N°d'ordre :
N° de série :

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

---

UNIVERSITE ECHAHID HAMMA LAKHDAR - EL OUED
FACULTÉ DES SCIENCES EXACTES
Département D'Informatique

Mémoire de Fin D'étude
Présenté pour l'obtention du Diplôme de

MASTER ACADEMIQUE

Domaine: **Mathématique et Informatique**
Filière : **Informatique**
Spécialité : **Systèmes Distribués et Intelligence Artificielle**

Présenté par :

- **Moussaoui Karima**
- **Zegueb Nadjia**

Thème

Identification Une Personne Par L'Empreinte Digitale En Utilisant La Méthode SVM.

Soutenue le **29-05- 2017** Devant le jury:

M.	Abbesse Messaoude	MCA	Président
M.	Ghattas Chourouk	MAA	Rapporteur
M.	Gharbi Kadour	MAA	Encadreur

Année Universitaire:2016-2017

Dédicaces

A mes chers parents ma mère « Safia » et mon père « Messaoud » pour leur patience, leur amour, leur soutien et leur encouragement …

A mon cher mari Ms Youcef Henka …

A mes chers frères Ms Boubaker ,Dr Hamza ,Zakaria,Mohamed el-Sasi …

A mes chères sœurs Khadidja ,Meriem …

A La famille de mon mari à chacun en son nom…

A mes chères amis houda ,kenza ,saida, khadidja ,Nadjia,…

Je dédie ce modeste travail

MOUSSAOUI Karima

Dédicaces

Je dédie ce mémoire:

À mes très chers parents pour leur soutient durant toute ma vie d'étudiant et sans eux je ne serai jamais devenu ce que je suis

À mon très cher mari Hocine

À mes chers filles Kater ennada et hibat errahmene

À mon frère et mes sœurs

À tous les professeurs et enseignants qui m'ont suivi durant tout mon cursus scolaire et qui m'ont permis de réussir Dans mes études.

À tous mes amis pour leurs soutiens et leurs encouragements.

Et a toute ma famille.

NADJIA ZEGUEB

Remerciements

Avant tous, Nous remercions dieu le tout puissant de nous avoir donné le courage et la patience pour réaliser ce travail malgré toutes les difficultés rencontrées.

Nous remercions infiniment tous qui nous a aidé de prés et de loin d'avoir compléter ce travail et dépasser tous les obstacles surtout notre enseignant

Gharbi Kadour

qui n'a pas cessé de nous donner les conseils et les bonnes orientations et nous prive pas de son temps

Résumé

Ce travail entre dans le cadre générale de la biométrie ; plus précisément il consiste à mettre au point un système de reconnaissance de l'empreinte digitale en utilisant des techniques évoluées du traitement de l'image et de la discrimination multi classes.

D'une manière générale, dans un système de Reconnaissance de l'empreinte digitale on trouve trois modules : prétraitement, génération de vecteur caractéristiques et classification.

Concernant l'étape de classification, nous avons choisi le classifieur machine à vecteurs de support multi classes (SVM).

Mots clés : Biométrie, Empreinte digitale, Reconnaissance, Classification, SVM

Abstract

This work is part of the overall biometrics field; specifically it consists to develop a fingerprint recognition system using advanced techniques of image processing and multi-class discrimination.

Generally, in a fingerprint recognition system, one can find three main modules: preprocessing, feature vector generation and classification.

Concerning the classification phase ,we have chosen the engine classifier linked to vectors of multi classes support(SVM).

Keyword: Biometrics, Fingerprint, Recognition, Classification ,SVM.

ملخص

هذا العمل هو جزء من مجال القياسات الحيوية الشاملة وعلى وجه التحديد هو تطوير لنظام التعرف على بصمات الأصابع باستخدام تقنيات متقدمة لمعالجة الصور والتمييز متعدد الأصناف.

عموما، في نظام التعرف على بصمات الأصابع، يمكن نجد ثلاث وحدات رئيسية: المعالجة المسبقة، وتوليد ناقلات الميزات والتصنيف.

وفيما يتعلق بمرحلة التصنيف، فقد اخترنا المصنف مكائن الدعم الإتجاهي متعدد الأصناف (SVM)

الكلمات المفتاحية : القياسات الحيوية، بصمة الأصبع، التعرف ، التصنيف، SVM

Table de Matières

Introduction générale……..……..…1

Chapitre I :La Biométrie

1.Introduction……..……..…....5
2.Définition de la biométrie…….…….....5
3.Pourquoi la biométrie ?..... 6
4.Etat de l'art des techniques biométriques…….…… 6
4.1.Biométrie morphologique (physique)…….....7
4.2. Biométrie comportementale…….....9
4.3. Biométrie biologique…….....11
5. Architecture d'un système biométrique…….……..11
5.1. Module d'apprentissage…….……...12
5.2. Module de reconnaissance…….....12
5.3. Module d'adaptation…….……...13
6. Evaluation des performances des Systèmes biométriques…….…13
7. Domaine d'applications de la biométrie…….....15
7.1. Application commerciales…….....15
7.2. Applications de gouvernement…….....15
7.3. Applications juridiques …….…...15
8.Les applications de la biométrie…….....16
9.Avantages et inconvénients de l'identification par empreinte digitale…….....17
10.Conclusion…….....17

Chapitre II :Le technique de reconnaissance d'empreinte digitale

1.Introduction…….....19
2. Historique…….....19
3. Définitions de l'empreinte digitale…….....19
4.Caractérisation d'une empreinte digitale et description du motif…..…….....20
4.1. Motif…….....20
5. Les classes de l'empreinte digitale…….…….....24

6. Structure d'un système complet de reconnaissance d'empreintes……..…..25
6.1. Principe général…….....25
6.2. L'acquisition de l'empreinte……...…...26
6.3.L'extraction de la signature …….....26
6.4. Le stockage et la phase d'appariement…….....27
6.5. Le prétraitement de l'image…….....27
6.5.1.Niveau de gris …….....27
6.5.2.Amélioration de l'image…….....28
7. La reconnaissance d'Empreinte…….....29
7.1. Approche Classique d'extraction de minuties…….....30
7.1.1. La binarisation…….....31
7.1.2.La Squelettisation (amincissement)…….....32
7.1.3.L'extraction des minuties…….....….32
7.1.4. Les problèmes rencontrés lors de l'extraction des minuties…….….34
7.1.5. Elimination des fausses minuties…….....35
7.1.6.Traitement des terminaisons détectées…….....36
7.1.7.Traitement des bifurcations détectées…….37
7.2.Approche d'extraction directe à partir de l'i mage en niveau de gris….....37
8.Conclusion…….....38

Chapitre III :Support Vecteur Machine

1.Introduction…….....40
2.Apprentissage statistique et SVM…….....40
3.Classification et Risque……..……40
3.1 Risque structurel…….....42
3.2 Classification binaire…….....43
3.3 Classification linéaire…….....43
4 Les SVM…….....…....45
4.1 Formulation de SVM : cas liniaires…….....45
4.2 Problèmes non linéairement séparables…….……47
4.3 Les SVM Multiclasses…….....48
4.3.3.1 M_SVM directes……..……...49
4.3.2 L'approche indirecte…….....50

4.4 Architecture du classifieur SVM proposée……53
4.5.Avantages et inconvénients de SVM…….54
5. Conclusion…….54

Chapitre IV :Reconnaissance d'Empreinte Digitale

1.Introduction …….....56
2.Schéma générale se notre système……...56
3. Les Méthodes d'Extraction De Caractéristiques…….....57
3.1. Approche statistique…….....57
3.2. Approche géométrique…….....57
4. Classification……...58
4.1. Apprentissage…….....59
4.2. Décision……...59
5. Méthode de classification choisi ……59
5.1. Classification des empreintes digitale (svm)……....59
5.2. Classifier SVM……..……....60
5.2.1.phase d'apprentissage……...……..61
5.2.2.phase de classification ……...……...62
5.3.Algorithme générale de SVM…….……..63
5.4.Algorithme de résolution ……...……...63
5.5.Classification multi classes ……...64
6.Bibliothèque LIBSVM …….……...64
7.Résultat et Bilan…….....67
7.1.Choix de langage de programmation……...……..67
7.2.Interface de Notre Système ……..68
7.3.Résultat de Notre Application……69
8.Conclusion……..……...69
Conclusion générale……..70
Bibliographies……...…….71

Liste des figures

Figure 1.1:illustration de la diversité des techniques biométriques……..7
Figure 1.2:lignes principales et secondaires (texture) crête ridules……..……..……..9
Figure 1.3: Architecture d'un système biométrique……..……..12
Figure1.4:graphe démonstratif EER représente la marge d'erreur autorisée par un système..15
Figure 1.5: Applications biométriques…….……...…..16
Figure 2.1: Empreinte digitale…….....20
Figure 2.2 : Type d'empreinte Arch……...……...21
Figure 2.3 : Type d'empreinte Boucle à droite……..……..21
Figure 2.4 : Type d'empreinte Tourbilon…….……...……21
Figure 2.5:Type d'empreinte deux spirales……..…….……..……...21
Figure 2.6: des familles des dessins digitaux…….....22
Figure 2.7: Point singulier globaux :des deltas……..…….…....22
Figure 2.8: Point singulier locaux :différentes type de minuties……...…..23
Figure 2.9:Types de minuties possibles(stries en noir)…….…….....23
Figure 2.10: Les formes des crêtes à la zone centrale de l'empreinte……..…....25
Figure 2.11: Architecture générale d'un système complet de reconnaissance d'empreintes.....25
Figure 2.12: Le filtre Gaussien……..……...……...29
Figure 2.13: Schéma général des différentes étapes d'un système de reconnaissance…….....30
Figure2.14: Extraction des minuties par Binarisation…… 30
Figure 2.15:Processus de binarisation/amincissement……..…32
Figure 2.16: Type de minuties……..33
Figure 2.17: l'extraction de minuties……33
Figure2.18: Des images d'empreintes de différentes qualités……..…34
Figure2.19: Exemple de minuties détectées, segment trop court (a), branche parasite (b), vraie terminaison (c), vraie bifurcation (d), triangle (e), pont (f), ilot (g), segment trop court (h)....35
Figure 2.20: Structures de fausse minutie……...……..…36
Figure 2.21 :Empreinte Avant le élimination de fausse minutie..…....……...……..36
Figure 2.22: Empreinte Après le élimination de fausse minutie..……...…..……...36
Figure2.23: Validation des terminaisons détectées : Vraie terminaison(a), Branche parasite (b)Segment trop court (c)…….…... 37

Figure2.24:Définitions associées à une bifurcation lors de la phase de validation……....37
Figure 3.1:Sur-apprentissage et complexité……..…….41
Figure 3.2 : Complexité de la classe des fonctions f……..……..….43
Figure3.3: Hyperplan séparateur entre 2 classes…….……..……..44
Figure 3.4 : Il peut exister plusieurs hyperplans séparant 2 classes…….……..…...…….44
Figure 3.5 :Les SVM trouvent l'hyperplan optimal pour la généralisation (un nouveau vecteur est bien classe dans le cas b)……..…….……..…….…....45
Figure 3.6 : Illustration de l'effet du changement d'espace par une fonction noyau..…...….47
Figure 3.7: Architecture du système en stratégie Un-contre-Tous…….…....51
Figure 3.8: Frontières de décision linéaires dans la stratégie Un-contre-Tous pour un problème à trois classes les zones d'ambigüités sont hachurées……..……..51
Figure 3.9: Classification de trois classes linéairement séparables par une SVM « un-contre-un » zone d'ambigüité est hachurée……..……....52
Figure 3.10: Architecture détaillée du classifieur SVM implémenté……..…..……53
Figure 4.1: schéma de notre système……... ….. 56
Figure 4.2: Code Source en java d'extraction de caractéristiques de l'empreinte digitale ….. 58
Figure 4.3 : Illustration des deux phases utilisées d'un classifieur SVM……..….. 61
Figure 4.4 : Illustration de la relation entre l'application utilisateur et le package SVM…...65
Figure 4.5 : Relation en détail entre l'application utilisateur et le package libsvm 2.83….....66
Figure 4.6 :interface de notre système……..…….....68
Figure 4.7: Résultat de notre système……..……69

Liste des tables

Tableau 2.1: Les différents types de minuties……...19
Tableau 3.1:Les types de noyaux……..….43

Introduction générale

Aujourd'hui, la croissance internationale des communications (déplacement physique, transaction financière, accès aux services…) implique le besoin de s'assurer de l'identité des individus. La lutte contre les fraudes d'informations personnelles, continue et les constructeurs de distributeurs automatiques s'engagent sur la voie de nouvelles technologies comme la biométrie. Il y a donc un intérêt grandissant pour les systèmes d'identification et d'authentification biométriques.

La biométrie utilise pour identifier une personne, ses propres caractéristiques physiques qui ne peuvent pas être changées, ni perdues ni encore volées.
En effet, les caractéristiques physiques d'un individu sont universelles (exister chez tous les individus), uniques (permettre de différencier un individu par rapport à un autre), permanentes (ne changent pas au fil du temps), enregistrables (collecter les caractéristiques d'un individu avec l'accord de celui-ci), mesurables (permettre une comparaison future).

Les systèmes biométriques peuvent être généralement divisés en trois catégories :

a) Analyses biologiques : Odeur, sang, salive, urine, ADN…

b) Analyses comportementales : La dynamique de la signature (la vitesse de déplacement du stylo, les accélérations, la pression exercée), la façon d'utiliser un clavier d'ordinateur (la pression exercée, la vitesse de frappe), la voix, la manière de marcher (démarche)...

c) Analyses morphologiques : empreintes, forme de la main, traits du visage, dessin du réseau veineux de l'œil…. Ces éléments ont l'avantage d'être stables dans la vie d'un individu et ne subissent pas les effets dû au stress par exemple, que l'on retrouve dans l'identification comportementale.

Le système biométrique utilise le matériel pour capturer les informations biométriques et le logiciel pour les gérer et les maintenir. Parmi toutes ces techniques, l'utilisation de l'empreinte digitale comme un moyen d'identification et d'authentification, est celle qui est la plus courante. La force de ce procédé tient du fait que l'utilisation de l'empreinte digitale est généralement plus facile d'acception par la communauté, et qu'elle est une des plus efficaces et des moins coûteuses. La raison principale de l'utilisation d'empreinte dans le système identification ou vérification est que l'empreinte est unique et reste invariable avec l'âge.

Objectifs

Les travaux effectués dans le cadre de ce mémoire portent donc sur la Reconnaissance de l'Empreinte Digitale (RED), qui est le moyen le plus utilisé.

Cependant, la plupart des systèmes d'identification des empreintes digitales consiste à extraire en premier lieu de l'empreinte à étudier tous les points de minuties, et ensuite comparer ces points avec ceux des modèles enregistrés dans une base de données pour trouver le modèle qui présente le plus de corrélation avec l'empreinte étudiée. Mais Cette approche présente quelques difficultés comme par exemple:

- Il est très difficile d'extraire les minuties d'une image d'empreinte digitale bruitée .Ce problème est très fréquent dans la pratique.
- Harmoniser les points de minuties est un passage obligé dans ce système, cependant, le nombre de minuties extraits de chaque empreinte digitale n'est pas uniforme et n'est pas cohérente. Cela rend le temps de calcul de l'étape d'identification très long.
- Pour identifier une empreinte digitale à partir d'un vecteurs caractéristique enregistré, nous utiliserons les techniques d'apprentissages statistiques à partir d'exemples; en particulier, nous avons opté pour les machines à vecteurs de support multi classes(SVM), et ce pour leur robustesse et leur grande capacité de généralisation à partir d'exemples.

Organisation de mémoire

Afin de présenter la méthodologie adoptée pour la RED, nous proposons dans le premier chapitre un panorama sur la biométrie et les empreintes digitales. Nous montrons que ces systèmes atteignent actuellement des performances satisfaisantes, notamment grâce aux progrès réalisés ces dernières années dans le domaine de la reconnaissance de l'empreinte
digitale.

Le deuxième chapitre est consacré à l'étude des différentes étape de reconnaissance de l'empreinte digitale (prée-traitement, extraction de minuties…).

Le troisième chapitre expliquera en détails le méthode utilisée pour l'identification (SVM) , Ensuit donnant les avantages et les inconvénients de cette méthode .

Le quatrième chapitre présente le dernier étape dans ce system la classification , ensuit nous présentons notre application et les fonctions comme résultat de notre travailler.

Ce mémoire est terminé par une conclusion générale mettant en relief les résultats obtenus ainsi que des perspectives à réaliser à long terme.

Chapitre I:

La Biométrie

1.Introduction

De plus en plus, notre société éprouve le besoin de se contrôler. que ce soit pour garantir la sécurité des gens dans les lieux publics ou pour éviter le détournement ou le vol d'informations sensibles ce qui pose un grand problème pour les personnes, les entreprises et les gouvernements dans leur quête de protection de données contre le vol.

Il existe traditionnellement deux manières d'identifier un individu :

- La première méthode est basée sur une connaissance (knowledge-based). Cette connaissance correspond par exemple au mot de passe utilisé au démarrage d'une session Unix ou au code qui permet d'activer un téléphone portable.
- La seconde méthode est basée sur une possession (token-based). il peut s'agir d'une pièce d'identité, une clef, un badge, etc. ces deux modes d'identification peuvent être utilisés de manière complémentaire afin d'obtenir une sécurité accrue comme pour la carte bleue.

Cependant, elles ont leurs faiblesses respectives. dans le premier cas, le mot de passe peut être oublié par son utilisateur ou bien deviné par une autre personne. on estime ainsi qu'une personne sur quatre écrits directement sur sa carte bleue son code secret afin de ne pas l'oublier. Dans le second cas, le badge (ou la pièce d'identité ou la clef) peut être perdu ou volé.

L'apparition de l'ordinateur et sa capacité à stocker les données et traiter les caractéristiques biométriques par certain ordre de processus automatisés à l'aide des dispositifs comme des modules de balayage, ont permis la création des systèmes biométriques informatisés qui envahit notre quotidien depuis quelques années. l'utilisation de la biométrie qui permet de vérifier que l'usager est bien la personne qu'il prétend être, s'est répandue énormément dans la vie quotidienne et trouve de nombreuses applications.

De nombreux travaux de recherche ont été menés et en cherche toujours de nouvelles méthodes, devant cette déferlante, il était nécessaire de faire le point sur ce qu'est exactement la biométrie, quelles techniques existent vraiment et leur degré de fiabilité pour ensuite détailler les plus utilisées dans ce chapitre.

2.Définition de la biométrie

On peut la définir comme suit :

- Le terme de **biométrie** est originaire d'une contraction des deux anciens termes grecs :« bios » qui signifie : la vie et « metron » qui se traduit par : mesure.

- La **biométrie** est la science d'établir l'identité d'une personne par l'analyse mathématique basée sur les attributs morphologiques (empreinte digitale, visage…etc.) ou comportementales (la démarche, la dynamique de frappe au clavier, la voix…etc.) ou biologiques (salive, ADN…etc.) liés à un individu. Ces caractéristiques sont appelées modalités biométriques qui doivent être fiables, infalsifiables, universelles, uniques pour chaque individu, permanentes, enregistrables et finalement mesurables [1].

3.Pourquoi la biométrie ?

Les arguments pour la biométrie se résument en 2 catégories:

Praticité : Les mots de passe comme les cartes de crédit, les cartes de débit, les cartes d'identité ou encore les clés peuvent être oubliés, perdus, volés et copiés.

En plus, aujourd'hui tous et chacun doivent se rappeler une multitude de mots de passe et avoir en leur possession un grand nombre de cartes. De son côté la biométrie serait immunisée contre ce genre de maux en plus qu'elle serait simple et pratique, car il n'y a plus ni cartes ni mots de passe à retenir.

La biométrie serait capable de réduire, sans l'éliminer, le crime et le terrorisme car, à tout de moins, elle complique la vie des criminels et des terroristes[2].

4.Etat de l'art des techniques biométriques

Dans les mesures de la biométrie ils existent trois types principaux (morphologiques, comportementales et biologiques) et cela pour obtenir des informations concernant les traits personnels[1].

La diversité de modalité biométrique comme la figure 1.1 illustre apparaît continûment de nouvelle, dans ce qui suit nous ne décrivons que les modalités les plus communes à savoir le visage, la parole…etc.

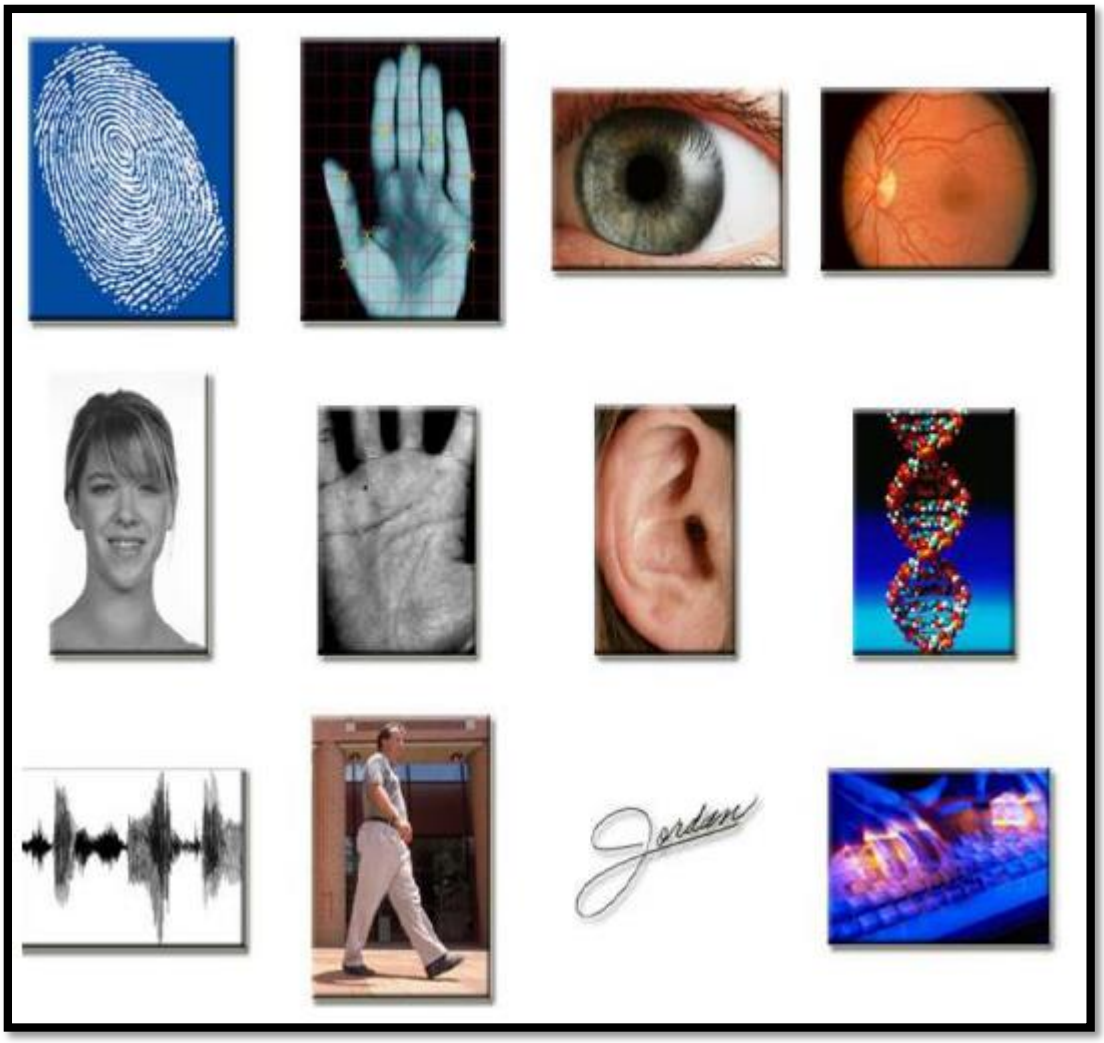

Figure 1.1:illustration de la diversité des techniques biométriques.

4.1.Biométrie morphologique (physique)

A. Empreinte digitale :L'identification à l'aide des empreintes digitales est la technique biométrique que la plupart des gens connaissent. Il s'agit de la plus vieille technique biométrique[1], les lecteurs d'empreintes digitales scannent puis relèvent des éléments permettant de différencier les empreintes. ces éléments sont appelés minuties.

Les minuties sont des changements de continuité de l'empreinte digitale. existe plusieurs types de minuties : lac, bifurcation, delta ou impasse...etc.

généralement une quarantaine sont extraites de la zone scannée. statistiquement il est impossible de trouver douze points identiques chez deux individus.

Ce type de système est utilisé par les institutions financières pour leurs employés et leurs clients. il se retrouve également dans les hôpitaux, les écoles, les aéroports, les cartes d'identité, les passeports, les permis de conduire et de nombreuses autres applications.

Son prix est faible, la taille du lecteur biométrique d'empreinte digital n'est pas volumineuse et le système reste très simple à mettre en place.

B. Visage : Le visage est sujet à une variabilité tant naturelle (vieillissement, par exemple) que volontaire (des produits de beauté, chirurgie esthétique, grimaces...etc.). Cette réalité demeurera un défi pour des systèmes d'identification de visage. L'individu doit être positionné devant la caméra ou peut être en mouvement à une certaine distance, le système retire certaines caractéristiques essentielles, uniques, et invariables comme (yeux, nez, le haut des joues, les coins de la bouche…etc.) selon le système utilisé.
La reconnaissance du visage est utilisée comme système de surveillance ou d'identification par les autorités ou les corps policiers principalement dans les lieux publics. elle est parmi les techniques les plus acceptables, mais elle nécessite un arrière-plan simple et fixe pour que le résultat soit précis[1].

C. L'iris :L'iris est la partie colorée de l'œil qui entoure la pupille noire. l'acquisition de l'iris est effectuée au moyen d'une caméra pour pallier aux mouvements inévitables de la pupille. Son inspection attentive révèle de nombreuses structures détaillées uniques et indépendantes du code génétique de l'individu et pratiquement ne varient pas pendant la vie. Environ 250 caractéristiques sont capturées, l'identification par l'iris est presque infalsifiable s'accroit en popularité ces dernières années dans le secteur financier pour les employés et les clients, dans les institutions carcérales, dans les aéroports et c'est une technique qui continuera sans doute à être employée couramment, mais elle est relativement désagréable pour l'utilisateur car l'œil doit rester grand ouvert et il est éclairé par une source lumineuse pour assurer un contraste correct toutefois la fraude étant néanmoins possible en utilisant des lentilles.

D. Empreintes des articulations des doigts –FKP-: Chaque doigt possède trois articulations et trois os qui sont appelés la phalange proximale, la phalange médiane et la phalange distale.la première articulation est l'endroit où le doigt se joint à la main appelé la phalange proximale. le deuxième joint est l'articulation inter phalangienne proximale, ou conjointe PIP : l'articulation du doigt et la surface arrière du doigt, il est également connu sous le nom dos de la main. Les modèles de peau inhérents à la surface extérieure autour de l'articulation de la phalange de doigt de l'individu, à une grande capacité à discriminer

des individus différents. Tel motif d'image du doigt est unique et peut être l'obtention ligne, hors ligne pour l'authentification.

L'extraction d'éléments de jointure pour l'identification dépend de l'utilisateur. Certains chercheurs extraient les caractéristiques pour l'authentification qui sont représentés sur la figure 1.2 .

Les caractéristiques sont centre de phalangienne, ligne en forme de U autour de la phalange. Le nombre de lignes, la longueur et l'espacement entre les lignes. 'Knuckle' motifs pliage et les bavures comme moyen d'identification photographique. Ces caractéristiques sont uniques et peuvent être utilisé pour l'identification.

Figure 1.2 :lignes principales et secondaires (texture) crête ridules.

4.2. Biométrie comportementale

A. Écriture (signature) : La vérification par signature comme technique est parmi les premières utilisées dans le domaine de la biométrie.

Elle se base généralement sur le fait que l'utilisateur signe avec un stylo électronique sur une palette graphique. Il y-a plusieurs systèmes concurrents dans ce domaine analysant les caractéristiques spécifiques d'une signature comme précision géométrique, variations de vitesse, pression exercée sur le crayon, le mouvement, les points et les intervalles de temps où le crayon est levé….etc.

Ces données sont enregistrées pour comparaison ultérieure. Certains systèmes ne font qu'enregistrer l'image statique de la signature pour comparaison [1].

L'acceptation de cette technique est très bonne car la signature est un geste commun pour tout le monde, ces systèmes sont utilisés dans les compagnies pharmaceutiques, les prisons, les services postaux et les banques, mais cette technique n'est pas très précise car

la signature peut être affectée par des facteurs physiques et émotionnels au même temps il y-a des incohérences de certaines personnes en signant leur nom dynamiquement et graphiquement aussi la falsification est possible en passant par une phase d'apprentissage.

B. Voix : La voix humaine est une caractéristique biométrique intéressante, puisqu'elle dépond des facteurs comportementaux et physiologiques.

Initialement une table de référence de la voix d'une personne doit être construite. Pour ce faire, celle-ci doit lire une série de phrases ou de mots à plusieurs reprises.

Les caractéristiques physiologiques de la voix d'un individu comme le débit, la force (pitch), la dynamique et la forme des ondes produites sont uniques et invariantes mais les caractéristiques comportementales changent avec le temps, selon les conditions sanitaires (mal de gorge) et des états émotionnels…etc.

Ce qui diminue l'exactitude du taux d'identification. Ces systèmes sont utilisés par les corps policiers, les agences d'espionnage et en téléphonie.

La capture de la voix est relativement facile à effectuer à l'aide d'un microphone. Mais ce moyen est sensible à un grand nombre de facteurs tels que le bruit, la fatigue, le stress ou la maladie peuvent altérer la voix. Aussi la fraude est possible par enregistrement. Ce qu'il ne rend pas un système complètement fiable.

C. Démarche : Il s'agit de reconnaître un individu par sa façon de marcher et de bouger, en analysant les déformations des jambes et bras au niveau des articulations.

La démarche serait en effet étroitement associée à la musculature naturelle, donc, elle est très personnelle, l'intérêt de cette technologie réside que l'identification de démarche se situe dans la capacité d'identifier un individu à distance[1].

Elle peut, aussi, détecter les comportements suspects (par vidéo surveillance), on l'utilise pour le contrôle d'accès aux bâtiments ou aux zones réglementées mais elle est facilement modifiable par l'individu.

D. Dynamique de frappe au clavier : Un tel système est peu coûteux, mais pas celui-ci car il ne nécessite pas de matériel d'acquisition autre que le clavier de l'ordinateur. Il s'agit d'un dispositif logiciel qui calcule la durée entre frappes, fréquence des erreurs où son temps de relâchement « Software Only », cette mesure est capturée environ mille fois par seconde, elle est appliquée au mot de passe qui devient ainsi beaucoup plus difficile à

«imiter », lors de la mise en place de cette technique il est demandé à l'utilisateur de saisir son mot de passe une dizaine de fois de suite.

Ce dispositif biométrique est utilisé comme méthode de vérification pour le commerce électronique et comme mécanisme de contrôle d'accès à des bases de données.

Il est facilement accepté par l'utilisateur, le but principal de cette technique est de renforcer la sécurité à des coûts moins élevés.

4.3. Biométrie biologique

A. La rétine :cette mesure biométrique se base sur le fait que les vaisseaux sanguins d'une rétine sont différents d'une personne à une autre et stables durant la vie.

L'utilisateur doit placer son œil à quelques centimètres d'un orifice de capture situé sur le lecteur de rétine. Un faisceau lumineux traverse l'œil jusqu'aux vaisseaux sanguins capillaires de la rétine. Le système localise et capture ainsi environ 400 points de référence. Cette technique demande la collaboration étroite de la part de l'utilisateur, car il doit placer son œil devant la caméra[1].

Cette technologie est la plus complexe à falsifier, mais probablement à cause de son coût élevé elle n'est pas utilisée que dans les cas où la sécurité est primordiale, notamment dans le domaine militaire , le secteur spatial (NASA) et par des agences d'espionnage comme la CIA. L'analyse biométrique de la rétine est la technologie la plus difficile à mettre en œuvre, aussi elle trouve peu de faveur au sein de la communauté parce qu'elle présente un gène pour les utilisateurs (rester sans cligner les yeux pendant quelque instants).

B. Structure des veines : On a longtemps considéré que le modèle des veines dans l'anatomie humaine peut être unique aux individus. En conséquence, il y a eu de diverses réalisations du balayage de veine au cours des années, du balayage de **main**, au balayage de **poignet** et, plus récemment, au balayage de **doigt**. Cette technique utilise un «scanner du réseau veineux palmaire», pour être identifié il faut placer la surface concernée au-dessus du lecteur. Il s'agit, ici, d'analyser le dessin formé par le réseau des veines pour en garder quelques points caractéristiques[1].

5. Architecture d'un système biométrique

Il existe toujours au moins deux modules dans un système biométrique :

Le module d'apprentissage et celui de reconnaissance [3].Le troisième module (facultatif) est le module d'adaptation. Pendant l'apprentissage, le système va acquérir une ou plusieurs mesures biométriques qui serviront à construire un modèle de l'individu.

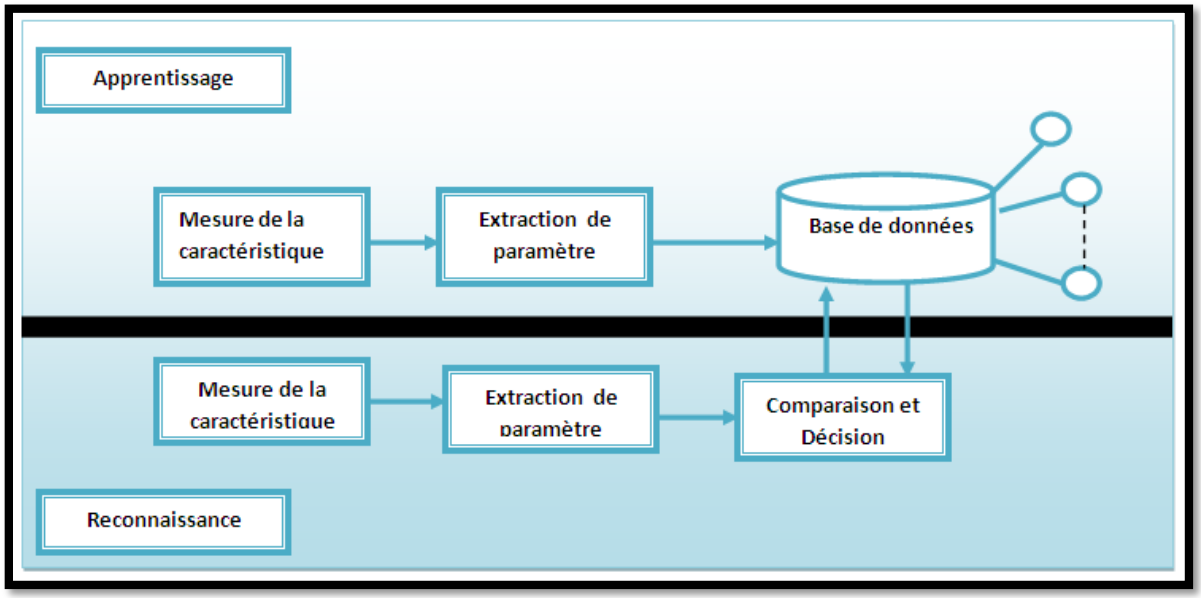

Figure 1.3: Architecture d'un système biométrique.

5.1.Module d'apprentissage

Au cours de l'apprentissage, la caractéristique biométrique est tout d'abord mesurée grâce à un capteur ; on parle d'acquisition ou de capture.

En général, cette capture n'est pas directement stockée et des transformations lui sont appliquées. en effet, le signal contient de l'information inutile à la reconnaissance et seuls les paramètres pertinents sont extraits.

5.2. Module de reconnaissance

Au cours de la reconnaissance, la caractéristique biométrique est mesurée et un ensemble de paramètres est extrait comme lors de l'apprentissage. Le capteur utilisé doit avoir des propriétés aussi proches que possibles du capteur utilisé durant la phase d'apprentissage.

Si les deux capteurs ont des propriétés trop différentes, il faudra en général appliquer une série de prétraitements supplémentaires pour limiter la dégradation des performances.

La suite de la reconnaissance sera différente suivant Le mode opératoire du système identification ou vérification.

En mode identification, le système doit deviner l'identité de la personne. Il répond donc à une question de type : « Qui suis-je ? ». Dans ce mode, le système compare le signal mesuré avec les différents modèles contenus dans la base de données (problème de type 1:n).

En général, lorsque l'on parle d'identification, on suppose que le problème est fermé, c'est-à-dire que toute personne qui utilise le système possède un modèle dans la base de données.

5.3. Module d'adaptation

Pendant la phase d'apprentissage, le système biométrique ne capture souvent que quelques instances d'un même attribut afin de limiter la gêne pour l'utilisateur. Il est donc difficile de construire un modèle assez général capable de décrire toutes les variations possibles de cet attribut.

De plus, les caractéristiques de cette biométrie ainsi que ses conditions d'acquisition peuvent varier. l'adaptation est donc nécessaire pour maintenir voire amélioré la performance d'un système utilisation après utilisation.

6. Evaluation des performances des Systèmes biométriques

Une question qui se pose souvent dans ce domaine est la suivante :

« Quelle est la meilleure technique biométrique? »

La réponse naturellement est qu'il n'y a aucune meilleure technique biométrique en termes absolus, tout dépend de la nature précise de l'application et des raisons de son exécution. L'International Biometric Group [IBG]– à effectuer une étude basée sur quatre critères d'évaluation :

1) Intrusivité: l'existence d'un contact direct entre le capteur utilisé et l'individu à reconnaitre.

2)Fiabilité :ce critère influe sur la reconnaissance de l'utilisateur par le système.

3) Coût :doit être raisonnable.

4) Effort :déployer par l'utilisateur lors de la saisie de mesures biométriques.

Au même temps, on peut mesurer la performance d'un système biométrique par deux indices : le FAR et le FRR.

- Le FAR: Ce taux représente le pourcentage d'individus reconnus par le système biométrique alors qu'ils n'auraient pas dû l'être. le système classe alors deux caractéristiques provenant de deux personnes différentes comme appartient à la même personne (indique la probabilité qu'un utilisateur soit reconnu comme quelqu'un d'autre).

$$\mathbf{FAR} = \frac{Nombre\,des\,imposteurs\,acceptés}{Nombre\,totale\,d'accès\,imposteurs}$$

- Le FRR: Ce taux représente le pourcentage d'individus censés être reconnus par le système mais qui sont rejetés. le système classe alors deux caractéristiques biométriques provenant de la même personne comme provenant de deux personnes différentes (indique la probabilité qu'un utilisateur connu soit rejeté).

$$\mathbf{FRR} = \frac{Nombre\,des\,clients\,rejetés}{Nombre\,totale\,d'accès\,clients}$$

Ces deux indices sont liés : une diminution du FAR entraîne systématiquement une augmentation du FRR(et inversement). Il s'agit d'adapter le système en fonction du niveau de sécurité souhaitée.

La statistique la plus simple pour mesurer la performance d'un algorithme est de calculer le point d'équivalence des erreurs.

- **Le EER:** Il est fréquemment utilisé pour donner un aperçu de la performance d'un système, ce taux est calculé à partir des deux premiers critères et constitue un point de mesure de performance courant. ce point correspond à l'endroit où FRR = FAR, c'est-à-dire le meilleur compromis entre les faux rejets et les fausses acceptations comme représente la figure 1.4. seuls des systèmes qui produisent des taux EER faibles sont capables d'être déployés en mode identification. Ainsi, les protocoles d'évaluation diffèrent dans le mode identification et le mode vérification.

$$\mathbf{EER = FAR = FRR} = \frac{FAR*X+FRR*Y}{X+Y}$$

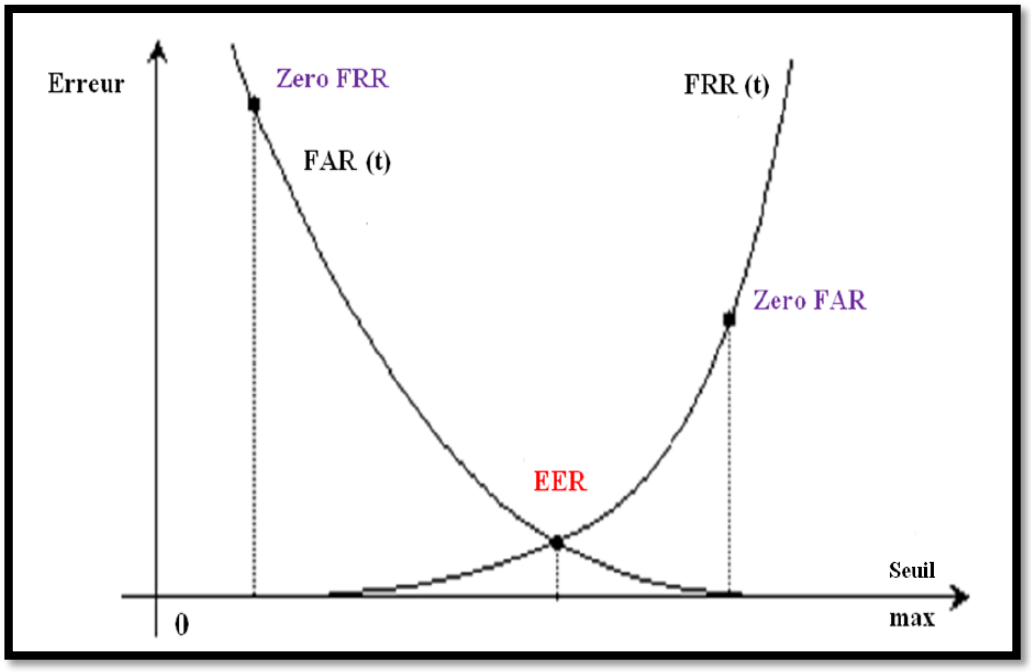

Figure1.4:graphe démonstratif EER représente la marge d'erreur autorisée par un système.

7 .Domaine d'applications de la biométrie

Aujourd'hui, les principales applications sont la production de titres d'identité, le contrôle d'accès à des sites sécurisés, le contrôle des frontières, l'accès aux réseaux, systèmes d'information et stations de travail, le paiement électronique, la signature électronique et même le chiffrement de données. Cette liste n'est pas exhaustive, et de nouvelles applications vont très certainement voir rapidement le jour[4].

Les techniques biométriques sont appliquées dans plusieurs domaines et leur champ d'application couvre potentiellement tous les domaines de la sécurité où il est nécessaire de connaître l'identité des personnes. Les applications peuvent être divisées en trois groupes principaux :

7.1. Application commerciales :telles que l'accès au réseau informatique, la sécurité de données électroniques, le commerce électronique, l'accès d'internet, l'ATM, la carte de crédit, le contrôle d'accès physique, le téléphone portable, le PDA, la gestion des registres médicales, l'étude de distances, etc.…

7.2. Applications de gouvernement :telles que la carte nationale d'identifications, le permis de conduite, la sécurité sociale, le contrôle de passeport, etc.…

7.3. Applications juridiques : telles que l'identification de cadavre, la recherche criminelle, l'identification de terroriste, les enfants disparus, etc.

Figure 1.5: Applications biométriques.

8.Les applications de la biométrie

Contrôle d'accès aux locaux:

- Salles informatiques.

- Sites sensibles (service de recherche, site nucléaire).

Equipements de communication:

- Terminaux d'accès.

- Téléphones portables.

Systèmes d'informations:

- Lancement du système d'exploitation,

- Accès au réseau.

- Transaction (financière pour les banques, données entre entreprises).

Machines & Equipements divers:

- Distributeur automatique de billets.

- Lieu sensible (club de tir, police).

- Contrôle des adhérents dans les clubs privés.

- Contrôle des temps de présence.

Etat/Administration:

- Fichier judiciaire.

- Services sociaux (sécurisation des règlements).

- Système de vote électronique.

9. Avantages et inconvénients de l'identification par empreinte digitale

Avantages :

La biométrie par l'empreinte digitale est la technologie :

- Le moins cher ;
- Plus rapide ;Plus pratique ;
- Le moyen le plus fiable pour l'identification.
- Il n'y a qu'une chance sur 17 milliards de trouver deux empreintes avec plus de 17 points de similitude.
- Les voitures, les téléphones cellulaires, les PDA, les ordinateurs personnels et des dizaines de produits et appareils utilisant les empreintes digitales sont de plus en plus.

Inconvénients :

- Cette technologie est ressentie comme intrusive.
- Certaines personnes peuvent créer de "faux doigt" en utilisant l'empreinte digitale d'une autre personne (sachant que l'empreinte est stockée dans la base de données du lecteur d'empreinte digitale).
- Le gros inconvénient est le manque d'hygiène, les traces de doigts se succèdent sur ce lecteur et ainsi les microbes se dispersent sur tout le lecteur ce qui rend celui-ci très sale.[5]

10. Conclusion

Dans ce premier chapitre, nous avons présenté le cadre de ce mémoire, aussi, nous avons mis en relief quelques notions et définitions de base liées à la biométrie et sa diversité technologique, les différents modes et modules des systèmes biométriques. Nous avons, aussi, donné un aperçu sur les techniques de mesure et leurs performances, ainsi, que les domaines d'applications.

Chapitre II:

Technique de Reconnaissance d' Empreinte Digitale

1. Introduction

Plusieurs caractéristiques humaines ont été exploitées par la biométrie pour l'identification et la vérification automatique des individus, les empreintes digitales sont un outil d'identification rapide, fiable et moins onéreux que certains autres. l'utilisation de l'empreinte digitale comme moyen d'identification d'une personne n'est pas nouvelle. c'est la technique biométrique la plus ancienne et la plus mature. les empreintes ont formellement été acceptées comme identificateur de personnes valide dès le début du siècle. elles ont d'abord étaient utilisées dans les milieux juridiques, avant de devenir une technique d'authentification effective.

2. Historique

Les premières traces d'utilisation d'empreintes digitales ont été découvertes en Egypte et datent de l'époque des pyramides il y a plus de 4000 ans. Les Chinois ont aussi utilisé très tôt ce moyen pour signer les documents officiels (le plus vieux document signé date du troisième siècle avant Jésus Christ) mais ils ne savaient sûrement pas que les empreintes étaient uniques pour chaque personne et permettaient ainsi une identification fiable. C'est en 1856 que l'anglais William Hershel [6], après avoir utilisé les empreintes en guise de signature sur la population indienne qu'il dirigeait, commença à comprendre que les empreintes étaient uniques et constantes dans le temps. En 1888 le britannique Francis Galton publia une étude sur les empreintes digitales où il établit leurs caractéristiques (unicité, empreintes fut adoptée officiellement en Angleterre dans le système judiciaire. Cette technique fut ensuite largement développée dans les enquêtes criminelles et permit de résoudre un bon nombre d'affaires. De nos jours les empreintes sont toujours largement utilisées et reconnues comme méthode d'identification fiable.

3. Définitions de l'empreinte digitale

Une empreinte digitale est le dessin formé par les lignes de la peau des doigts, des paumes des mains, des orteils ou de la plante des pieds. e dessin se forme durant la période fœtale. Il existe deux types d'empreintes : l'empreinte directe (qui laisse une marque visible) et l'empreinte latente (saleté, sueur ou autre résidu déposé sur un objet). elles sont uniques et

immuables, elles ne se modifient donc pas au cours du temps (sauf par accident comme une brûlure par exemple).[7]

Une empreinte digitale peut être divisée en quatre zones : la zone centrale (le centre de figure), la zone basale (la partie basse), la zone distale (la partie haute) et les zones marginales (les côtés).[8]

Figure 2.1: empreinte digitale

4. Caractérisation d'une empreinte digitale et description du motif

L'empreinte digitale est unique pour chaque individu et garde la même forme tout au long de la vie. Elle subit des transformations homothétiques ou des distorsions modélisables par des similitudes dues à la croissance. Lorsque l'épiderme est altéré, celui-ci se régénère de façon identique.

- Cette unicité est donc une opportunité pour identifier un individu, mais alors qu'est-ce qui caractérise l'empreinte digitale?

4.1. Motif

En regardant les images d'empreintes ci-dessous, on s'aperçoit que les lignes foncées localement parallèles appelées stries ou crêtes caractérisent la forme de l'empreinte. On peut répertorier trois grandes familles d'empreintes : arches ou tentes (Figure 2.2), boucles à droite (Figure 2.3) ou boucles à gauche, spires ou verticilles ou tourbillons (Figure 2.4).

Ces trois types d'empreintes regroupent 95% des doigts humains :

- 30% pour les spirales(tourbillon),
- 60% pour les boucles et
- 5% pour les tentes.

Figure2.2 **Figure2.3** **Figure2.4**

Des dessins beaucoup plus rares sont par exemple des doubles boucles imbriquées. Empreinte assez rare: deux spirales

Figure 2.5

Chaque empreinte possède un ensemble de points singuliers globaux (les *centres* et les *deltas*) et locaux (les *minuties*). Les centres correspondent à des lieux de convergences des stries, tandis que les deltas correspondent à des lieux de divergence [9]

Les éléments qui permettent de différencier deux empreintes digitales ayant le même motif sont :

- **D'une part les points singuliers globaux :**
  - Noyau ou centre :lieu de convergences des stries.[4]la forme du centre de figure permet de rattacher l'empreinte à une des familles des dessins digitaux : les arcs, les tentes pures, les composites, les boucles ou les verticilles.[10]

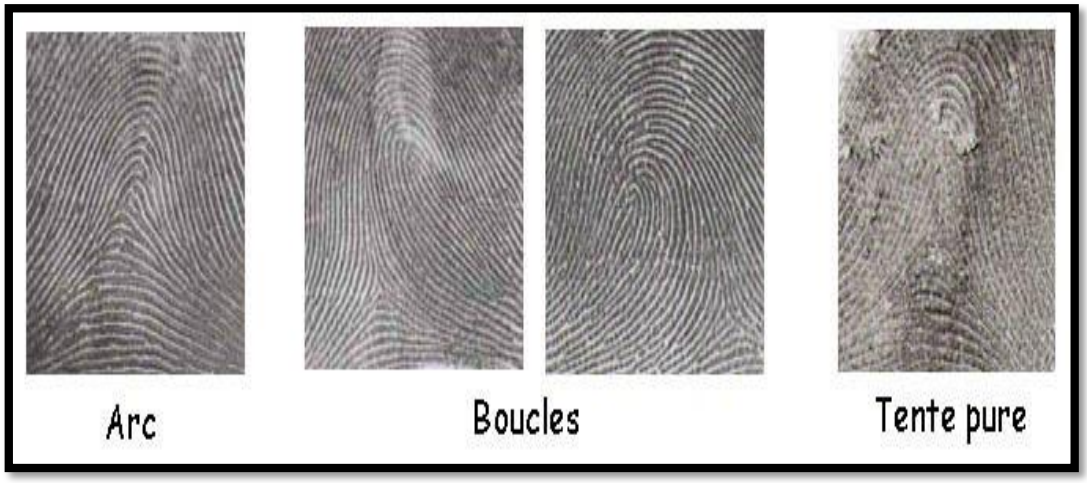

Figure 2.6: des familles des dessins digitaux

- Delta :Un delta est un point de convergence entre les zones centrale, basale et marginales. Le delta peut être en forme de triangle ouvert ou fermé, ou formé directement par une crête.[10]Voici quelques exemples de deltas :

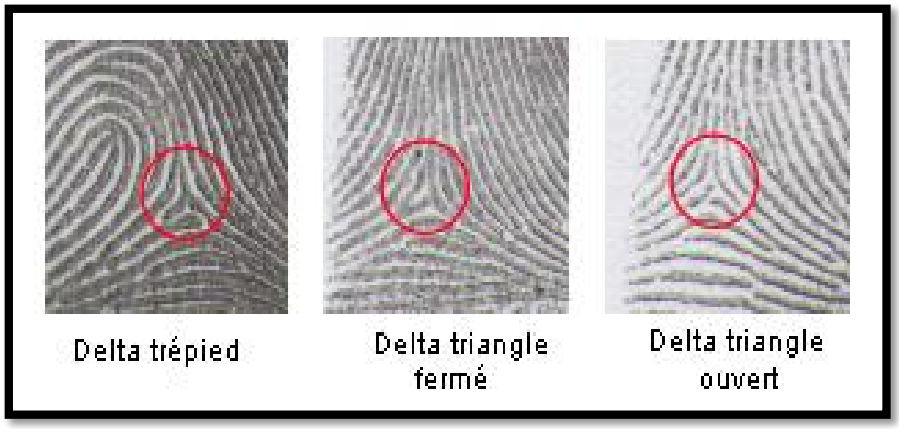

Figure 2.7: des deltas

➢ D'autre part les points singuliers locaux:

- Les minuties : points d'irrégularité se trouvant sur les lignes capillaires.[9]

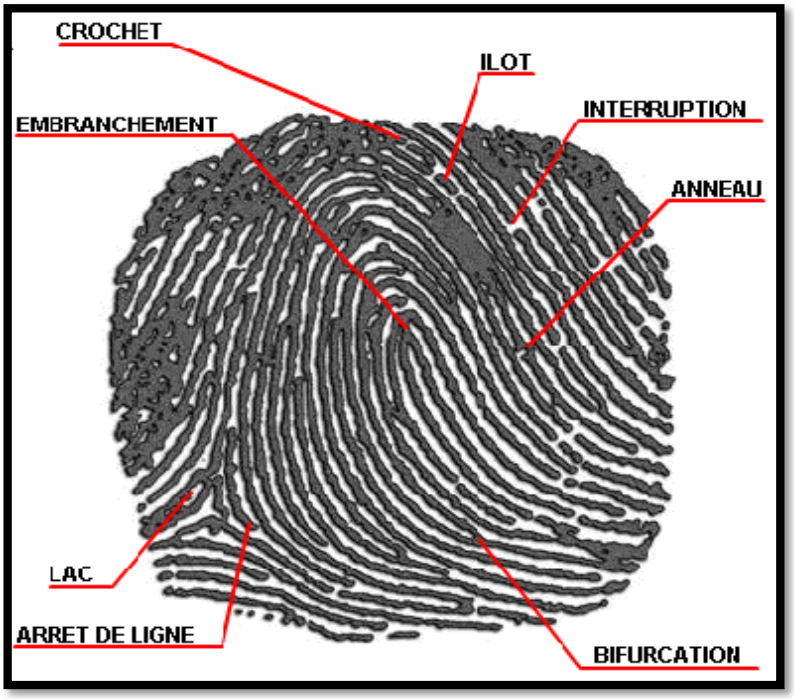

Figure 2.8:différentes type de minuties

Plusieurs études ont montré l'existence de seize types de minuties différentes (Figure 2.8) mais en général les algorithmes ne s'intéressent qu'aux bifurcations et terminaisons qui permettent d'obtenir les autres types par combinaison.

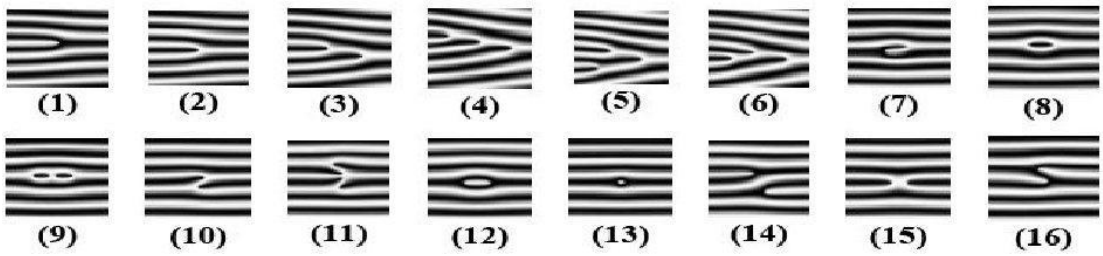

Figure 2.9: Types de minuties possibles (stries en noir)

1. termination	2. bifurcation simple
3. bifurcation double	4. bifurcation triple 1
5. bifurcation triple 2	6. bifurcation triple 3
7. crochet	8. boucle simple
9. boucle double	10. pont simple
11. point jumeau	12. intervalle
13. point isolé	14. traversée
15. croisement	16. tête bêche

Tableau 2.1: Les différents types de minuties[11]

5. Les classes de l'empreinte digitale

Français Galton (1822-1916) ont été faites les premières études scientifiques sur les classifications des empreintes digitales, ces études ont affiné par Edward Henry (1850-1931), il est classé les empreintes en cinque classes : arc, arc tendu, boucle à gauche, boucle à droite, et spire.

- **Classe 1**: il contient en maximum un Delta et au moins une crête montre une courbure élevée, est une classe poubelle.
- **Classe 2**:il contient un Delta à droite et des boucles situé en côté à gauche de l'empreinte.
- **Classe 3**: il contient un Delta à gauche et des boucles situé en côté à droite de l'empreinte.
- **Classe 4**: il contient un Delta à gauche et d'autre à droite avec un centre spirale.
- **Classe 5**: il contient trois Delta autour de forme besace
- **Classe 6**: il contient des empreintes invisibles.[12]

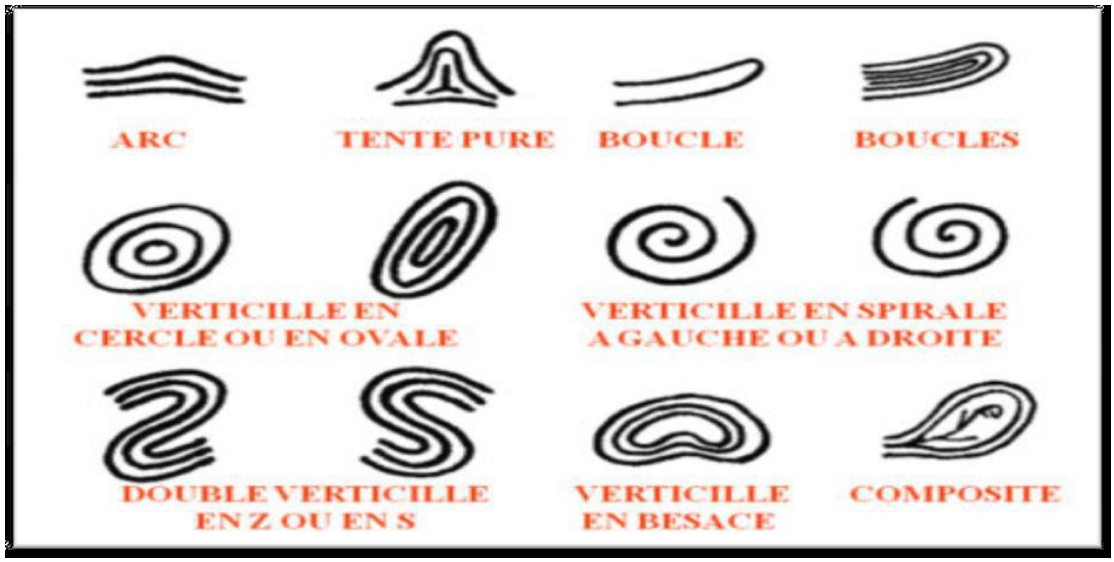

Figure 2.10: Les formes des crêtes à la zone centrale de l'empreinte.[12]

6.Structure d'un système complet de reconnaissance d'empreintes

6.1. Principe général

Un système automatique complet de reconnaissance d'empreintes digitales est une chaîne de processus qui à partir du doigt d'un utilisateur en entrée renvoie un résultat en sortie, permettant ainsi à l'utilisateur d'accéder ou non à des éléments nécessitant une protection .La réalisation d'un tel système a fait l'objet de très nombreuses recherches et des méthodes très différentes de traitement ont été proposées [13]. Cependant ces système répondent toujours à la même structure (Figure 2.11).

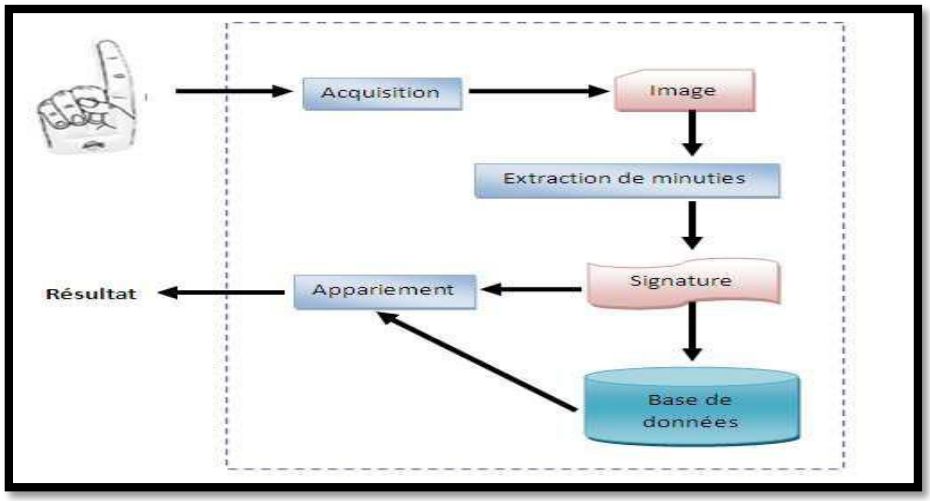

Figure 2.11: Architecture générale d'un système complet de reconnaissance d'empreintes.

La première phase permet d'obtenir une image de l'empreinte de l'utilisateur (**acquisition**), laquelle va subir un prétraitement pour extraire l'information utile de l'image (**signature**) suivi éventuellement d'un traitement supplémentaire permettant d'éliminer de possibles fausses informations qui se seraient glissées entre temps dans la chaîne de traitement. Ensuite si l'utilisation du système consiste juste à créer une base de données (**stockage**) la signature est éventuellement compressée puis stockée dans la base de données au moyen d'une technique d'archivage (**classification**).

Pour un système d'identification l'ensemble des empreintes présentes dans la base de données pouvant correspondre à celle de l'utilisateur (**modèle identique**) sont désarchivées et comparées (**appariement**) une à une avec celle de l'utilisateur, si une éventuelle correspondance est trouvée des informations personnelles concernant l'utilisateur sont renvoyées par le système. Dans le cas d'un système de vérification il n'y a qu'une seule comparaison et un résultat binaire est renvoyé, permettant l'acceptation ou le rejet de l'utilisateur.[13]

6.2. L'acquisition de l'empreinte

La première phase d'un système de reconnaissance consiste à obtenir une image de l' empreinte du doigt. Longtemps le seul moyen existant a été l'utilisation du papier et de l'encre ce qui a rendu la tâche de reconnaissance très lourde. En effet la qualité de l'image était plutôt mauvaise (plusieurs acquisitions étaient nécessaires) et l'extraction de la signature était effectuée visuellement par un expert (processus très long et pénible). Heureusement avec le développement de l'informatique et de la microélectronique de nouveaux moyens d'acquisition ont fait leur apparition, permettant ainsi d'accélérer la chaîne de traitement en l'automatisant (un capteur dédié fournit directement une image numérique).

6.3.L'extraction de la signature

La reconnaissance d'empreinte est basée sur l'extraction de la signature . La signature d'une empreinte digitale correspond à l'information utile nécessaire à l'identification fiable de la personne ou à l'archivage dans la base de données. Elle permet de caractériser de manière unique la personne.

Un extracteur de minuties cherche des terminaisons de stries et des bifurcations dans les empreintes. Si les stries sont bien déterminées, alors l'extraction de minuties est une tâche relativement simple. Cependant, dans la pratique, il n'est pas toujours possible d'obtenir une carte parfaite de stries. Donc la performance des algorithmes d'extraction de minuties dépend fortement de la qualité des images des empreintes digitales d'entrée.

6.4. Le stockage et la phase d'appariement

Pour les systèmes disposant de grosses bases de données, l'identification peut poser problème en temps de calcul si la signature d'entrée doit être comparée avec les signatures présentes dans la base donnée. C'est pourquoi, un processus de classification et de déclassification est nécessaire pour limiter les temps de recherche [14].

Lorsqu'une image est stockée, un groupe spécifique lui est attribué en fonction de ses caractéristiques. Lors de l'identification, on désarchive l'ensemble des signatures de la base correspondant au groupe de l'empreinte nécessitant l'identification. Puis chacune des images désarchivées est comparée avec celle de l'utilisateur. Ceci permet de réduire sensiblement le temps de recherche en limitant le nombre d'images à comparer. Parmi les différentes techniques existantes on distingue principalement : l'extraction des singularités de l'image (la position des centre et delta permet de déterminer la classe de l'empreinte) [15]. La phase d'appariement est l'étape critique du système, elle reçoit en entrée deux signatures issues de deux acquisitions différentes d'empreinte et renvoie en sortie un résultat binaire indiquant si oui ou non les deux signatures proviennent de la même empreinte [16].

6.5. Le prétraitement de l'image

Lors de l'acquisition de l'empreinte l'image obtenue contient souvent beaucoup ayant des origines diverses:

- Les substances parasites présentes sur le doigt (encre, graisse, saletés).
- La personne (cicatrices, métiers manuels, âge).
- L'environnement où se produit l'acquisition (température de l'air, degré d'humidité).

La reconnaissance d'une empreinte digitale est directement liée à la qualité de l'image obtenue au moyen du capteur. Ainsi dans la plupart des cas, un prétraitement est nécessaire pour améliorer la qualité de l'image. Pour limiter les calculs des étapes suivantes du système. Une opération de filtrage utilisant les caractéristiques locales de l'empreinte est ensuite appliquée à l'image de manière à améliorer sa qualité en éliminant le bruit.

6.5.1.Niveau de gris

Le niveau de gris est une image de profondeur k=8 bits, chaque pixel prendre l'une des valeurs entre de l'intervalle [0 …..255], tel que le zéro représente le noir et 255 représente le blanc. Dans les applications professionnelles 8 bits n'est pas suffisants, donc il y'a d'autre type d'image de niveaux de gris de profondeur k=14 bits ou k=16 bits.[12]

6.5.2.Amélioration de l'image

A. Filtres de Gabor

Filtre de Gabor est directement lié aux ondelettes de Gabor, puisqu'ils peuvent être conçus pour le nombre de *érosions et de rotations*. Cependant, généralement l'expansion n'est pas appliquée pour des ondelettes de Gabor, puisque ceci exige le calcul des ondelettes de bi orthogonal, qui peuvent prendre très du temps. Par conséquent, habituellement, une batterie de filtres se composant de Gabor filtre avec de diverses balances et des rotations est créée.

Les filtres sont enlacés avec le signal, ayant pour résultat un prétendu espace de Gabor. Ce processus est étroitement lié aux processus dans le cortex visuel primaire. les relations entre les activations pour un endroit spatial spécifiques sont très distinctives entre les objets dans une image. En outre, des activation importantes peuvent être extraites à partir de l'espace de Gabor afin de créer une représentation clairsemée d'objet.

$$h(x,y:\phi,f)=exp\left\{-\frac{1}{2}\left[\frac{(x\cos\phi)^2}{\delta_x^2}+\frac{(y\sin\phi)^2}{\delta_y^2}\right]\right\}\cos(2\pi f x\cos\phi),$$

Où ∅ est l'orientation et f est la fréquence.$\delta_x$et$\delta_y$sont les constance d'enveloppe Gaussien pour l'axis x et y respectivement[17].

Appliquant ce filtre pour une image, trois paramètres doivent être spécifiés : la fréquence de crêtes et de vallées, l'orientation de crêtes, la normalisation d'image :

$$\varepsilon(i,j)=\begin{cases}255, si R(i,j)=0\\ \sum_{u=-w/2}^{w/2}\sum_{v=-w/2}^{w/2}h(u,v,O(i,j),F(i,j))G(i-u,j-v), \text{sinon}\end{cases}$$

Où *w=11* est la taille de filtre Gabor.

Le résultat de cette étape est une image en niveau de gris. Elle est convertie à l'image binaire par un seuil moyen.

B. Filtre gaussien :

Le filtre Gaussien est un filtre de traitement d'image appliqué par convolution (utilise un masque (matrice) appliqué à chaque pixel) Ce type de filtre est utilisé pour diminuer le bruit ou appliquer un flou sur une image. Les résultats de filtre gaussien applique à une image d'empreinte digitale, est résumé dans le tableau suivant :

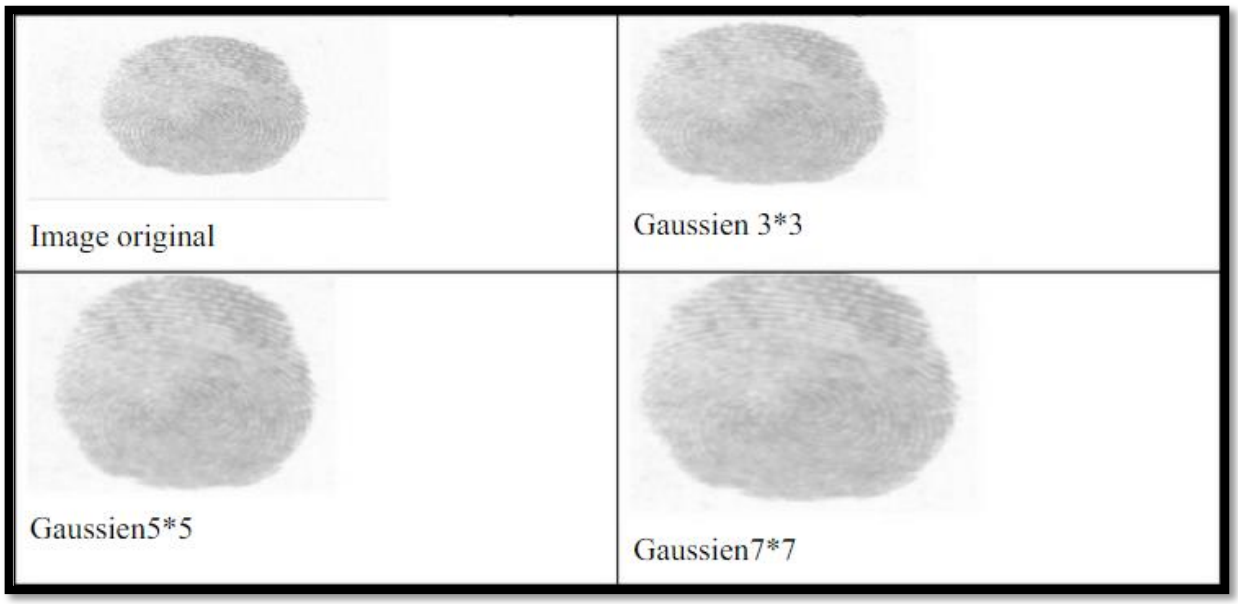

Figure 2.12: Le filtre Gaussien**[18]**

7.La reconnaissance d'empreinte

Nous présentons ici les deux étapes suivantes permettant de réaliser un système complet de reconnaissance d'empreinte digitale (voir Figure 2.15):

- **La phase d'extraction :** l'ensemble des minuties de l'empreinte (signature) est extrait à partir de l'image filtrée de l'empreinte. Pour cela nous étudierons deux méthodes:

  - **La méthode classique** consiste à extraire l'information sur un squelette binaire (noir et blanc) de limage filtrée.
  - **La méthode directe** consiste à extraire les minuties directement sur limage filtrée.

- **L'appariement :** c'est l'étape de reconnaissance qui consiste à calculer le degré de similarité entredeux signatures pour décider si oui ou non elles peuvent être considérées comme identiques.

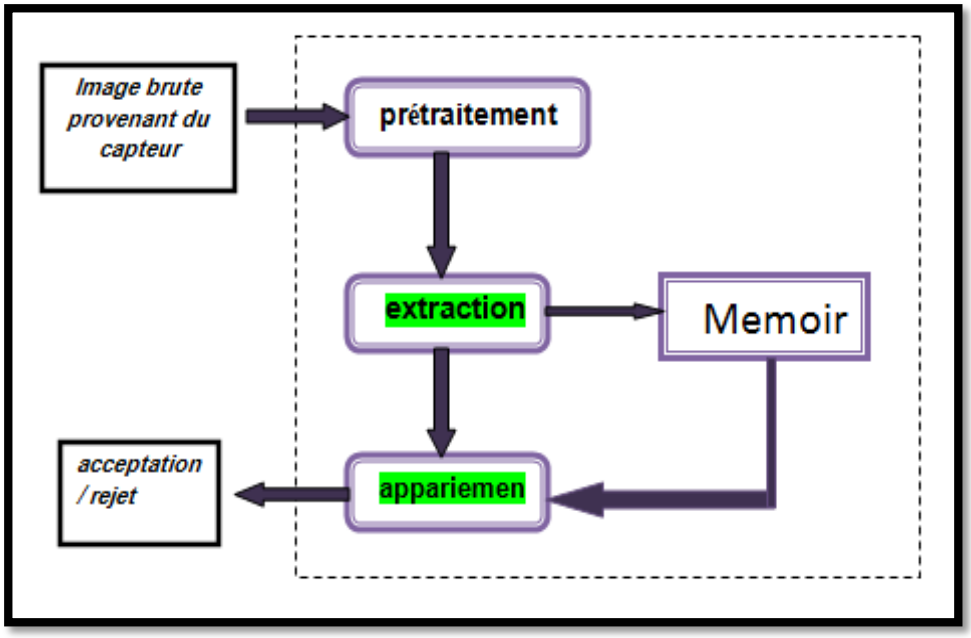

Figure 2.13: Schéma général des différentes étapes d'un système de reconnaissance.

7.1.Approche Classique d'extraction de minuties

Dans cette approche, l'image en niveau de gris est convertie en une image binaire. Puis elle sera amincie (squelettisée) afin de diminuer l'épaisseur des stries (Figure2.16), à ce stade, les minuties seront bien visible et facile à détecter. [46]

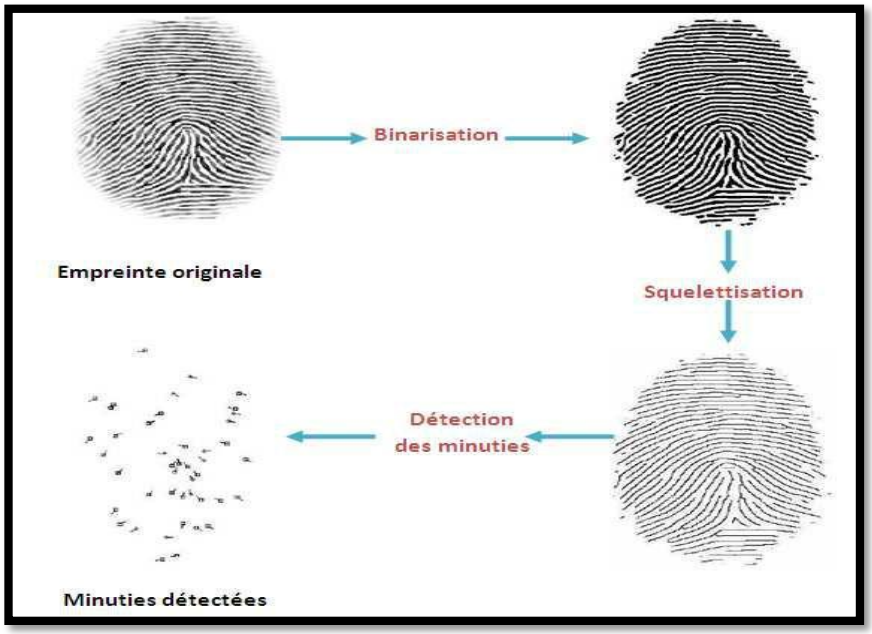

Figure2.14: Extraction des minuties par Binarisation [46]

7.1.1. La Binarisation

La Binarisation repose sur le choix d'un seuil global T. les pixels dont le niveau de gris est en dessous du seuil deviennent noirs, et ceux au dessus deviennent blancs.

$$I_T(x,y) = \begin{cases} 1 & \text{si } I(x,y)>T \\ 0 & \text{si } I(x,y)\leq T \end{cases}$$

Dans le cas d'une empreinte digitale, le but de la binarisation est de repérer les crêtes. Il existe des méthodes de binarisation optimales qui détermine le seuil T en se basant sur la distribution des niveaux de gris [46]. Cependant, Le contraste dans une image d'empreinte digitale peut varier considérablement à travers les différentes régions constituant l'image. Par conséquent, un seuil unique n'est pas suffisant pour une segmentation correcte. Ainsi, une binarisation adaptative est souvent préférée pour ce type d'images où le seuil T se détermine dans un voisinage local. Néanmoins, ces techniques manquent leurs efficacité si l'image de l'empreinte digitale est de mauvaise qualité, la solution étant d'utiliser la structure des crêtes pour réaliser la segmentation/binarisation [46].

7.1.2.La Squelettisation (amincissement)

La squelettisation consiste à réduire une forme en un ensemble de courbes, appelées squelettes. C'est un outil d'analyse de forme non-scalaire, qui conserve les propriétés topologiques de la forme d'origine ainsi que ses propriétés géométriques, selon la méthode employée. La squelettisation est une méthode qui a été développée à l'origine dans les années soixante par Harry Blum, en vue de créer un nouveau descripteur de formes. Son but est de limiter la perte d'information [46].

Image Originale [6] Image Binarisée Image squelettisée

Figure2.15: Processus de binarisation/amincissement

La squelettisation réduit l'épaisseur des crêtes en un seul pixel, ce qui facilite et simplifie l'étape de détection des minuties.

7.1.3.L'extraction des minuties

L'extraction des minuties consiste à calculer le nombre de connexion **CN** de chaque pixel blanc avec ses 08 voisins [46].

$$CN = 0.5 \sum_{i=1}^{8} |P_i - P_{i+1}|$$

Cette étape va extraire des minuties dans l'image sortie d'étape dernière. Cette image se compose des crêtes que son diamètre est un pixel. En général, le minutie est catégorisé par plusieurs types : le crête de bifurcation, la fin de ligne, le lac, l'île…

Pour chaque pixel *I(i,j)*, on traite un fenêtre de 3 x 3 autourde *(i,j)*, cinq cas différents sont reconnus :

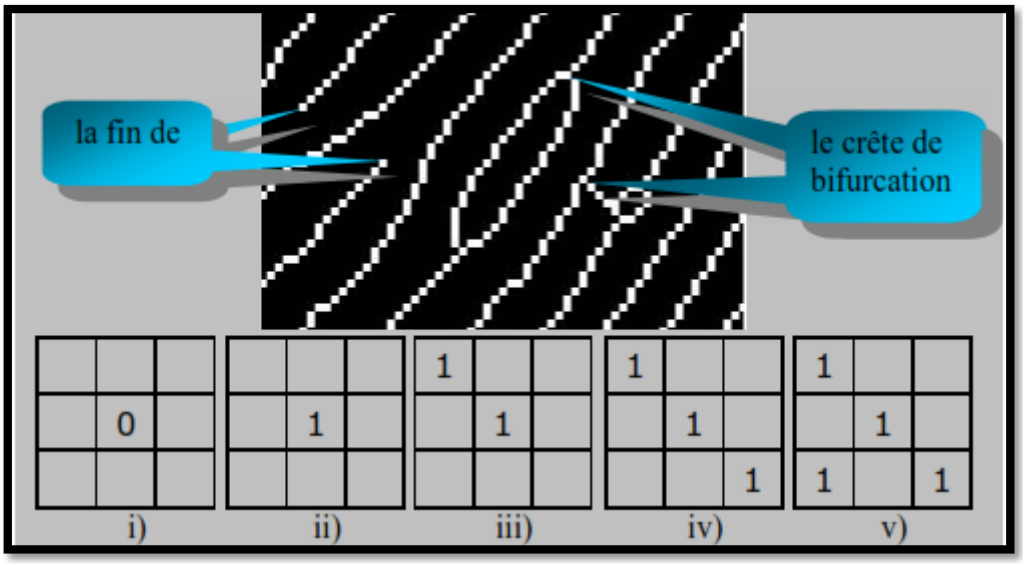

Figure 2.16: Type de minuties

i.I(i,j)=0, c'est la fond d'empreinte digitale.

ii.I(i,j)=1 et aucune crête voisine. C'est une île (marquer pour une minutie).

iii.I(i,j)=1 et un seul crête voisine. C'est un pixel à la fin de ligne (marque pour unes minutie).

iv.I(i,j)=1 et deux crêtes voisines. C'est un ligne continu et pas de minutie. Donc, on supprime cette crête.

v.I(i,j)=1 et trois crêtes voisines. C'est un la crête de bifurcation (marquer pour une minutie).

Le cas de I(i,j)=1 et plus de trois crêtes voisines ne compte pas car les crêtes d'empreinte digitale ne sont jamais transversales.[17]

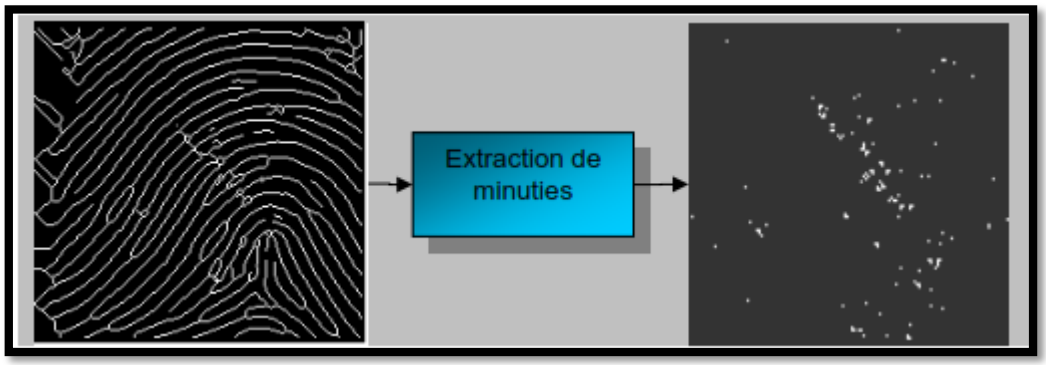

Figure 2.17: l'extraction de minuties

A la fin de cette étape, on obtient une liste de minuties.

7.1.4. Les problèmes rencontrés lors de l'extraction des minuties

La performance d'un système de reconnaissance d'empreinte digitale dépend de l'extraction des minuties et leurs appariements. Cependant, la mauvaise qualité de l'image en entrée engendre les problèmes suivant [46] :

- Création de fausses minuties.
- Ignorance de varies minuties.
- Problème de localisation (position et direction).

La qualité de l'empreinte rencontrée durant la vérification est très incertaine, elle varie sur une grande portée. La plus grande partie est endommagée par l'état de l'épiderme (Figure2.18) :

- Les crêtes se cassent par la présence de blessures ,de coupures.
- Des empreintes très sèches donnent des crêtes fragmentées.

Figure2.18: Des images d'empreintes de différentes qualités.

La qualité décroit de la gauche vers la droite. (a) image de bonne qualité avec un bon contraste (b) distinction insuffisante sur le centre de l'image (c) un empreinte sèche. [Bel,09].Les fausses minuties engendrent par la suite, l'échec de l'algorithme d'appariement.

7.1.5. Elimination des fausses minuties

Des solutions ont été proposées dans [46] pour éliminer les fausses minuties produites au cours du processus de binarisation et squelettisation. L'objectif étant de ne conserver que les vraies minuties. Pour cela, La distance entres les minuties a été utilisée comme moyen pour éliminer l minuties voisines qui ne répondent pas aux critères. En effet, la distance entre deux minuties voisines est toujours supérieure à un certain seuil et pratiquement, il est extrêmement rare de trouver deux vraies minuties très proches. Par contre on a approximativement une concentration locale de plusieurs fausses minuties.

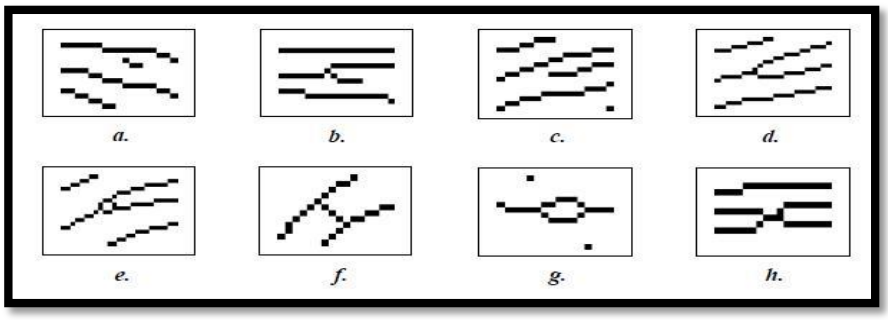

Figure2.19: Exemple de minuties détectées, segment trop court (a), branche parasite (b), vraie terminaison (c), vraie bifurcation (d), triangle (e), pont (f), ilot (g), segment trop court (h)

Lorsque le taux est calculé à des distances entre les arêtes parallèles D, et sur la base de cette valeur, il a été disposé.

1.Si la distance entre une bifurcation et une Terminaison est inférieure à D et les deux minuties Sont dans la même arête …(cas m1) elle sera supprimée.

2. Si la distance entre deux bifurcations est inférieure que D et ils sont dans la même arête, sera supprimée le deux bifurcations (m2, m3, m4, m6, m7Cas).

3. Si deux terminaisons se trouvent à une distance D et Leurs directions coïncident avec un petit angle variation. Et ils suffisent à la condition qu'aucun Toute autre terminaison est située entre les deux Terminaisons. Les deux terminaisons sont alors Considérée comme une fausse minutie dérivée d'un Crête et sont enlevés (cas m4, m5, m6).

4.Si deux terminaisons sont situées dans une arête courte Avec une longueur inférieure à D, retirez les deux Terminaisons (m7).[19]

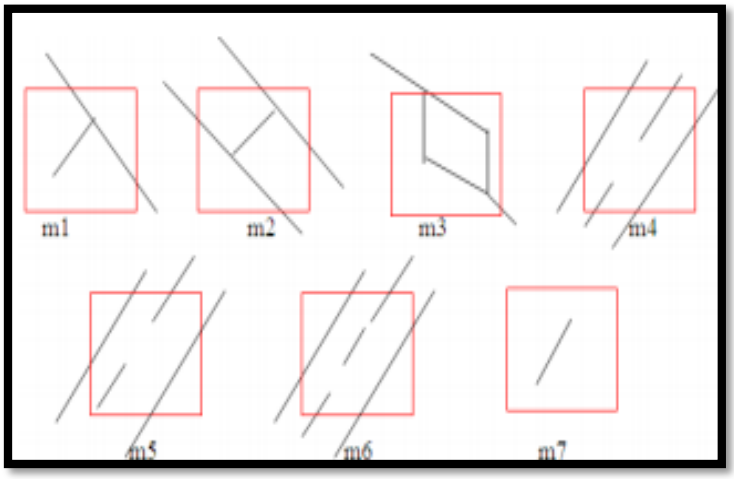

Figure 2.20: Structures de fausse minutie [19]

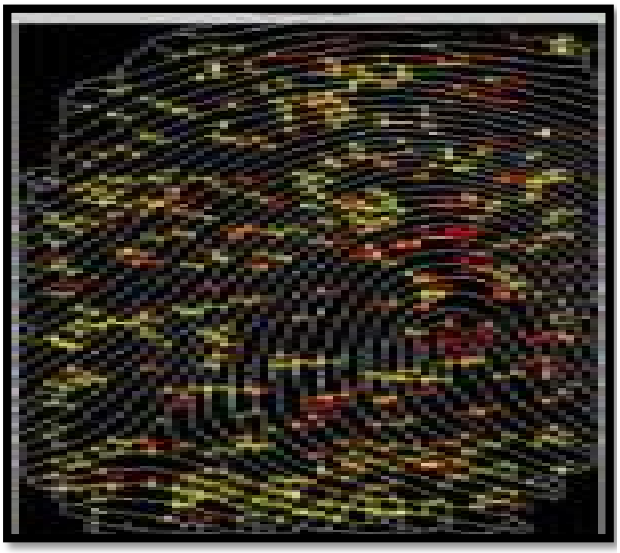

Figure 2.21 : Avant

Figure 2.22 :Après

7.1.6.Traitement des terminaisons détectées

Un point $T(x_T,y_T)$ est considéré ,ce point est une terminaison si (CN(T)=1), afin d'éliminer les fausses minuties nous devons vérifier si celui-ci se situe au bord de l'image car, la majorité des fausses terminaisons se trouvent aux bord de l'image. Pour les terminaisons restantes *T*, on parcourt la strie qui lui est associée sur une distance maximum K jusqu'à atteindre le point A (d=**K**, Figure2.24) [46].

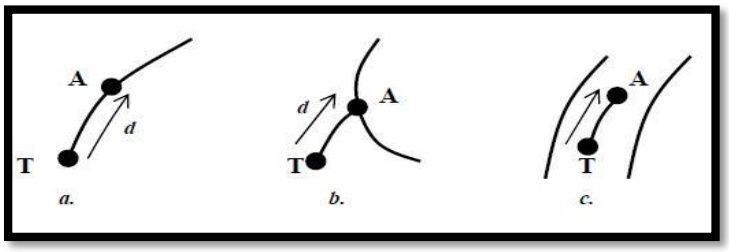

Figure2.23: Validation des terminaisons détectées : Vraie terminaison(a), Branche parasite (b) Segment trop court (c)

7.1.7.Traitement des bifurcations détectées

Lorsque l'on détecte un point B candidate pour le titre de *Bifurcation* (CN(B)=3), on parcourt les trois stries qui lui sont associées sur une distance maximum de K jusqu'à atteindre trois points A_1, A_2 et A_3 (Figure2.24) [46]

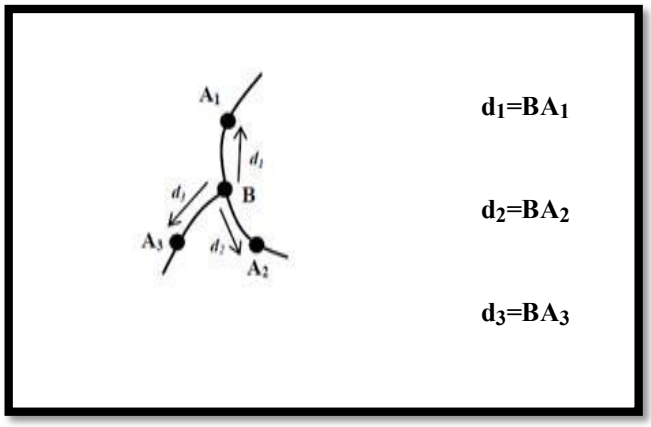

Figure2.24:Définitions associées à une bifurcation lors de la phase de validation

7.2.Approche d'extraction directe à partir de l'i mage en niveau de gris

Pratiquement, toutes les approches d'extraction de minuties existantes se basent sur le processus de binarisation-squelettisation, peu d'approches d'extraction à partir de l'image en niveau de gris ont été proposées :

Une utilisation des réseaux de neurones pour la détection des minuties a été introduite par M.T. Leung et al [46].

Une autre approche basée sur la localisation des maximums locaux par un suivi des lignes le long du flot directionnel des crêtes a été proposée par Maio et Maltoni [46].

Contrairement à l'approche classique basée sur l'extraction des minuties par le processus de binarisation-squelettisation et qui nécessite un traitement à posteriori pour éliminer les fausses minuties.

8.Conclusion

Dans ce chapitre, nous avons décrit les différents prétraitements couramment utilisés dans les systèmes de reconnaissance des empreintes digitales. Le module de prétraitement englobe principalement deux étapes pour l'amélioration de la qualité de l'image : binarisation, et squelettisation. Enfin, nous avons présenté les méthodes d'extraction des caractéristiques.

Dans notre travail, la recherche de méthodes d'extraction des caractéristiques nécessitées une attention particulière pour pouvoir générer des vecteurs caractéristiques qui permettent de discriminer au mieux des images empreintes digitales provenant de personnes différentes. Et faciliter ainsi la tache de classification.

Dans le chapitre suivant, nous présentons donc les méthodes optées de la classification.

*Chapitre **III:***

Méthode support vecteur machine

1.Introduction

Parmi les méthodes à noyaux, inspirées de la théorie statistique de l'apprentissage de Vladimir Vapnik, les Machines à Vecteurs de Support (SVM) constituent la forme la plus connue. SVM est une méthode de classification binaire par apprentissage supervisé, elle fut introduite par Vapnik en 1995. Cette méthode est donc une alternative récente pour la classification.

Elle repose sur l'existence d'un classificateur linéaire dans un espace approprié. Puisque c'est un problème de classification à deux classes, cette méthode fait appel à un jeu de données d'apprentissage pour apprendre les paramètres du modèle. Elle est basée sur l'utilisation de fonctions dites noyau (kernel) qui permettent une séparation optimale des données. Dans la présentation des principes de fonctionnements, nous schématiserons les données par des « points » dans un plan.[20]

2. Apprentissage statistique et SVM

La notion d'apprentissage étant importante, nous allons commencer par effectuer un rappel. L'apprentissage par induction permet d'arriver à des conclusions par l'examen d'exemples particuliers. Il se divise en apprentissage supervisé et non supervisé. Le cas qui concerne les SVM est l'apprentissage supervisé. Les exemples particuliers sont représentés par un ensemble de couples d'entrée/sortie. Le but est d'apprendre une fonction qui correspond aux exemples vus et qui prédit les sorties pour les entrées qui n'ont pas encore été vues. Les entrées peuvent être des descriptions d'objets et les sorties la classe des objets donnés en entrée .[20]

3. Classification et Risque

Effectuer une classification consiste à déterminer une règle de décision capable, à partir d'observations externes, d'assigner un objet à une classe parmi plusieurs. Ceci revient à

construire une approximation de la fonction f de X vers Y à partir d'un ensemble d'apprentissage constitué de couples (xi, yi) ,qu'on suppose suivre une distribution de probabilité $P(x, y)$ inconnue, tels que f classifie correctement des exemples inconnus. Les exemples inconnus sont supposés suivre la même distribution de probabilité $P(x, y)$ que ceux de l'ensemble d'apprentissage. La meilleure fonction est celle obtenue en minimisant le risque théorique : [21]

$$R[f] = \int (f(x) - y)^2 \, dP(x, y) \tag{3.1}$$

Malheureusement, le risque théorique ne peut pas être directement minimisé dans la mesure où la distribution de probabilité sous-jacente $P(x, y)$ est inconnue. Aussi, on va chercher une fonction de décision proche de celle optimale à partir de l'information dont on dispose, c'est à dire l'ensemble d'apprentissage formé de M couples (x_i, y_i) avec

$xi = (x_1 , x_2 ,..., x_d)$: Vecteur des caractéristiques

Et $yi \in \{C_1, C_2 ,...,Cn\}$: sa classe d'appartenance à l'une des n classes possibles

Pour ce faire, on approxime le minimum du risque théorique par le minimum du risque empirique qui s'écrit : [21]

$$R_{emp}[f] = \frac{1}{n}\sum_i (f(x_i) - y_i)^2 \tag{3.2}$$

Il est possible de donner des conditions au classifieur pour qu'asymptotiquement (quand), le risque empirique (3.2) converge vers le risque (3.1). Cependant, si l'ensemble d'apprentissage est fini, la minimisation du risque empirique donné par l'équation 3.2 ne garantit pas un minimum pour le risque réel figure (3.1).

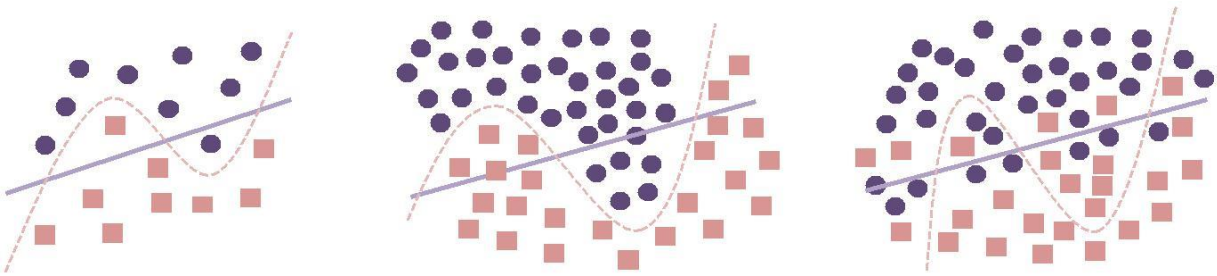

Figure 3.1:Sur-apprentissage et complexité

Problème du sur-apprentissage : Etant donné un petit ensemble d'apprentissage (schéma de gauche), deux frontières de discrimination (représentées par les lignes continues et discontinues) sont possibles. La ligne discontinue est plus complexe mais minimise davantage le risque empirique. Seul un ensemble d'exemples plus grand permet de déterminer la meilleure des deux frontières de décision. S'il s'agit de la ligne discontinue, alors la ligne continue n'est pas suffisamment discriminante (schéma du milieu) ; s'il s'agit de la ligne continue, alors la ligne discontinue ne convient pas et caractérise un sur-apprentissage (schéma de droite).

Ce n'est qu'en 1979, que Vapnik [21] a trouvé une solution à ce problème en mettant au point le principe de la minimisation du risque structurel qui évite le sur-apprentissage des données à la convergence de la procédure d'apprentissage. C'est le principe de base des machines à Vecteurs de support : **SVM**

3.1.Risque structurel

Le risque structurel permet de quantifier la complexité des modèles f, en bornant le risque théorique en ajoutant au risque empirique une borne de confiance (dimension de Vapnik-Chervonenkis ou VC [21]).

$$R[f] \leq R_{emp}[f] + \sqrt{\frac{h\left(\ln\left(\frac{2n}{h}\right)+1\right)+\ln\frac{4}{\delta}}{n}} \tag{3.3}$$

Le paramètre h est appelé VC ou dimension de Vapnik-Chervonenkis. Il mesure la richesse (complexité) de l'ensemble de fonctions solutions ; et il est lié au nombre maximum de points pouvant être correctement séparés quel que soit leur étiquetage. Pour un problème de classification binaire, h est le nombre maximum de points séparables selon 2h configurations par l'ensemble de fonctions solutions. En minimisant la borne VC, on minimise aussi le risque théorique figure (3.2).

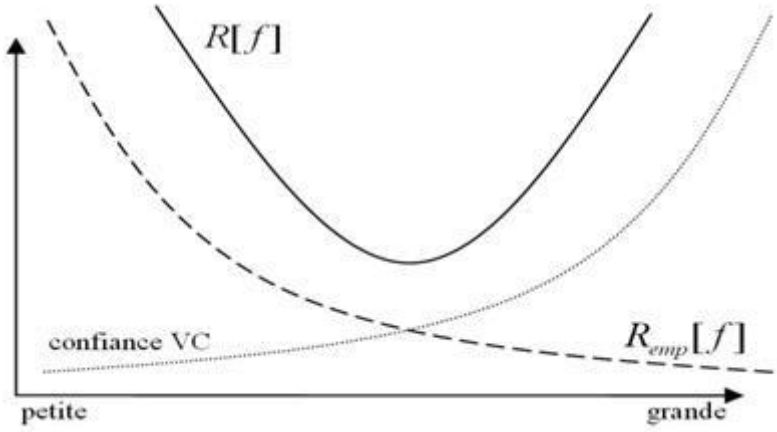

Figure 3.2 : Complexité de la classe des fonctions f

La limite supérieure du risque réel (erreur de généralisation) définie dans l'équation 3.3 constitue un principe essentiel de la théorie des Machines à Vecteurs de Support.

3.2.Classification binaire

La classification binaire (problèmes de deux-classes), l'étiquette y peut prendre seulement deux valeurs distinctes. Y = {−1, +1}. Dans ce cas-ci, le problème de classification consiste à trouver une surface de séparation simple S qui divise l'espace X en deux demi-espaces, chacun d'eux correspondant à une classe[22].

La surface de séparation peut être décrite par une fonction à valeurs réelles h en tant que :

$$S = \{x : h(x) = 0\} \tag{3.4}$$

La fonction de décision f correspondante, classe alors un exemple X selon le côté de S où il se trouve. Mathématiquement cela se traduit par :

$$f(x) = \text{signe}(h(x)) \tag{3.5}$$

3.3.Classification linéaire

Dans le cas de la classification linéaire, la surface de séparation S est un hyperplan défini par[23] :

$$S = \{x : h(x) = w^T x + b = 0\} \tag{3.6}$$

Entrainer un classifieur, avec un ensemble d'apprentissage $\{(xi, yi), i = 1: M\}$ consiste à trouver le modèle f:

$$f(x_i) = \text{signe}(w^T xi + b) = yi,\ i = 1,\ldots, M \tag{3.7}$$

Ce qui est équivalent à :

$$y_i\ (w^T x_i\ + b) > 0,\ i = 1,\ldots, M. \tag{3.8}$$

Si un tel hyperplan existe, alors l'ensemble d'apprentissage est linéairement séparable figure (3.3).

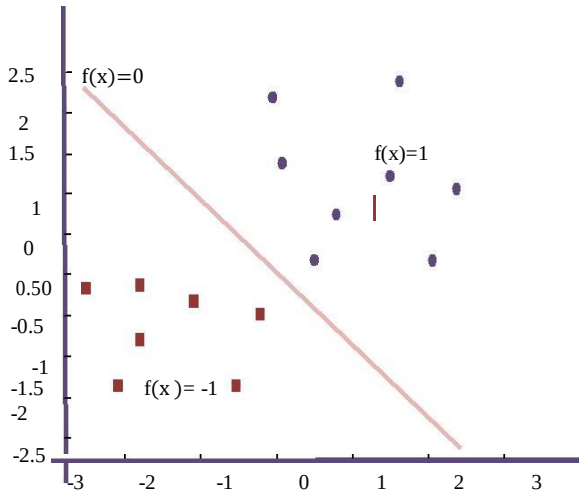

***Figure 3.3** : Hyperplan séparateur entre 2 classes*

Il est évident qu'il existe une multitude d'hyperplans (figure 3.4) qui séparent les deux classes, le problème qui se pose alors, est comment trouver le plan optimal, c'est-à-dire celui qui minimise le risque d'erreur.

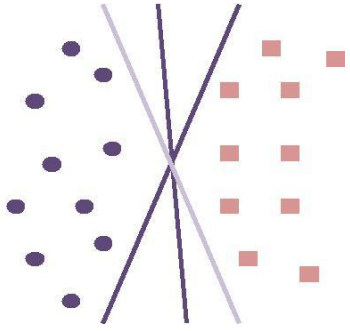

Figure 3.4 : Il peut exister plusieurs hyperplans séparant 2 classes

4. Les SVM

L'hyperplan optimal est celui dont la marge est maximale (figure 3.5). La marge est définie par les points les plus proches de l'hyperplan, ces points sont appelés les vecteurs de support.

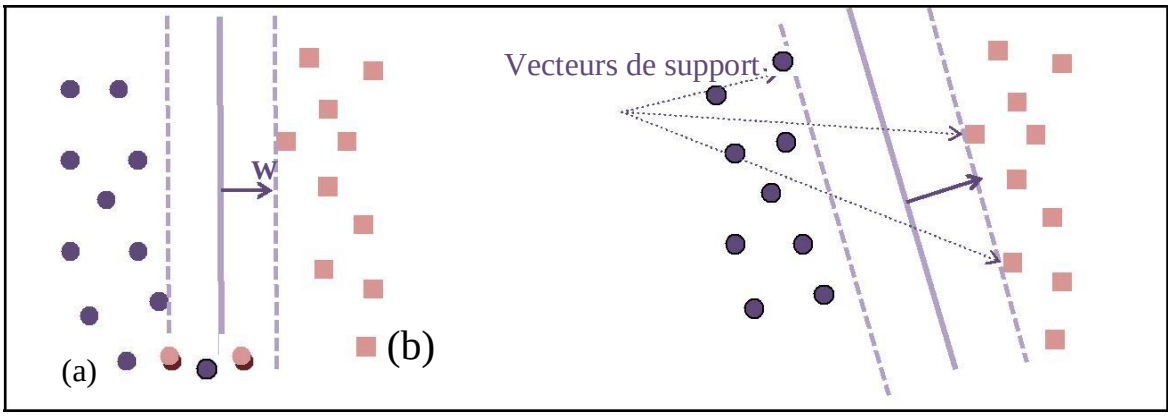

Figure 3.5 : Les SVM trouvent l'hyperplan optimal pour la généralisation (un nouveau vecteur est bien classe dans le cas b.

Normalisée, la marge vaut $2/\|w\|$ avec $|w^T x + b| = 1$ (3 .9)

A noter aussi que la marge et la dimension VC = $\sqrt{\frac{h\left(\ln\frac{2n}{h}+1\right)+\ln\frac{4}{\delta}}{n}}$ sont corrélées c'est-à-dire qu'en augmentant la marge, on réduit *h*, par conséquent on réduit aussi le risque structurel (3.3)

4.1 .Formulation de SVM : cas liniaires

La formulation d'une SVM pour le cas linéaire se fait comme suit :

Maximiser la marge $2/\|w\|$, et bonne classification, ce qui est équivalent au problème d'optimisation sous contraintes suivant [23]:

$$\begin{cases} \min\limits_{w,b} \ \frac{1}{2}\|w\|^2 \\ \text{Sous contraintes : } \forall i, \quad y_i(w^T x_i + b) \geq 1 \end{cases} \qquad (3.10)$$

Le problème peut être résolu en utilisant les multiplicateurs de Lagrange α, et en introduisant les contraintes dans la fonction objectif, le problème primal devient :

$$\begin{cases} L(w,\alpha,b) = \frac{1}{2}\|W\|^2 - \sum_{i=1}^{M} \alpha_i \left[y_i \left(w^T x_i + b \right) - 1 \right] \\ \forall i \ \alpha_i > 0 \end{cases} \tag{3.11}$$

M étant le nombre d'exemples d'apprentissage.

On utilise ensuite les conditions Karush Khun Tucker (KKT) [5] pour exprimer les variables primales en fonction des variables duales :

$$\begin{cases} \dfrac{\partial L(w,\alpha,b)}{\partial w} = w - \sum_{i=1}^{M} \alpha_i \ y\, x_i = 0 \Rightarrow w = \sum_{i=1}^{M} \alpha_i y_i x_i \\ \dfrac{\partial L(w,\alpha,b)}{\partial b} = \sum_{i=1}^{M} \alpha_i y_i = 0 \end{cases} \tag{3.12}$$

En substituant (3.12) dans (3.11), on peut reformuler le problème dual comme suit:

$$\max L(\alpha) = \sum_{i=1}^{M} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{M} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \tag{3.13}$$

$$\text{Sous contraintes} \begin{cases} \sum_{i=1}^{n} \alpha_i y_i = 0 \\ \forall i, \quad \alpha_i \geq 0 \end{cases}$$

$\langle x_i , x_j \rangle$ représente le produit scalaire de deux vecteurs x_i et x_j

La fonction de décision est exprimée alors par l'expression suivante :

$$f(x,\alpha_i,b) = \sum_{i \in SV} \alpha_i y_i \langle x, x_i \rangle + b \tag{3.14}$$

A partir d'un ensemble d'apprentissage $\{(x_i, y_i)\}$, l'apprentissage d'une MVS consiste à déterminer les paramètres de la fonction f (*les* α_i *et* b) les meilleurs au sens de la maximisation de la marge. Les α_i sont non nuls pour un sous ensemble des points d'apprentissage que l'on appelle les vecteurs supports (VS) et sont nuls pour les autres. Par conséquent, un avantage de cet algorithme est son espacement réduit puisque seulement un petit ensemble des exemples d'apprentissage sont finalement retenus pour le classifieur.

4.2. Problèmes non linéairement séparables

Dans la réalité, la plupart des tâches de reconnaissance de formes sont complexes, comme par exemple le cas des arythmies cardiaques. Par conséquent le besoin de classifieurs non linéaires capables de mettre en application des surfaces de séparation plus complexes est nécessaire.

Par ailleurs, même si l'algorithme des SVM sous sa forme initiale revient à chercher une frontière de décision linéaire entre deux classes, ce modèle peut être facilement étendu au cas non linéaire. Le principe consiste à projeter les données initiales dans un espace des caractéristiques (feature space), éventuellement de plus grande dimension que l'espace des entrées, afin de rendre linéairement séparable le jeu de données, et de construire ensuite un hyperplan optimal séparant les deux classes dans cet espace (figure 3.7). L'objectif est de trouver une surface de séparation linéaire dans l'espace de caractéristiques qui corresponde à une surface non-linéaire dans l'espace d'entrée. [24]

Considérons l'application $\varphi : X \rightarrow F$

$x \rightarrow \varphi(x)$

Il suffit alors d'appliquer l'algorithme des SVM linéaire dans et non plus dans ()

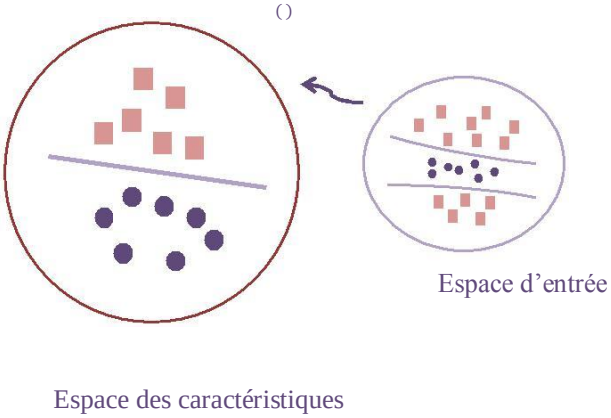

Figure 3.6: Illustration de l'effet du changement d'espace par une fonction noyau.

Les données non linéairement séparables dans l'espace de départ le sont à présent dans l'espace des caractéristiques.

En appliquant la technique de construction de l'hyperplan dans cet espace on obtient la fonction de classification suivante[24] :

$$classe(x) = sign\left[\sum_{i \in SV} \alpha_i y_i \langle \Phi(x), \Phi(x_i) \rangle + b\right] \quad (3.15)$$

On note que pour calculer cette fonction de classification on n'utilise que le produit scalaire dans l'espace de caractéristiques qui peut également être exprimé dans l'espace d'entrée par ce qu'on appelle le noyau K(x,) symétrique vérifiant les conditions de Mercer [24]. Un noyau est intuitivement une mesure de similarité définie entre deux objets d'un même ensemble

On remplace donc le produit scalaire par le noyau *K*

$$K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle \tag{3.16}$$

On obtient la fonction de classification suivante qui n'utilise même pas la transformation Φ :

$$classe(x) = sign\left[\sum_{i \in SV} \alpha_i y_i \langle \Phi(x), \Phi(x_i) \rangle + b\right] \tag{3.15}$$

Parmi les noyaux les plus utilisés on trouve :

Fonction noyau	**type**	**paramètres**
linéaire	$K(x,y) = x.y$	
polynomiale	$(x,y) = ((x.y) + c)^d$	Puissance d
RBF	$K(x,y)=\exp\langle -\frac{\|x-y\|}{2\delta^2} \rangle$	Ecart type δ^2
sigmoide	$K(x,y) = \tanh(a(x.y) - b))$	Les conditions de Mercer ne sont vérifiées que pour certaines valeurs de a et b

Tableau 3.1:Les types de noyaux

Les valeurs des paramètres de noyaux affectent directement la complexité de la frontière de décision du classifieur. Aussi, ses valeurs influencent le nombre de vecteurs de support du classifieur. On peut trouver une description plus détaillée sur les noyaux dans les références [21].

4.3.Les SVM Multiclasses

Le SVM est un classifieur binaire, qui ne traite habituellement que des données étiquetées par rapport à deux classes d'appartenance {-1,1}Ce principe ne s'appliquent pas, du moins pas directement, à des ensembles de fonctions de X dans {1,2,…,N},N > 2, où N désigne le nombre de classes.

Pour la plupart des modèles d'apprentissage, l'extension au cas multi classes est facile et semble parfois même naturelle. Les réseaux de neurones, les modèles linéaires généralisés et

les arbres de décision en sont des exemples. En revanche, la supériorité des SVM ne cache pas la grande difficulté de leur adaptation aux problèmes de discrimination multi classes. Depuis la première extension proposée par Vapnik [21], plusieurs chercheurs se sont attachés à utiliser les SVM dans des applications à plusieurs classes : Les M-SVM. Les approches utilisées jusqu'à nos jours sont diverses et elles peuvent être réparties en deux catégories.

- La première catégorie de méthodes, qu'on qualifie d'indirectes, consistent à fusionner plusieurs classifieurs binaires, les résultats obtenus par chaque classifieur binaire sont ensuite combinés pour donner le résultat final, c'est le cas pour les approches un-contre-un, un-contre-tous.

- La deuxième catégorie de méthodes, sont désignées par directes, elles essayent de résoudre un problème d'optimisation général ; Il n'est pas prouvé toutefois que celles-ci minimisent le risque structurel [25]. La recherche demeure, cependant, très active sur ce sujet [24].

4.3.1.M_SVM directes

Elles consistent à résoudre le problème multi classes en une seule étape sans le décomposer en une collection de sous-problèmes binaires. Cette méthode revient à résoudre un unique problème d'optimisation quadratique conformément à ce qui est fait lorsqu'il s'agit de deux classes. Cette approche inclut quatre modèles : le modèle de Weston et Watkins (WW) [[26], le modèle de Crammer et Singer (CS) [27,28], le modèle de Lee, Lin et Wahba (LLW) [39] [30], et le modèle de M-SVM^2 de Guermeur [30]. Ces modèles partagent le même principe de base [28,31] qui fait intervenir N fonctions de décisions et consiste à résoudre dans le cas où les données sont non séparables, le problème d'optimisation quadratique suivant, au sein d'une même fonction objectif (qui est une généralisation du cas bi-classes équation 4.15):

$$\min_{w,b,\xi_i^m} \left(\frac{1}{2}\sum_{m=1}^{N} w_m . w_m + C \sum_{i=1}^{d} \sum_{m \neq y_i} \xi_i^m\right) \quad (3.17)$$

Sous les contraintes

$$\begin{cases} w_{y_i} . x_i + b_{y_i} \geq w_m . x_i + b_m + 2 - \xi_i^m \\ \xi_i^m \geq 0 \, , \mathrm{i} = 1, \ldots \mathrm{d}, \mathrm{m} = 1, \ldots \mathrm{N}, \mathrm{m} \neq \mathrm{y_i} \end{cases} \quad (3.18)$$

La résolution de ce problème est bien évidemment plus complexe que dans le cas bi-classes, mais des travaux de recherches sont très actifs dans ce sens [26,30] .

Cette approche devient donc vite couteuse en temps de calcul, particulièrement quand on a un grand volume de données, ou bien quand les données sont non linéairement séparables [21,24], comme c'est le cas en détection d'arythmies cardiaques. Par conséquent, nous avons opté pour l'utilisation d'une des méthodes indirecte. Cependant une étude théorique s'avère nécessaire afin de dégager la méthode indirecte la plus adéquate à notre problème de discrimination.

4.3.2. L'approche indirecte :

Nous présentons ici les méthodes indirecte**s,** appelées aussi méthodes de décomposition. Elles consistent à subdiviser le problème multi classes initial en une collection de sous-problèmes bi-classes. Parmi les méthodes de décomposition les plus utilisées, on peut citer :

L'approche "un contre tous"

L'approche "un contre tous" [24] est la plus simple et la plus ancienne des méthodes de décomposition. Elle consiste à utiliser un classifieur binaire par catégorie. Le $k^{ième}$ classifieur est destiné à distinguer la catégorie d'indice k de toutes les autres (figure 3.6). Durant l'apprentissage, tous les exemples appartenant à la classe considérée sont étiquetés positivement (+1) et tous les exemples n'appartenant pas à la classe sont étiquetés

négativement (-1). A la fin de l'apprentissage, nous disposons de modèles correspondant aux hyper-plans (Wi, bi) tels que (figure 3.7). Pour affecter un exemple, on le présente donc à classifieurs, et la décision s'obtient en appliquant en général le principe

"Winner-Takes-All (WTA)" : l'étiquette retenue est celle associée au classifieur ayant renvoyé la valeur la plus élevée.

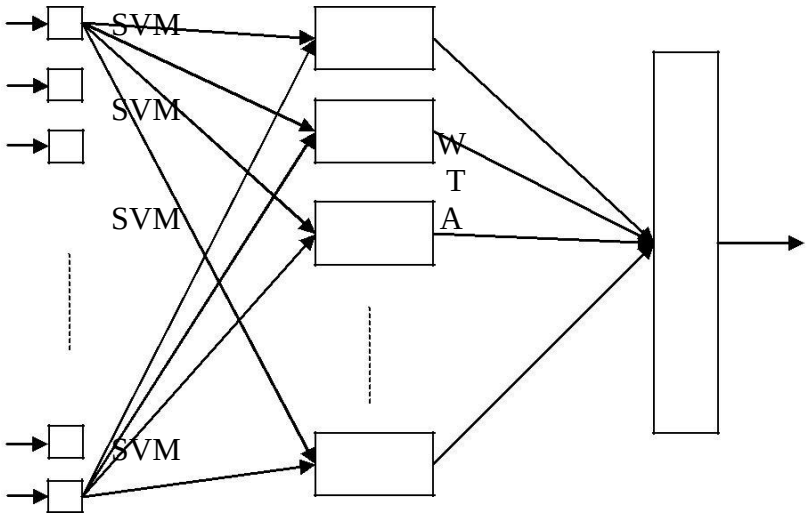

Figure 3.7: Architecture du système en stratégie Un-contre-Tous

Dans [26], les auteurs soutiennent la thèse selon laquelle cette approche, aussi simple soit-elle, lorsqu'elle est mise en ouvre avec des SVM correctement paramétrées, obtient des performances qui ne sont pas significativement inférieures à celles des autres méthodes de décomposition et des SVM multi-classes actuelles. Il convient cependant de souligner qu'elle implique d'effectuer des apprentissages aux répartitions entre catégories très déséquilibrées, ce qui soulève souvent des difficultés pratiques [24], comme par exemple des zones d'ambiguïtés qui correspondent au cas où plusieurs classifieurs binaires renvoient la même fonction de décision (figure 3.9).

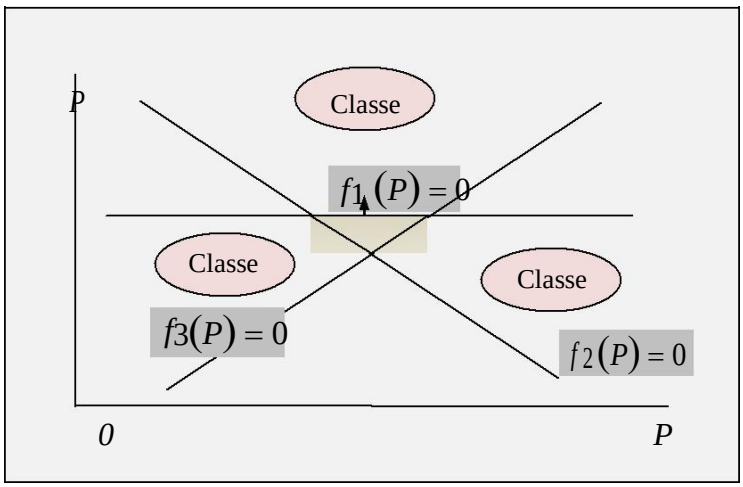

Figure 3.8: Frontières de décision linéaires dans la stratégie Un-contre-Tous pour un problème à trois classes les zones d'ambiguités sont hachurées

Il est à noter aussi qu'il est possible de convertir les sorties brutes des SVM binaires en probabilité à posteriori [32] en utilisant une sigmoïde de la forme :

$$g(x)=1/1+\exp(Ax+B)$$

Le paramètre A permet d'ajuster la pente de la sigmoïde à la marge de la SVM. Notons aussi que ses paramètres sont appris sur un ensemble de validation [33].

L'approche « un contre un »

Une autre méthode de décomposition très naturelle est la méthode "un contre un". Ordinairement attribuée à Knerr et ses co-auteurs [21]. Dans ce cas, chaque classe est discriminée d'une autre ; Par consequent $\frac{N(N-1)}{2}$ fonctions de décision, sont apprises (au lieu de dans le cas de l'approche un contre tous) ; Chacune d'entre elles constitue la réponse d'un classifieur binaire. classificateur Cij dit que x appartient à la classe yi, le vote pour la classe yi est incrémenté de 1. Cette stratégie assigne x à la classe qui a reçu le plus grand nombre de votes (WTA).

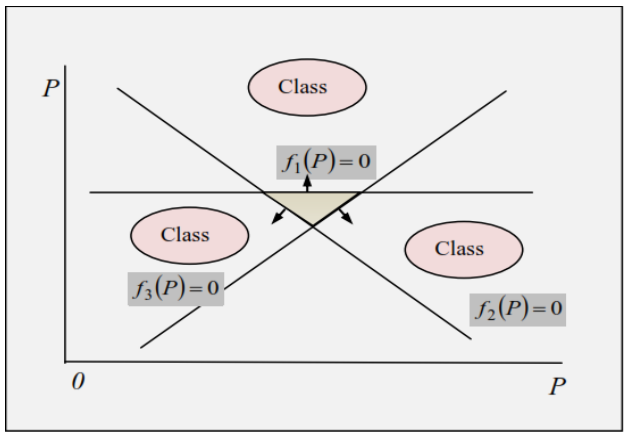

Figure 3.9: Classification de trois classes linéairement séparables par une SVM « un-contre-un » zone d'ambigüité est hachurée.

Même si le modèle SVM « un-contre-un » utilise plus de classifieurs SVM binaires ($\frac{N(N-1)}{2}$) que la méthode « un-contre tous » (N) cependant ce modèle est plus simple et plus rapide à implémenter, et donne en général de meilleurs performances en généralisation. En effet, la tache de chaque classifieur binaire est plus simplifiée puisque il n'a à traiter qu'un sous-

ensemble restreint et équilibré de l'ensemble de données d'apprentissage. On note aussi que la zone d'ambigüité est réduite (figure3.9).

4.4. Architecture du classifieur SVM proposée

Comme nous l'avons vu dans la section 3.4 de ce chapitre, les SVM sont proposées initialement pour traiter des problèmes de classification linéaires binaires. Leur extension aux problèmes non linéaires multi-classes, qui représentent le cas de la majorité des applications réelles est actuellement un domaine de recherches très actif. Plusieurs approches ont été proposées dans ce sens tel que : un contre un, un contre tous, méthodes directes,… ; Cependant, des études [34] comparatives récentes ont prouvé que la méthode « un contre tous» est meilleure du point de vu généralisation et temps d'exécution, par rapport aux autres approches, nous avons utilisé SVMs binaire puisque chaque classifieur binaire n'a à discriminer qu'entre deux classes. Aussi, nous avons choisi d'implémenter ce classifieur pour traiter notre problématique figure 3.10.

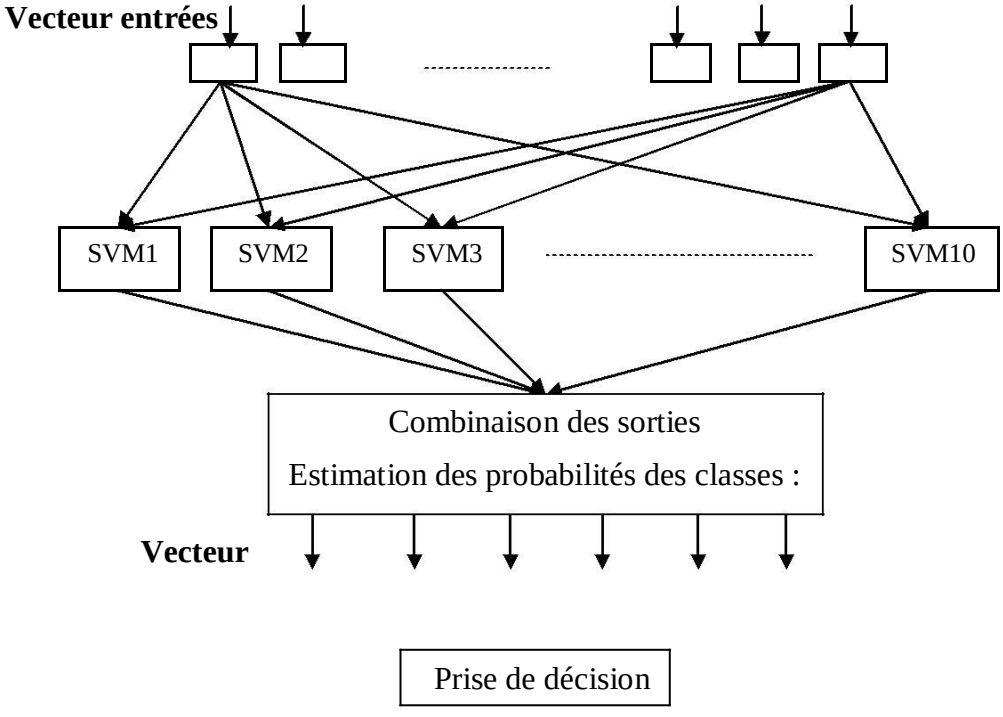

Figure 3.10: Architecture détaillée du classifieur SVM implémenté.

4.5.Avantages et inconvénients de SVM

Avantages

- Les SVM possèdent des fondements mathématiques solides.
- Les exemples de test sont comparés juste avec les supports vecteur et non pas avec tout les exemples d'apprentissage.
- Décision rapide. La classification d'un nouvel exemple consiste à voir le signe de la fonction de décision f(x).[47]

Inconvénients

- Classification binaire d'où la nécessité d'utiliser l'approche un-contre-un.
- Grande quantité d'exemples en entrées implique un calcul matriciel important.
- Temps de calcul élevé lors d'une régularisation des paramètres de la fonction noyau.[47]

5. Conclusion

Dans ce chapitre, nous avons présenté les principales notions de la théorie de l'apprentissage et le principe de minimisation du risque structurel. Nous avons rappelé la théorie de classifieur SVMs en passant par le principe de séparation à vaste marge, Comme nous nous intéressons à l'approche un contre tous (OAA).

Nous concluons enfin, que le SVMs sont suffisamment performants mais théoriquement, le temps d'apprentissage qu'ils nécessitent constitue un inconvénient majeur.

Chapitre IV:

Reconnaissance de L'empreinte Digitale

1.Introduction

Dans ce chapitre ,nous allons parler sur la méthode d'extraction de caractéristique et la phase de classification comme dernier étape dans le système de reconnaissance de l'empreinte digitale. En plus de nous présentons l'interface de notre application accompagné de quelques éclaircissements.

2.Schéma Générale de Notre Système :

On a présenté le système de caractérisation et de classification automatique des individus par la figure suivante :

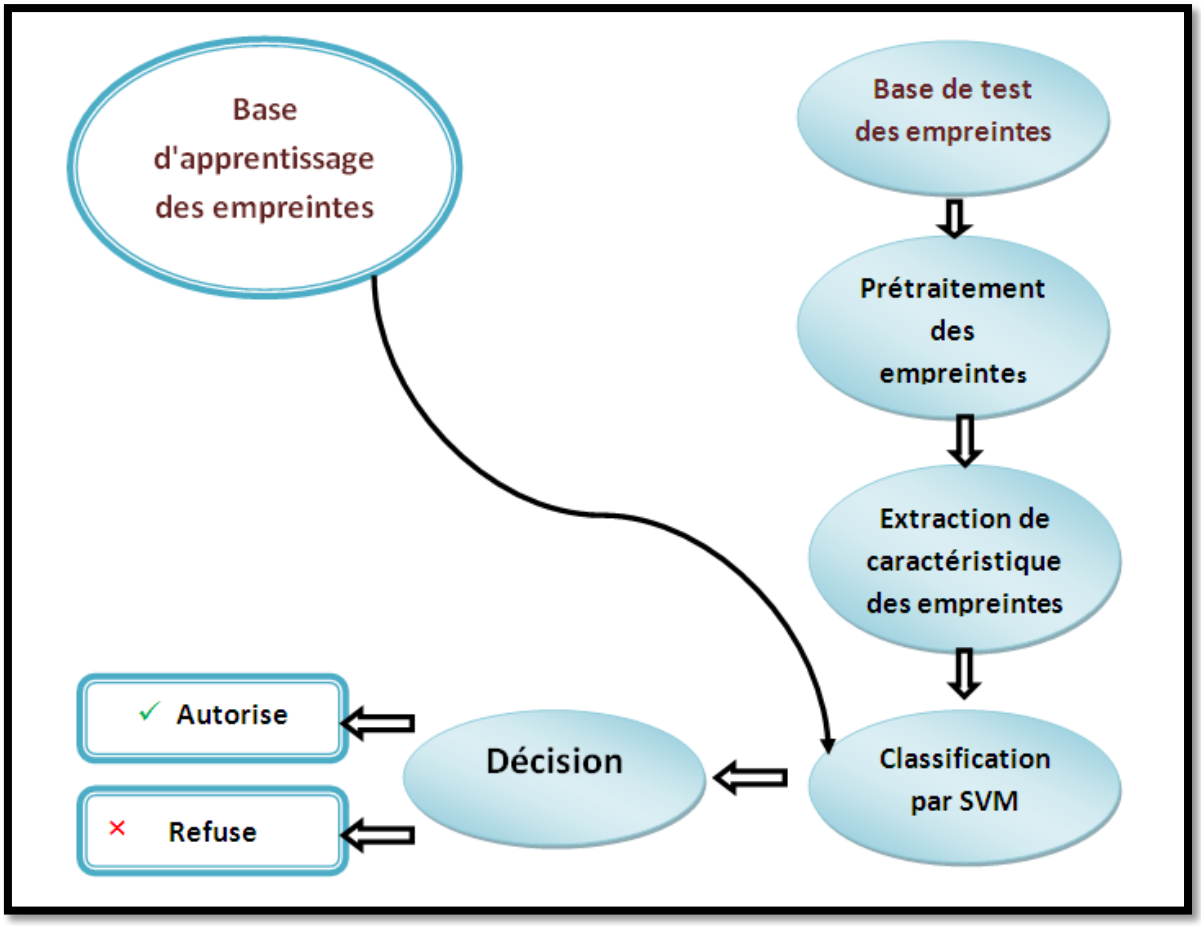

Figure 4.1: schéma de notre système

3. Les Méthodes d'Extraction De Caractéristiques

La génération des caractéristiques est une étape cruciale dans tout système de reconnaissance. Généralement on peut distinguer deux approches:

- Approche statistique.
- Approche géométrique.

3.1. Approche statistique [43]

Les caractéristiques statistiques représentent la densité et la distribution des pixels dans une image. Parmi les caractéristiques statistiques les plus utilisés, on peut citer:

- **Moyenne :** représente le nombre des pixels noirs sur le nombre des pixels de la fenêtre.

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

- **Variance :** mesure donc la dispersion autour de la moyenne.

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2$$

- **Entropie :** mesure la quantité d'information contenue dans le champ des données extraites. Elle est calculée selon l'équation :

$$E = -\sum_i P_i \,.\log(P_i)$$

Où:

est la probabilité d'occurrence des pixels.

Une entropie faible informe sur l'uniformité de la zone alors qu'une entropie forte informe sur l'hétérogénéité.

3.2.Approche géométrique

Plusieurs méthodes ont été proposées pour générer un vecteur caractéristique en tenant compte de la géométrie de la forme [44]. Nous pouvons citer : les concavités, moments géométriques ... [45].

Dans notre cas, nous avons choisi le méthode Moyenne (expliqué dans II .7.1.3.L'extraction des minuties) pour ses propriétés intéressantes.et la facilité d'utilisation dans le compte afin de générer vecteur caractéristique

```
public class ExtractFeatured {
    public ExtractFeatured() {

    }

    public int[][] getFeatured(Pixel[][] pi) {
        int[][] res = new int[pi.length][pi[0].length];
        for (int i = 1; i < pi.length - 1; i++) {
            for (int j = 1; j < pi[0].length - 1; j++) {
                res[i][j] = getFeatured(pi, i, j);
            }
        }
        return res;
    }

    public int getFeatured(Pixel[][] pi, int i, int j) {
        if (pi[i][j].mBlue == 255)
            return 0;
        int numberPoint = 0;
        for (int m = i - 1; m <= i + 1; m++) {
            for (int n = j - 1; n <= j + 1; n++) {
                if (pi[m][n].mBlue == 0)
                    numberPoint++;
            }
        }
        if (numberPoint == 2)
            return 1;
        if (numberPoint ==4)
            return 2;
        return 0;
    }
}
```

Figure 4.2: Code Source en java d'extraction de caractéristiques de l'empreinte digitale

4.Classification

Dans la littérature, on trouve une variété de classifieurs tels que: KPV, réseaux de neurones, et MMC…etc. quelque soit le classifieur choisit, le système de reconnaissance garde qu'il faut faire deux phase importantes:

- Apprentissage, et
- Décision.

Suivant le type de classifieur on note une différence en terme de vecteur caractéristique, temps d'apprentissage, temps d'exécution ou de reconnaissance et par conséquent, une variation dans la certitude de classification [35,36].

4.1.Apprentissage

En intelligence artificielle, l'apprentissage est représenté par deux courants- numérique et symbolique - qui exploitent respectivement et majoritairement des formalismes statistiques et logiques. C'est principalement l'aspect numérique que nous considérerons ici [37]. Les exemples particuliers sont représentés par un ensemble de couples d'entrée/sortie. Le but est d'apprendre une fonction qui correspond aux exemples vus et qui prédit les sorties pour les entrées qui n'ont pas encore été vues. Ce qui nécessite de bien choisir:

- De bons exemples,
- La fonction noyau et les paramètres adéquats,…etc.

L'apprentissage est une phase cruciale car les résultats de décision se basent sur les paramètres fixés durant cette phase [38].

4.2.Décision

Effectuer une classification consiste à déterminer une règle de décision capable, à partir d'observations externes, d'assigner un objet à une classe parmi plusieurs. Le cas le plus simple consiste à discriminer deux classes [39].

5.Méthode de classification choisi

Dans ce travail, nous avons choisis d'utiliser un classifieur SVM dans le but d'avoir un compromis de choix entre: un vecteur caractéristique de taille important, temps d'apprentissage plus au moins réduit, et un temps d'exécution ou de reconnaissance raisonnable, également, une précision de classification est ciblée.

5.1.Classification des empreintes digitale (svm)

Dans ce travail , nous avons utilisé les techniques de classification basées sur les SVM (Support Vector Machine) principalement en raison de leurs fortes capacités en généralisation. Cette méthode a été développée par Vapnik [40]. Le but est alors de classer un objet x a l'aide d'une marge maximale associée a un sous-ensemble de la base

d'apprentissage dont les éléments sont les vecteurs de support et d'une fonction noyau. cette dernière permet d'opérer un changement de repère dans un espace de plus grande dimension afin d'arriver a un problème de séparation linéaire des données, lorsque initialement les données ne sont pas linéairement séparables.

Soit un ensemble d'apprentissage A ={(x_1,y_1),…,(xm,ym)} compose de m couples (vecteur d'attributs, classe) avec xi ∈{-1,1}.L'algorithme des SVM projette les vecteurs xi dans un espace de travail H a partir d'une fonction non linéaire φ:R^n→H. L'hyperplan optimal de séparation des deux classes dans l'espace H est ensuite recherche. Cet hyperplan (w, b) matérialise la frontière de séparation entre les deux classes. La classe y d'un nouvel exemple x est définie par :

$$y = < \mathbf{w}, \phi(\mathbf{x}) > + b = \sum_{i \in SV} \alpha_i^* y_i K(\mathbf{x_i}, \mathbf{x}) + b$$

avec α_i^*∈IR et K(., .)est la fonction noyau .L'hyperplan de séparation est optimal s'il maximise la distance qui le sépare des exemples lui étant le plus proche .Cette distance est appelée marge du classifieur. Les valeurs α_i^*maximisant le critère d'optimalité sont calcules en maximisant

$$-\frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j K(\mathbf{x_i}, \mathbf{x_j}) + \sum_{i=1}^{m} \alpha_i$$

sous les contraintes$\forall_{i=1}^{m}$: $\sum_{i=1}^{m} yi\alpha i = 0, 0 \leq \alpha i \leq C$ avec C le coefficient de pénalisation.

L'algorithme SVM est initialement conçu pour des problèmes de classification à 2 classes.

Dans cet travailler, étant donne que nous avons 5 classes, nous avons utilise l'approche 1contre1 avec le critère de vote majoritaire pour la sélection de la classe finale.

5.2.Classifieur SVM

Un classifieur SVM est chargé de prendre le vecteur de caractéristiques (buffirication , términisation , lac…etc) généré par le module précédent comme une donnée d'entrée, rechercher un hyperplan séparateur qui sépare les exemples dans la phase d'apprentissage et faire une décision de classification dans la phase d'identification. Dans le module SVM, il y a deux phases : une pour l'apprentissage et l'autre pour la classification. La figure suivante représente la relation entre ces deux phases.

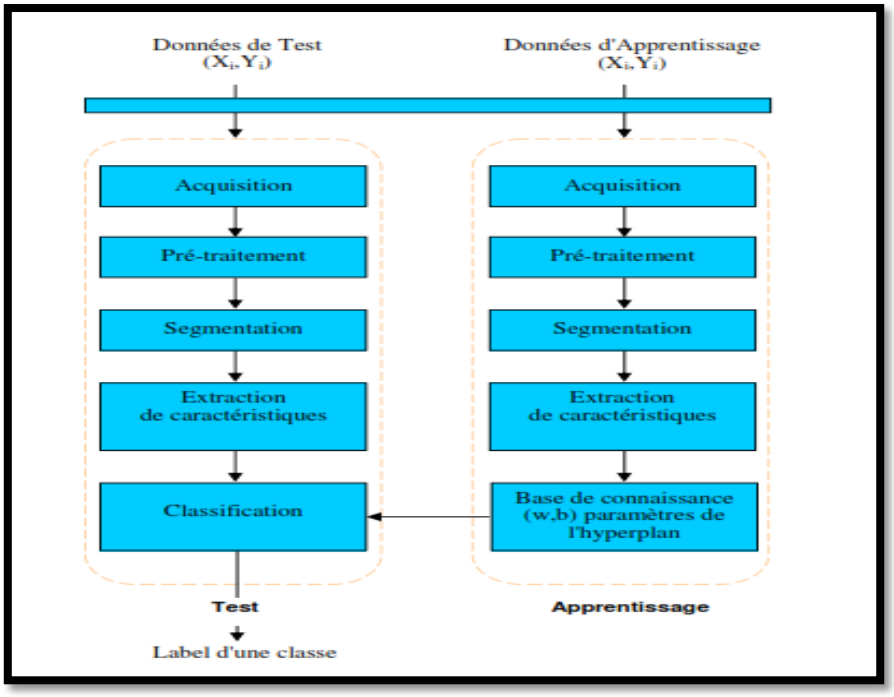

Figure 4.3 : Illustration des deux phases utilisées d'un classifieur SVM.

Où :

Xi : représentent le vecteur caractéristique d'un caractère ;
Yi :l'étiquette d'une classe **.**

5.2.1.phase d'apprentissage

A présent nous devons résoudre le problème dual :

$$\text{Min } \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j K(x_i,x_j) \qquad (1)$$

Sous les contraintes :

$$0 \le \alpha_i \le c \qquad (2)$$

$$\sum_i \alpha_i = 1 \qquad (3)$$

Faire converger ce problème d'optimisation revient à trouver une solution respectant les conditions de Karush-Kuhn-Tucker (KKT). En effet, on sait d'une part que ces conditions sont satisfaites à la solution de tout problème d'optimisation(convexe ou non), et d'autre part que ce sont des conditions nécessaires et suffisantes pour un problème convexe, ce qui est le cas pour les SVM [41].

Pour le problème primal évoqué précédemment, ces conditions sont d'une part :

$$\frac{\partial L}{\partial w} = 0 \Leftrightarrow w = \sum_i \alpha_i \phi(x_i)$$

$$\frac{\partial L}{\partial b} = 0 \Leftrightarrow \sum_i \alpha_i = 1$$

$$\frac{\partial L}{\partial \xi_i} = 0 \Leftrightarrow \alpha_i = c, \beta_i \leq c .$$

Ce qui nous garantissons comme on l'a vu précédemment que l'on a équivalence entre les problèmes primal et dual ; et d'autre part :

$$\alpha_i(\langle w, \phi(x_i)\rangle - b + \xi_i) = 0$$

Avec toujours les contraintes de positivité sur les (α_i,$\beta_i$etε_i) Qu'est ce que cela donne pour nos trois cas de figure évoqués précédemment (sous l'hypothèse que les différentes contraintes sont vérifiées et qu'on n'a par conséquent que la dernière condition à vérifier) ?

Si α_i=0 alors les conditions sont vérifiées.

Si 0<= α_i<=c alors on peut démontrer que l'on a β_i=0 et comme (w,$\emptyset(x_i)$)=b les conditions sont vérifiées.

Si α_i=c alors ε_i= b-(w,$\emptyset(x_i)$)($\neq$ 0) et les conditions sont vérifiées. Que l'on a C'est donc grâce à de la que l'on va définir l'algorithme d'optimisation : on va chercher *b* et des coefficientsα_i($i \in$ [1,l] (où *l* le nombre des exemples)) tels que l'ensemble des points de l'ensemble d'apprentissage vérifie les conditions de KK.[41]

5.2.2.phase de classification

Dans la phase précédente nous résolvons un problème dual. Nous connaissons alors *w et b* et nous pouvons définir la fonction de décision pour une nouvelle donnée *x* :

$$f(x) = signe(\langle w.\phi(x)\rangle - b .$$

5.3.Algorithme générale de SVM

Avant la discussion des détails de l'algorithme SVM, Nous notons que lors de l'implémentation, on est contraint de considérer une précision pour le zéro numérique.

Ceci est de toute façon vrai pour toute implémentation. Ici nous avons considéré que deux données sont égales si elles diffèrent de moins de10-3. Ce seuil proposé par Platt dans la présentation de l'algorithme est raisonnable car il revient dans le cas des SVM à considérer pour une marge théorique de 1, une marge effective comprise entre 0.999 et 1.001. Cette précision nous sert en particulier lorsque l'on recherche les données violant les conditions de KKT et pour contrôler l'optimisation des couples de multiplicateurs de Lagrange. Notons que cette valeur a son influence sur la vitesse de convergence de l'algorithme : plus elle est faible et plus l'algorithme est lent à converger. Cette précision est appelée *zéro_ tolérance* dans nos programmes [41].

5.4.Algorithme de résolution

Entrée:

Q,b,y,Cp,Cn, et an des points initial faisable alpha;

l: la taille des vecteurs et matrices;

eps: critère d'arrêt.

Sortie:

alpha: vecteur contenant la solution;

obj: la valeur objective;

Variables locales:

G,G_bar: vecteurs utiliser pour les calcules du gradient;

alpha_status: vecteur contenant l'état de alpha;

active_size: variable contenant la taille du vecteur active_set;

active_set: vecteur contenant l'ensemble de travail courant;

Début

1. Initialisation des paramètres : taille de l'échantillon (l), paramètre c, largeur de bande du noyau (gaussien), précision numérique eps (zéro_tolérance).
2. Initialiser le vecteur alpha_status (vecteur contenant l'état d'alpha).
3. Initialiser la variable active_size et le vecteur active_set (utilisés durant l'opération de shrinking).
4. Initialiser le gradient.
5. Commencer l'optimisation:
   a. Faire l'opération de shrinking.
   b. Reconstruire tout le gradient.

c. Réinitialiser active_size et choisir un ensemble de travail.
d. Mis à jour d'αi, α

6. Calculer la valeur objective.
7. Affichage des résultats.j, G, G_bar et alpha_status.

Fin.

L'algorithme présenté ci-dessus ne montre que les grandes étapes de résolution des SVM; pour plus d'informations il faut consulter l'ad-doc jointe avec le package libsvm-2.83 [42].

5.5.Classification multi classes

Pour une classification multi-classes nous utilisons une stratégie de vote: chaque classification binaire est considérée à être un vote, où les votes sont considérés comme un cas pour les points des données X. la classe choisi est celle qui a le nombre max de votes [42].

6.Bibliothèque LIBSVM

La bibliothèque LIBSVM est développée dans le but de simplifier l'utilisation des SVM comme un outils, dans ce travail nous utilisant la version 2.83 qui est présentée par le package java libsvm-2.83.

La relation entre une application utilisateur et le package libsvm-2.83 peut être vu selon un modèle de trois couches:

- **Application:** qui représente l'application utilisateur,
- **Interface:** qui permet la communication entre l'application et les modules de calcule, et
- **Calcule:** qui permet de réaliser les calcules nécessaire.

La figure suivante illustre cette relation.

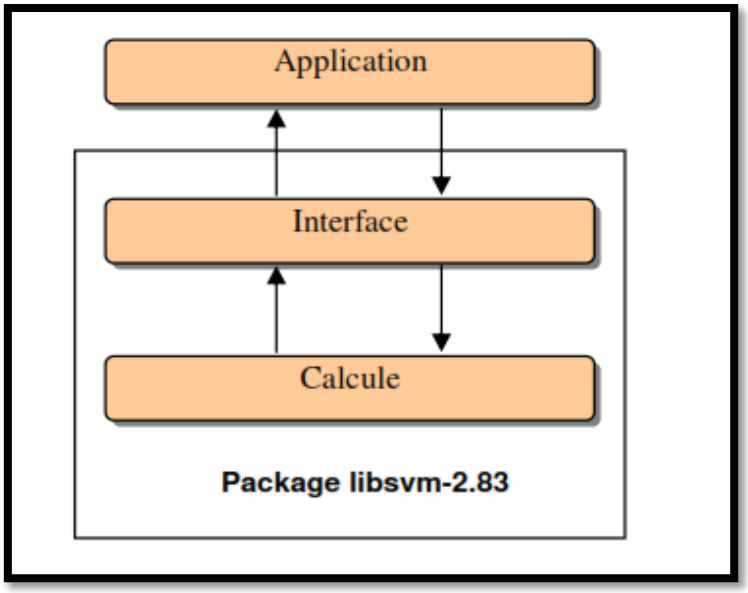

Figure 4.4 : Illustration de la relation entre l'application utilisateur et le package SVM.

La couche interface est présentée par les deux modules:

- svm_train(CarVec.txt) : pour la phase d'apprentissage, et
- svm_predict (CarVec.txt, CarVector.model, classLabels.txt) : pour la phase de test (prédiction).

Où:

CarVec.txt: fichier contenant les caractéristiques extraites.

CarVector.model: fichier contenant les paramètres (w,b) après la phase d'apprentissage.

classLabels.txt: fichier contenant les étiquettes des classes après la phase de test ou décision.

Autant que la couche calcule est présentée par les modules:

- svm: comme module principale,
- svm_node, svm_parameter, et svm_problem: comme des modules Complémentaire

Les modules de la couche calcule sont invisible a l'utilisateur du package. Un utilisateur ne peut se communiquer avec le package qu'à travers les deux modules de la couche interface présentée ci-dessus.

La relation entre l'application et le package libsvm-2.83 ainsi que les différents modules du package est présentée en détail par la figure suivante:

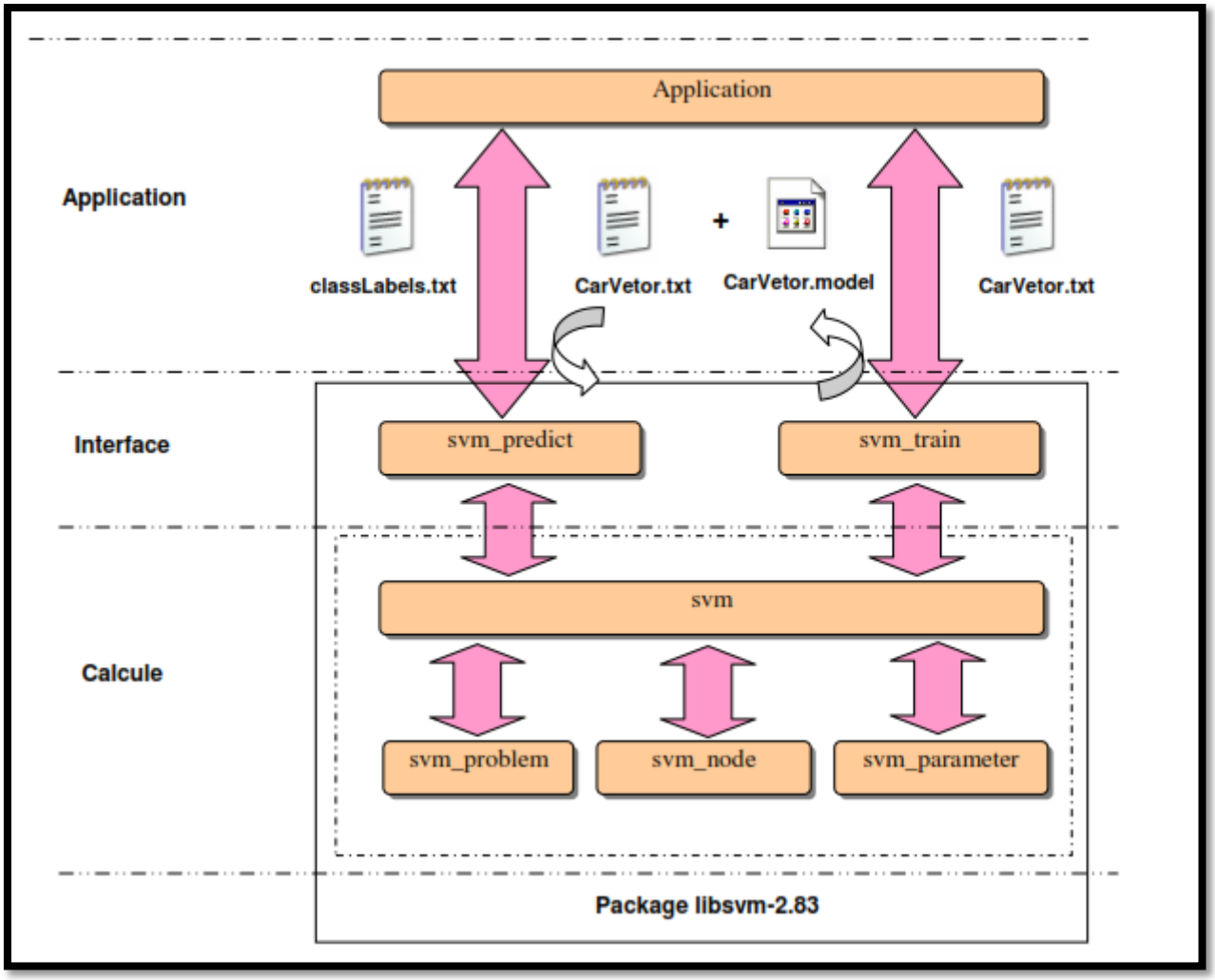

Figure 4.5 : Relation en détail entre l'application utilisateur et le package libsvm-2.83.

Premièrement, dans la phase d'apprentissage la couche application génère un fichier texte (CarVector.txt) contenant les vecteurs caractéristiques (Xi) et les étiquettes des classes (Yi), en utilisant ce fichier comme entrée au module svm_train nous obtenant un fichier de modèle (CarVector.model) qui contient les paramètres (w,b) nécessaire à la phase de test ou prédiction.

En revanche, dans la phase de test on utilise le fichier de modèle généré précédemment plus un fichier texte (CarVector.txt) généré par l'application nous obtenant comme résultat un autre fichier texte (classLabels.txt) qui contient les étiquettes des classes.[42]

7.Résultat et Bilan

Cette section, présentera le choix du langage de programmation, l'interfaces du système.

7.1.Choix de langage de programmation

Dans ce travail, nous avons choisis comme environnement de programmation le langage JAVA qui possède une richesse et offre une grande simplicité de manipulation d'images, soit en acquisition ou en génération des fichiers images.

Ce langage possède des avantages très intéressants tel que :

- La portabilité des logiciels ;
- La réutilisation de certaines classes déjà développées ;
- La possibilité d'ajouter à l'environnement de base des composants fournis par l'environnement lui même ;
- La quasi-totalité de contrôle de Windows (boutons, boites de saisies, listes déroulantes, menus …etc.) qui sont représentés par classes;

7.2.Interface de Notre Système

La forme suivant représente l'interface de notre application

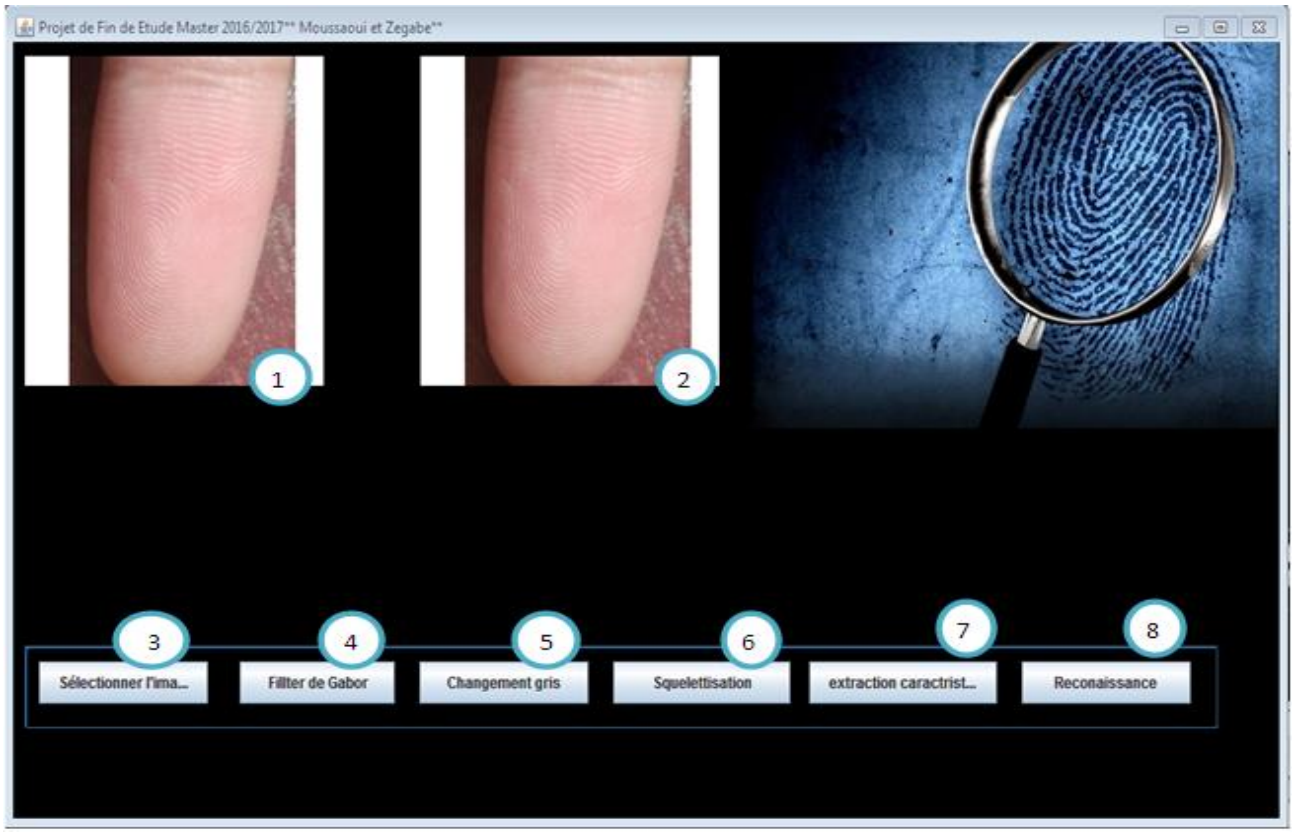

Figure 4.6: interface de notre système

1) Image d'empreinte digitale témoin
2) Image d'empreinte digitale traitée
3) Botton pour sélectionner image d'empreinte a partir le data set
4) Botton pour appliquer le filtre Gabor
5) Botton pour changement l'empreinte en gris
6) Botton pour squeletter l'image d'empreinte
7) Botton pour extraire le caractéristiques de l'empreinte
8) Botton pour identifier l'empreinte trouvé dans la base d'apprentissage ou non

7.3.Résultat de Notre Application:

Après avoir passé toutes les étapes(filtre Gabor, squelettes l'empreinte ,extrait le caractéristiques…), nous obtenons le résultat suivant:

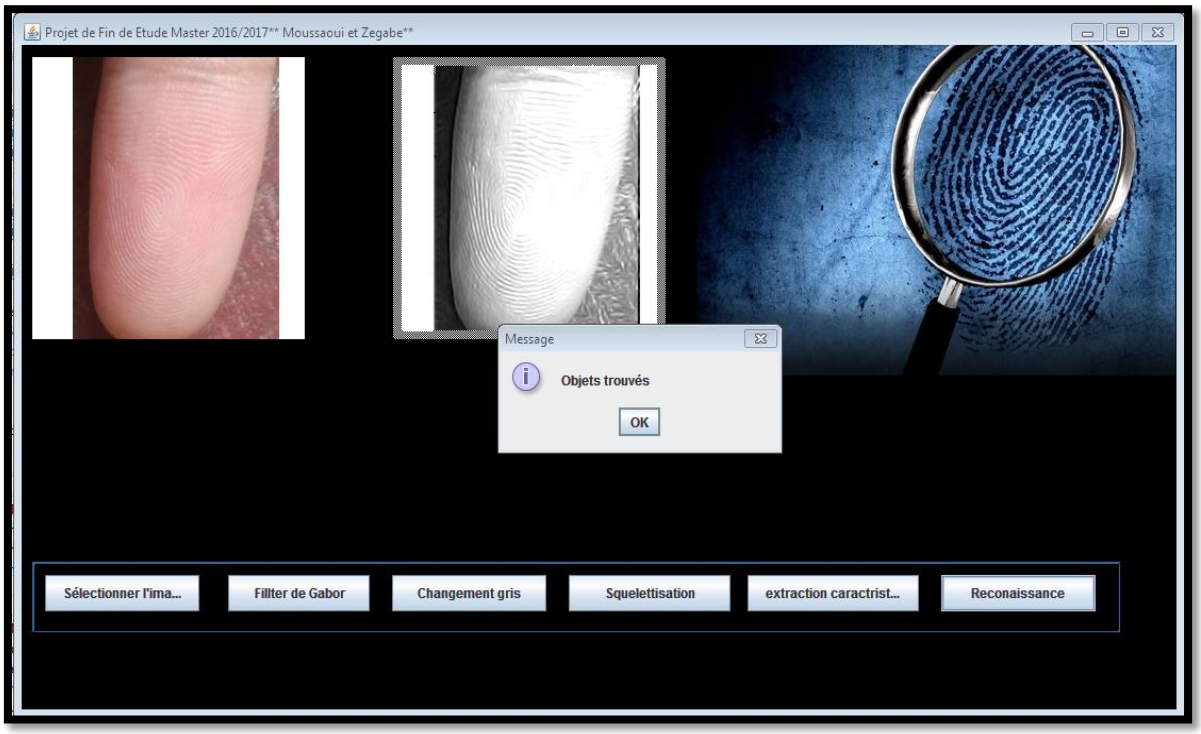

Figure 4.7: Résultat de notre système.

8.Conclusion

Nous avons présenté dans ce chapitre notre application et quelque illustration sur comment faire le extraction et le classification Avec le classifier SVM .

Conclusion générale

L'objectif de ce travail est la mise en œuvre le méthode d'extraction de caractéristiques pour la reconnaissance des empreintes digitales en utilisant les SVM.

Nous avons donc décrit dans ce mémoire les différentes étapes nécessaires à la construction d'un système de reconnaissance des empreintes digitales à savoir : les pré-traitements, l'extraction des caractéristiques et la classification.
Dans le domaine de la reconnaissance des empreintes digitales, les caractéristiques peuvent être décrites comme étant un moyen permettant de distinguer une empreinte d'une classe, d'une autre empreinte d'une autre classe. Dès lors, il est nécessaire de définir des caractéristiques significatives lors du développement d'un système de reconnaissance. Les caractéristiques sont généralement définies par expérience ou par intuition. Plusieurs primitives peuvent être extraites.

Nous avons essayé de proposer une méthode pour extraire les propriétés pertinentes. De nombreuses études ont été proposées dans la littérature. Notre objectif est focalisée sur les caractéristiques statistiques (moyenne), qui représente la densité et la distribution des pixels dans image. L'évaluation de ces caractéristiques a été faite par le classifieur SVM.

Au cours de ce mémoire, nous avons rappelé les différentes techniques de reconnaissance des empreintes digitales. Nous avons pu constater que la reconnaissance des empreintes digitales a connu ces dernières années des progrès très importants, permettant désormais de faire face à la variabilité des empreintes digitales entre les individus. Dans ce travail, nous nous sommes focalisés sur la reconnaissance des empreintes digitales par apprentissage.

En l'état actuel, les performances de notre système sont très encourageantes. Cependant une étude plus approfondie pour choisir l'algorithme de reconnaissance le plus performant et le mieux adaptés, une implémentation d'un système en temps réel, le fusion entre classifieurs ayant pour but de minimiser l'erreur(EER),reconnaissance complète sur l'identité des personne(en ajoutent base données) n'ayant pas pu être étudiées faute de temps et de moyens. feront l'objet de travaux futurs associés a ce projet.

Bibliographies

[1]: mémoire master académique (Identification des personnes par les articulations des doigts),Université Kasdi Merbah Ouargla,MOULAY BRAHIM OUSSAMA et ARBAOUI MOHAMED IBRAHIM,Mai 2015.

[2]: Mémoire Présenté Comme Exigence Partielle De La Maîtrise En Science Politique. La Biométrie, Sa Fiabilité Et Ses Impacts Sur La Pratique De La Démocratie Libérale,Université DuQuébec À Montréal,Frédéric Massicotte, Novembre 2007.

[3]: A. Jain, R. Bolle, S. Pankanti, « Biometrics: Personal Identification in Networked Society », Kluwer, New York, 1998.

[4]: Thése (Etude et mise au point d'un procédé biométrique multimodale pour la reconnaissance des individus) ,Université des Sciences et de la Technologie d'Oran Mohamed Boudiaf,BENCHENNANE Ibtissam,2016.

[5]: Mémoire (Empreinte digitale),NOUIRA HEJER et BEN HAJ SASSI AMNA

, 2016/2017.

[6]:W.J. Babler, Embryologic Development of Epidermal Ridges and Their Configurations, Dermatoglyphics:Science in transition. Birth defects, New York, Wiley-Liss, pp. 95-112, 1991.

[7]: http://perso-laris.univ-angers.fr/~cottenceau/ArduinoCottenceau1112.pdf

[8]: https://www.police-scientifique.com/empreintes-digitales/type-de-dessin-et-classification.

[9]: Véronique Messéant ,Patrick Nizou ,Nathalie Villain,"Les empreintes digitales", Master Didactique des Mathématiques ,Université Paris VII/Juin2006.

[10]: https://www.police-scientifique.com/empreintes-digitales/type-de-dessin-et-classification

[11]: A.K. Jain, S. Prabhakar and S. Pankanti, "Twin Test: On Discriminability of Fingerprints", Proc. 3rd International Conference on Audio- and Video-Based Person Authentication,, pp. 211-216, Sweden, June 6-8, 2001.

[12]: Mémoire de fin d'études Pour l'obtention du diplôme de Master en Informatique ,*Option: Génie Logiciel (G.L)*, université Abou Baker blkaid-Tlemcin "Reconnaissance des Empreintes Digitales", Ben Hamed Amina ,Medjadji Omar, 17 juillet 2015.

[13]:N. Yager and A. Amin, "Fingerprint verification based on minutiae features: a review", *Pattern Analysis and Applications*, Vol. 7, No. 1, pp. 94-113, April 2004.

[14]:L.C. Ern and G. Sulong, "Fingerprint Classification Approaches: An Overview", *International Symposium on Signal Processing and its Applications*, Kuala Lumpur, Malaysia, 13-16 August, 2001.

[15]:S. Pankanti, S. Prabhakar and, A.K. Jain, On the Individuality of Fingerprints, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 8, pp. 1010-1025, August 2002.

[16]:D.P. Mital and E.K. Teoh, "An automated Matching Technique for Fingerprint Identification", *Proceedings of 22nd International Conference on Industrial Electronics, Control, and Instrumentation*, Vol.2,pp. 806-911, 1996.

[17]: LE Duc Bao," Authentification des empreintes digitales dans un système BioPKI ", L'Institut de la Francophonie pour l'Informatique, Travail d'intérêt personnel encadré, Hanoi le 20 janvier 2007.

[18]: https://www.scribd.com/document/226023026/Biometrie-d-Empreinte-Digitale

[19]: Fingerprint Recognition Using Minutiae Extraction (Manisha Balkishan,2013).

[20]: Mohamadally Hasan ,Fomani Boris :" SVM machine à vecteurs de support ou séparateur à vaste marge ". BD Web, ISTY3,Versailles St Quentin, France, janvier 2006.

[21]:V. Vapnik. "An overview of statistical learning theory". IEEE Transactions on Neuralb Networks, 10(5), September 1999. V. N. Vapnik. Statistical learning theory. Wiley,New York, (1998)**.**

[22]:B. E. Boser, I. Guyon, and V. Vapnik. "A training algorithm for optimal margin classifiers". In Proc. of the 5th Annual ACM Workshop on Computational Learning Theory (COLT), Pittsburgh, PA, USA, pages 144 -152, 1992.

[23]:C. Cortes and V. Vapnik, "Support-vector network", Machine Learning, vol. 20, pp. 273–297, 1995.

[24]: Guermeur .Y, "SVM Multiclasses, Théorie et Applications. Habilitation à Diriger des Recherches de l'Université Nancy I, 28 novembre 2007.

[25]: Laanaya .H, Martin .A, Aboutajdine.D, Khenchaf.A, "Classification des sédiments marins par fusion de classifieurs binaires SVM". CMM'06- caractérisation du milieu marin, 16 - 19 Octobre 2006.

[26]: Platt J.C, Probabilities for SV Machines. In A.J. Smola, P.L. Bartlett, B. Schlkopf, and D. Schuurmans, editors, Advances in Large Margin Classifieers, chapter 5, pages 61-73. The MIT Press, Cambridge, MA, 2000.

[27]:K. Crammer and Y. Singer, "On the learnability and design of output codes for multiclass problems," Comput. Learing Theory, pp. 35–46, 2000.

[28]: Crammer K. and Singer Y., 2001. On the algorithmic implementation of multiclass kernel-based vector machines. Journal of Machine Learning Research, 2 :265-292.

[29]:W.A. Barrett, "A survey of face recognition algorithms and testing results", Conference Record of the Thirty-First Asilomar Conference on Signals, Systems & Computers, pp.301-305, 1997.

[30]:Y. Guermeur, “Combining discriminant models with new multi-class svms,” PattAnalysis and Applications, vol. 5, no. 2, pp. 168 – 179, 2002.

[31]: N.E. Ayat, M. Cheriet, L. Remaki, and C.Y. Suen. Kmod- a new support vector machine kernel for pattern recognition. application to digit image recognition. In Proceedings of the IEEE International Conference on Document Analysis and Recognition, pages 1215–1219, Seattle,USA, Sept. 2001.

[32]: Tabbone, S., and Wendling, L, "Binary shape normalization using the Radon transform", Discrete Geometry for Computer Imagery, Lecture Notes in Computer Science, Springer Verlag, Vol. 2886, pp. 184-193. 2003.

[33]:P. Milanfar, "A Model of the Effect of Image Motion in the Radon Transform Domain",IEEE Transactions on Image processing, vol. 8, no. 9, September 1999.

[34]:N.Michael, "Artificial Intelligence: A guide to Intelligent Systems",Addison-Wesley394p, 2002.

[35]: M. Eldawy :" A Survey of Classification techniques ". Presentation, May 2006.

[36]: H. Oulhadj, J. Lemoine, E. Petit, H. Wehbi :" Combinaison d'algorithmes pour la reconnaissance des chiffres et des lettres batons dans un environnement multiscripteur d’ecriture courante mixte ". Laboratoire d’Etude et de Recherche en Instrumentation, Signaux et Systèmes , Université Paris XII, France, May 1999.

[37]: P. Gallinari, H. Zaragoza, M. Amini :" Apprentissage et données textuelles ". LIP6, Université Paris 6, 4 Place Jussieu, 75252 Paris cedex 05, France.

[38]: Mohamadally Hasan,Fomani Boris :" SVM machine à vecteurs de support ou séparateur à vaste marge ". BD Web, ISTY3,Versailles St Quentin, France, janvier 2006.

[39]:P. Mahé :" Noyaux pour graphes et Support Vector Machines pour le criblage virtuel de molécules ". Rapport de stage, DEA MVA 2002/2003,Septembre 2003.

[40]: T. Paquet, L. Heutte, Y. Lecourtier :"Problématique de la Reconnaissance de l'Ecriture".ASTI'2001 des Sons, des Images et des Documents à leur Interprétation, France, 2001.

[41] :P. Mahé, L. Ait-Ali :" Projet d'apprentissage statistique SVM pour l'apprentissage non supervisé". DEA MVA, Février 2003.

[42]: Chih-Chung Chang et Chih-Jen Lin :" LIBSVM: a Library for Support Vector Machines ". Technical_Report, Septembre 2006.

[43]: H. O. Nyongesa, S. Al-Khayatt, S. M. Mohamed and M. Mahmoud, " Fast Robust.

[44]: E. Liu, H.Zhao, "Fingerprint segmentation based on an AdaBoost classifier ",Higher Education Press and Springer-Verlag Berlin Heidelberg ,2010.

[45] P.Dargenton, "Contribution à la segmentation et la reconnaissance de l'écriture manuscrite", Thèse de Doctorat, Institut national des sciences appliquées de Lyon, France, p 227, 1994.

[46]:Mémoire de fin d'études Pour l'obtention du diplôme magister en spécialité: informatique ,option: ingénierie des données et connaissances, université d'Oran Ahmed ben Bellah, "caractéristique biométrique pour l'identification" ,Fatima Boukraa 2015/2016.

[47]: http://dspace.univ-tlemcen.dz/bitstream/112/322/11/ChapitreII.pdf