acquérir, elle résulte d'un geste spontané et propre à chaque individu. D'autre part, la mise en place d'un système de vérification de signatures est moins coûteuse et plus simple que celle des systèmes biométriques basés, par exemple, sur l'identification de l'iris ou du visage.

Nous avons choisi d'articuler notre étude autour de trois chapitres principaux : Le premier chapitre est consacré à la présentation générale de la biométrie ainsi que sur la performance des systèmes biométriques, Les différentes modalités comportementales et/ou physiologiques, système biométrique.

Le deuxième chapitre est consacré a la signature manuscrite en générale, les type de signature , types de faux signature, Principes de fonctionnement et extraction de caractéristiques

Le troisième chapitre est une exposition de quelque différentes bases de donné basons sur les bases de données GPDS960 et MCTY-75 et finalement une explication de la partie expérimentale de nôtre projet.

Chapitre I

Biométrie générale

I.1 Définition

Le terme "biométrie" provient des mots grecs, «bios» qui veut dire la vie et du mot «métrique» qui veut dire mesure. La biométrie est la science permettant l'identification d'individus à partir de leurs caractéristiques physiologiques ou comportementales[1] .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. [2]

I.2 caractéristique d'une modalité biométrique

Un certain nombre de caractéristiques sont utilisées dans diverses applications. Chaque trait biométrique a ses avantages et ses inconvénients, c'est pourquoi, le choix de la technique pour une application particulière dépend d'une variété de questions en plus de sa performance. Jain et al [3] ont identifié sept facteurs déterminant la convenance des traits physiques ou comportementaux pour être utilisés dans une application biométrique : [4]

- **Universalité** : toute personne ayant accès à l'application doit posséder le trait.
- **Unicité** : le trait doit être suffisamment différent d'une personne à une autre.
- **Permanence** : le trait biométrique d'une personne doit être suffisamment invariant au cours d'une période de temps.
- **Mesurabilité** : il devrait être possible d'acquérir et de numériser les données biométriques à l'aide d'un dispositif approprié.
- **Performance** : la précision de la reconnaissance et les ressources nécessaires pour atteindre la précision que doit satisfaire les contraintes imposées par l'application.
- **Acceptabilité** : les individus qui vont utiliser cette application doivent être disposés à présenter leurs traits biométriques au système.
- **Contournement** : il s'agit de la facilité avec laquelle le caractère d'un individu peut être imité en utilisant des objets (par exemple : faux doigts dans le cas de traits physique et le mimétisme, dans le cas de traits de comportement).

I.3 Les différentes modalités

I.3.1 L'analyse morphologique

Elle est basée sur l'identification de traits physiques particuliers qui, pour toute personne, sont uniques et permanents. Cette catégorie regroupe l'iris de l'œil, le réseau veineux de la rétine la forme de la main, les empreintes digitales, les traits du visage, les veines de la main, etc.

a) Empreintes digitales

A l'heure actuelle la reconnaissance des empreintes digitales est la méthode biométrique la plus utilisée. Les empreintes digitales sont composées de lignes localement parallèles présentant des points singuliers (minuties) et constituent un motif unique, universel et permanent. Pour obtenir une image de l'empreinte d'un doigt, les avancées technologiques ont permis d'automatiser la tâche au moyen de capteurs intégrés, remplaçant ainsi l'utilisation classique de l'encre et du papier. Ces capteurs fonctionnant selon différents mécanismes de mesure (pression, champ électrique, température) permettent de mesurer l'empreinte d'un doigt fixe positionné sur ce dernier (capteur matriciel) ou en mouvement (capteurs à balayage).

L'image d'empreinte d'un individu est capturée à l'aide d'un lecteur d'empreinte digitale puis les caractéristiques sont extraites de l'image puis un modèle est créé. Si des précautions appropriées sont suivies, le résultat est un moyen très précis d'authentification.

Les techniques d'appariement des empreintes digitales peuvent être classées en deux catégories : les techniques basées sur la détection locale des minuties et les techniques basées sur la corrélation. L'approche basée sur les minuties consiste à trouver d'abord les points de minuties puis trace leurs emplacements sur l'image du doigt (figure1.1)

Figure 1.1 : reconnaissance par empreinte digitale

Cependant, il y a quelques difficultés avec cette approche lorsque l'image d'empreinte digitale est d'une qualité médiocre, car l'extraction précise des points de minutie est difficile. Cette méthode ne tiens pas en compte la structure globale de crêtes et de sillons.

Les méthodes basées sur la corrélation sont capables de surmonter les problèmes de l'approche fondée sur les minuties. Ces méthodes utilisent la structure globale de l'empreinte, mais les résultats sont moins précis qu'avec les minuties. De plus, les techniques de corrélation sont affectées par la translation et rotation de l'image de l'empreinte. C'est pour cela que les deux approches sont en général combinées pour augmenter les performances du système. [2]

b) L'iris

L'identification d'iris est devenue une technique biométrique populaire. Elle est généralement reconnue qu'étant peut-être la technique la plus précise. En conséquence, c'est une technique utile que ce soit pour l'assortiment linéaire aux fins de vérification individuelle d'identité, ou un assortiment un à- plusieurs aux fins d'identifier un iris particulier parmi plusieurs dans une grande base de données. En outre, l'exécution opérationnelle relative de l'identification d'iris peut être très bonne. Dans des réalisations antérieures, le défaut d'acquisition d'image de qualité appropriée dans de vraies conditions de fonctionnement pouvait être un problème, également pour la possibilité d'acquérir des modèles référentiels de bonne qualité. Cependant, la technique a rapidement évolué et de tels problèmes sont rarement rencontrés aujourd'hui. Les lecteurs d'identification d'iris ont tendance à être un peu plus chers que ceux pour certaines autres techniques, en grande partie en raison de leur complexité relative. En outre, l'installation et le commandement peuvent être un peu plus exigeants, particulièrement en ce qui concerne le placement environnemental et l'accommodation pour une large gamme d'individus de taille physique différente. Toutefois, de tels soucis de déploiement peuvent être surmontés et peuvent être considérés insignifiants pour des applications où l'exactitude et la performance de l'identification d'iris est exigée. En termes simples, la technique implique la localisation de l'iris dans un visage humain, le séparant de la pupille et de la sclérotique, divisant l'iris évident en segments et analysant chaque segment en conséquence. De cette analyse, un code relativement sophistiqué d'iris peut être dérivé et comparé à une référence précédemment stockée. La quantité de détails représentée dans le code d'iris permet un niveau important de confiance en entreprenant les comparaisons, même en recherchant dans des bases de données très grandes. Ceci est facilité par la quantité de l'information disponible qui peut être dérivée d'un iris typique, et l'unicité relative de l'iris dans la population humaine. En effet, même les iris gauche et droit du même individu ont tendance à être distincts et des iris

sont considérés comme invariables durant toute la vie, une fois fixés peu de temps après la naissance. L'identification par iris s'accroit en popularité ces dernières années et c'est une technique qui continuera sans doute à être employée couramment.

Figure 1.2 : Photo d'iris

c) Le visage

L'identification par visage a été disponible comme technique biométrique pendant longtemps, bien qu'elle soit probablement juste pour indiquer que les réalisations primaires ont laissé à désirer en termes d'exactitude et fiabilité de comparaison. Cependant, la technique a beaucoup d'applications potentielles, et le développement continu a assuré qu'il a rapidement mûri dans une technique opérationnelle viable. Typiquement, la technique implique la métrique des et entre caractéristiques distinctes dans le visage, se fondant moins sur des facteurs d'une nature transitoire tels que la coupe de cheveux ou l'utilisation des produits de beauté. Néanmoins, le visage humain est sujet au changement avec le temps et cette réalité demeurera un défi pour des systèmes d'identification de visage, comme le changement d'expression, la maladie, la vieillesse et d'autres facteurs normaux. En outre, les facteurs humains et environnementaux joueront presque toujours un très grand rôle dans l'efficacité d'un système d'identification de visage, dans un scénario donné de déploiement. En conséquence, l'identification de visage peut tout à fait ne pas égaler l'exactitude fournie par certaines autres techniques. Cependant, elle se prête aisément aux applications où le visage est déjà employé dans un contexte de vérification d'identité. De même, la capacité de comparer avec une image stockée, peut-être d'une source différente, semblera attrayante dans quelques applications de secteur public. L'identification de visage a été parfois employée en même temps qu'une autre biométrie afin d'augmenter la confiance en procédé de vérification d'identité. Le visage et l'empreinte digitale sont une combinaison populaire dans ce contexte. Tout en n'offrant pas les niveaux superlatifs de l'exactitude ou de l'exécution opérationnelle, l'identification de visage

néanmoins demeure une technique populaire, et une de celles qui tireront bénéfice sans doute d'un développement ultérieur. [5]

Figure 1.3 image de visage

d) La géométrie de la main

La géométrie de la main est une technologie biométrique récente. Comme son nom l'indique, elle consiste à analyser et à mesurer la forme de la main, c'est-à-dire mesurer la longueur, la largeur et la hauteur de la main d'un utilisateur et de créer une image 3-D. Des LEDs infrarouges et un appareil-photo numérique sont utilisés pour acquérir les données de la main. Cette technologie offre un niveau raisonnable de précision et est relativement facile à utiliser. Cependant elle peut être facilement trompée par des jumeaux ou par des personnes ayant des formes de la main proches. Les utilisations les plus populaires de la géométrie de la main comprennent l'enregistrement de présence et le contrôle d'accès. Par contre, les systèmes de capture de la géométrie de la main sont relativement grands et lourds, ce qui limite leur utilisation dans d'autres applications comme l'authentification dans les systèmes embarqués : téléphones portables, voitures, ordinateurs portables, etc.

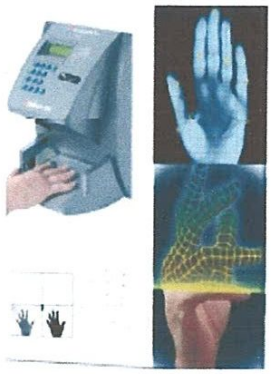


Figure 1.4 : Dispositif de reconnaissance par géométrie de la main

e) La rétine.

Le balayage rétinien est une technique biométrique primaire, développée au début pour le contrôle d'accès dans les environnements militaires. Son exécution donnait de très bons résultats sous certaines conditions. Cependant, sa rentabilité était en général plutôt ennuyeuse, au moins en ce qui concerne les réalisations primaires, bien qu'elles soient améliorées dans des essais postérieurs. C'est principalement parce que son utilisation, à l'origine, imposait une fixation d'un dispositif binoculaire et d'aligner sa vision sur une cible chose que beaucoup de personnes ont, au début, eu du mal à faire- particulièrement ceux dont la vision est altérée. En outre, beaucoup d'utilisateurs n'ont pas beaucoup apprécié l'idée du contact physique avec l'interface binoculaire. En conséquence, alors que l'utilisation dans un environnement militaire commandé a pu être acceptable (en grande partie parce que de tels utilisateurs n'ont eu aucun choix dans la matière) la technique trouvait peu de faveur au sein de la communauté intégrale. La technique d'exploration rétinienne impliquait de balayer les modèles de veine de la rétine avec un faisceau actionné bas brillant à l'intérieur de l'œil : une fonction intrusive qui n'a pas été typiquement considérée comme une proposition attrayante par les utilisateurs potentiels. En outre, les premières versions des modules de balayage rétiniens étaient excessivement chères pour n'importe qui en dehors des militaires. Les versions qui ont suivi sont devenues beaucoup moins coûteuses et étaient un peu mieux considérées en termes de connectivité, intégration de systèmes et interface utilisateurs [5]

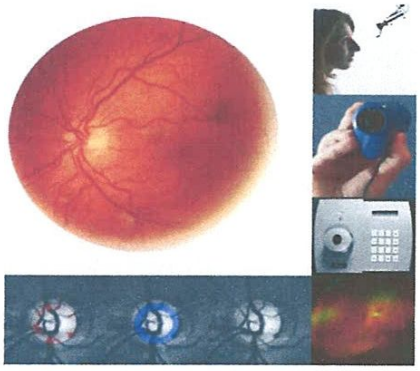


Figure 1.5 : image de rétine

I.3.2 L'analyse comportementale

Elle se base sur l'analyse de certains comportements d'une personne. Cette catégorie regroupe la reconnaissance vocale, la dynamique de frappe au clavier, la dynamique de la signature,

l'analyse de la démarche, etc. Il existe, par ailleurs, une autre catégorie qui est l'étude des traces biologiques telles que : l'ADN, le sang, la salive, l'urine, l'odeur, etc.

a) La voix

De tous les traits humains utilisés dans la biométrie, la voix est celle que les humains apprennent à reconnaître dès le plus jeune âge. Les systèmes de reconnaissance de locuteur peuvent être divisés en deux catégories : les systèmes dépendant du texte prononcé et les systèmes indépendants du texte. Dans le premier cas, l'utilisateur est tenu d'utiliser un texte (un mot ou une phrase) fixe prédéterminé au cours des séances d'apprentissage et de reconnaissance. Alors que, pour un système indépendant du texte le locuteur parle librement sans texte prédéfini.

Cette dernière catégorie est plus difficile, mais elle est utile dans le cas où l'on a besoin de reconnaître un locuteur sans sa coopération. La recherche sur la reconnaissance de locuteur est en pleine croissance, car elle ne nécessite pas de matériel cher, puisque la plupart des ordinateurs personnels de nos jours sont équipés d'un microphone. Toutefois, la mauvaise qualité et le bruit ambiant peuvent influencer la vérification et par suite réduire son utilisation dans les systèmes biométriques. Dans un système de reconnaissance de locuteur le signal est premièrement mesuré puis décomposé en plusieurs canaux de fréquences passe-bande. Ensuite, les caractéristiques importantes du signal vocal sont extraites de chaque bande.

Parmi les caractéristiques les plus communément utilisées sont les coefficients Cepstraux. Ils sont obtenus par le logarithme de la transformée de Fourier du signal vocal dans chaque bande. Finalement, la mise en correspondance des coefficients Cepstraux permet de reconnaître la voix. Dans cette étape, généralement on fait appel à des approches fondées sur les modèles de Markov cachés, la quantification vectorielle, ou la déformation temps dynamique.

Figure 1.6 : image de voix

b) La dynamique de frappe

C'est une autre technique primitive dans laquelle un énorme apport en temps et en effort a été investi, notamment par quelques grandes compagnies de technologie de l'information. L'idée

d'identifier un individu par sa dynamique particulière de frappe était clairement attrayante parmi les perspectives de la technologie de l'information et des réseaux.

Tandis qu'il semblait possible de déterminer une signature dynamique individuelle de frappe dans des conditions soigneusement contrôlées, les utilisateurs réels sous de réelles conditions de fonctionnement n'étaient pas aussi cohérents qu'on le voudrait dans la manière d'utiliser un clavier afin de mettre en application cette technologie. En outre, en utilisant les claviers standards, il n'y avait pas vraiment une richesse d'information individualiste avec laquelle travailler. Après beaucoup de recherches et quelques démonstrations intéressantes, l'idée de la dynamique de frappe en tant que technique biométrique comportementale viable semblait se faner, particulièrement quand d'autres techniques ont été vues accomplir de bons progrès. [5]

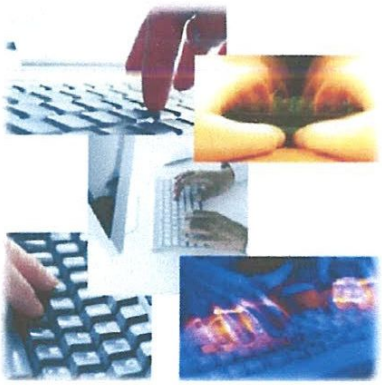


Figure 1.7 : image de frape

c) La démarche

L'attraction potentielle de l'identification de démarche se situe dans la capacité d'identifier un individu à distance. Cependant, il y a des défis sérieux à surmonter à cet égard. L'idée qu'un individu marche typiquement avec une démarche unique est intéressante et, sous des conditions de laboratoire, le concept de l'identification de démarche peut être démontré. Cependant, la vie réelle est pleine de désaccords dynamiques qui rendent l'exécution d'un tel système particulièrement difficile.

En plus des complexités de comparaison, il y a des facteurs tels que l'occasion de saisir l'image mobile d'un individu en isolement et dont le détail est suffisant pour pouvoir entreprendre une telle comparaison. La création d'un modèle fiable est également quelque chose qui présente de vrais défis. L'identification de la démarche représente un exemple intéressant de la recherche biométrique conduite par une condition perçue : dans ce cas-ci, pour identifier un individu à une distance au-delà de laquelle la biométrie de contact et à bout-portant ne peuvent fonctionner. C'est peut-être une idée attrayante pour des applications

militaires et de très haute sécurité, mais il est douteux que l'identification par la démarche deviendra une technique biométrique courante. [5]

Figure 1.8 : image de la démarche

d) La signature

La vérification par signature comme technique était parmi les premières utilisées dans le domaine de la biométrie. Il y avait plusieurs systèmes concurrents dans ce domaine. Elle semblait être une application évidente de la biométrie car il y avait tant de processus familiers qui avait utilisé la signature comme moyen de vérification d'identité En outre, la signature biométrique, du moins en théorie, fournissait une profondeur d'analyse autre que celle de la mesure de la dynamique inhérente dans son écriture, la précision géométrique de la signature. Dans des tests indépendants, la vérification de la signature a donné une raisonnable présentation d'elle-même. Cependant, dans les situations réelles, l'utilisation des tablettes graphiques disponibles dans le marché et les systèmes adéquats n'était souvent pas une chose aussi aisée. En outre, il est intéressant, en termes proportionnels, de voir les incohérences de certaines personnes en signant leur nom dynamiquement et graphiquement. Tandis qu'un observateur humain peut tolérer ces incohérences tant que la signature est correcte, l'algorithme de vérification automatique de la signature prenait un temps important, particulièrement quand il essayait de fonctionner avec un niveau de tolérance serré. En conséquence, la vérification par signature biométrique reste une technique traditionnelle, bien qu'il puisse y avoir des applications où elle peut s'avérer utile. [5]

Figure 1.9 : image de signature

I.4 Système biométrique

Un système biométrique est essentiellement un système de reconnaissance de formes qui fonctionne en acquérant des données biométriques à partir d'un individuel, extrayant un ensemble de caractéristiques à partir des données acquises, et comparant ces caractéristiques contre la signature dans la base de données. Selon le contexte d'application, un système biométrique peut fonctionner en mode de vérification ou mode d'identification:

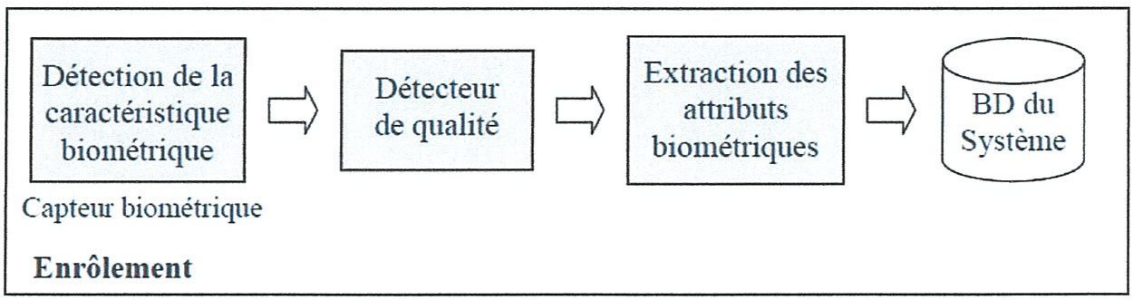
I.4.1 L'identification

Elle permet d'établir l'identité d'une personne à partir d'une base de données, le système biométrique pose et essaye de répondre à la question, "qui est la personne X ? ", il s'agit d'une comparaison du type un contre plusieurs (1:N).

I.4.2 La vérification

Le système biométrique demande à l'utilisateur son identité et essaye de répondre à la question, "est-ce la personne X? ". Dans une application de vérification l'utilisateur annonce son identité par l'intermédiaire d'un mot de passe, d'un numéro d'identification, d'un nom d'utilisateur, ou toute combinaison des trois. Le système sollicite également une information biométrique provenant de l'utilisateur, et compare la donnée caractéristique obtenue à partir de l'information entrée, avec la donnée enregistrée correspondante à l'identité prétendue, c'est une comparaison un à un (1:1). Le système trouvera ou ne trouvera pas d'appariement entre les deux. La vérification est communément employée dans des applications de contrôle d'accès et de paiement par authentification [6].

Les schémas d'un système de vérification et d'un système d'identification sont illustrés dans la figure 10; le processus d'enrôlement, qui est commun à ces deux tâches est également illustré. Le module d'enrôlement correspond à l'enregistrement biométrique des individus dans la base de données du système. Pendant la phase d'enrôlement, la caractéristique biométrique d'un individu est capturée par un lecteur biométrique pour produire une représentation numérique



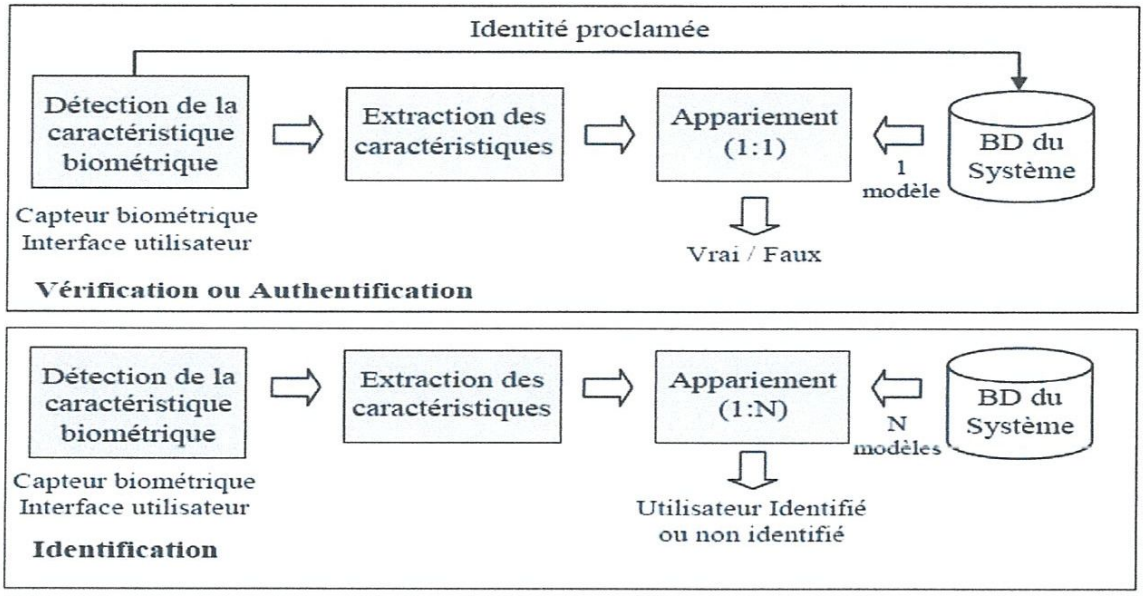


Figure 1.10 : Schéma de fonctionnement d'un système biométrique. Diagrammes des processus d'enrôlement, de vérification et d'identification

Le processus de vérification est établi pour authentifier une personne se présentant à un point d'accès. Durant la phase de vérification, le nom de l'utilisateur ou le code PIN (*Personal Identification Number*) est entré par l'intermédiaire d'un clavier. Le lecteur biométrique capte la biométrie de l'individu et la convertit à un format numérique, qui est traité ultérieurement par la fonction d'extraction pour produire une signature compacte qui représente l'identité de l'utilisateur. Cette signature numérique (appelée aussi vecteur caractéristique) est présentée à la fonction d'appariement, qui la compare avec le modèle proclamé par l'utilisateur et qui est extrait de la base de données (BD) du système à partir de son code PIN.

Le processus d''identification ne nécessite pas de code PIN. Le système compare la signature biométrique saisie avec les modèles de tous les utilisateurs dans la base de données du système ; la sortie est généralement l'identité d'un utilisateur enrôlé ou un message du genre «utilisateur non Identifié» si l'individu n'est pas enregistré dans la base de données du système. L'identification dans les grandes bases de données est ainsi coûteuse en termes de complexité et efficacité de calcul. Les techniques de classement et d'indexation s'imposent pour limiter le nombre de modèles qui doivent être comparés avec la signature de l'utilisateur.

La procédure d'enrôlement est opérée généralement d'une façon semi automatique suivant un mode hors ligne. En effet, l'enrôlement peut être supervisé par un expert ou un agent de police qui peut orienter le processus d'acquisition quand un criminel est enrôlé, par exemple. D'autre part, les processus d'authentification et d'identification s'opèrent en ligne. Ces

procédures de reconnaissance en ligne doivent s'exécuter rapidement car une réponse immédiate est imposée dans la majorité de ces applications. [7]

I.5 performance des systèmes biométriques

La performance mesure l'efficacité et la fiabilité d'un système biométrique dans un contexte d'utilisation donné .nous présentons les différentes pour quantifier la performance d'un système biométrique

Les mesure des taux d'erreur son divisées en deux classes qui sont : les taux d'erreur de systèmes de vérification et les taux d'erreur de systèmes d'identification .

I.5.1 Taux d'erreur de systèmes vérification

- **FRR** (False Rejection Rate) **et FAR** (Acceptance Rate)

Afin de décider si une modalité biométrique de test est authentique ou pas, le système compare le score de l'échantillon de test à un seuil de décision. Si ce score est un score de similarité, s'il dépasse le seuil, alors le système accepte l'identité proclamée, sinon il la rejette (voir Figure 1.11) Inversement, dans le cas ou le score est un score de dissimilarité, si ce dernier est inferieur au seuil alors le système accepte l'identité proclamée, sinon il la rejette.

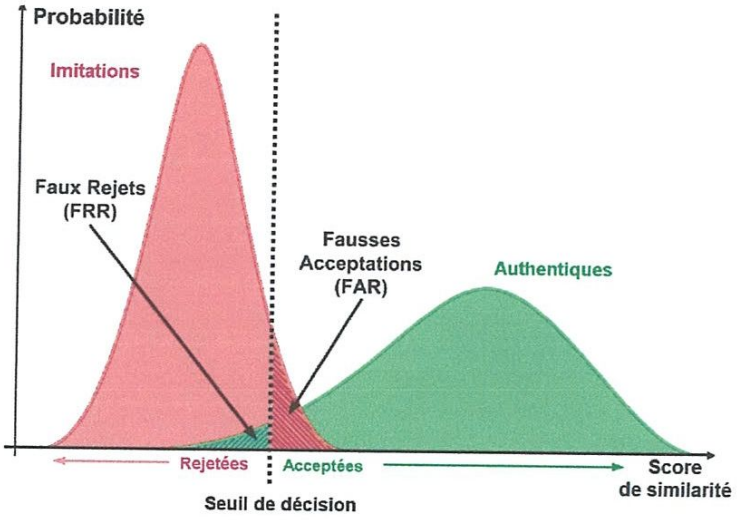


Figure 1.11: Distribution des scores des signatures authentiques et des imitations.

Lorsqu'un système fonctionne en mode vérification, celui-ci peut faire deux types d'erreurs. Il peut rejeter un utilisateur l'légitime et dans ce premier cas on parle de faux rejet. Il peut aussi accepter malencontreusement un imposteur et on parle dans ce second cas de fausse acceptation. La performance d'un système se mesure donc en se basant sur son taux de faux rejet (False Rejection Rate ou FRR) et son taux de fausse acceptation (False Acceptance Rate

ou FAR). La Figure 1.11 illustre le FRR et le FAR à partir des distributions des scores authentiques et imposteurs.

Ainsi, pour une valeur du seuil quelconque, nous calculons les taux d'erreur correspondant à ces deux types d'erreur. Le taux de faux rejets est le pourcentage des données de test authentiques qui ont été rejetées :

$$\text{FRR}=100*\frac{\text{nombre de tests authentiques rejet'es}}{\text{nombre total de tests authentiques}}*(en\%)$$

De même, le taux de fausses acceptations est le pourcentage des imitations qui ont ´été acceptées :

$$\text{FAR}=100*\frac{\text{nombre d'imitations accept'ees}}{\text{nombre total d'imitations}}*(en\%)$$

Le taux de faux rejets et celui de fausses acceptations dépend du seuil de sécurité, et sont inversement proportionnels. Plus la valeur du seuil sera grande, plus il y aura de faux rejets et moins de fausses acceptations, et inversement, plus la valeur du seuil sera petite, moins il y a aura de faux rejets et plus de fausses acceptations. Le choix de la valeur du seuil à utiliser dépend principalement de la finalité du système de vérification. Cette valeur est choisie de manière à faire un compromis adéquat entre la sécurité et l'utilité du système.

- **EER** (Equal Error Rate)

Souvent, on caractérise un système de verification de signature en-ligne par exemple par un point ou les deux courbes, FAR en fonction du seuil et FRR en fonction du seuil, se croisent. En ce point, les taux d'erreur FAR et FRR sont´égaux, ils sont donc représentées par une valeur unique qui est le taux d'erreur égal (Equal Error Rate ou EER). En fait : EER = FAR = FRR.

La valeur du EER est un indicateur de performance d'un système de verification, indépendamment des exigences d'une application spécifique (rapport entre FRR et FAR). Plus cette valeur est faible, meilleur est le système. L'inconvénient de l'EER est qu'il ne nous permet de comparer les systèmes qu'en un seul point de fonctionnement, et donc ne permet pas d'´evaluer le système `a des niveaux de sécurité différents (faible FAR par exemple).

Afin d'´evaluer un système de verification indépendamment du seuil, nous faisons varier la valeur du seuil intrinsèquement sur un intervalle donné, et pour chaque valeur du seuil, nous calculons le taux de fausses acceptations et le taux de faux rejets.

- **Courbe DET** (Detection Error Tradeoff)

Pour obtenir une représentation compacte des performances d'un système de vérification au travers d'une seule courbe, nous utilisons souvent les courbes DET (Detection Error Tradeoff)

pour représenter les FRR en fonction des FAR, comme illustré sur la Figure1.12 L'échelle est basée sur une distribution normale pour rendre la courbe plus lisible et plus exploitable. Cette représentation graphique est très utilisée pour comparer différents systèmes de vérification qui ont des performances similaires. [8]

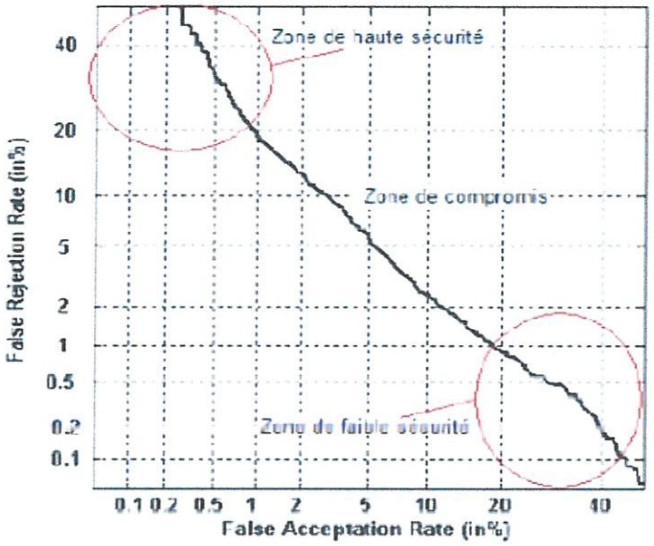


Figure1.12 : Exemple de courbe DET.

➢ TVA (Taux de Vraie Acceptation) TFA (Taux de Fausse Acceptation)

D'autres mesures d'erreurs sont nécessaires pour évaluer les systèmes d'authentification. En effet, différents cas peuvent se produire lors de la phase de reconnaissance (rejet/acceptation) :

a) les utilisateurs peuvent être acceptés de façon justifiée, quantifiés par le TVA (Taux de Vraie Acceptation)
b) les imposteurs peuvent être acceptés par erreur, quantifiés par le TFA (Taux de Fausse Acceptation)
c) les utilisateurs peuvent être rejetés à tord, on parle alors de TFR (Taux de Faux Rejet)
d) les imposteurs peuvent être rejetés, on parle de TVR (Taux de Vrai Rejet)

Souvent l'évaluation de la performance de la méthode est réalisée à l'aide des deux situations d'erreurs de classement, c'est-à-dire en utilisant le TFA et le TFR. Ces deux taux sont liés. En jouant sur les paramètres du système, notamment en faisant varier le seuil de sécurité, ils sont modifiés de façon importante. Ainsi le choix du niveau de TFA et de TFR dépend de l'application étudiée et du niveau de sécurité souhaitée. Si, par exemple, les concepteurs souhaitent une relative facilité d'utilisation c'est-à-dire, s'ils ne désirent pas que l'utilisateur

recommence plusieurs fois le processus d'authentification (surtout s'il est contraignant), on choisira un niveau de sécurité correspondant à un faible TFR, et donc avec une erreur plus grande en TFA ce qui pourrait permettre à quelques imposteurs de tromper le système . A l'inverse, s'ils désirent un système ultra sécurisé, les concepteurs imposeront alors un seuil de sécurité correspondant à un TFA proche de zéro ce qui peut faire monter le TFR aux alentours de 20 % ou plus. Le risque est alors d'empêcher une partie des utilisateurs authentiques d'accéder à la zone sécurisée.

Pour comparer de façon plus simple les méthodes d'authentification, les chercheurs utilisent fréquemment le Taux d'Erreur Egale (TEE) qui correspond à la valeur pour laquelle les deux taux d'erreur sont identiques (Figure 1.13)

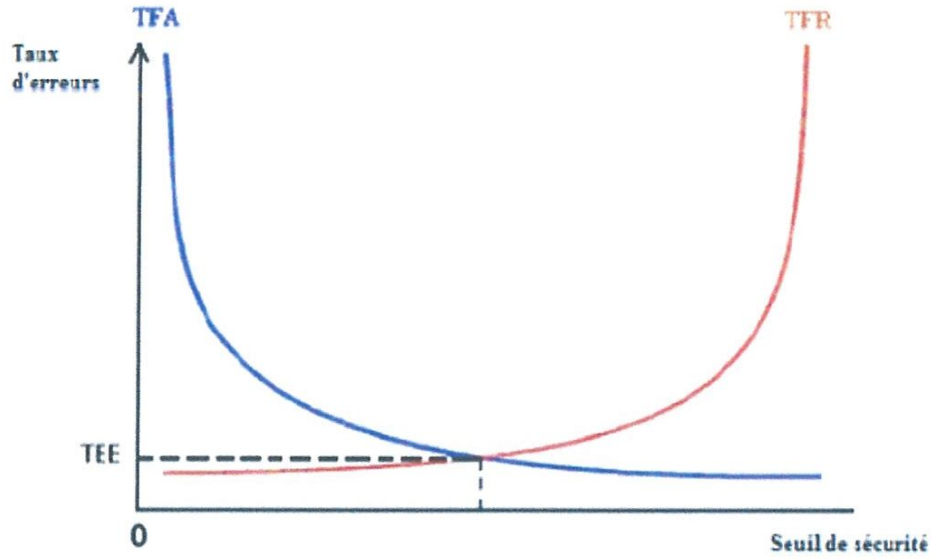


Figure 1 .13 **:** TFR, TFA et TEE

➢ **Courbe ROC** (ROC (*Receiver Operating Characteristic curve)*

Pour comparer les performances de deux systèmes biométriques il est aussi possible d'utiliser une représentation graphique appelée courbe ROC (*Receiver Operating Characteristic curve)* [9] montrant l'évolution d'un des taux en fonction de l'autre (Figure 14). Sur cette courbe chaque point a pour ordonnée le TFA et pour abscisse le TFR. Cette courbe est obtenue en faisant varier le seuil de sécurité sur une plage de valeurs prédéfinie. La comparaison de deux systèmes de performance se fait ensuite en comparant l'aire sous la courbe ROC des deux systèmes.

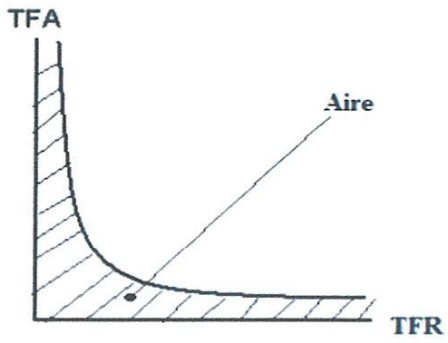


Figure 1.14 : Courbe ROC

1.5.1 Taux d'erreur de systèmes d'identification

➢ **Taux d'identification**

L'indicateur de performance utilisé pour un problème d'identification est le Taux de personne Bien Classée (T.B.C.) appelé aussi le taux d'identification

$$\text{Taux d'identification} = \frac{nembre\ des\ image\ bien\ classé}{nembre\ des\ image\ teste}$$

➢ **CMC** (*Cumulative Match Characteristics*)

Pour comparer la performance de méthodes biométriques, on peut également utiliser la courbe CMC (*Cumulative Match Characteristics*), qui indique pour un entier *n* la probabilité que le système retourne le bon identifiant pour une observation dans les *n* premières réponses fournies par le système d'identification, On trace alors la courbe (CMC) qui représente la probabilité que le bon choix se trouve parmi les *N premiers* [10].Comme l'illustre la figure 1.12

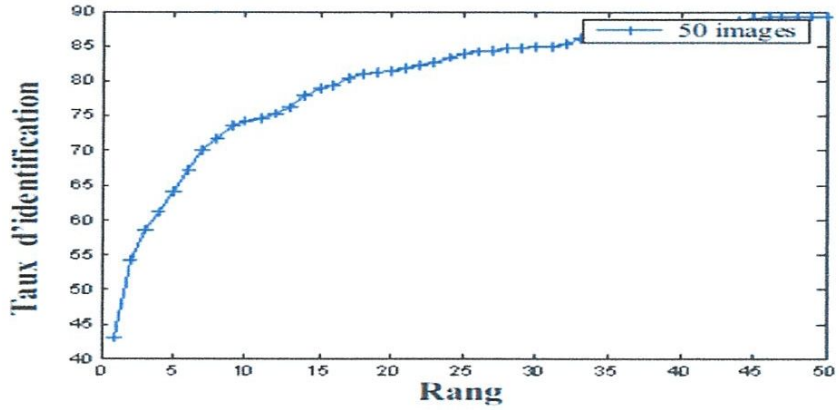


Figure 1. 15 Courbe CMC.

Conclusion

Dans ce chapitre, nous avons passé en revue les principales technologies biométriques. Ensuite, nous avons présenté dans un premier temps l'architecture de base d'un système biométrique pour se focaliser, dans un deuxième temps, sur la reconnaissance de signature manuscrite.

Ces dernières années, de nombreux systèmes pour la reconnaissance biométrique ont été proposés. Dans le chapitre suivant nous avons présenté le cadre de ce mémoire : la vérification hors ligne de la signature manuscrite

Chapitre II

La signature manuscrite

II.1 Introduction

Le terme signature provient du nom latin « *signum* », qui réfère au signe, à la marque, à l'empreinte, de même qu'au sceau et au cachet, une signification multiple qui recoupe l'évolution historique connue par la signature.[11] Les tentatives doctrinales de fournir une définition de la signature sont abondantes, mais ne présentent pas une grande diversité, la plupart d'entre elles s'articulant autour des finalités poursuivies par la signature.

Soutenant que la signature est un résultat de la coutume, nous ne pouvons faire référence qu'à des catégories coutumières. Ceci dit, nous remarquons que traditionnellement, trois éléments sont susceptibles de constituer une signature, à savoir le nom, le prénom ou les initiales du signataire.

II.2 Particularités du La signature manuscrite

La signature est avant tout un trait. Ce trait se caractérise par des paramètres physiques comme la largeur, la forme et la qualité qui en est l'aspect visuel.[12] Comme le trait résulte d'un certain dynamisme, il porte aussi une conduite spécifique, une pression et une tension.

Parmi les plus importantes caractéristiques du seing, il faudrait sans doute mentionner son caractère personnel. Tel que l'écrit le professeur Quickenborne[13], la signature implique « *un graphisme personnel, essentiellement propre au signataire.*».

Comme signe personnel, la signature indique des caractéristiques d'un individu faisant en sorte qu'on puisse le reconnaître.[14]. Selon cette vision que Syx traite de «dogmatique »,[15] le caractère personnel de la signature est assuré par la combinaison de son contenu traditionnel (le patronyme) et une caractéristique biométrique de la personne, qu'est son écriture.[16] Il existe la croyance empiriquement vérifiée, mais sans fondement théorique, que l'écriture est un paramètre strictement personnel et que deux individus ont nécessairement des écritures différentes.[17] Cependant, notons encore ici, que le caractère personnel de la signature n'est pas une donnée absolue. Une autre particularité de la signature manuscrite est son caractère inchangeable.

II.3 L'Authentification par signature manuscrite

La signature manuscrite est depuis plusieurs siècles le moyen le plus répandu pour manifester sa propre volonté. Elle est aujourd'hui, le demeurera sans doute dans le futur, le moyen biométrique d'authentification le plus utilisé [18].

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 :

- Le support et le stylet utilisés
- L'importance du document sur lequel on appose la signature
- 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 :

- ***Fonction d'identification*** : Le système d'authentification doit être suffisamment fiable pour être reconnu comme moyen de non répudiation.
- ***Fonction d'adhésion au contenu*** : Culturellement le fait d'apposer sa signature manuscrite signifie que l'on adhère au contenu indépendamment du support.
- ***Fonction de garantie de l'intégrité*** : La garantie de l'intégrité du document peut être assurée par une fonction de hachage.
- ***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.
- ***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.

II.4 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[18].

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.

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.

II.5 Type de signatures

Étant donné les différences d'origine et de culture entre les signataires à travers le monde, on distingue quatre types de signatures :

II.5.1 Le type américain

Ces signatures s'apparentent à l'écriture cursive. TI est donc possible, pour un expert qui traite ce type de signatures d'utiliser un texte manuscrit écrit par le signataire pour comparer la forme de l'écriture (figure 2.1).

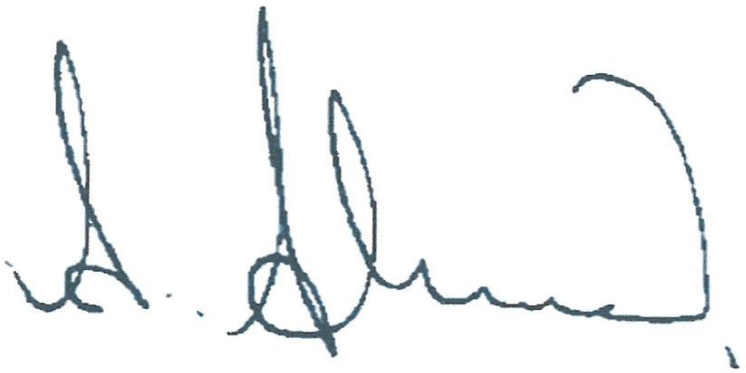


Figure 2 .1 : exemple d'une signature de type américain

II.5.2 Le type européen

Ces signatures possèdent une composante graphique importante qw oblige à un traitement global des signatures (figure 2.2)

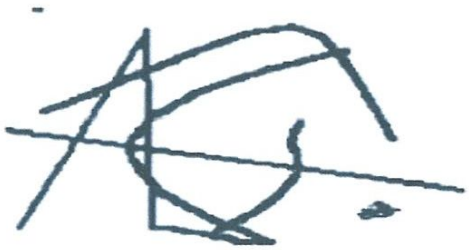


Figure 2.2 : exemple d'une signature de type européen

II.5.3 La signature de type Arabe

qui possède en partie les caractéristiques de l'écriture cursive propre à la culture Arabe et qui peut être également lisible ou très personnalisée au même titre que la paraphe.

Figure 2.3 : exemple d'une signature de type arabe

II.5.4 la signature de type Asiatique

Ce type de signature est facilement distinguable des autres types mentionnés précédemment

Figure 2.4 : exemple d'une signature de type asiatique

II.6 Fausses signatures

Nous consacrons ce paragraphe à l'énumération des types de faux. Suivant le type de faux recherché, l'expert utilise des traitements spécifiques, tel que les proportions de la signature, le nombre de parties des signatures et les projections.

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. 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.

II.6.1 Les faux aléatoires

Les signatures des personnes autres que le signataire présumé sont appelées des faux aléatoires. Pour tester les systèmes, c'est le faux le plus simple à simuler. En effet, il est difficile d'obtenir des banques de faux réels en quantité suffisante alors qu'il suffit de prendre les signatures des autres personnes dans une base de signatures. Les faux aléatoires font partie des faux grossiers et on suppose que les résultats obtenus sur ce type de faux sont identiques à ceux que l'on obtiendrait sur des faux grossiers.

II.6.2 Les faux simples

Le faussaire fabrique une signature à partir du nom du signataire sans imiter un original. Ce faux est généralement peu ressemblant surtout pour les signatures de type européen de nature graphique; il est pourtant plus difficile à détecter que des faux grossiers. Ce type de faux est intéressant d'un point de vue technique (validation) pour les systèmes traitant les signatures de type américain de nature cursive car il permet de tester les systèmes d'authentification avec des faux plus ressemblants que les faux grossiers.

Pour les faux simples, grossiers et aléatoires le tracé est spontané, les caractéristiques pseudo-dynamiques sont donc peu discriminantes, par contre la ressemblance est faible et les caractéristiques sont assez discriminantes. Ces faux sont considérés comme les plus faciles à détecter par un système automatique car ils ne nécessitent pas d'expertise et de plus, on les rencontrent couramment.

Pour les autres faux, leur aspect étant très ressemblant, il est nécessaire d'utiliser des caractéristiques pseudo-dynamiques (trais hésitant, pression, vitesse, pleins et déliés, retouches, arrêts de stylo, ...). Ces caractéristiques sont plus ou moins apparentes suivant le stylo utilisé, elles s'avèrent difficiles à extraire automatiquement.

II.6.3 Les faux par calque

Le faux par calque reproduit fidèlement l'image d'une signature authentique sur un document par l'utilisation d'un moyen de recopiage par exemple : une copie par transparence, avec du carbone, par photocopie. Ce faux est évidemment très difficile à détecter même pour les experts.

II.6.4 Les faux par imitation

On distingue deux types de faux par imitation

➢ Les faux par imitation servile

Dans le cas d'un faux par imitation servile, le faussaire doit posséder un exemplaire de la signature authentique. Ce faux quoique ressemblant à l'original présente des différences dans les espacements, dans les inclinaisons. De plus le tracé est lent et hésitant d'où des variations visibles de la pression.

➢ Les faux par imitation libre

Pour ce faux le faussaire étudie soigneusement la signature authentique et s'entraîne à la reproduire de mémoire jusqu'à être satisfait du résultat. Ce sont, de l'avis des experts, les faux les plus difficiles à détecter car ils sont très ressemblants et le tracé est spontané. Ils diffèrent des originaux par les proportions relatives des éléments de la signature et par l'alternance des pleins et de déliés.

II.6.5 Les faux par déguisement

Le faux par déguisement est particulier car il correspond à une signature faite par le signataire présumé (d'origine) mais déguisée délibérément dans le but de pouvoir renier celle-ci

ultérieurement. Ces 'signatures sont généralement ressemblantes malgré une perte d'harmonie dans le tracé et des changements de vitesse.

II.6.6 Les faux grossiers

Dans le cas d'un faux grossier, le faussaire n'essaie pas de faire un faux ressemblant à un original. Ce faux est fréquent et il est le plus facile à détecter. Ce sont ces faux que nous espérons pouvoir détecter facilement à l'aide de la méthodologie développée dans ce mémoire.

Figure 2 .5 : Les différents types de faux

II.7 Variabilité des signatures manuscrites

II.7.1Variation 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 [19], 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.

II.7.2Variation 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.

II.7.3Gestion 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 [20].

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 [20].

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

II.8 Les systèmes de vérification de la signature

II.8.1 Le système hors ligne :

Dans un système dit "hors-ligne", la signature est d'abord réalisée sur un support papier puis numérisée de façon différée a l'aide d'un scanner ou d'une camera numérique. Dou le terme "hors-ligne". La signature est alors assimilée a une image en niveaux de gris. Ainsi, seulement la forme de la signature est disponible, et seules les données statiques décrivant la géométrie de la signature sont prises en compte. C'est le cas notamment pour les systèmes de vérification de chèques, de contrats ou de formulaires administratifs.

En mode "hors-ligne", on ne dispose pas de paramètres représentant la dynamique de la signature. Afin de rendre les systèmes d'authentification plus fiables, cette dernière peut être générée de façon indirecte par le biais de certaines informations Par exemple, l'épaisseur du trait ou la variation d'intensité du niveau de gris dans la signature décrivent les différentes coulées d'encre sur le papier, et peuvent donc être des indicateurs de la pression exercée par le signataire sur le papier [21].

II.8.2 Le système en ligne

Dans un système dit " 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. les chercheurs dans ce domaine, remarqué, lors de vos 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 [22].

Dans ce mémoire, nous nous sommes plus intéressé au problème de la vérification qu'à celui de l'identification

II.9 Principes de fonctionnement

Classiquement, la conception d'un système d'authentification nécessite d'apporter des solutions à cinq problèmes [22] :

- Acquisition des données
- Prétraitement
- Extraction des caractéristiques et/ou parties stables
- Comparaison (et donc décision)
- Evaluation des performances

Dans le cas ou le système est hors ligne on suive les étapes suivante

II.9.1 Acquisition d'image

La signature est digitalisée par un scanneur et elle est transformée en une image. C'est l'entrée de ce système. Cette étape est assez simple mais très importante car elle influence sérieusement les étapes suivantes. Il y a deux paramètres important :

- **Résolution** : la résolution normale est 300 dpi. Pourtant, quand la taille de l'écriture est petit, il faut augmenter la résolution.
- **Niveau d'éclairage** : si on ajuste le scanneur pour que l'image soit plus claire, le bruit est réduit mais des traits minces disparaissent aussi.

II.9.2 Prétraitement

Le prétraitement consiste à préparer les données issues du capteur à la phase suivante. Il s'agit essentiellement de réduire le bruit superposé aux données et essayer de ne garder que l'information significative de la forme représentée. Le bruit peut être dû aux conditions d'acquisition (éclairage, mise incorrecte du document, ...) ou encore à la qualité du document d'origine.

Parmi les opérations de prétraitement généralement utilisées on peut citer : l'extraction des composantes connexes, le redressement de l'écriture, le lissage, la normalisation et la squelettisation .effets de certaines opérations de prétraitement.

a) La binarisation

La binarisation c'est le passage d'une image en couleur ou définie par plusieurs niveaux de gris en image bitonale (composée de deux valeurs 0 et 1) qui permet une classification entre le fond (image du support papier en blanc) et la forme (traits des gravures et des caractères en noir).

Pour des images de niveaux de gris, on peut trouver dans [19] une liste des méthodes de binarisation, proposant des seuils adaptatifs (ex. s'adaptant à la différence de distribution des

niveaux de gris). [20] proposent une solution pour les images d'adresses postales. La recherche du seuil passe par plusieurs étapes : binarisation préliminaire basée sur une distribution de mixture multimodale, analyse de la texture à l'aide d'histogrammes de longueurs de traits, et sélection du seuil à partir d'un arbre de décision.

b) Lissage

L'image des caractères peut être entachée de bruits dus aux artefacts de l'acquisition et à la qualité du document, conduisant soit à une absence de points ou à une surcharge de points. Les techniques de lissage permettent de résoudre ces problèmes par des opérations locales qu'on appelle opérations de bouchage et de nettoyage [23].

L'opération de nettoyage permet de supprimer les petites tâches et les excroissances de la forme. Pour le bouchage il s'agit d'égaliser les contours et de boucher les trous internes à la forme du caractère en lui ajoutant des points noirs.

Plusieurs autres techniques similaires sont utilisées dont la méthode statistique, une méthode basée sur la morphologie mathématique [24].

c) Normalisation

Après la normalisation de la taille, les images de tous les caractères se retrouvent définies dans une matrice de même taille, Pour faciliter les traitements ultérieurs.

Le principe de la normalisation est d'essayer de normaliser localement différentes parties du mot, de manière à augmenter la ressemblance d'une image à une autre.

Cette opération introduit généralement de légères déformations sur les images. Cependant certains traits caractéristiques tels que la hampe dans les caractères **(Ø Ù á Ç par exemple)** peuvent être éliminées à la suite de la normalisation, ce qui peut entraîner à des confusions entre certains caractères [25].

II.9.3 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.

Il existe plusieurs type de caractères come Texture, géométriques, Locales, Globale, et combinaison et dans notre thèse on base sure la texture utilisons les deux méthodes :

- La méthode dite « LBP »
- La matrice de cooccurrence

a) LBP (Local binary pattern)

Ahonen et al. [26] ont appliqué une représentation LBPH (histogramme LBP) pour faire de la reconnaissance avec de très bons résultats sur la base de données. Dans leur méthode, l'image du visage est d'abord divisée en petites régions à partir desquelles les histogrammes LBP sont extraits et concaténées en une seule fonction histogramme représentant la texture locale et la forme globale de signature. La reconnaissance est effectuée en utilisant un classificateur plus proches voisins.

L'opérateur LBP a été proposé initialement par Ojala et al. [27] dans le but de caractériser la texture d'une image. Le calcul de la valeur LBP consiste pour chaque pixel à seuiller ses huit voisins directs avec un seuil dont la valeur est le niveau de gris du pixel courant. Tous les voisins prendront alors une valeur 1 si leur valeur est supérieure ou égale au pixel courant et 0 si leur valeur est inférieure (figure 2.6). Le code LBP du pixel courant est alors produit en concaténant ces 8 valeurs pour former un code binaire. On obtient donc, comme pour une image à niveaux de gris, une image des valeurs LBP contenant des pixels dont l'intensité se situe entre 0 et 255.

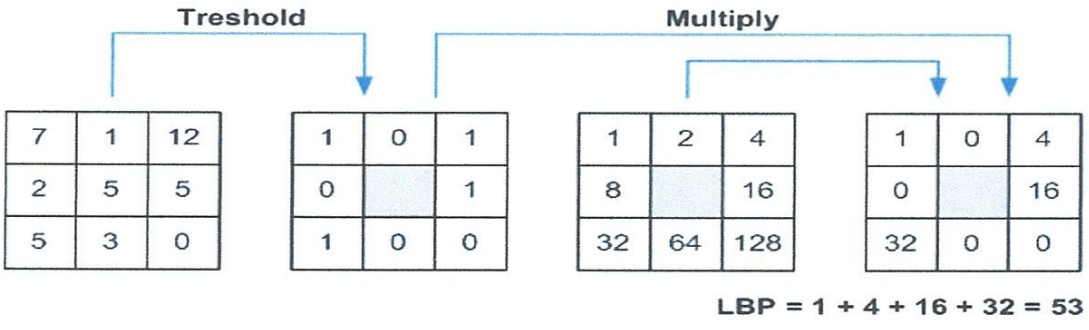


Figure 2.6 : Opérateur de LBP

Le LBP a été étendu ultérieurement en utilisant des voisinages de taille différente. Dans ce cas, un cercle de rayon R autour du pixel central est considéré Les valeurs des P points échantillonnées sur le bord de ce cercle sont prises et comparées avec la valeur du pixel central. Pour obtenir les valeurs des P points échantillonnés dans le voisinage pour tout rayon R, une interpolation est nécessaire. On adopte la notation (P;R) pour définir le voisinage de P points de rayon R d'un pixel.

Soient g_c le niveau de gris du pixel central, $g_p(p = 1...P)$ les niveaux de gris de ses voisins, l'indice LBP du pixel courant est calculé comme :

$$LBP_{P,R}(x_c, y_c) = \sum s(g_P, g_c)2^{P-1}$$

$$\begin{cases} 1 & si \quad x \geq 0 \\ 0 & si \quad x < 0 \end{cases}$$

Où $(x_c; y_c)$ sont les coordonnées du pixel courant, $LBP_{P;R}$ est le code LBP pour le rayon R et le nombre de voisins P. L'opérateur LBP obtenu avec P = 8 et R = 1 (LBP8;1) est très proche de l'opérateur LBP d'origine. La principale différence est que les pixels doivent d'abord être interpolés pour obtenir les valeurs des points sur le cercle (voisinage circulaire au lieu de rectangulaire).

Une autre extension à l'opérateur d'origine est le LBP uniforme. Un code LBP est uniforme s'il contient au plus deux transitions de bits de 0 à 1 ou vice-versa lorsque la chaine binaire est considéré circulaire. Par exemple, 00000000, 00011110 et 10000011 sont les codes uniformes. L'utilisation d'un code LBP uniforme, noté LBPu2 a deux avantages.

Le premier est le gain en mémoire et en temps calcul. Le deuxième est que LBPu2 permet de détecter uniquement les textures locales importantes, comme les spots, les fins de ligne, les bords et les coins. (figure 2.7 pour des exemples de ces textures particulières).

OJALA a constaté que seuls 58 des 256 motifs LBP sont uniformes mais expérimentalement, il a être constaté que 90% des patterns rencontrés dans les images sont uniformes.

Figure 2.7 : Primitives texture différente détectée par le LBP .

b) Matrices de cooccurrences

La deuxième caractéristique utilisée pour la comparaison entre les images est la texture. Nous utiliserons les matrices de cooccurrences.

[28] proposent d'extraire des statistiques à partir d'une matrice de cooccurrence (GLCM) afin de caractériser une texture. Les matrices de cooccurrence d'Haralick, encore appelées matrices de dépendance spatiale des niveaux de gris, utilisent les statistiques d'ordre2 et permettent de déterminer la fréquence d'apparition d'un motif formé de deux pixels séparés par une certaine distance d dans une direction particulière θ. Cette matrice est une matrice carrée de taille n^2 où n correspond au nombre des niveaux de gris de l'image.

La normalisation de cette matrice produit une distribution de probabilité. Pour une image I, une matrice de cooccurrence des niveaux de gris normalisés P_υ pour un vecteur de séparation donné $\upsilon = (\upsilon_x, \upsilon_y)$ est défini par ses composantes $P_\upsilon(i,j)$ données par l'expression suivante:

$$P_{\upsilon'}(i,j) = \frac{\#\{(a.b) : I(a.b) = i.I(a+\upsilon_x, b+\upsilon_y) = j\}}{\#\{I\}}$$

Avec # la fonction cardinal, I(a,b) le niveau de gris de l'image I aux coordonnées (a,b) et $i, j \in \{1,2,.....,n\}$

[29] suggère de calculer des matrices de cooccurrence avec différentes directions et d'en faire la moyenne. De cette façon une matrice de cooccurrence peut être un descripteur de textures invariant à la rotation.

Pour chaque image, 4 matrices de cooccurrences seront calculées sur les images en niveaux de gris pour une distance=1 et pour quatre direction (0, 45, 90 et 135 degrés).

$$G = \begin{vmatrix} P(1,1) & P(1,2) & \cdots & P(1,N_g) \\ P(2,1) & P(2,2) & \cdots & P(2,N_g) \\ \vdots & \vdots & \ddots & \vdots \\ P(N_g,1) & P(N_g,2) & \ldots & P(N_g,N_g) \end{vmatrix}$$

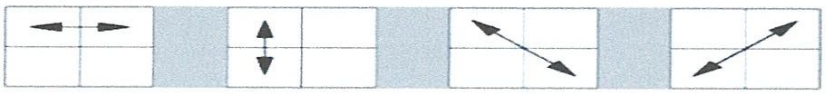


Figure 2.8 : Méthode de calcule de la matrice de cooccurrence

Cinq étapes sont nécessaires ici pour obtenir une version simple mais fonctionnelle pour calculer une distance entre les textures :

- **Conversion de l'image en niveaux de gris**

Nous calculerons les textures uniquement sur l'image en niveau de gris pour simplifier la tache. Pour convertir les images en niveaux de gris (*NdG*) il suffit d'utiliser (pour chaque pixel) la fonction suivante : *NdG = rouge*0.299 + vert*0.587 + bleu*0.114*

- **Réduction des niveaux de gris de l'image**

Comme pour les couleurs, il est inutile de conserver tous les niveaux dc gris dans lc calcul dcs textures. Nous réduirons la quantification de l'image pour la faire passer de 256 a T niveaux de gris (T=8, 16 ou 24). Vous choisissez la valeur de T ou pouvez ajouter une option en argument au programme. Pour réduire les niveaux de gris, il suffit de diviser chaque pixel de l'image par T.

- **Calcul des matrices de cooccurrences**

4 matrices de cooccurrences par image doivent être calculées, pour une distance=1 et pour des directions=0, 45, 90 et 135 degrés. Normalement, une seule fonction est nécessaire pour le calcul, si on passe en paramètres dx et dy, les distances x et y (pour direction=0 degré, dx=1 et dy=0 ; pour direction=45 degrés, dx=1 et dy=1 ; pour direction=90 degrés, dx=0 et dy=1 ; pour direction=135 degrés, dx=-1 et dy=1).

- **Calculs des paramètres sur les matrices de cooccurrences**

A partir de cette matrice, il est possible de calculer plusieurs caractéristiques. Parmi les plus fréquemment utilisées, nous retrouvons: la moyenne, la variance, le contraste, l'énergie (second moment angulaire), l'entropie, la corrélation, l'homogénéité et l'uniformité. Nous présentons dans le tableau ci-dessous les caractéristiques d'Haralick que nous utilisons par la suite dans notre travail. Il existe 14 paramètres de Haralike possibles sur les matrices de cooccurrences.

Le seconde moment anguler	$\sum_i \sum_j p(i,j)^2$
Contraste	$\sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \right\}, \lvert i-j \rvert = n$

Corrélation	$\frac{\sum_i \sum_j p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$
Somme de la variance carrés	$\sum_i \sum_j (i-\mu)^2 p(i,j)$
Moment différentielle inverse	$\sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,p)$
Somme moyenne	$\sum_{i=2}^{2N_g} ip + y(i)$
Somme variance	$\sum_{i=2}^{iN_g} (i - f_8)^2 p_{x+y}(i)$
Somme entropie	$-\sum_{i=2}^{2N_g} p_{i+y}(i) \log\{p_{i+y}(i)\} = f_8$
Entropie	$-\sum_i \sum_j p(i,j) \log(p(i,j))$
Déférence variance	$-\sum_{i=0}^{N_g-1} i^2 p_{i-y}(i)$
Déférence entropie	$-\sum_{i=0}^{N_g-1} p_{x-y}(i) \log\{p_{x-y}(i)\}$
Info mesure de correlation1	$\frac{HXY - HXY1}{max\{HX, HY\}}$
Info mesure de correlation2	$(1 - Exp[-2(HXY2 - HXY)])^{\frac{1}{2}}$
Max. Correltion Coeff .	$Q(i,j) = \sum_k \frac{P(i,k)p(j,k)}{p_x(i)p_y(k)}$

II.9.4 Classification

Les méthodes de classification ont pour but d'identifier les classes auxquelles appartiennent des objets a partir de certains traits descriptifs. Elles s'appliquent a un grand nombre d'activités humaines et conviennent en particulier au problème de la prise de décision automatisée. La procédure de classification sera extraite automatiquement a partir d'un ensemble d'exemples. Un exemple consiste en la description d'un cas avec la classification correspondante. Un système d'apprentissage doit alors, a partir de cet ensemble d'exemples, extraire une procédure de classification, il s'agit en effet d'extraire une règle générale a partir

des données observées. La procédure générée devra classifier correctement les exemples de l'échantillon et avoir un bon pouvoir prédictif pour classier correctement de nouvelles descriptions. Dans ce chapitre, une définition du concept d'apprentissage automatique est donnée. Nous introduisons dans les sections suivantes, les concepts de la classification supervisée ainsi qu'un état de l'art des méthodes et algorithmes usuels en apprentissage automatique. Finalement, nous présentons l'approche d'ensembles de classificateurs et nous établissons un état de l'art sur les principaux algorithmes ensemblistes.

a) L'apprentissage

Il s'agit lors de cette étape d'apprendre au système les propriétés pertinentes du vocabulaire utilisé et de l'organiser en modèles de références.

L'idéal serait d'apprendre au système autant d'échantillons que de formes d'écritures différentes, mais cela est impossible à cause de la grande variabilité de la signature qui conduirait à une explosion combinatoire de modèles de représentation. La tendance consiste alors à remplacer le nombre par une meilleure qualité des traits caractéristiques, [28]. L'apprentissage consiste en deux concepts différents : l'entraînement et l'adaptation. L'entraînement consiste à enseigner au système la description des caractères tandis que l'adaptation sert à améliorer les performances du système.

Certains systèmes permettent à l'utilisateur d'identifier un caractère lorsqu'ils échouent à le reconnaître et ils utilisent l'entrée de l'utilisateur à chaque fois que le caractère est rencontré [30].

b) Reconnaissance et décision

La décision est l'ultime étape de reconnaissance. A partir de la description en paramètres du caractère traité, le module de reconnaissance cherche parmi les modèles de référence en présence, ceux qui lui sont les plus proches.

La reconnaissance peut conduire à un succès si la réponse est unique (un seul modèle répond à la description de la forme du caractère). Elle peut conduire à une confusion si la réponse est multiple (plusieurs modèles correspondent à la description). Enfin elle peut conduire à un rejet de la forme si aucun modèle ne correspond à sa description. Dans les deux premiers cas, la décision peut être accompagnée d'une mesure de vraisemblance, appelée aussi score ou taux de reconnaissance [24].

II.9.4 les méthodes de classification utilisées dans un système de vérification de signature

a) Séparateurs à vastes marges (SVM) :

Pour un problème de classification binaire, les méthodes d'apprentissage statistique classiques [31, 32, 33] visent à déterminer une fonction de décision de la forme

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i x_i^{\mathrm{T}} x$$

On a pour cela recours `a un ensemble d'apprentissage de la forme $x_1, x_2, \ldots, x_n \in \mathrm{IR}^d$ ou chaque élément est étiquetée selon $y_i = \pm 1$. Une fonction de décision non-linéaire est aisément obtenue, si nécessaire, en remplaçant le produit scalaire $x_i^T x$ par une fonction noyau $K(x_i, x)$

Un problème de classification multi-classe peut être résolu en traitant préalablement plusieurs sous-problèmes de classification binaire, puis en combinant le résultat des différentes fonctions de décision. Chaque sous-problème définit alors une fonction de décision de la forme

$$f_k(x) = \sum_{i=1}^{n} \alpha_{i,k} y_{i,k} x_i^T x$$

Bien qu'utilisant les mêmes données $x_1, x_2, \ldots, x_n$ pour l'apprentissage, chaque sous-problème de classification binaire définit ses propres étiquettes $y_{1,k}, y_{2,k}, y_{3,k}, \ldots, y_{n,k}, \in \{+1, -1\}$.

L'algorithme SVM est basé sur une fonction cout à la solution, au prix d'une optimisation par programmation quadratique. Le problème d'optimisation ici revisité est donné par

$$\min_{W,b,\xi} \frac{1}{2}\|W\|_f^2 + \gamma_i \sum_{i=1}^{n} \xi_i$$

Sous contrainte : $y_i^T (W^T x_i + b) \geq 1 - \xi_i$

$\xi_i \geq 0, \quad \forall_i$

pour i = 1, 2,..., n, ou γ est le paramètre de régularisation et ξi est une variable d'écart autorisant à un petit nombre de données d'apprentissage à violer la règle de grande marge.

b) Least square support vector machines

Dans notre travail pour modéliser chaque signature, le classifier LS-SVM (Least Square Support Vector Machine) a été utilisé. SVM standard a été introduit dans le cadre de la théorie de l'apprentissage statistique et minimisation du risque structurel. LS-SVM sont des reformulations à SVM standard. Une seule équation linéaire doit être résolu dans le processus d'optimisation, qui non seulement simplifie le processus, mais aussi d'éviter le problème de minima locaux dans SVM. Le LS-SVM est définie dans son espace de poids primal par

$$\hat{Y}(x) = w^T \varphi(x) + b$$

où φ (x) est une fonction qui mappe l'espace d'entrée en un espace de caractéristique dimensionnelle élevée, x est le vecteur à M dimensions, w et b les paramètres du modèle. Étant donné N d'entrée-sortie d'apprentissage $(x_i, y_i) \in R^M \times R$, LS-SVM cherchent w et b qui minimisent

$$\min_{wbe} j(w,e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^{N} e_I^2$$

Avec $y = w^T \varphi(x^i) + b + e^i, 1 \leq i \leq N$

Dans notre cas, nous utilisons en fonction noyau RBF gaussien c.-à-d. Les paramètres du modèle LS-SVM sont la largeur de la gaussienne C et le facteur de régularisation γ . La méthode de formation pour l'estimation de w et b peut être trouvée dans [36]. Dans ce travail, les paramètres méta (γ, C) est choisi d'un faon empirique. Le LS-SVM formés pour chaque signataire utilise les mêmes (γ, C) paramètres.

c) Les réseaux de neurones

Un réseau de neurones est un graphe orienté pondéré. Les nœuds de ce graphe sont des automates simples appelés neurones formels. Les neurones sont dotés d'un état

interne, l'activation, par lequel ils influencent les autres neurones du réseau. Cette activité se propage dans le graphe le long d'arcs pondérés appelés liens synaptiques [34].

En OCR, les primitives extraites sur une image d'un caractère (ou de l'entité choisie) constituent les entrées du réseau. La sortie activée du réseau correspond au caractère reconnu. Le choix de l'architecture du réseau est un compromis entre la complexité des calculs et le taux de reconnaissance [36].

Par ailleurs, le point fort des réseaux de neurones réside dans leur capacité de générer une région de décision de forme quelconque, requise par un algorithme de classification, au prix de l'intégration de couches de cellules supplémentaires dans le réseau.

d) La méthode du plus proche voisin

L'algorithme KNN (K Nearest Neighbors) affecte une forme inconnue à la classe de son plus proche voisin en le comparant aux formes stockees dans une classe de références nommée prototypes. Il renvoie les K formes les plus proches de la forme à reconnaître suivant un critère de similarité. Une stratégie de décision permet d'affecter des valeurs de confiance à chacune des classes en compétition et d'attribuer la classe la plus vraisemblable (au sens de la métrique choisie) à la forme inconnue [24, 23].

Cette méthode présente l'avantage d'être facile à mettre en œuvre et fournit de bons résultats. Son principal inconvénient est lié à la faible vitesse de classification due au nombre important de distances à calculer.

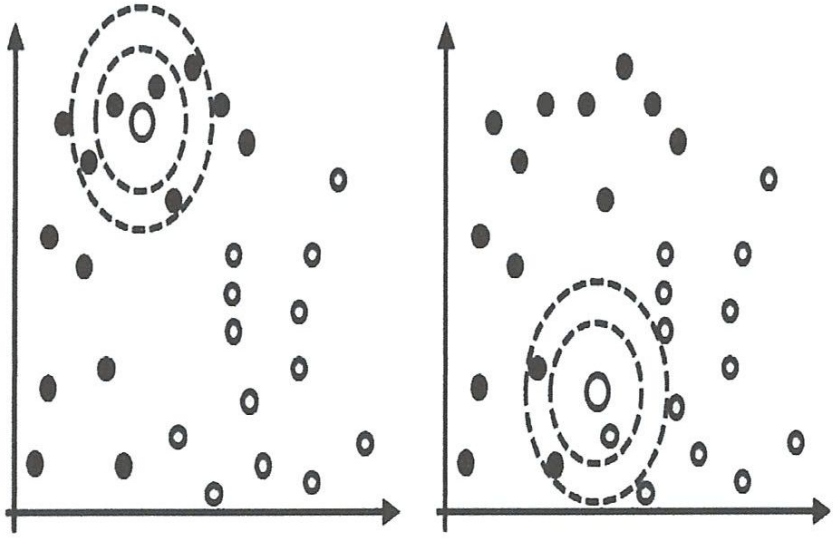


Figure 2.9 : Exemple de classification avec les KNN

Conclusion

Dans ce chapitre on a défini la signature manuscrite et les types de signature dans le monde. Ensuite on a parlé sur les systèmes d'authentification de la signature, leur type et les problèmes qu'on peut les trouver dans les opérations d'authentification de la signature manuscrite. Nous avons discuté aussi les avantages de la signature manuscrite, le principe de fonctionnement de ce système et les méthodes de l'extraction des paramètres caractéristiques. Enfin, une revue des différentes méthodes de classification dans un système de vérification de la signature sont présentées

Chapitre III

Partie expérimentale

III.1 Introduction

Aujourd'hui, le développement d'un système d'identification/vérification de signature s'affronte avec le problème de la faible disponibilité des bases de données. Il est vraiment difficile de faire une comparaison entre les différentes approches présenté sur la littérature du fait de l'utilisation de bases de données propres dans chaque travail et que ces bases de données ne sont pas disponible en publique. Il ya juste quelques corpus publics. Certains d'entre eux sont résumés dans le tableau 1.

Nom de a base	signataires	authentique	faux
GPDS signature [1]	160	24	30
SVC2004 [2]	40	20	20
MCYT-100 [3]	100	25	25
MCYT-75 [4]	75	15	15

Tableau 1 : Les bases de Signatures manuscrits disponibles

La conception et la construction d'une base de signature hors ligne implique un processus long et complexe dans lequel les aspects tels que :

- la variabilité de la surface de dessin
- les changements de stylo
- les différences entre la session
- le nombre de signataires

- le nombre de véritables signes par personne
- la procédure de créé les faux signatures et le nombre de faux par personne, etc

doivent être pris en compte. Malheureusement, il est difficile de construire un corpus qui a examiné l'ensemble ci-dessus.

Nous avons utilisé deux bases de données pour tester notre système. Les deux bases ont été scannées à 600 dpi, ce qui garantit une représentation suffisante. La différence principale entre eux est le stylo utilisé. Dans la MCYT base, tous les signataires, authentiques, et faussaire sont signés avec le même stylo sur la même surface. Par contre, dans la base de données GPDS, tous les utilisateurs ont signé avec leurs propres stylos sur différents surfaces.

III.1.1 La base de données GPDS 960

Le Groupe de Traitement numérique du signal (GPDS) de L'Université « de Las Palmas de Gran Canaria (ULPGC-Espagne) » a consacré des efforts à la création d'une base de données de signature manuscrite hors-ligne impliquant 960 individus pour les signatures authentiques et 1920 personnes pour faux.

La base GPDS-960 contient 960 personnes, 24 signatures authentiques et 30 imitations par personne. Donc, il ya 960 x 24 = 23 049 signatures authentiques et 960 x 30 = 28 800 faux. Les signatures authentiques ont été prises en une seule session. Les signataires rempli un formulaire avec 24 boîtes de taille différente. Les répétitions de chaque signatures authentique et imitation ont été captées à l'aide d'un propre plume de chaque personne sur des feuilles blanc format A4, comportant deux tailles de boîte différent :

- La première case est de 5cm de largueur et 1.8cm de hauteur
- La deuxième zone est de 4,5 cm de largueur et 2,5 cm de hauteur

Les falsifications ont été effectuées par 1920 personnes différentes à la 960 déjà mentionné. Chaque faussaire rempli un formulaire avec 15 boîtes. Chaque forme de falsificateur contient 5 images de différentes signatures authentiques choisis au hasard. Le falsificateur imita 3 fois chacun des cinq signes.

La Figure 3.1 montre des exemples de signatures authentiques et imitées de trois personnes différents: les deux signatures sur la gauche sont authentiques, et celle sur la droite est une imitation.

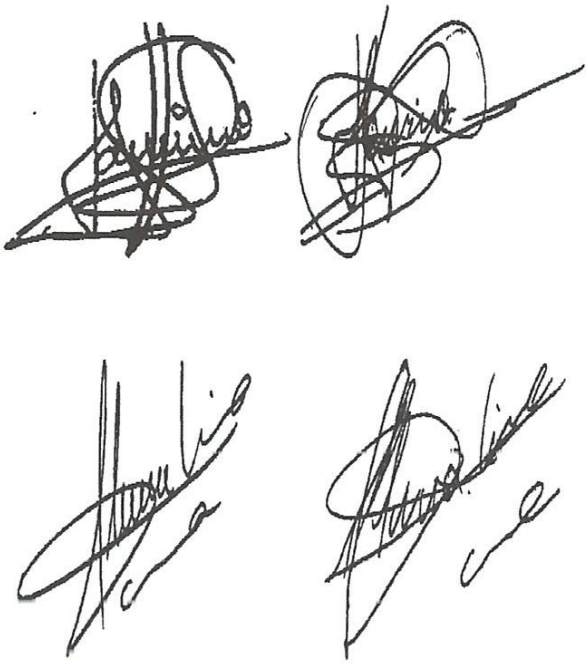


Figure 3.1 : les signatures authentiques et imitées de la base GPDS 960

III.1.2 La base de données MCTY-75

La création de cette base a été motivée par le manque de grandes bases biométriques publiques utilisables pour l''evaluation des performances des systèmes de reconnaissance. Dans ce contexte, le projet espagnol MCYT [91], achève à la fin de 2003, avait pour mission l'acquisition d'une base biométrique bimodale comprenant entre autre la signature manuscrite en-ligne et hors ligne.

La base MCYT complète contient 330 signataires. Par contre, nous ne disposons que d'une sous-partie de la base qui est gratuite , dénommée la base MCYT-75.

Le corpus signature *de MCYT* -75 contient 2250 signatures de 75 individus. Chaque classe individuelle consistent 30 signatures 15 signatures authentiques et 30 imitation. Totalement elle forme une base de données de signature de 1125 (c.-à-d. 75 x15) authentique et 1125 (c.-à-d. 75 x15) ont contrefaçons en différé signatures.

La Figure 3.2 montre des exemples de signatures de la base MCYT

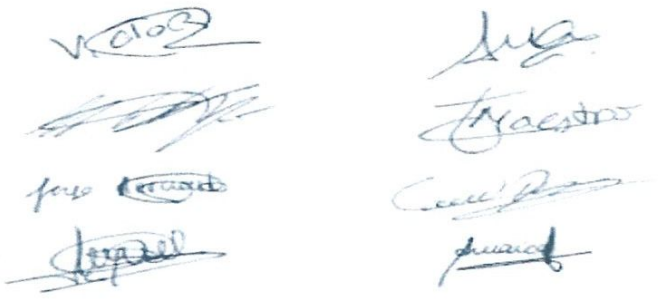


Figure3.2 : Des signatures prise de la base MCYT

III.2 Expérimentations et résultats

Nous présentons dans cette partie deux expériences. La première expérience montre l'apport des différents vecteurs caractéristiques de la signature. On à appliqué les méthodes d'extraction de caractéristique qui base sur la texture et que sont les LBP et la matrice de concurrence. La deuxième expérience montre l'influence de deux paramètres du classifieur SVM (le paramètre de régularisation C, et le paramètre sigma du noyon RBF).

Les tests ont été menés sur les deux bases de signatures (GPDS et MCYT). Durant cette expérience, nous utilisons seulement 20 personnes de chaque base. Toutes les images de signature ont été en niveau de gris.

Pour l'évaluation des performances, le protocole d'expérimentation que nous avons adopté est comme suit : 120 tirages déterministe sont réalises sur les signatures authentiques et les imitations. Chaque tirage contient 07 signatures authentiques et 07 imitations utilisées comme signatures d'apprentissage du classifier SVM dE référence. Pour le test, les 03 signatures authentiques restantes et les 03 imitations sont utilisées dans le cas de MCYT-20. Pour GPDS-20, les tests sont exécutés sur les N1 signatures authentiques restantes et les N2 imitations. Les taux de fausses acceptations et de faux rejets sont calcules sur le premier tirage.

La séparation des signatures authentiques en signatures d'apprentissage à partir de chaque base a été faite par programmation.

Pour le classifieur SVM le score de similarité est défini comme étant un étiquette (+1,-1). La qualité du SVM construit à partir des mêmes signatures authentiques n'est pas toujours identique à cause de l'influence du choix de paramètres. Ainsi, la performance du système de

vérification change en testant sur les différents SVMs, construits à partir des mêmes signatures authentiques.

Des expériences ont été effectuées en utilisant différentes valeurs des paramètres P, R pour LBPriu2. D'abord, les valeurs ont été établies à R =1 et P = 8. Puis ils ont été mis à la valeur R=2 et P = 16. Enfin, une combinaison des deux paires a été utilisée.

Donc pour le LBP on à utiliser :

- le LBP de rayon R = 1 et le P = 8 ($LBP_{8,1}$)
- le LBP de rayon R = 2 et le P = 16 ($LBP_{16,2}$)
- la combinaison $LBP_{8,1}$ plus le $LBP_{16,2}$

Une fois la matrice des vecteurs caractéristiques est estimée, nous devons résoudre un problème de classification de deux classe (authentiques ou faux). Pour modéliser chaque signature, le classifier « Support Vector Machine (LS-SVM) » a été utilisé.

Une fois la matrice des vecteurs caractéristiques est estimée, nous devons résoudre un problème de classification de deux classe (authentiques ou faux). Pour modéliser chaque signature, le classifier « Support Vector Machine » a été utilisé. Le code a été écrit en utilisant Matlab avec la bibliothèque « LS-SVMLab 1.5 », développée par l'équipe de J.A.K. Suykens de l'Université K.U.Leuven

Le tableau 1 présente les résultats obtenus sur la base MCYT(base01).

Base 01		C=10 Sig=0,01	C=10 Sig=0,1	C=10 Sig=0 ,2	C=10 Sig=0,3	C=10 Sig=1	C=10 Sig=2	C=20 Sig=1	C=100 Sig=1
LBP81	FRR	15%	15%	15%	15%	13%	17%	13%	13%
	FAR	33%	33%	33%	33%	23%	22%	25%	25%
LBP162	FRR	13%	13%	13%	13%	13%	12%	13%	12%
	FAR	33%	33%	33%	33%	25%	23%	25%	25%
LBP81+LBP162	FRR	15%	15%	15%	15%	15%	13%	15%	13%
	FAR	33%	33%	33%	33%	33%	20%	18%	33%

Le tableau 2 présente les résultats obtenus sur la base GPDS (base02).

Base 02		C=10 Sig=0,01	C=10 Sig=0,1	C=10 Sig=0 ,2	C=10 Sig=0,3	C=10 Sig=1	C=10 Sig=2	C=20 Sig=1	C=100 Sig=1
LBP81	FRR	96%	96%	96%	96%	100%	100%	100%	90%
	FAR	4%	4%	4%	4%	0%	0%	0%	7%
LBP162	FRR	87%	87%	87%	87%	100%	100%	100%	100%
	FAR	11%	11%	11%	11%	0%	1%	0%	0%
LBP 62 LBP81+	FRR	93%	93%	93%	93%	100%	100%	100%	100%
	FAR	6%	6%	6%	6%	0%	0%	0%	0%

Pour la matrice de cooccurrences deux vecteurs caractéristiques de taille différents sont utilisé pour l'évaluation

- la matrice de cooccurrence avec 8 paramètres caractéristiques.

Le tableau3 présente les résultats obtenus sur la base01

Base 01		C=10 Sig=0,01	C=10 Sig=0,1	C=10 Sig=0 ,2	C=10 Sig=0,3	C=10 Sig=1	C=10 Sig=2	C=20 Sig=1	C=100 Sig=1
GLMC8	FRR	12%	12%	12%	12%	18%	8%	8%	18%
	FAR	23%	23%	23%	23%	20%	17%	17%	22%
LBP+GLMC	FRR	15%	15%	15%	15%	15%	13%	15%	13%
	FAR	33%	33%	33%	33%	33%	20%	18%	33%

Le tableau4 présente les résultats obtenus sur la base02

Base 02		C=10 Sig=0,01	C=10 Sig=0,1	C=10 Sig=0 ,2	C=10 Sig=0,3	C=10 Sig=1	C=10 Sig=2	C=20 Sig=1	C=100 Sig=1
GLMC8	FRR	79%	79%	79%	79%	80%	80%	85%	
	FAR	22%	22%	22%	22%	20%	20%	15%	
LBP+GLMC	FRR	57%	57%	57%	57%	57%	57%	50%	35%
	FAR	37%	37%	37%	37%	35%	35%	50%	65%

Discussion

Une première expérience a été effectuées en utilisant différentes valeurs P et R. Tableau 1 et 2 présentent les résultats obtenus à l'aide de 7 échantillons authentiques et faux pour la formation du classifier. Comme on peut le voir les meilleurs résultats ont été obtenus en utilisant la combinaison LBP8,1 +LBP 16,2. C'est logique, parce que le nouveau vecteur de caractéristiques de longueur 10+18 = 28 comprend des informations sur les premier et deuxième anneaux de pixels autour d'un pixel central. Ces Tableaux montrent aussi des informations plus détaillées sur les résultats obtenus avec LBP 8,1 et LBP16,2.

Une Deuxième expérience a été effectuée en utilisant des vecteurs caractéristiques calculés à partir de la matrice de cooccurrence (GLCM). Tables 3 et 4 présentent les résultats pour GLCM et la combinaison LPB+GLMC

L'analyse des résultats de la base 1 confirme l'efficacité de la modélisation proposée par rapport à l'approche linéaire notamment dans le cas des fausses signatures par imitation.
Les meilleurs résultats enregistrés sont de 13% et8% pour le FRR, 20% et respectivement pour le FAR lors de l'utilisation d'une combinaison de LBP81+LBP162 et l'utilisation de GLCM8

par contre Les mauvaises résultats sont de la base 2 par **87%** et **57%** pour le FRR et **11% , 37%** pour le FAR lors de l'utilisation d'une combinaison GLCM8+LBP8+LBP162 et l'utilisation de l'LBP162

Conclusion

Dans ce chapitre, nous avons considéré nos travaux en vérification hors ligne de la signature manuscrite. utilisons deux base de donnée (GPDS et MCYT)

Deux types d'extraction de paramètre ont été présentés. Dans les deux types nous avons exploré la matrice de cooccurrence et le LBP pour la description des images des signatures ainsi que des caractéristiques texturale. Le SVM a été utilisés pour la classification.

Les différentes expérimentations ont montré le rôle de la vérification dans le cas de rejet de signatures authentiques FRR et dans le cas de l'acceptation de fausses signatures FAR. La considération de manière simplifiée de l'aspect bidimensionnel de l'image signature a permis d'affiner la description et d'améliorer les performances du système.

Conclusion générale

Dans ce mémoire, nous avons abordé le problème de l'authentification par signature manuscrite hors-ligne. La signature manuscrite est une modalité biométrique largement utilisée dans la vie quotidienne ce qui motive l'intérêt de ce travail. Nous nous sommes intéressés à construire un système de vérification basé sur l'information de niveau de gris.

Dans un premier temps, nous avons calculé plusieurs vecteurs caractéristiques à partir des images en niveau de gris tel que LPB et GLMC. Ensuite Le SVM d'un utilisateur est construit à partir de ces vecteurs caractéristiques d'enrôlement. Dans un deuxième temps Les signatures de test seront aussi normalisées par les méthodes d'extractions des paramètres avant d'être introduites dans le SVM. Les paramètres de régularisation et de noyau détermine la qualité du classifier SVM construite.

La performance du système est présentée en référence aux deux bases de données de signature contenant des fausses signatures par imitation expérimenté à partir de 20 individus, y compris les faux qualifiés. Les résultats expérimentaux pour faux qualifiés (tableaux 1, 2, 3 et 4) montrent que l'utilisation d'informations de niveau de gris permet d'obtenir performances raisonnable du système. Les performances sont acceptables mais restent encore insuffisantes pour une sécurité totale d'un système global de vérification

BIBLIOGRAPHIE

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

[2] BOUDJELLAL Sofiane thèse de Magister en Electronique Option : Télédétection Détection et identification de personne par méthode biométrique »

[3] Anil. K. Jain, R. Bolle, And S. Pankanti, « Biometrics: Personal Identification In Networked Society » Kluwer Academic Publishers, 1999.

[4] Anil. K. Jain, P. Flynn, A. Ross, « *Handbook of Biometrics* », Springer, 2007.

[5] Julian Ashbourn, « Guide To Biometrics For Large-Scale Systems », Springer 2011.

[6] Florent Perronnin, Jean-Luc Dugelay, « Introduction à la biométrie : Authentification des individus par traitement audio-vidéo », Institut Eurocom, Multimédia Communications Département, Revue Traitement du signal, Vol. 19, N° 4, 2002.

[7] Anis chaar thèse « Nouvelle approche d'identification dans les bases de données biométriques basée sur une classification non supervisée » le mardi 6 Octobre 2009 au laboratoire IBISC.

[8] Houmani Nesma thèse de doctorat télécom & management Sud paris dans le cadre de l'école doctorale S&I en co-accréditation avec l'université d'evry-Val d'Essonne« Analyse de La qualité des signatures manuscrites en ligne par la mesure d'entropie » le 13 janvier 2011

[9] Bergadano F., Gunetti D., Picardi C. (2002), User authentication through keystroke dynamics. *ACM Transactions on Information and System Security (TISSEC)*, 5, 4, pp. 367-397

[10] P.J. Phillips, H. Hyeonjoon, S. Rizvi, P. Rauss. The FERET Evaluation Methodology for Face-Recognition Algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vo. 22, No. 10, Octobre 2000

[11] Delphine MAJDANSKI, *La signature et les mentions manuscrites dans les contrats*, Bordeaux, 2000, note 22, p.44

[12] Alain BUQUET, *La signature du sceau à la clé numérique, histoire, expertise, interprétation*, Paris, Édition Service Gutenberg XXIème siècle, 2000,note 15, p. 65

[13] Marc VAN QUICKENBORNE, « Quelques réflexions sur la signature des actes sous seing privé », *R.C.J.B.*, 1985, 68 note 25, 81

[14] Étienne DAVIO, « Questions de certification, signature et cryptographie », dans Cahiers du CRID, Namur, CRID, *Internet face au droit*, 1997, 69 note 26, 67

[15] Etienne WÉRY, Thibault VERBIEST, *Le droit de l'Internet et de la société de l'information: droits européen, belge et français*, Bruxelles, Larcier, 2001

[16] Arnaud-F. FAUSSE, *La signature électronique*, Paris, Dunod, 2001, p. 348

[17] Critères communs, « Specification of Protection Profiles », source: http://www.commoncriteria.org/cc/part1/part1b.html, visitée en mai 2002

[18] Thèse pour obtenir le grade de docteur de l'université de tours « authentification par signature manuscrite sur support nomade » par : Matthieu Wirotius le 10 novembre 2005

[19] O. D. Trier and T. Taxt. Evaluation of binarization methods for document images, On Pattern Analysis and Machine Intelligence, vol. 11, n. 12, pp. 312- 314, December 1995.

[20] Y. Liu and S. Srihari. Document image binarization on texture features, On Pattern Analysis and Machine Intelligence, vol. 19, n.5, pp. 540>544, May 1997.

[21] Thèse présentée pour l'obtention du grade de docteur de télécom et management sudparis « analyse de la qualité des signatures manuscrites enligne par la mesure d'entropie » soutenue le 13 janvier 2011 par Houmani Nesma

[22] These pour obtenir le grade de docteur de l'université de tours « authentification par signature manuscrite sur support nomade » par : Matthieu Wirotius le 10 novembre 2005

[23] P. Burrow : «Arabic handwriting recognition ». Master of science thesis. School of Informatics, university of Edinburg, England, 2004.

[24] « Utilisation des modèles de Markov cachés planaires en reconnaissance de l'écriture arabe imprimée ». Thèse de doctorat, spécialité Génie Electrique, Université des sciences, des Techniques et de médecine de Tunis II, 1999.

[25] T. Steinherz, E. Rivlin, N. Intrator: «Off-line cursive word recognition: a survey ». International journal on document analysis and recognition, 2(2), pp. 90>110, 1999.

[26] T. Ahonen, A. Hadid, and M. Pietikainen. Face recognition with local binary patterns. ECCV, pages 469-481, 2004

[27] Ojala, T., Pietik ainen, M., Harwood, D. : A comparative study of texture measures with classi_cation based on feature distributions. Pattern Recognition 29 (1996) 51-59

[28] Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, (6), 610-621.

[29] Haralick, R. M. (1979). Statistical and structural approaches to texture.*Proceedings of the IEEE*, *67*(5), 786-804.

[30] B. Al-Badr, S.A. Mahmoud : «Survey and bibliography of Arabic optical text recognition ». Signal processing, vol. 41, pp. 49-77, 1995.

[31] V. N. Vapnik, *Statistical Learning Theory*. Wiley-Interscience, 1998.

[32] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, pp. 293–300, 1999.

[33] R. Rifkin, "Everything old is new again : A fresh look at historical approachesin machines learning," in *PhD thesis, MIT*, 2002.

[34] L. G. Abril, C. Angulo, F. Velasco, and J. A. Ortega, "A note on the biasin svms for multi classification," *IEEE Transactions on Neural Networks*,vol. 19, no. 4, pp. 723–725, 2008.

[35] D. Sch¨olkopf and A. J. Smola, *Learning with Kernels . Support Vector Machines, Regularization, Optimization, and Beyond (Adaptive Computation and Machine Learning)*. The MIT Press, 2001.

[36] L. Souici, Z. Zmirli, M. Sellami : « Système connexionniste pour la reconnaissance de l'arabe manuscrit ». 1ères journées scientifiques et techniques (JST FRANCIL), pp. 383-388, Avignon, France, 1997.