



جامعة العربي التبسي - تبسة
Université Larbi Tébessi - Tébessa

République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la
recherche scientifique

Université Larbi Tébessi - Tébessa

Faculté des Sciences Exactes et des Sciences de la
Nature et de la Vie

Département : Mathématiques et Informatique



كلية العلوم الدقيقة وعلوم الطبيعة و الحياة
FACULTÉ DES SCIENCES EXACTES
ET DES SCIENCES DE LA NATURE ET DE LA VIE

Mémoire de fin d'études

Pour l'obtention du diplôme de MASTER

Domaine : Mathématiques et Informatique

Filière : Informatique

Option : Systèmes d'information

Thème

***Vers des caractéristiques profondes de l'image
Pourquoi ? et quand ?***

Présenté Par :

BOUAZDIA Fayçal

Devant le jury :

| | | | |
|------------------------|-----|--------------------------|--------------|
| Dr. Gattal A. Eldjalil | MCA | Université Larbi Tébessi | Président |
| Mr. Nouioua Tarek | MAA | Université Larbi Tébessi | Examineur |
| Mr. Souahi. M.S | MAA | Université Larbi Tébessi | Encadreur |
| Dr. Meraoumia. A | MCA | Université Larbi Tébessi | Co-Encadreur |

Date de soutenance : 14/09/2020

Résumé

Dans tous les systèmes de reconnaissance de formes, l'étape d'extraction de caractéristiques est l'étape la plus importante étant donné son impact significatif sur la précision du système. Malheureusement, malgré les nombreuses méthodes artisanales de mise en œuvre de cette étape, ils ont atteint leurs limites pour représenter avec précision ces caractéristiques. Récemment, des recherches connexes ont été consacrées à l'extraction de caractéristiques profondes à l'aide de techniques dites d'apprentissage profond. Dans ces méthodes, l'image peut être représentée par des caractéristiques à plusieurs niveaux, dans lesquelles nous pouvons extraire les caractéristiques qui ne peuvent pas être obtenues par des méthodes artisanales. Par conséquent, ces méthodes reposent généralement sur des filtres convolutifs multicouches. Dans la pratique, les méthodes profondes souffrent d'une capacité mémoire limitée et de ralentissements du processeur, ce qui conduit inévitablement à la recherche obligatoire de technologies simples et faciles qui peuvent être exploitées dans la plupart des appareils actuellement disponibles. Dans ce contexte, nous avons proposé dans ce travail une méthode d'extraction de caractéristiques basée sur le principe du deep learning, qui se caractérise par sa simplicité et légèreté. Nous avons évalué les performances de notre méthode à l'aide d'un système biométrique, qui est un système de reconnaissance de formes typique et les résultats expérimentaux ont montré que notre méthode a une grande précision par rapport à de nombreuses méthodes artisanales actuelles et à certaines méthodes basées sur l'apprentissage profond dans le Littérature.

Mots clés : *Reconnaissance de formes, Biométries, Empreinte palmaire, Empreinte de veines de la paume, Extraction de caractéristiques, LBP, CBD, DCTNet.*

Abstract

In all pattern recognition systems, the feature extraction step is the most important step given its significant impact on the accuracy of the system. Unfortunately, despite the many handcrafted methods of implementing this step, they have reached their limits in accurately representing these features. Recently, related research has been devoted to extracting deep features using so-called deep learning techniques. In these methods, the image can be represented by features at several levels, in which we can extract the features that cannot be obtained by hand-crafted methods. Therefore, these methods usually rely on multi-layered convolutional filters. In practice, deep methods suffer from limited memory capacity and processor slowdown, which inevitably leads to the mandatory search for simple and easy technologies that can be exploited in most devices currently available. In this context, we have proposed in this work a feature extraction method based on the principle of deep learning, which is characterized by its simplicity and lightness. We have evaluated the performance of our method using a biometric system, which is a typical pattern recognition system, and the experimental results have shown that our method has high accuracy compared to many current hand-crafted methods and to some methods based on deep learning in Literature.

Index term: *Pattern recognition, Biometrics, Palmprint, Palm-vein, Feature extraction, LBP, CBD, DCTNet.*

ملخص

في جميع أنظمة التعرف على الأنماط، تعد خطوة استخراج الميزات أهم خطوة نظراً لتأثيرها الكبير على دقة النظام. لسوء الحظ، على الرغم من العديد من الأساليب اليدوية لتنفيذ هذه الخطوة، فقد وصلوا إلى حدودهم في تمثيل هذه الميزات بدقة. في الآونة الأخيرة، تم تكريس الأبحاث ذات الصلة لاستخراج الميزات العميقة باستخدام تقنيات التعلم العميق. في هذه الطرق، يمكن تمثيل الصورة بميزات على عدة مستويات، حيث يمكننا استخراج الميزات التي لا يمكن الحصول عليها بالطرق اليدوية. لذلك، تعتمد هذه الطرق عادةً على مرشحات تلافيفية متعددة الطبقات. في الممارسة العملية، تعاني الأساليب العميقة من محدودية سعة الذاكرة وتباطؤ في المعالج، مما يؤدي حتماً إلى البحث الإلزامي عن تقنيات بسيطة وسهلة يمكن استغلالها في معظم الأجهزة المتوفرة حالياً. في هذا السياق، اقترحنا في هذا العمل طريقة استخراج الميزات بناءً على مبدأ التعلم العميق الذي يتميز ببساطته وخفته. لقد قمنا بتقييم أداء طريقتنا باستخدام نظام القياسات الحيوية، وهو نظام نموذجي للتعرف على الأنماط، وأظهرت النتائج التجريبية أن طريقتنا تتميز بدقة عالية مقارنة بالعديد من الأساليب الحرفية الحالية وبعض الأساليب القائمة على التعلم العميق في المؤلفات.

الكلمات المفتاحية: التعرف على الأنماط، البيومتري، بصمة كف اليد، بصمة وريد كف اليد، إستخلاص الميزات، LBP, CBD, DCTNet.

Remercîments

Tout d'abord j'adresse ma gratitude sincère à notre Dieu

Tout puissant de m'avoir tous donné.

*Je remercie mes encadreurs Mr. Mohamed Salah SOUABI
et Dr. Abdallah MERAOUZIA, pour m'avoir dirigé tout
au long de mon travail, avec beaucoup d'efforts, de
patience, d'expérience et de précieux conseils.*

*Ils furent toujours disponibles malgré leurs charges et
malgré les précautions prises contre la pandémie de
Covid-19.*

*J'exprime aussi mes vifs remerciements aux membres de
jury d'avoir accepté d'évaluer ce travail.*

*De plus, j'aimerais remercier tous le staff du département
des Maths et Informatique – Université de Tébessa.*

Merci à tous

BOUAZDIA Fayçal

Dédicace

Je dédie ce modeste travail à :

Mes parents, que Dieu les protège,

Qui m'ont toujours aimé inconditionnellement.

Ma chère femme et à mes petites princesses

✍ Tasnime (Mon âme),

✍ Djana (Ma joie)

✍ et Nouha (ma petite nounours).

Mes chers frères et ma sœur et leurs familles.

*Tous mes amis et tous ceux qui ont contribué un jour
à mon éducation et à ma formation.*

BOUAZDIA Fayçal

TABLE DES MATIERES

| | |
|---|------------|
| <i>Résumé</i> | i |
| <i>Abstract</i> | ii |
| <i>ملخص</i> | iii |
| Remerciements..... | iv |
| Dédicace | v |
| Table des matieres | vi |
| Liste des figures | ix |
| Liste des tableaux | xi |
| Introduction | 1 |
| Chapitre I : Reconnaissance de formes | |
| Introduction | 4 |
| I.1. Aperçu général..... | 4 |
| I.1.1. Présentation générale..... | 4 |
| I.1.2. Besoin et nécessité | 5 |
| I.1.3. Principales difficultés | 6 |
| I.2. Domaines d'applications | 7 |
| I.2.1. Signaux unidimensionnels (temporels) | 7 |
| I.2.2. Signaux bidimensionnels (images)..... | 8 |
| I.2.3. Signaux multidimensionnels (vidéo)..... | 10 |
| I.3. Système de reconnaissance de formes | 11 |
| I.3.1. Espaces de représentation..... | 11 |
| I.3.2. Structure de base | 12 |
| I.4. Evaluation de performance | 18 |
| I.4.1. Taux de reconnaissance..... | 19 |
| I.4.2. Taux d'erreurs | 20 |
| Conclusion | 21 |
| Chapitre II : Discrimination de caractéristiques | |
| Introduction | 22 |
| II.1. Type de caractéristiques | 22 |

| | | |
|---|---|-----------|
| II.1.1. | Caractéristiques de type statistique | 23 |
| II.1.2. | Caractéristiques de type structurel | 24 |
| II.2. | Techniques d'extraction de caractéristiques..... | 25 |
| II.2.1. | Techniques classiques (Hand-crafted) | 25 |
| II.2.2. | Techniques classiques apprises (Learned Hand-crafted) | 26 |
| II.2.3. | Techniques modernes (Deep learning)..... | 26 |
| II.3. | Rôle d'apprentissage..... | 28 |
| II.3.1. | Principes de l'apprentissage | 28 |
| II.3.2. | Schéma de l'apprentissage | 29 |
| II.3.3. | Types d'apprentissage | 29 |
| II.4. | Vers une approche profonde..... | 30 |
| II.4.1. | Influence des caractéristiques sur la classification..... | 30 |
| II.4.2. | Adaptation des caractéristiques..... | 31 |
| II.5. | Reconnaissance biométrique | 32 |
| II.5.1. | Biométrie | 32 |
| II.5.2. | Modalités biométriques | 32 |
| II.5.3. | Fonctionnement d'un système biométrique | 33 |
| | Conclusion..... | 35 |
| Chapitre III : Résultats expérimentaux | | |
| | Introduction | 36 |
| III.1. | Modalité biométrique utilisée | 36 |
| III.1.1. | Biométries de la main..... | 37 |
| III.1.2. | Base d'images | 37 |
| III.2. | Extraction des caractéristiques..... | 38 |
| III.2.1. | Prérequis théoriques | 38 |
| III.2.2. | Méthode proposée | 41 |
| III.3. | Résultats et discussions..... | 46 |
| III.3.1. | Performance des systèmes proposés | 46 |
| III.3.2. | Etude comparative..... | 52 |
| | Conclusion..... | 55 |
| Conclusion générale | | 56 |
| Glossaire..... | | 57 |
| Annexe 1 | | 59 |

| | |
|--|-----------|
| Méthodes classiques d'extraction de caractéristique..... | 59 |
| Bibliographie..... | 64 |

LISTE DES FIGURES

| | |
|--|----|
| Fig.I. 1: Variabilité et distorsion des formes. | 6 |
| Fig.I.2: Signal de la parole..... | 7 |
| Fig.I.3: Signal électro-cardiographique (ECG)..... | 7 |
| Fig.I.4: Diagnostics des pannes | 8 |
| Fig.I.5: Reconnaissance des caractères..... | 8 |
| Fig.I.6 : Images médicales | 9 |
| Fig.I.7 : Contrôle qualité industriel..... | 9 |
| Fig.I.8 : Images satellitaires..... | 10 |
| Fig.I.9 : Télésurveillance | 10 |
| Fig.I.10 : Suivi d'objets | 11 |
| Fig.I.11 : Espaces de représentation pour la reconnaissance des gestes..... | 12 |
| Fig.I.12: Illustration de la réduction d'information. | 12 |
| Fig.I.13 : Schéma général d'un système de reconnaissance de formes..... | 13 |
| Fig.I.14 : Exemples de capteurs pour l'acquisition de données. | 13 |
| Fig.I.15 : réduction du bruit basée sur un filtre médian. | 14 |
| Fig.I.16 : Correction des distorsions géométriques d'une image..... | 15 |
| Fig.I.17 : Augmentation du contraste par la méthode d'égalisation d'histogramme..... | 15 |
| Fig.I.18 : Normalisation de la taille | 15 |
| Fig.I.19 : Extraction de la région d'intérêt | 16 |
| Fig.I.20 : Extraction de caractéristiques | 17 |
| Fig. II.1 : Exemple de la méthode statistique..... | 23 |
| Fig. II.2 : Exemple de la méthode structurelle | 24 |
| Fig. II.3 : Système de reconnaissance de formes basé sur l'apprentissage profond par convolution..... | 27 |
| Fig. II.4 : Schéma de l'apprentissage..... | 29 |
| Fig. II.5 : Schéma fonctionnel d'un système de reconnaissance biométrique..... | 33 |
| Fig.III.1: Images typiques de l'empreinte de la paume dans la base de données CASIA | 38 |
| Fig.III.2 : Schéma fonctionnel de la méthode d'extraction de caractéristiques proposée (DFTNet). Un exemple de structure DFTNet à 2 stages avec 2 et 3 filtres de convolution dans le premier et le deuxième stage, respectivement..... | 41 |

| | |
|--|----|
| Fig.III.3 : Performances du système biométrique en un seul stage. (a) Système biométrique basé sur DCTNet, (b) Système biométrique basé sur DSTNet, (c) Système biométrique basé sur DFT ^{PH} Net, et (d) Système biométrique basé sur DFT ^{MD} Net. | 48 |
| Fig.III.4 : Performances du système biométrique en deux stages. (a) Système biométrique basé sur DCTNet, (b) Système biométrique basé sur DSTNet, (c) Système biométrique basé sur DFT ^{PH} Net, et (d) Système biométrique basé sur DFT ^{MD} Net. | 49 |
| Fig. A.1 : Extraction de caractéristiques utilisant TP-LBP..... | 60 |
| Fig. A.2 : Extraction de caractéristiques utilisant FP-LBP..... | 61 |
| Fig. A.3 : Formation de la matrice de projection et du dictionnaire pour l'encodage des caractéristiques..... | 62 |
| Fig. A.4 : Extraction de caractéristiques utilisant CBD..... | 63 |

LISTE DES TABLEAUX

| | |
|--|----|
| Tableau III.1 : Performances du système biométrique (DFTMDNet) sous les différentes bandes..... | 51 |
| Tableau III.2 : Meilleur système biométrique (DFT ^{MD} Net) pour chaque bande spectrale | 51 |
| Tableau III.3 : sélection des paramètres CBD sous les trois bandes spectrales | 53 |
| Tableau III.4 : Comparaisons de performances de DFT ^{MD} Net avec les méthodes classiques | 53 |
| Tableau III.5 : Comparaisons de performances de DFT ^{MD} Net avec l'état de l'art | 54 |

Introduction Générale

Introduction

Récemment, les machines peuvent remplacer des tâches manuelles par des technologies de vision par ordinateur pour résoudre plusieurs problèmes. Par exemple, dans le secteur industriel, ces technologies peuvent automatiser le processus pour réduire les coûts, éliminer les travaux dangereux, améliorer les conditions de travail et effectuer des opérations qui ne peuvent être contrôlées manuellement. Pour la mise en œuvre de telles technologies, les données à traiter doivent être acquises par plusieurs capteurs. Fondamentalement, les images sont l'aspect essentiel de représentation de données largement utilisée en robotique, télédétection, biométrie et imagerie médicale... Dans toutes ces applications, les algorithmes de traitement de données doivent être embarqués dans leurs propres machines pour accéder aux informations utiles dans les images pour effectuer la tâche souhaitée. En général, ces algorithmes incluent des modules de reconnaissance de formes qui cherchent à extraire des caractéristiques (dites aussi descripteurs) discriminantes dans les images, qui ont un impact significatif sur les performances de la vision par ordinateur (Gudivada, 2020).

Les images étant des phénomènes aléatoires, deux acquisitions indépendantes de la même scène donnent deux descripteurs différents. En effet, pour réduire l'effet de ces différences, la machine doit apprendre toutes les descripteurs possibles afin de reconnaître la scène en toutes circonstances (ex: image bruitée, image déformée, etc.), cette technique est appelée "*apprentissage automatique*" ("*Deep Learning*" en anglais) (Srinidhi, 2019). Dans ce domaine, les chercheurs ont déployé beaucoup d'efforts qui ont finalement abouti au développement de nombreuses techniques utiles. Les performances de toutes ces techniques reposent principalement sur le vecteur de caractéristiques discriminant extraits de l'image, par des méthodes d'extraction de caractéristiques. En effet, le rôle de cette tâche est de donner une représentation réduite et discriminante aux images, afin que la machine puisse distinguer les différents motifs dans un temps raisonnable (Hamouda, 2019).

Généralement, les méthodes d'extraction de caractéristiques sont basées sur trois aspects, à savoir *la texture, les lignes et l'apparence (forme)*. En effet, plusieurs techniques sont utilisées dans ces trois aspects en fonction de l'application envisagée (Philippe Bolon, 2017). La *transformée en cosinus discrète (DCT)*, la *transformée en ondelettes discrète (DWT)* et le *modèle binaire local (LBP)* peuvent extraire efficacement les informations de texture, tandis que les techniques de détection des contours et de seuillage adaptatif peuvent représenter les lignes d'image. Enfin, *l'analyse en composantes principales (PCA)*, *l'analyse en composantes indépendantes (ICA)* et *l'analyse discriminante linéaire (LDA)* donnent un aperçu général de la forme de l'image. Malheureusement, de nombreuses études ont montré que la majorité de ces méthodes, généralement appelées méthodes classiques (*hand-crafted methods*), ont atteint leurs limites, notamment dans les applications traitant un grand nombre d'images. En général, ces méthodes peuvent fournir des vecteurs de caractéristiques avec un taux de corrélation très élevé entre les classes (interclasses) et un faible taux de corrélation au sein de la même classe (intra-classe), ce qui conduit à de graves erreurs de classification. A cet effet, de nombreux efforts ont été faits pour accéder à des techniques capables de pénétrer profondément dans l'image pour rechercher les meilleures caractéristiques représentatives. Ces caractéristiques, qui doivent nécessairement augmenter la corrélation intra-classe et réduire la corrélation interclasse, sont appelées caractéristiques profondes et sont extraites par des méthodes basées sur le *deep learning* (Li Deng, 2019).

Les "*techniques d'apprentissage profond basées sur la convolution*" nous permettent d'analyser des images à plusieurs niveaux pour extraire des caractéristiques profondes qui peuvent augmenter la précision des vecteurs de caractéristiques, ce qui à son tour augmente la précision du système de reconnaissance de formes (Vincent Dumoulin, 2018). Dans ce mémoire, nous examinons l'efficacité des méthodes d'extraction de caractéristiques basées sur le *deep learning*. Pour illustrer davantage ces techniques, une nouvelle méthode basée sur la convolution a été proposée, à savoir *DFT^{MD}Net deep learning*, basée sur la *Transformée de Fourier Discrète (DFT)*. De plus, un système de reconnaissance biométrique, qui est un prototype de système de reconnaissance de formes, a été utilisé. Comme toutes les méthodes d'extraction de caractéristiques basées sur le *deep learning*, la méthode utilisée analyse la texture de la modalité biométrique à plusieurs niveaux, les résultats expérimentaux démontrent, après une comparaison sérieuse, l'efficacité de notre méthode par rapport aux méthodes classiques (avec ou sans apprentissage). De plus, en comparant avec plusieurs travaux sur la base de données multi-spectrale de l'empreinte palmaire de CASIA (CASIA Palmprint Database, s.d.),

notre méthode prouve sa supériorité du point de vue de la précision d'identification. Enfin, il est à noter que les performances de cette méthode peuvent être améliorées en augmentant le nombre de niveaux d'analyse, en particulier dans les machines puissantes.

Le manuscrit est organisé en trois chapitres : Le **premier chapitre** présente un état de l'art sur les systèmes de reconnaissance de formes, ce chapitre comprend également certaines métriques utilisées pour évaluer les performances de ces types de systèmes. Le **deuxième chapitre** est consacré à l'étude des différentes caractéristiques qui peuvent être tirées de l'image pour former un vecteur bien représentatif, ce chapitre présente un exemple de système de reconnaissance de formes qui est un système d'identification biométrique. Le **troisième chapitre** démontre la faisabilité et l'efficacité de notre méthode proposée à travers la description détaillée des fonctionnalités et des principales caractéristiques du système proposé pour l'identification des personnes basée sur les empreintes palmaires multi-spectrales. Enfin, une conclusion générale avec des perspectives possibles, sont données à la fin de ce mémoire.

Chapitre 1

Reconnaissance de formes

Résumé

La reconnaissance de formes vise à identifier des structures spécifiques (formes ou objets) dans des signaux complexes (tels que des images, de la parole et des signaux industriels) et à les classer en un nombre limité de catégories pertinentes pour l'application envisagée, le tout automatiquement, à l'aide d'un ordinateur numérique. Les algorithmes RDF sont appliqués dans de multiples domaines d'activité tels que, par exemple, la vision par ordinateur, la reconnaissance vocale, l'authentification de signature, la reconnaissance de caractères, l'analyse de documents, l'identification de personne. Ce chapitre présente un état de l'art sur les systèmes de reconnaissance de formes. Il comprend également certaines mesures utilisées pour évaluer les performances de ces types de systèmes.

Chapitre I

Reconnaissance de formes

Introduction

Nous appelons le processus de recherche de l'ensemble de formes auquel appartient une forme particulière "*classification*", tandis que le processus de recherche de formes dans la base qui ont les mêmes propriétés que cette forme est appelé "*recherche*". Ces deux processus relèvent de ce que l'on appelle la reconnaissance de formes (RdFs). Ce processus est l'une des composantes du domaine de l'intelligence artificielle qui entre dans le processus de production de robots qui ont la capacité d'apprendre, de raisonner et d'interagir avec le monde extérieur, et donc la capacité d'identifier des objets. Dans ce chapitre, nous nous intéressons plus particulièrement aux systèmes de RdFs, qui constituent souvent un maillon essentiel des systèmes de traitement de l'information. Nous y discuterons également des différentes métriques d'évaluation de la performance de ces systèmes.

I.1. Aperçu général

Aujourd'hui, la diffusion croissante des moyens informatiques a conduit à une généralisation de la technologie au sein de la société, de sorte qu'il devient très rare de trouver un domaine dans lequel l'ordinateur n'interfère pas. Ainsi, pour que cette diffusion soit plus efficace, toutes les sciences concernées cherchent à simplifier les interactions entre l'homme et les machines pour les rendre accessibles à tous types d'utilisateurs et en toutes circonstances.

I.1.1. Présentation générale

En informatique, la RdF (Edwin Hancock, 2020) est la discipline scientifique dont le but est la recherche, l'identification et la classification de formes ou d'objets en un certain nombre de

catégories ou de classes, en fonction de certaines de leurs caractéristiques discriminantes. Selon l'application, ces objets peuvent être des images ou des signaux ou tout type de mesure à classer. Nous ferons référence à ces objets en utilisant le terme générique anglais "*patterns*". Ces caractéristiques sont généralement interprétées dans un contexte décisionnel, à l'aide d'algorithmes d'apprentissage automatique (*machine learning*), pour permettre la classification ou le regroupement (*clustering*) d'objets.

La RdF a une longue histoire, mais avant les années 1960, elle était principalement le résultat de recherches théoriques dans le domaine des statistiques. Rapidement, l'avènement des calculateurs numériques (ordinateurs) a augmenté la demande d'applications pratiques de la reconnaissance de formes, qui à leur tour ont créé de nouvelles exigences pour de nouveaux développements théoriques (Khodaskar & Ladhake, 2014).

I.1.2. Besoin et nécessité

Le domaine de la reconnaissance de formes (RdF) est maintenant devenu l'un des principaux domaines dans lesquels de plus en plus de chercheurs travaillent pour trouver des algorithmes capables de résoudre de nombreux problèmes de calcul que les humains peuvent souvent résoudre de manière intuitive. En effet, avec sa rapidité et sa précision de calcul, la machine dépasse considérablement le cerveau humain. Grâce aux progrès scientifiques et techniques, les chercheurs tentent d'imiter certains travaux humains à l'aide de machines. Sans aller plus loin que l'idée révolutionnaire de remplacer complètement les humains par une machine, il existe encore de nombreux cas particuliers où une machine peut se voir confier un simple travail de reconnaissance effectué par des humains (Chakladar D, 2019). Les moteurs de cette substitution se situent à trois plans principaux: plan efficacité, plan social et économique.

1. Efficacité : Malheureusement, l'être humain est incomparable avec la machine du point de vue de la rapidité et de la précision du calcul. Par exemple, la plupart des gens ne peuvent pas effectuer d'opérations mathématiques simples (addition, soustraction, multiplication et division) avec trois ou quatre chiffres après la virgule, sans parler des exposants, des logarithmes, des fonctions trigonométriques, etc. Même si les opérations sont simples, l'aspect psychique et moral de l'être humain peut affecter la vitesse et la précision du calcul. A l'heure actuelle, personne ne pouvait être trouvé semblable à une machine en termes de vitesse et de précision (Chakladar D, 2019). Concrètement, rien n'encourage ni ne donne même l'espoir que cela puisse se produire.

2. Social : L'être humain est souvent confronté à des tâches monotones, ennuyeuses, répétitives tel que l'ouvrier qui, à longueur de journée, doit placer les mêmes vis dans les mêmes trous, ou dangereuses tel que la réparation d'une centrale nucléaire. Une machine ou un robot peuvent effectuer des tâches similaires.

3. Economique : Dernièrement, les technologies de reconnaissance de formes ont tellement évolué et sont entrées dans presque tous les domaines qu'elles ont obtenu des résultats remarquables pour les humains qu'ils n'auraient pas obtenus sans une machine. Cette évolution peut certainement apporter de grands bénéfices dans le domaine économique en réduisant les coûts de production, par exemple en automatisant le contrôle qualité dans une ligne de production ou le tri sur la ligne d'assemblage.

I.1.3. Principales difficultés

Bien que les humains puissent détecter et identifier des formes dans une scène sans beaucoup de peine, la construction d'un système automatique qui accomplit une telle tâche est un sérieux défi. Ce défi est encore plus grand lorsque les conditions d'acquisition des sujets varient fortement (Belahcen, 2013).

La RdF se doit de résoudre les premiers problèmes liés au codage des formes, à leur paramétrisation et à leur discrimination. Cette tâche est souvent difficile pour plusieurs raisons. D'abord, les formes appartiennent à un monde physique dont la transcription numérique est très complexe à cause de l'absence de capteurs adaptés à toutes les situations. De plus, la nature des formes et leur apparence varient d'un échantillon à l'autre, même au sein d'une même famille ce qui accroît les dimensions de l'espace de représentation et les temps de décision.

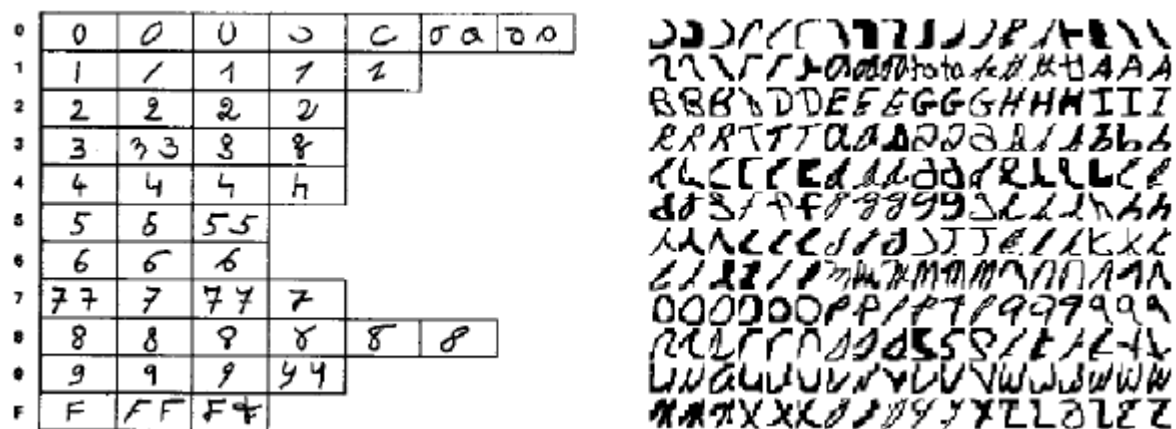


Fig.I. 1: Variabilité et distorsion des formes.

Il est essentiel que l'information transmise à l'ordinateur soit la plus fiable vis à vis de la forme à traiter, et ceci n'est pas a priori une chose simple car il s'agit de passer d'un objet dans un monde physique continu à une forme numérique dans un monde discret. Ce passage, même s'il se fait par la voie de capteurs spécialisés, impose le plus souvent des simplifications de la représentation, donc un appauvrissement de l'information et une "mise à plat" des propriétés de la forme. Comme montre la Fig. I.1, la difficulté de la RdF provient de la variabilité des formes et de la distorsion (déformation, bruit) des données

I.2. Domaines d'applications

La RdF répond aux exigences élevées de la vision par ordinateur dans de nombreuses applications. La RdF est susceptible de couvrir tous les domaines où il est nécessaire de connaître le type de forme. En fonction des applications à automatiser, les données brutes peuvent être des signaux unidimensionnels (simples), bidimensionnels ou multidimensionnels (Srinidhi, 2019).

I.2.1. Signaux unidimensionnels (temporels)

Dans ces types d'applications, la forme ou l'ensemble des formes est représenté comme un signal, généralement temporel, dans une dimension (unidimensionnelle).

1. Signaux audio: Deux informations peuvent être extraites du signal vocal, la première concerne la connaissance de «ce que dit» une personne donnée, c'est-à-dire un système de reconnaissance vocale. La seconde est de savoir «qui parle», donc l'identité de la personne qui parle, c'est-à-dire un système d'identification du locuteur. Le premier peut être utilisé pour la bureautique et la commande vocale, tandis que la reconnaissance vocale est destinée au commerce et à la banque. La Fig. I.2 donne un exemple de signal vocal.



Fig.I.2: Signal de la parole.

2. Signaux biomédicaux : Dans le domaine biomédical, il existe de nombreux dispositifs d'acquisition de signaux (pour le diagnostic) dont la sortie est sous une dimension. Les signaux ECG et EEG sont parmi les signaux les plus couramment utilisés pour RdF afin d'analyser, d'interpréter et de connaître l'état de santé du patient. Les résultats prouvent que le système RdF simule les capacités des médecins à inférer de nombreuses maladies. La Fig. I.3 donne un exemple de signal ECG.



Fig.I.3: Signal électro-cardiographique (ECG).

3. Signaux industriels : Dans le domaine industriel, il existe de nombreux capteurs qui émettent en permanence des signaux afin de détecter automatiquement la panne et de diagnostiquer sa source. Le système RdF analyse tous les signaux résultants et les compare à l'état normal pour diagnostiquer les défauts, puis prendre une décision telle que l'arrêt direct de l'opération pour la protection et la sécurité. La Fig. I.4 donne un exemple de signal de diagnostic industriel.

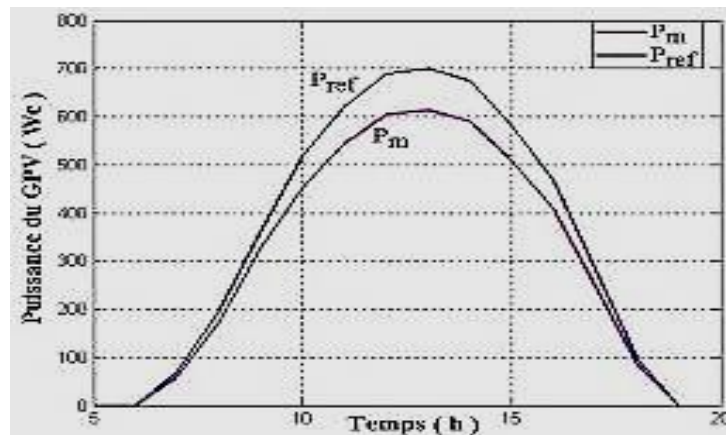


Fig.I.4: Diagnostics des pannes

I.2.2. Signaux bidimensionnels (images)

Dans ces types d'applications, la forme ou l'ensemble des formes est représenté sous la forme d'une ou plusieurs images, en deux dimensions (bidimensionnelles).

1. Reconnaissance de l'écriture: Les systèmes de reconnaissance de caractères ont été l'une des premières applications à utiliser la RdF. En effet, dans ces systèmes, il est possible d'accomplir une tâche de reconnaissance de l'écriture manuscrite, afin de convertir du texte manuscrit tapé sur tout type de support en texte numérique (éditable par un traitement de texte), ou de déterminer l'identité de l'écrivain, dans ce cas le système devient un système biométrique qui utilise des modalités comportementales. La Fig. I.5 montre un exemple d'image d'une écriture manuscrite

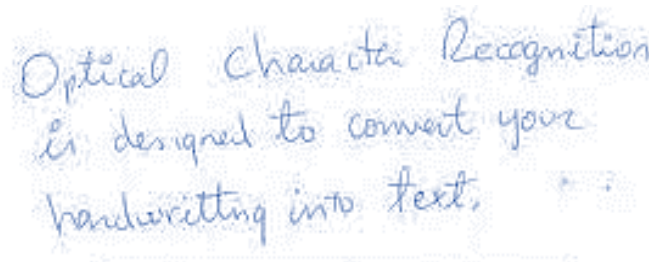


Fig.I.5: Reconnaissance des caractères

2. Imagerie médicales : Du diagnostic à la chirurgie assistée par ordinateur en passant par la simulation, les techniques d'imagerie sont une aide constante pour le médecin, par exemple dans la détection des cellules cancéreuses du cerveau. Lors de l'étape de segmentation, nous recherchons les images pour différentes régions du cerveau: matière blanche, matière grise, liquide cébrospinal, nous cherchons ensuite à détecter d'éventuelles tumeurs bénignes ou malignes à partir de caractéristiques extraites de régions de l'image. Evidemment, il ne s'agit que de fournir une aide au diagnostic validé par la suite par le médecin car le coût associé à un mauvais diagnostic est important. La Fig. I.6 montre un exemple d'images IRM.

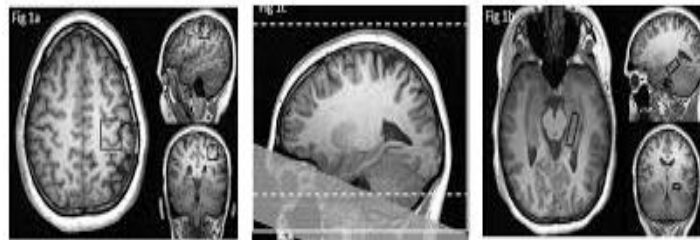


Fig.I.6 : Images médicales

3. Contrôle de qualité industriel : Dans les usines cherchant à contrôler la qualité de leurs produits, il existe des systèmes basés sur la vision par ordinateur pour surveiller les processus de production. Ces systèmes sont équipés d'un RdF pour traiter et analyser les images capturées à partir de la ligne de production afin de vérifier l'exactitude du processus de fabrication. Habituellement, ces systèmes fonctionnent en temps réel. La Fig. I.7 montre un exemple d'image pour le contrôle de la qualité industrielle.



Fig.I.7 : Contrôle qualité industriel

4. Imagerie satellitaires : Parmi les domaines les plus populaires de la RdF, on trouve l'interprétation d'images aériennes ou satellitaires conduisant à la surveillance militaire ou aux prévisions agricoles, ressources foncières et prévisions météorologiques. La Fig. I.8 montre un exemple d'une image aérienne.



Fig.I.8 : Images satellitaires

I.2.3. Signaux multidimensionnels (vidéo)

Dans ces types d'applications, la forme ou l'ensemble de formes est représenté comme une séquence d'images, ou vidéo, en trois dimensions (multidimensionnelles).

1. Télésurveillance : La télésurveillance (vidéosurveillance) est un sujet qui a connu des progrès considérables au cours de la dernière décennie, en particulier avec les développements technologiques tels que les caméras à capteurs de profondeur. Le RdF est utilisé ici pour la reconnaissance automatique de la scène qui regroupe toutes les techniques visant à identifier et à traduire les mouvements pour détecter les différentes anomalies de la scène. La Fig. I.9 montre un exemple de télésurveillance.



Fig.I.9 : Télésurveillance

2. Suivi d'objets : La détection d'objets en mouvement dans les flux vidéo est une étape essentielle pour de nombreux algorithmes de vision par ordinateur. Cette tâche devient plus complexe lorsque la caméra utilisée est en mouvement. Le type de caméra utilisé pour détecter les objets en mouvement dépend du domaine d'application et est soit statique, soit en mouvement. Une séquence vidéo capturée par une caméra statique a des caractéristiques différentes de celle capturée par une caméra mobile. En outre, l'apparence, la forme et la position des éléments en mouvement varient en fonction de leurs mouvements, occultations et poses. Un système efficace basé sur la RdF peut utiliser cette vidéo pour faire suivre un objet en mouvement comme par exemple dans le domaine militaire dont l'avion de chasse suit une

fusée, dont les deux sont en mouvement. La Fig. I.9 montre un exemple de reconnaissance de marche.

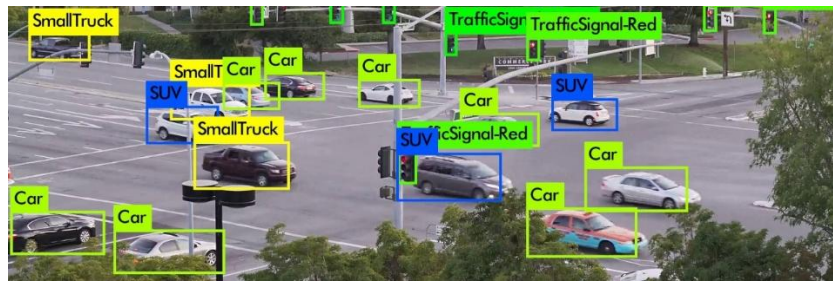


Fig.I.10 : Suivi d'objets

I.3. Système de reconnaissance de formes

Dans un système de reconnaissance de formes, les données brutes initiales passent par plusieurs étapes de traitement jusqu'à ce que la décision soit prise. Chacune de ces étapes a des sorties représentées dans un espace différent de celui dans lequel les entrées sont représentées. Le but de tout cela est d'obtenir les caractéristiques discriminantes de la forme qui sont généralement plus petites que les données d'origine en raison de la suppression d'informations inutiles.

I.3.1. Espaces de représentation

Le but du processus de RdF est la réduction de la quantité de données à manipuler, partant de l'information originale qui fait partie de l'espace d'observation ou monde réel (souvent une image ou un signal) jusqu'à arriver à sa description symbolique, dans l'espace d'interprétation ou espace des catégories, en passant par l'espace de représentation ou espace des caractéristiques où en extrait les primitives pertinentes. A partir de cette description, on déduit que dans le cas général, trois espaces sont impliqués dans un processus de RdF, il s'agit (Alice Ahlem Othmani, 2015):

1. Espace d'observation : également appelé "*espace de données*", il contient les informations d'origine. Un système RdF doit convertir des informations du monde réel en une représentation compatible avec les outils de classification utilisés, par exemple une image numérique. Cet espace contient de nombreuses informations redondantes.

2. Espace caractéristique : également appelé "*espace de représentation*", permet de représenter les caractéristiques. Il contient l'information jugée pertinente.

des formes à partir des données. Lorsque cette phase est terminée, le système sera alors prêt à fonctionner pour reconnaître des formes inconnues. Un système RdF comprend quatre modules (voir Fig. I.13) dont certains sont communs à la phase d'apprentissage et à la phase de classification : acquisition, extraction de caractéristiques, comparaison (mesure de similarité) et décision. L'acquisition et l'extraction des caractéristiques ont lieu pendant l'apprentissage et pendant la classification (Geeks for geeks, 2020). Cependant, la comparaison et la décision sont utilisées pendant la phase de classification.

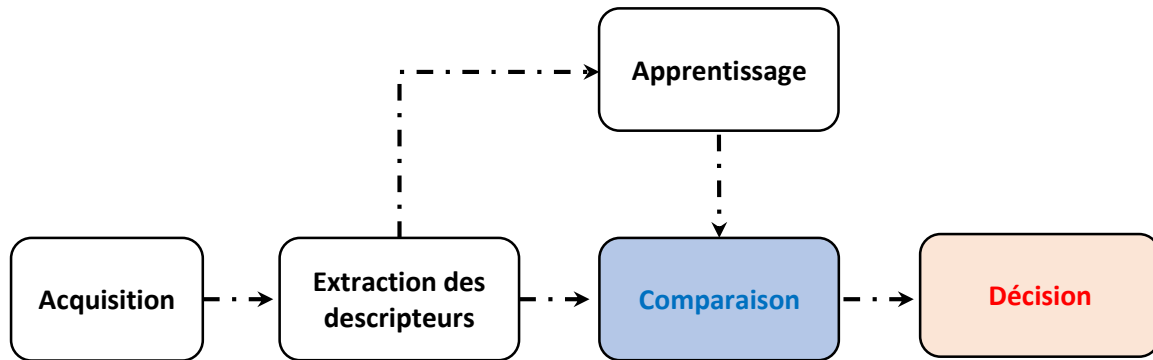


Fig.I.13 : Schéma général d'un système de reconnaissance de formes.

1. Acquisition des données : Dans une application, les données acquises peuvent être de type et de nature différente et dépendent de ce que l'on recherche. Dans la partie de l'acquisition des données, suivant la nature de donnée, un capteur est nécessaire pour acquérir le signal puis faire la conversion en un format numérique pour qu'il soit traitable par un système informatique (voir Fig. I.14). Le but de cette étape est d'obtenir une représentation des formes à traiter qui soit manipulable en machine.



Fig.I.14 : Exemples de capteurs pour l'acquisition de données.

Dans le cas où les données sont disponibles sur un support souvent papier, les capteurs physiques sont des scanners (Fig.I.14.(a)) ou des caméras numériques (Fig.I.14.(b)) (*Dispositif à Transfert de Charges (CCD)*), et dans le cas où la saisie effectuée en ligne le capteur est un écran tactile (Fig.I.14.(c)) ou une tablette graphique munie d'un stylo électronique (Fig.I.14.(d)). Pour capturer le signal vocal un microphone (Fig.I.14.(e)) doit être utilisé, et la numérisation ensuite se fait à l'aide d'un convertisseur analogique-numérique.

Les signaux (ou images) obtenus (données brutes) des capteurs ne pouvant être utilisés directement par les systèmes RdFs (par l'étape d'extraction caractéristique), ils doivent subir un prétraitement dont le rôle est de configurer et modifier les données brutes afin de préparer les pour l'étape d'extraction des caractéristiques. Le traitement du signal et de l'image sont les principales sources des méthodes de prétraitement : filtrage, amélioration du contraste, extraction de contour ou de squelette, segmentation, normalisation, etc. Les techniques de prétraitement sont différentes, dans ce qui suit, nous discutons des techniques les plus utilisées.

- **Réduction du bruit** : Le bruit se manifeste par la présence d'informations résiduelles qui perturbent les données appartenant à la forme. Elle est souvent due aux capteurs (conditions d'acquisition, appareils de mesure). L'atténuation du bruit est un traitement préliminaire souvent difficile à réaliser car pas forcément bien localisé. Dans le cas des images, l'une des méthodes de réduction du bruit les plus courantes est le filtrage médian. La Fig. I.15 montre un exemple de réduction du bruit basée sur un *filtre médian* (Todd Rearick et al, 2019).



Fig.I.15 : réduction du bruit basée sur un filtre médian.

- **Correction des erreurs**: Lors de l'acquisition de la forme, des erreurs peuvent être commises, elles sont de plusieurs types (Loren Peter Lugosch, 2018) :

- **Erreurs dues à l'équipement d'entrée**: telles qu'un mauvais réglage du capteur, sa mauvaise qualité ou son inadaptation à l'acquisition de la forme.
- **Erreurs dues à l'environnement**: telles que bruit ambiant, mauvais éclairage, positionnement des objets.
- **Erreurs dues à l'objet lui-même**: survenant lorsque l'objet n'a pas les mêmes caractéristiques que tous les autres objets de la même famille.

La correction des erreurs est fortement liée aux types d'erreurs et de formes considérées, par exemple, dans le cas d'images, la correction des distorsions géométriques se fait soit en connaissant les fonctions de déformation du dispositif de prise de vue, soit en tenant compte de

la disposition géométrique de quelques points de référence connus dans l'objet. La Fig. I.16 montre un exemple correction des distorsions géométriques d'une image.



Fig.I.16 : Correction des distorsions géométriques d'une image

- **Homogénéisation des données** : Cette opération consiste à débarrasser la forme d'informations redondantes, superflues et inutiles pour l'application (Jose G. Moreno-Torres et al, 2009). L'une des techniques utilisées consiste à augmenter le contraste d'une image ce qui améliore la qualité des données afin de faire ressortir les détails significatifs et d'éliminer le reste. La Fig. I.17 montre un exemple utilisant l'*égalisation d'histogramme* (V. Stimper et al, 2019).

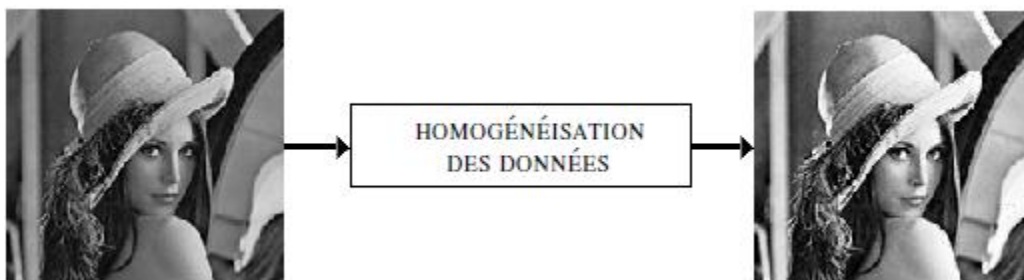


Fig.I.17 : Augmentation du contraste par la méthode d'égalisation d'histogramme

- **Normalisation de la taille**: La taille d'une forme (exemple de visage, voir Fig.I.17) peut varier d'une image à l'autre, ce qui peut causer une instabilité des caractéristique. Une technique de prétraitement consiste à ramener les formes à la même taille.

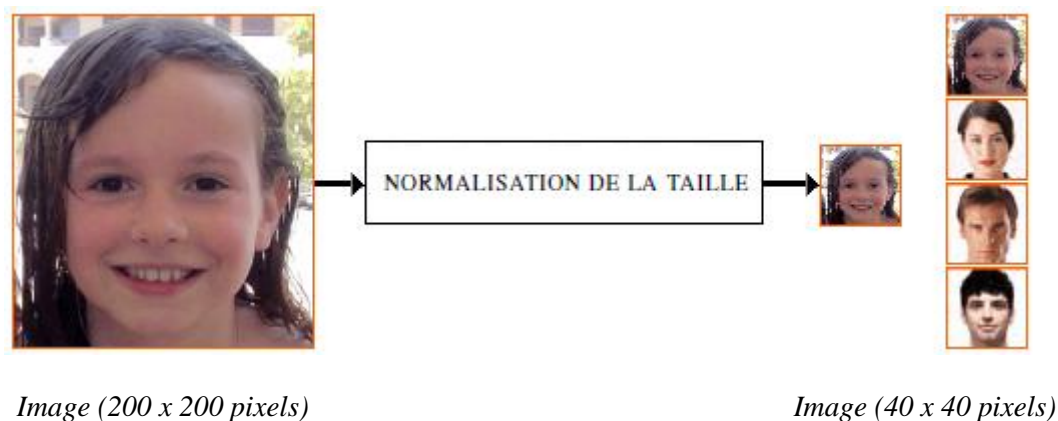


Fig.I.18 : Normalisation de la taille

- **Segmentation:** par exemple l'extraction d'une région d'intérêt (*Region of Interest-ROI*) (Mehmet Umut Demircin et al, 2011). Cependant, étant donné que les images obtenues (données brutes) ne sont pas directement exploitables par les systèmes de reconnaissance, elles doivent subir un prétraitement au cours duquel une ROI est extraite, voir Fig. I.19.

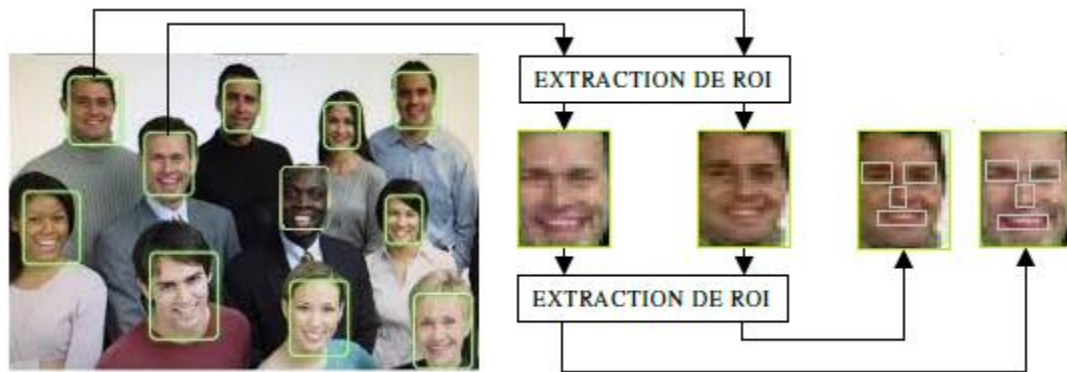


Fig.I.19 : Extraction de la région d'intérêt

Une ROI est une région d'intérêt dans une image et peut être utilisée comme point de départ pour de nombreux algorithmes de traitement. Par conséquent, la qualité de l'algorithme utilisé pour détecter les ROIs conditionne souvent la qualité du résultat de toute la chaîne de traitement que l'on souhaite appliquer à une image. De plus, le fait que les mêmes ROIs (ou plus ou moins) puissent être détectées sur deux images différentes mais représentant la même scène, est une propriété importante et généralement requise pour tous les algorithmes de détection de ROI. Dans le domaine de la RdF, la signification des ROIs dépend de l'application. Elle peut correspondre aux zones des yeux ou aux zones autour de la bouche dans l'image du visage, comme elle peut correspondre à l'iris dans l'image de l'œil.

2. Extraction des caractéristiques : Le processus RdF utilise un critère spécifique pour reconnaître la forme. Pour cela, il est nécessaire d'exploiter et d'interpréter un certain nombre d'indices (ou de caractéristiques) extraits d'une forme inconnue afin de produire une hypothèse sur sa classe. D'une manière générale, l'extraction de caractéristiques transforme le signal d'origine en une description numérique ou symbolique dans un espace abstrait, selon un formalisme prédéfini. Un système RdF peut être considéré comme une boîte noire avec une entrée (la forme) et une sortie (la représentation de la forme).

La phase d'extraction de caractéristiques (Shifei Ding et al, 2011) prend l'image prétraitée et extrait uniquement les informations pertinentes pour former une nouvelle représentation des données (vecteur de caractéristiques). Idéalement, cette nouvelle représentation est censée être

unique pour chaque forme et relativement invariante. La Fig. I.20 montre un exemple d'extraction des caractéristiques

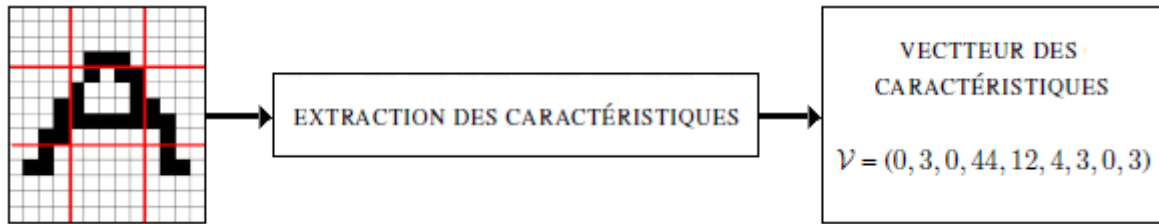


Fig.I.20 : Extraction de caractéristiques

Sur cette figure, il est clair que la taille du vecteur dépend du nombre de blocs, donc un bon choix de ce nombre (généralement empirique) peut conduire à un bon résultat de classification. En pratique, cette tâche n'est pas facile à réaliser avec une précision exacte. Ceci est principalement dû au choix des caractéristiques extraites qui doivent être représentatives de la forme à reconnaître. Ce choix est important car il conditionne toute la méthodologie utilisée pour la classification. Pour cela, le but de cette étape est d'extraire l'information pertinente, qui discrimine les classes, qui est supposée stable pour les formes d'une même classe.

La caractéristique (également appelée descripteur) est une information qui peut être mesurée sur la forme à reconnaître. Cependant, pour un problème spécifique de RdF, de nombreux types de caractéristiques pourraient être utilisés pour représenter et identifier des formes. Quelques exemples sont présentés ci-dessus:

- **Reconnaissance de caractères** : les caractéristiques utilisables peuvent provenir de la densité de points, des moments, du rapport hauteur/largeur, des transformations mathématiques (Fourier, Ondelette, Hadamard), elles peuvent aussi provenir de squelettes ou de contours.
- **Analyse de textures** : comme la télédétection et l'analyse de scène, les caractéristiques utilisables peuvent provenir de la matrice de cooccurrence, des descripteurs de Fourier, du spectre de puissance, des moments.
- **Analyse de formes d'onde** : telles que le signal sismique, l'ECG et la parole, les caractéristiques utilisables peuvent provenir du spectre de puissance et des fonctions d'approximation.

Les caractéristiques peuvent être sujettes au bruit ou aux erreurs, elles peuvent également être continues ou discrètes. Ils sont choisis pour résoudre un problème spécifique et conduisent à une faible erreur de classification. Cela sera possible si les caractéristiques sont

discriminantes : c'est-à-dire si elles varient peu entre les formes d'une classe donnée et diffèrent autant que possible entre les formes de classes distinctes. De plus, les caractéristiques doivent être invariantes en ce qui concerne la rotation, et la translation.

3. Comparaison : Dans cette étape, toutes les caractéristiques extraites sont comparées à celles enregistrées dans la base de données système (vecteurs de référence) afin de déterminer le degré de similitude (obtenu par la mesure de similarité). Le processus de comparaison se fait généralement avec une métrique dont l'utilité est de pouvoir comparer les similitudes et les différences entre deux vecteurs. Cette opération est importante dans le domaine de la classification. Il est plus probable que deux vecteurs similaires appartiennent à la même classe que deux vecteurs différents. En fait, l'utilisation de ces métriques est presque toujours une étape essentielle de la classification automatique. Plusieurs métriques basées sur la probabilité, la corrélation ou la distance peuvent être utilisées (Ricco Rakotomalala, 2017).

4. Décision : La décision (ou classement) qui consiste à associer une description symbolique, par exemple une étiquette, à l'objet en fonction du degré de similitude entre les caractéristiques extraites et celles stockées.

I.4. Evaluation de performance

Bien que les systèmes de reconnaissance de forme promettent d'être très efficaces, nous ne pouvons actuellement pas garantir leur robustesse en pratique dans un contexte d'utilisation spécifique et des utilisateurs cibles. De plus, plusieurs facteurs affectent la fonctionnalité de ces systèmes tels que (Zulfiqar Ali et al, 2017):

- **Le manque de stabilité:** Les systèmes RdF sont moins précis car ils donnent des réponses en termes de pourcentage de similarité (entre 0% → 100%, 100% n'étant presque jamais atteint). Cette variation des résultats peut être due à des conditions d'acquisition de mesures différentes (cas de changements d'éclairage pour un système de reconnaissance faciale) ou à l'utilisation de différents capteurs lors de la phase d'apprentissage et de classification. En raison de cette variation, la plupart des systèmes sont vulnérables. Ce manque de stabilité peut augmenter le taux d'erreur du système.

- **Le manque de précision :** Les vecteurs de caractéristiques extraits des différentes formes peuvent être relativement similaires, par exemple, le cas de jumeaux identiques dans la reconnaissance biométrique de personnes. Ce manque d'unicité peut donc augmenter le taux d'erreur de certains systèmes RdF.

Pour contrôler la fonctionnalité des systèmes de RdF, l'étape de test permet d'évaluer leurs performances pour un apprentissage donné. Il s'agit d'une étape importante car elle peut mettre en cause le choix des caractéristiques ou le choix de la méthode d'apprentissage. En effet, il est difficile de trouver a priori les caractéristiques pertinentes et la méthode d'apprentissage la plus adaptée au problème posé d'où l'utilité de procéder par itérations successives. Ces itérations consistent à extraire des caractéristiques jugées utiles pour résoudre le problème de reconnaissance et à tester les performances du système avec cet ensemble de caractéristiques. Comme les performances souhaitées du système ne sont pas atteintes, il suffit alors de retrouver une nouvelle famille de caractéristiques ou de combiner les caractéristiques extraites avec de nouvelles caractéristiques. Cependant, pour la conception d'un système RdF, les conditions suivantes doivent être respectées (Joel Harp, 2009):

1. L'évaluation est une composante essentielle du RdF.
2. Il existe des mesures permettant de comparer les performances des systèmes entre eux, à condition que :
 - Les données soit les mêmes (d'où l'importance des bases de données publiques) ;
 - Les hypothèses de travail sont les mêmes.
3. L'évaluation ne doit pas porter sur les données d'apprentissage: les données doivent être séparées.
 - Par exemple 50% pour l'apprentissage, 50% pour les tests ;
 - Ou, de manière cyclique, utilisez 80% pour l'apprentissage et évaluez les performances avec les 20% restants.

La réponse d'un système RdF est généralement un score de similitude entre le vecteur de caractéristiques acquis (image test) et le modèle enregistré (modèle de référence) après apprentissage, dans la base de données du système. Pour une entrée donnée, un système peut générer les réponses suivantes:

- **Un rejet:** pour indiquer que le système n'a pas pu identifier cette entrée.
- **Une reconnaissance:** dans ce cas, il identifie correctement l'entrée, et lui attribue sa classe appropriée.
- **Une substitution:** le système attribue une autre classe à l'entrée.

I.4.1. Taux de reconnaissance

La performance d'un système de reconnaissance de formes peut être mesurée en calculant les trois taux suivants (Wu Chou, 2010) :

$$T_C = \frac{N_{DC}}{N_{TD}} \quad (I.1)$$

Où T_C représente le taux de reconnaissance et N_{DC} et N_{TD} représentent respectivement le nombre de décisions correctes et le nombre total de décisions.

$$T_R = \frac{N_{DN}}{N_{TD}} \quad (I.2)$$

Où T_R représente le taux de rejet et N_{DN} le nombre de décisions non univoques.

$$T_E = \frac{N_{DE}}{N_{TD}} \quad (I.3)$$

Où T_E représente le taux d'erreur et N_{DE} le nombre de décisions univoques erronées.

On peut écrire :

$$T_C = 1 - T_R - T_E \quad (I.4)$$

I.4.2. Taux d'erreurs

Deux types d'erreur peuvent être faites dans le cas d'une décision binaire (appartenance ou non à une classe) : il peut rejeter une forme appartient à la classe, il s'agit donc de faux rejets. Il peut aussi accepter une forme n'appartient pas à la classe et il s'agit ici de fausses acceptations. En effet, le taux de fausse acceptation (*False Acceptance Rate-FAR*) et le faux taux de rejet (*False Rejection Rate-FRR*) sont donnés par:

Le FAR détermine la probabilité qu'un système reconnaisse une forme qui n'aurait normalement pas dû être reconnue. Il s'agit d'un rapport entre le nombre de formes qui ont été acceptées alors qu'elles n'auraient pas dû l'être et le nombre total de formes non autorisées qui ont tenté d'être acceptés :

$$FAR = \frac{N_{FA}}{N_I} \quad (I.5)$$

Où N_{FA} représente le nombre de fausses acceptations et N_I le nombre de tests des formes \notin à une classe.

Le FRR détermine le pourcentage de formes censées être reconnues par le système mais qui sont rejetées, le système indique la probabilité qu'une forme connue soit rejetée. Il s'agit d'un rapport entre le nombre de formes qui ont été rejetées alors qu'elles n'auraient pas dû l'être et le nombre total de formes autorisées qui ont tenté d'être acceptés :

$$FRR = \frac{N_{FR}}{N_A} \quad (I.6)$$

Où N_{FR} représente le nombre de faux rejets et N_A le nombre de tests des formes \in à une classe. La décision optimale est une question de compromis entre les FAR et le FRR. En général, les systèmes RdF sont évalués à un point où les taux FAR et FRR sont égaux. Ce point (appelé *Equal Error Rate-EER*) donne le meilleur compromis entre faux rejets et fausses acceptations.

Conclusion

La reconnaissance de formes (RdF) est une compétence commune chez tous les organismes vivants. Cependant, différentes créatures se reconnaissent de différentes manières afin de répondre à leurs besoins quotidiens. Un homme peut reconnaître une autre personne à son visage, à sa voix, à son écriture, etc. Mais un chien reconnaît une personne ou un autre chien à son odorat. Une personne aveugle reconnaît les objets en les touchant. Tous ces exemples sont à classer dans le domaine de la reconnaissance. L'objet inspecté lors du processus de reconnaissance est appelé une forme. Dans ce chapitre, nous avons présenté un état de l'art des systèmes de reconnaissance de formes. Où nous nous sommes concentrés sur l'architecture du système et les applications pouvant contenir ce type de système. Avant de conclure, les métriques utilisées pour évaluer les performances de ces types de systèmes ont été présentées.

Chapitre 2

Discrimination des caractéristiques

Résumé

Les technologies de reconnaissance de formes visent à développer de nouvelles techniques (en particulier celles destinées à la tâche d'extraction de caractéristiques) pour transformer les données d'origine en d'autres représentations qui illustrent honnêtement leur contenu et peuvent ensuite être utilisées pour analyser et comprendre les données originales. Ce chapitre est consacré à l'étude des différentes caractéristiques qui peuvent être tirées de l'image pour former un vecteur bien représentatif. Ce chapitre présente un exemple de système de reconnaissance de formes qui est un système d'identification biométrique.

Chapitre II

Discrimination de caractéristiques

Introduction

Dans tous les systèmes de reconnaissance de formes, une étape importante consiste à extraire les caractéristiques de l'image qui affectent de manière significative la précision du système. Malgré les nombreuses façons de mettre en œuvre des techniques classiques, elles ont atteint leurs limites pour représenter ces caractéristiques avec précision. Récemment, des recherches pertinentes ont été consacrées à l'extraction de caractéristiques profondes à l'aide de méthodes basées sur l'apprentissage profond. Dans ce chapitre, nous discuterons des méthodes d'extraction de caractéristiques distinctes de l'image pour former un vecteur hautement représentatif. Nous nous concentrerons sur la méthode basée sur le *deep learning*. Enfin, nous donnerons un exemple de système de reconnaissance de formes, qui est le système d'identification.

II.1. Type de caractéristiques

Dans les approches de RdF, les caractéristiques de la forme peuvent être gâchées par du bruit ou des erreurs, elles peuvent également être continues ou discrètes. En effet, ils sont choisis pour résoudre un problème précis et conduisent à une faible erreur de classification. Cela sera possible si les caractéristiques sont discriminantes (Brian S. Everitt, 2013): c'est-à-dire si elles varient peu entre les formes d'une classe donnée (faible variabilité intra-classe) et diffèrent autant que possible entre les formes de classes distinctes (forte variabilité interclasse). De plus, les caractéristiques doivent être invariantes en ce qui concerne la rotation, l'échelle et la translation. Il prendra également en compte les déformations les plus probables des formes et leurs répercussions dans la chaîne de traitement.

Principalement, nous distinguons deux approches de RdF, l'une statistique et l'autre structurelle (Georgy Gimel'farb et al, 2012). Il est également possible de les combiner pour

aboutir à une approche hybride combinant leurs avantages. Ces approches correspondent à différentes manières de représenter les caractéristiques des formes et à différentes méthodes de classification. Elles correspondent également à différentes manières d'aborder le problème du RdF. Dans les méthodes statistiques, l'extraction des caractéristiques produit des valeurs numériques qui sont comparées aux modèles statistiques caractérisant chaque classe. Alors que dans les méthodes structurelles (ou syntaxiques), l'extraction des primitives produit des valeurs symboliques et des relations qui sont soumises à une analyse structurelle ou syntaxique.

II.1.1. Caractéristiques de type statistique

Une approche classique en RdF est basée sur l'étude statistique des mesures qui ont été faites sur les formes à reconnaître. Elle repose principalement sur l'utilisation de modèles probabilistes pour les distributions de vecteur de caractéristiques dans les classes pour dériver des fonctions de classification. L'évaluation de ces distributions repose sur un ensemble de formes d'apprentissage dont la classification est généralement connue à l'avance. Si le classificateur est conçu de manière efficace, il fonctionnera bien sur les nouvelles formes.

Dans cette approche, nous supposons que les mesures faites sur la forme peuvent être exprimées par un vecteur $V = [x_1, x_2, x_3, \dots, x_d]$ de l'espace \mathbb{R}^d . Nous avons un ensemble d'apprentissage, c'est-à-dire un ensemble de tels vecteurs dont nous connaissons également la classe à laquelle il appartient. Le problème peut donc se résumer brièvement :

Étant donné un vecteur inconnu, obtenu par des mesures sur une forme, trouver la classe à laquelle on doit l'affecter ? En d'autres termes, reconnaître la forme inconnue en fonction de l'apprentissage effectuée. L'approche statistique (parfois appelée vecteur) consiste à représenter les formes par des vecteurs de caractéristiques puis à construire des classificateurs opérant sur ces vecteurs.

Exemple: Dans une image aérienne, il s'agit de reconnaître deux types de forme (routes et forêt) à partir de la moyenne et de l'écart type des intensités de pixels.

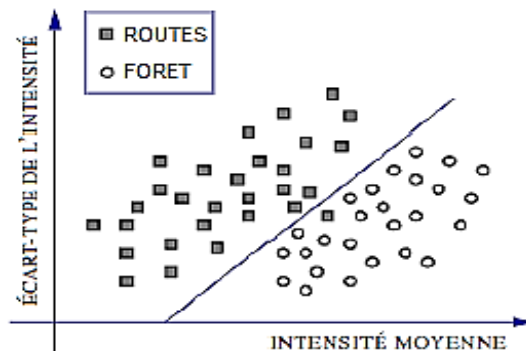


Fig. II.1 : Exemple de la méthode statistique.

Sur la Fig. II.1, la discrimination se fait par une fonction linéaire.

II.1.2. Caractéristiques de type structurel

La notion de structure révèle l'existence d'une décomposition de toute la structure en parties et relations entre ces parties. Les méthodes structurales de RdF adoptent des représentations faisant apparaître explicitement les relations entre éléments de la forme (pas de vecteurs à valeurs réelles). Ils sont évidemment plus adaptés pour traiter des structures visuelles où les composants sont liés par des relations spatiales.

Dans l'approche structurale dite primitive, une classe est décrite par une grammaire. La grammaire est constituée d'un ensemble de règles syntaxiques qui déterminent l'ensemble des formes admissibles dans cette classe et qui en principe ont des caractéristiques structurelles communes, par exemple, ligne verticale ou horizontale, cercle, rectangle, etc. Un exemple illustratif est donné dans la Fig. II.2.

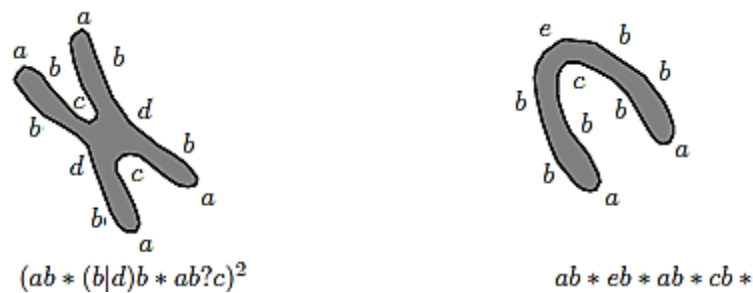


Fig. II.2 : Exemple de la méthode structurale

- Les images des chromosomes sont prétraitées par un détecteur de contours.
- Les contours sont décomposés en segments de lignes et de courbes codés par a, b, c, d et e .
- La décision revient à analyser la conformité d'une chaîne par rapport à une expression régulière (grammaire).

Il est à noter que :

- Le symbole ? signifie que nous acceptons n'importe lequel des primitives.
- Le symbole * indique que nous acceptons un nombre non spécifié de la même primitive.
- Le symbole $x|y$ signifie que nous acceptons x ou y
- Le symbole x^n signifie que nous acceptons n séquences de type x

Exemples: Les séquences $ab * c? d = abacbd; abbbbccd; abcad;$

L'hybridation entre les deux méthodes peut être utilisée. C'est une collaboration entre l'approche statistique et l'approche structurale, cela s'appelle l'approche statistico-structurale. La combinaison des deux techniques peut améliorer les performances des systèmes RdF. Ces

systèmes utilisent deux niveaux d'interprétation. Le premier extrait puis reconnaît les primitives des formes grâce à un classificateur statistique. Le second exploite les caractéristiques structurelles des formes puis reconnaît les formes à l'aide d'un classificateur structurel.

II.2. Techniques d'extraction de caractéristiques

L'extraction de caractéristiques d'une image consiste à transformer les valeurs brutes des pixels de l'image en informations plus significatives et utiles pouvant être utilisées dans une application de vision par ordinateur (Onur C. Hamsici et al, 2016). Ainsi, ces procédés ont été soigneusement conçus par des spécialistes (*Hand-crafted methods*) utilisant leurs connaissances et leurs expériences. Généralement, ces techniques sont appliquées à l'image sans connaissance préalable de son contenu. Ainsi, pour améliorer leurs performances, ces technologies ont reçu une connaissance préalable de leur contexte grâce à un apprentissage automatique (*Learned Hand-crafted methods*). Heureusement, avec l'avènement de "l'apprentissage profond convolutif", les réseaux de neurones eux-mêmes font automatiquement le travail de l'expert à travers les couches convolutives. En règle générale, toutes les méthodes d'extraction d'entités se divisent en trois catégories principales :

II.2.1. Techniques classiques (*Hand-crafted*)

L'image contient de nombreuses caractéristiques qui peuvent être utilisées pour analyser une scène donnée. Le type de caractéristique utilisée varie en fonction de l'application prévue. Par exemple, dans une image contenant une scène composée de plusieurs formes, plusieurs méthodes doivent être utilisées pour extraire chaque forme séparément. Par conséquent, il est nécessaire de connaître le type de scène pour choisir efficacement la méthode appropriée. Parce que ces techniques nécessitent une expertise, les caractéristiques obtenues sont alors considérées comme artisanales (*Hand-craft*). Une des faiblesses de ces méthodes est que les caractéristiques de l'image sont extraites selon un ensemble d'étapes préconçues et ne sont pas apprises dans le contexte auquel appartient l'image. Comme exemples de certaines méthodes d'extraction de caractéristiques artisanales, transformée en cosinus discrète (*Discrete Cosine Transform (DCT)*) (Gilbert Strang, 1999), Transformée en Ondelettes Discrète (*Discrete Wavelet Transform (DWT)*) (C. Lai and C. Tsai, 2017), motifs locaux binaires (*Local Binary Pattern (LBP)*) (Z. Guo et al, 2010), analyse en composantes principales (*Principal Component Analysis (PCA)*) (Jonathon Shlens, 2014), analyse en composantes indépendantes (*Independent Component Analysis (ICA)*) (Alaa Tharwat, 2018), Ondelettes de Gabor (*Gabor Wavelet*) (David Bařina, 2011) et autres.

II.2.2. Techniques classiques apprises (Learned Hand-crafted)

Pour améliorer l'efficacité de la méthode d'extraction de caractéristiques, de nombreux chercheurs ont tenté d'incorporer l'idée d'apprentissage dans des méthodes d'extraction de caractéristiques artisanales pour extraire des vecteurs de caractéristiques plus précis et plus discriminants. L'idée générale est de fournir des connaissances (informations) a priori à la méthode artisanale pour extraire des vecteurs de caractéristiques efficaces. À ce jour, la façon dont ces informations sont présentées varie d'un chercheur à l'autre. En fait, de nombreux travaux dans ce sens se sont concentrés sur la représentation du visage et ont relativement réussi. À titre d'exemples de telles méthodes, les motifs quantifiés locaux (Local Quantized Pattern (LQP)) (V. Khryashchev et al, 2015), le descripteur de visage discriminant (Discriminant Face Descriptor (DFD)) (Z. Lei et al, 2014), l'apprentissage des caractéristiques binaires locales orientés vers les contextes (Context-Aware Local Binary Feature Learning (CA-LBFL)) (Y. Duan et al, 2018) et autres.

II.2.3. Techniques modernes (Deep learning)

Les systèmes de RdF incorporent généralement des classificateurs avec un comportement d'apprentissage, tels que « machines à vecteurs de support » (*Support Vector Machine (SVM)*) (Zhang XD, 2020), « forêt aléatoire » (*Random Forest Transform (RFT)*) (Chelaoua R et Meraoumia A, 2019) et « fonctions de base radiale » (*Radial Basis Function (RBF)*) (Zuzana Majdisova et Vaclav Skala, 2017). Malheureusement, ces types de systèmes peuvent rencontrer des problèmes importants lorsque les vecteurs de caractéristiques d'apprentissage ont un taux de corrélation élevé, ce qui affecte considérablement la précision du système pendant la phase de classification. Afin de remédier à cet inconvénient, le système doit lui-même extraire les vecteurs de caractéristiques (une étape d'extraction appropriée), puis les vecteurs de caractéristiques obtenus servent de vecteurs d'apprentissage (base d'apprentissage) pour le classificateur utilisé. Le système reçoit donc des images brutes en entrée au lieu de vecteurs de caractéristiques pré-extraits.

En outre, l'analyse d'image à plusieurs niveaux peut fournir des vecteurs de caractéristiques précis car elle doit se comporter comme une méthode multimodale de sorte qu'un concept de fusion au niveau des caractéristiques soit appliqué lors du passage d'un niveau d'analyse à un autre. Bien sûr, la nécessité de décorrélérer les différents vecteurs de caractéristiques de différentes classes (interclasses) est plus importante pour obtenir des systèmes de reconnaissance de formes efficaces, ce qui nécessite l'utilisation de méthodes aussi profondes.

L'apprentissage profond présente de nombreux avantages par rapport aux techniques d'apprentissage automatique classiques ou à d'autres méthodes en dehors de l'apprentissage classique (Zhang D et al, 2016). Bien que de nombreux problèmes puissent être résolus par des méthodes artisanales (appries), il existe des cas où ces méthodes ne peuvent pas couvrir suffisamment la complexité de la tâche pour être vraiment efficaces. Parmi plusieurs architectures de réseau deep learning, celles basées sur la « convolution » ont montré une efficacité très remarquable (C. Jia et al, 2019).

En effet, le plus grand avantage de cette architecture est sa flexibilité, qui est généralement obtenue à partir du nombre de couches d'analyse qui peut être augmenté pour atteindre une complexité incroyable. Un système de RdF typique basé sur l'apprentissage profond est illustré à la Fig.II.3, qui montre ses deux parties principales, l'une pour l'analyse qui peut inclure plusieurs niveaux de convolution et l'autre pour la classification, qui est l'apprentissage automatique. Puisque notre travail est basé sur l'apprentissage profond, nous expliquerons dans ce qui suit leurs couches constitutives.

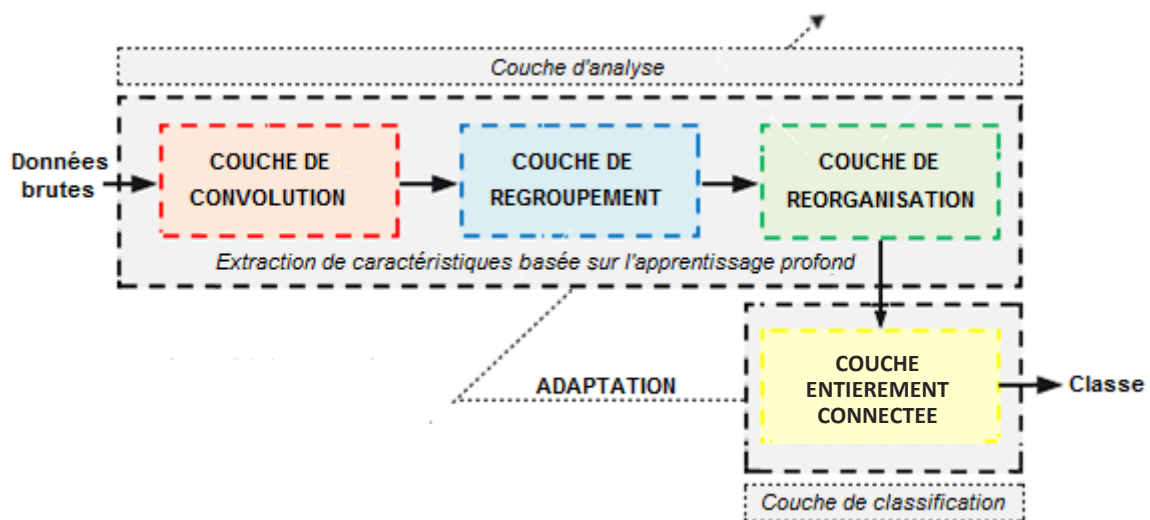


Fig. II.3 : Système de reconnaissance de formes basé sur l'apprentissage profond par convolution.

1. Couches d'analyse : La principale différence entre la technique artisanale et l'apprentissage profond réside dans la couche d'analyse. En fait, ces deux méthodes se correspondent dans la couche de classification, qui est généralement un classificateur intelligent. En apprentissage automatique, les caractéristiques de l'image sont extraites indépendamment du classificateur, ce qui peut affecter son résultat si la méthode d'extraction de caractéristiques utilisée est inappropriée (donne des vecteurs de caractéristiques avec une corrélation interclasse élevée). Au contraire, en apprentissage profond, les vecteurs de caractéristiques extraits sont automatiquement adaptés au classificateur. De plus, l'image est

analysée à plusieurs niveaux, ce qui permet de pénétrer dans la profondeur de l'image et d'en extraire les caractéristiques appropriées pour sa classification.

Comme le montre la Fig. II.3, la partie analyse se compose de trois sous-parties, qui sont *i*) la sous-partie «convolution», dans laquelle l'image est analysée à l'aide de plusieurs filtres de «convolution» afin de capturer uniquement les caractéristiques importantes, *ii*) la sous-partie «regroupement», dont l'objectif principal est de réduire le volume de données en éliminant les données moins importantes. Enfin, *iii*) la sous-section «réorganisation», qui nous permet de normaliser le vecteur de caractéristiques final pour qu'il soit approprié pour l'entrée de classificateur. Il est à noter que l'image peut être analysée sous plusieurs niveaux afin que chaque niveau puisse être composé d'une sous-partie de convolution suivie d'une sous-partie de regroupement. La longueur du vecteur de caractéristiques final est principalement liée à la sous-partie regroupement (mise en commun) du dernier niveau, qui est liée au nombre de niveaux d'analyse et au nombre de filtres convolutifs à chaque niveau.

2. Couches de classification : La deuxième couche est consacrée à la classification des images, basée sur les vecteurs de caractéristiques extraits de la première couche (couche d'analyse). Dans cette couche, nous pouvons utiliser n'importe lequel des classificateurs, qu'ils dépendent de l'apprentissage, par exemple : réseau de neurones artificiels (*Artificial Neural Network* (ANN) (Jake Frankenfield, 2020)), (SVM) ou (k-plus proches voisins (*K-Nearest Neighbours* (KNN) (Pdraig Cunningham et Sarah Jane Delany, 2020)). Par exemple, si le système de reconnaissance de formes profondes dépend de l'ANN, il s'appelle un «réseau neuronal convolutif» (*Convolutional Neural Network* (CNN) (U Raghavendra, 2018)).

II.3. Rôle d'apprentissage

Dans l'apprentissage, il s'agit de doter le système d'un ensemble de formes déjà connues. C'est cet ensemble d'apprentissage qui ajustera le système de reconnaissance afin qu'il soit capable de reconnaître les formes ultérieures de classe inconnue.

II.3.1. Principes de l'apprentissage

En général, un système RdF a de nombreux paramètres qui doivent être adaptés. Ils sont basés sur la spécificité des données et les conditions externes. Cependant, les systèmes de RdF sont conçus pour que les paramètres puissent être fixés pendant une phase d'apprentissage à partir des données d'apprentissage. Par conséquent, le rôle du module d'apprentissage consiste à caractériser chaque classe par les paramètres définissant la forme afin de bien distinguer les familles de formes homogènes. C'est une étape clé du système intelligent de classification.

II.3.2. Schéma de l'apprentissage

La décision nécessite une définition claire des connaissances que nous avons sur les formes à traiter. Cette définition est basée sur l'apprentissage, qui est chargé d'acquérir des connaissances et de les organiser en classes ou modèles de référence (voir Fig. II.4).

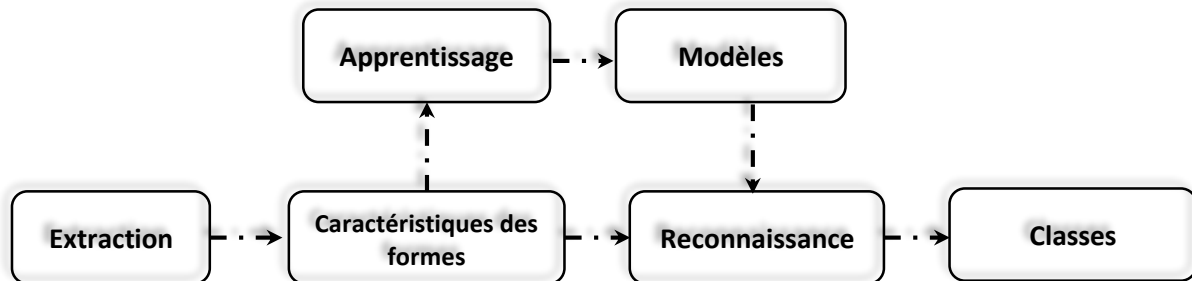


Fig. II.4 : Schéma de l'apprentissage.

Ceci nécessite donc d'apprendre à bien définir les classes de formes et leurs séparateurs (ou modèles) afin de bien distinguer les familles homogènes de formes et donc d'identifier les classes des nouvelles formes. Le rôle du concepteur système n'est ici pas négligeable. Il peut contribuer de diverses manières à un bon apprentissage, par exemple en proposant un bon choix de formes de référence, ou en donnant au système les bons critères pour modéliser sa connaissance de son environnement.

II.3.3. Types d'apprentissage

L'apprentissage consiste à apprendre les caractéristiques communes aux classes afin de distinguer les différentes classes. Nous constituons un ensemble d'apprentissage à partir d'exemples de différentes classes. Cependant, les méthodes d'apprentissage peuvent être regroupées en deux classes principales: les «méthodes supervisées» (avec superviseur) (Zhi-Hua Zhou, 2018) où le système connaît les classes auxquelles appartiennent les échantillons d'apprentissage et les «méthodes non supervisées» (sans superviseur) (Ghahramani Z, 2014) où le système ne connaît pas les classes des échantillons. Parfois, le système ne connaît même pas le nombre de classes comme par exemple *mean-shift* (Pulkit Madaan et Abhishek Maiti, 2019).

1. **Apprentissage supervisé :** Dans le cas d'un apprentissage supervisé (ou avec superviseur), un échantillon représentatif de toutes les formes à reconnaître est fourni au module d'apprentissage. Chaque forme est étiquetée par le superviseur, cette étiquette est utilisée pour indiquer au module d'apprentissage la classe dans laquelle le superviseur souhaite que la forme soit rangée. Cette phase d'apprentissage consiste à analyser les similitudes entre les éléments d'une même classe et les dissemblances entre les éléments de différentes classes pour en déduire la meilleure partition de l'espace de

représentation. Enfin, les paramètres trouvés sont stockés dans une table d'apprentissage à laquelle le module de décision se référera ensuite pour classer les formulaires qui lui sont présentés.

2. **Apprentissage non-supervisé :** Dans ce cas le système de reconnaissance est pourvu d'un grand nombre de formes non étiquetées. L'étape de classification se chargera d'identifier automatiquement les formes appartenant à la même classe. Dans ce cas, le nombre de classes peut être connu ou non, mais les classes des échantillons ne sont pas connues, ce que nous savons, c'est que l'ensemble d'échantillons représente ce que nous devons savoir. Une fois l'apprentissage terminé, le système doit être validé par une phase de test. Cette méthode d'apprentissage nécessite un grand nombre d'échantillons et des règles de construction précises et non contradictoires pour bien former les classes.

II.4. Vers une approche profonde

En plus de l'algorithme utilisé, l'efficacité de l'apprentissage dépend principalement de la base d'apprentissage, qui doit remplir les conditions suivantes:

- Il doit être informatif, ce qui signifie qu'il doit contenir la plupart des variations intra et interclasses;
- Elle devrait être assez grande

La base d'apprentissage contient des vecteurs de caractéristiques extraits de manière artisanale pour toutes les formes du contexte.

II.4.1. Influence des caractéristiques sur la classification

Pour préparer la base d'apprentissage d'un classificateur, une méthode d'extraction de caractéristiques doit être utilisée pour extraire des caractéristiques des formes. Cette méthode est choisie par un expert (ce qui signifie qu'il s'agit d'une méthode artisanale) après avoir fait un tour d'horizon des formes dans le contexte. Après avoir appliqué cette méthode à toutes les images de la base de données, plusieurs vecteurs sont produits. Pour que ces vecteurs soient efficaces, ils doivent remplir deux conditions:

- Faible variabilité des vecteurs de caractéristiques des formes de la même classe;
- Grande variabilité des vecteurs de caractéristiques de forme des différentes classes

Malheureusement, de nombreux facteurs (bruit, déformation, rotation,...etc.) affectent les données brutes (images d'entrée), ce qui peut conduire à une dé-corrélation des vecteurs de caractéristiques d'une même classe et une corrélation des vecteurs de caractéristiques des

différentes classes, entraînant de graves erreurs de classification. L'une des techniques utilisées par les chercheurs pour réduire l'impact de ce problème sur le processus de classification consiste à combiner (fusion de données (TAYEBI Abderrazeq et BOMEDIANA Touhami, 2018)) plusieurs vecteurs extraits de différentes manières pour tirer parti des propriétés positives de chaque méthode.

La fusion de données est utilisée pour traiter les informations provenant de plusieurs sources. Elle consiste à combiner des données de plusieurs sources afin d'obtenir une meilleure décision que celle obtenue à partir de chacune des sources considérées séparément. En général, la fusion de données est une opération consistante à combiner plusieurs données pour extraire de nouvelles informations plus représentatives de l'ensemble de données. Actuellement, la fusion de données devient de plus en plus importante dans de nombreux domaines et elle aide efficacement les scientifiques à extraire des informations de plus en plus pertinentes et précises.

Bien qu'il existe certaines solutions, comme la fusion des données, le problème demeure, en particulier lorsqu'il s'agit de grandes bases de données ou de domaine contenant des formes similaires comme la reconnaissance faciale, où l'apparence faciale d'un jumeau est très proche. Par conséquent, les chercheurs se sont concentrés sur les méthodes d'extraction de caractéristiques pour réduire et pourquoi pas éliminer cet effet.

II.4.2. Adaptation des caractéristiques

Le point sur lequel les chercheurs ont commencé est de savoir comment les vecteurs de caractéristiques peuvent distinguer chaque forme séparément. La réponse à cette question repose sur deux hypothèses:

- La méthode doit avoir la particularité de fusionner les données, à savoir la possibilité de traiter les données à plusieurs niveaux puis de les combiner pour obtenir un vecteur distinct.
- Le classificateur doit jouer le rôle de superviseur de la méthode d'extraction de caractéristiques en modifiant les paramètres de cette méthode en fonction du taux de classification le plus élevé (processus d'adaptation).

L'application de ces deux hypothèses a abouti à ce que l'on appelle maintenant l'apprentissage en profondeur qui, à tout le crédit, a donné des résultats étonnants, surtout si les niveaux d'analyse et la taille de l'apprentissage de la base de données étaient importants.

II.5. Reconnaissance biométrique

Pour évaluer la performance d'une technique d'extraction de caractéristiques basée sur l'apprentissage profond, elle est intégrée dans un système de reconnaissance de formes comme un système biométrique. Nous avons choisi ce type d'application pour deux raisons, la première est que ce système est une application typique de reconnaissance de formes, tandis que la seconde est que les performances du système biométrique sont sensibles à la méthode d'extraction de caractéristiques.

II.5.1. Biométrie

En général, la reconnaissance de l'identité d'une personne est cruciale dans une grande variété d'applications afin de ne fournir des services qu'à de véritables utilisateurs. De nos jours, grâce à l'avantage de la fiabilité et de la stabilité, les technologies biométriques se sont développées rapidement pour résoudre ce problème (Henniger O et al, 2017). Fondamentalement, cette technologie utilise des caractéristiques biométriques telles que les caractéristiques comportementales (par exemple, la voix, la signature, la démarche, la dynamique de frappe, etc.), les caractéristiques physiologiques (par exemple le visage, l'empreinte palmaire, les empreintes digitales, l'iris, etc.) et les caractéristiques biologiques (par exemple l'ADN, la rétine, etc.) de la personne pour déterminer précisément son identité.

La biométrie (Abhilash Sharma et al, 2015) est un domaine émergent où la technologie améliore notre capacité à identifier une personne. La protection des usagers contre la fraude ou le vol est un des buts de la biométrie. L'avantage de la reconnaissance biométrique est que chaque individu a ses propres caractéristiques physiques qui ne peuvent être changées, perdues ou volées. La méthode de reconnaissance biométrique peut aussi être utilisée en complément ou remplacement des mots de passe.

II.5.2. Modalités biométriques

Pour mettre en œuvre un système biométrique, une question évidente se pose lors de sa conception : quelle modalité biométrique allons-nous utiliser? La réponse à cette question nous permet de choisir la technique d'extraction de caractéristiques appropriée qui peut donner la meilleure précision. Ainsi, parmi plusieurs caractéristiques biométriques, celles basées sur des traits biologiques, comme l'ADN, offrent une grande précision, mais malheureusement, son acceptabilité par les utilisateurs est très faible. De plus, bien que les caractéristiques biométriques extraites du comportement humain, comme la voix, aient un taux d'acceptabilité plus élevé, mais leur précision est faible en raison de leur grande variabilité intra-classe. Enfin,

les caractéristiques biométriques basées sur des traits physiques, comme le visage, offrent un bon compromis entre l'acceptabilité de l'utilisateur et la précision du système. En effet, toutes les traits physiques ne se caractérisent pas par une acceptabilité élevée ainsi qu'une grande précision, par exemple, les traits du visage sont très acceptables par les utilisateurs mais peuvent conduire à une très faible précision. De plus, les traits d'empreintes digitales offrent une haute précision mais ne sont pas acceptables pour les utilisateurs.

II.5.3. Fonctionnement d'un système biométrique

Un système biométrique est essentiellement un système de reconnaissance de formes. L'objectif de ces types de systèmes est de classer des formes (modalités biométriques) en catégories (personnes) à partir d'observations (vecteur de caractéristiques) effectuées sur eux. En général, un système de reconnaissance biométrique de personne peut être décomposé en deux phases, la phase «d'enrôlement» (enregistrement ou apprentissage) et la phase de «reconnaissance» (classification). La Fig. II.5 illustre les deux phases de fonctionnement. Une question importante dans la conception d'un système biométrique est de savoir comment une personne est reconnue.

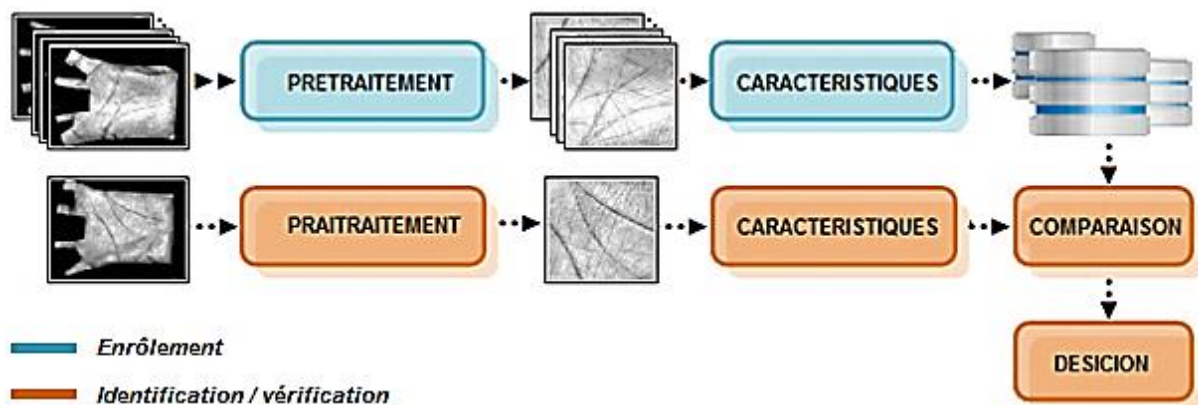


Fig. II.5 : Schéma fonctionnel d'un système de reconnaissance biométrique.

Selon le contexte de l'application, un système biométrique peut être soit un système de vérification, soit un système d'identification. Ce système procède d'abord par l'acquisition des modalités biométriques de la personne en question, puis extrait un ensemble de caractéristiques, enfin, il compare ces caractéristiques avec les modèles de la base de données.

1. Phases du système biométrique : Comme le montre la figure II.3, un système biométrique comprend deux phases distinctes : l'enrôlement (enregistrement) et la reconnaissance (vérification ou identification)

- **Phase d'enrôlement** : Avant d'utiliser les systèmes biométriques, tous les clients doivent s'inscrire dans la base de données du système. A cette phase, et après un processus de prétraitement de la modalité biométrique, le système extrait les caractéristiques biométriques et les stocke dans sa base de données. Ces caractéristiques sont stockées directement (enrôlement sous forme de tableau) si le système ne dépend pas du principe d'apprentissage, sinon, le système effectue la phase d'apprentissage sur les caractéristiques biométriques de tous les clients (enrôlement sous forme de modèle). Dans la phase de reconnaissance, le système se réfère à ces références pour comparaison. Lors de la phase d'enrôlement, les clients peuvent ajouter leurs informations personnelles.

- **Phase de reconnaissance** : À cette phase, le système pourra reconnaître ou vérifier l'identité de la personne en question. Au cours de cette étape, le système extrait les caractéristiques de la modalité biométrique capturée (les propriétés du capteur doivent être aussi proches que possible de celles du capteur utilisé lors de la phase d'enrôlement) puis effectue un processus de mesure de similitude qui change en fonction du mode de fonctionnement du système : vérification ou identification.

- **Mode de vérification** : La vérification est une comparaison un à un (1:1), c'est la confirmation de la validité d'une identité déclarée par la comparaison entre un vecteur de caractéristiques associé à une identité de vérification (proposée par l'utilisateur) et un vecteur de caractéristiques pré-stocké. Lors de la vérification, le système doit vérifier que l'identité de la personne est bien celle proposée par l'utilisateur, il suffit donc de la comparer avec un seul des vecteurs de caractéristiques présents dans la base de données. Dans ce cas, le système répond à la question « *Suis-je bien la personne que je prétends être ?* » par *oui* ou *non*.
- **Mode d'identification** : Lors de l'identification, le système fait une comparaison un à plusieurs (1:N) entre le vecteur de caractéristiques qui est extrait à partir de la personne en question et celles qui sont stockées dans la base de données lors de la phase d'enrôlement. Il répond donc à une question du type : « *Qui suis-je?* » par *accepter* ou *rejeter*. Lors de l'identification, deux modes de fonctionnement se produisent, l'identification en mode ensemble ouvert (*open-set identification mode*) où la personne n'est pas garantie d'exister dans la base de données et l'identification en mode ensemble fermé (*closed-set identification mode*) si la personne est supposée exister dans la base de données.

Conclusion

Les technologies de reconnaissance de formes visent à développer de nouvelles techniques pour transformer les données originales en d'autres représentations qui illustrent honnêtement leur contenu et peuvent ensuite être utilisées pour analyser et comprendre les données originales. En général, ces représentations sont très indispensables dans de nombreuses tâches de vision par ordinateur, telles que les systèmes biométriques, le contrôle et la mesure, et l'imagerie médicale et le diagnostique, pour automatiser la tâche et améliorer la précision des résultats. Dans ce chapitre, nous avons présenté les principales catégories de méthodes d'extraction de caractéristiques. Nous avons donné plus de détails sur les méthodes basées sur le *deep learning* basé sur la convolution, sur lesquelles repose notre contribution. Enfin, parce que notre méthode doit tester sur un système de reconnaissance biométrique, la dernière partie de ce chapitre a été consacrée à la biométrie.

Chapitre 3

Résultats Expérimentaux

Résumé

Afin de représenter efficacement les caractéristiques biométriques d'une empreinte palmaire, les images capturées par le dispositif de capture sont généralement analysées par la méthode d'extraction de caractéristiques. La nouvelle représentation produite par cette méthode doit être unique à chaque personne. En fait, ce chapitre vise à démontrer la faisabilité et l'efficacité de notre méthode proposée à travers la description détaillée des fonctionnalités et des principales caractéristiques du système proposé pour l'identification des personnes sur la base des empreintes multi-spectrales de la paume.

Chapitre III

Résultats expérimentaux

Introduction

L'extraction de caractéristiques est une tâche importante dans les applications de reconnaissance de formes en raison de la grande quantité d'informations existantes dans le signal (en particulier dans l'image) et de ses multiples domaines d'application qui couvrent l'imagerie médicale, la biométrie, la télédétection et la robotique. En raison de cette nécessité, un effort considérable a été fait par les chercheurs dans ce sens, conduisant dans de nombreux cas à des résultats de classification suffisamment excellents. En effet, tous les problèmes associés à la conception finale d'un système de reconnaissance de formes sont généralement liés à cette tâche. Dans ce chapitre, l'impact des techniques d'apprentissage profond (*Deep learning*) sur les performances des systèmes de reconnaissance de formes sera évalué. Pour une évaluation fiable, l'efficacité de la méthode d'extraction des caractéristiques profondes est comparée à celle des méthodes classiques (*Hand-crafted methods*) et classiques apprises (*Learned Hand-crafted methods*). Pour effectuer cette comparaison, un système de reconnaissance biométrique a été développé, qui est une application typique de reconnaissance de formes.

III.1. Modalité biométrique utilisée

Depuis la mise en place de la biométrie comme moyen efficace d'identifier les personnes, les chercheurs ont concentré leurs efforts sur l'exploration du corps humain afin de déterminer quels traits peuvent être utilisés comme biométrie. La biométrie extraite à partir de la main est maintenant une technologie mature largement utilisée dans un large éventail d'applications couvrant le secteur gouvernemental, la médecine légale et le secteur commercial, etc. Cette technologie offre une solution fiable, économique et facile à une gamme d'applications de contrôle d'accès. Par conséquent, les utilisateurs aiment toujours les systèmes d'accès manuel pour deux raisons principales, la première est que les informations extraites de la main ne sont pas considérées comme sensibles ou privées sauf pour les empreintes digitales, et la seconde

est la méthode d'acquisition qui n'est pas une source d'anxiété pour les utilisateurs, contrairement à l'iris ou de la rétine.

III.1.1. Biométries de la main

Fondamentalement, plusieurs caractéristiques biométriques peuvent être extraites de la main de la personne, telles que l'empreinte digitale (J. Galbally et al, 2019), l'empreinte des articulations du doigt (Zhang D et al, Advanced Biometrics, 2018), le réseau veineux du doigt (Shiny Chandra Bai et al, 2017), la géométrie de la main (Dvořák M et Drahanský M, 2019), l'empreinte palmaire (D. Palma et al, 2019) et le réseau veineux de la paume (H. Setiawan et E. M. Yuniarno, 2017). En effet, l'empreinte digitale a un taux d'acceptabilité très faible pour les utilisateurs. Jusqu'à présent, les empreintes des articulations du doigt, le réseau veineux du doigt et la géométrie de la main fonctionnent avec une faible précision. Heureusement, la riche information de la texture de l'empreinte palmaire et du réseau veineux de la paume constitue l'un des moyens puissants d'identification d'une personne. Ainsi, ces deux modalités biométriques peuvent être utilisées pour renforcer la sécurité des applications de contrôle d'accès en les intégrant dans des systèmes classiques.

III.1.2. Base d'images

Le développement de systèmes d'identification biométrique implique l'utilisation d'une base de données pour la phase d'évaluation. Ces dernières années, plusieurs bases de données publiques et gratuites ont été développées pour évaluer ces systèmes. Ces bases de données sont devenues des standards, ce qui permet de comparer les résultats des algorithmes développés avec ceux des algorithmes existants. Les différences dans les modalités biométriques sont étudiées en ayant plusieurs sessions à un intervalle spécifié. En général, les images capturées sont des images en couleur ou en niveaux de gris capturées avec la lumière visible. Ces deux représentations nous donnent essentiellement des informations sur la partie externe uniquement (la peau).

Des études ont montré que l'absorption de la lumière proche infrarouge dans la fenêtre spectrale est mauvaise pour l'eau, et que les tissus humains (peau, graisse sous-cutanée, etc.) sont constitués de 70% d'eau et que le plasma qui occupe 40% du sang est constitués de 95% d'eau, donc la lumière proche infrarouge pénètre profondément dans les couches de tissu humain (Chuang S, 2018). Cela nous permet de nous fournir des informations sur les couches de la peau, en particulier le réseau de veines sanguines, car cette lumière est absorbée dans le

sang par l'hémoglobine et par conséquent, le sang apparaît comme une zone sombre, tandis que la peau et la graisse apparaissent plus claires.

Dans nos travaux, nous avons utilisé la base de données multi-spectrale de l'empreinte palmaire de CASIA qui contient des images de la main acquise à l'aide d'un capteur sans contact à basse résolution. Il a un total de 7200 images obtenues de 100 personnes. Fig. III.1 montre un exemple des images typiques de l'empreinte de la paume dans CASIA.

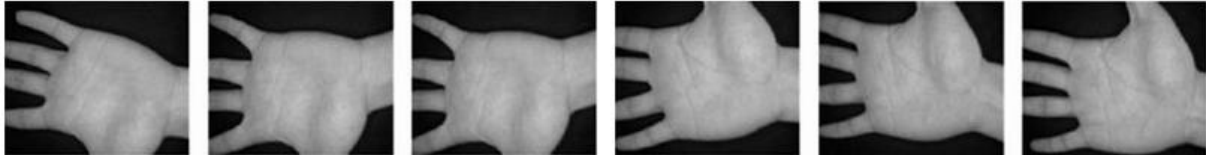


Fig.III.1: Images typiques de l'empreinte de la paume dans la base de données CASIA

Dans cette base de données, les images de chaque main sont collectées en deux sessions distinctes. L'intervalle de temps entre les deux sessions est supérieur à un mois, dans chaque session il y a trois échantillons. Chaque échantillon contient six images de paume qui sont capturées en même temps avec six spectres électromagnétiques différents. Les longueurs d'onde de la lumière correspondant aux six spectres sont respectivement de 460 nm, 630 nm, 700 nm, 850 nm, 940 nm et la lumière blanche. Entre deux échantillons, certains degrés de variation de la position de la main sont appliqués. Toutes les images de paume sont en basse résolution (<150 ppp) stockées sous forme d'images en niveaux de gris codées en 8 bits.

III.2. Extraction des caractéristiques

La méthode d'extraction de caractéristiques basée sur l'apprentissage profond (*Deep Learning*) permet l'analyse de l'image à plusieurs niveaux pour extraire les caractéristiques profondes ou cachées qui peuvent augmenter la précision des vecteurs de caractéristiques, ce qui augmente à son tour la précision du système de reconnaissance de formes. Dans ce travail, nous proposons une nouvelle méthode pour l'extraction de caractéristiques profondes. Cette méthode est basée sur la technique d'apprentissage profond convolutif et se caractérise par sa flexibilité car elle peut produire diverses caractéristiques possibles en fonction de paramètres spécifiques qui sont ajustables selon les besoins.

III.2.1. Prérequis théoriques

Dans les systèmes de reconnaissance de formes, la tâche d'extraction de la caractéristique lors de la conception prend généralement une grande importance en raison de son effet évident sur les taux de reconnaissance. Dans cette section, nous allons essayer de donner quelques

prérequis théoriques concernant la transformée de Fourier discrète (DFT) utilisée dans la technique d'extraction de caractéristiques proposée.

1. Transformée de Fourier discrète : La transformée de Fourier discrète (Olson T, 2017) est un procédé très connu dans le domaine du traitement du signal et de l'image qui permet de passer de la représentation spatiale du signal à la représentation spectrale. La DFT d'une image décorrèle les pixels de l'image et concentre les informations en coefficients basse fréquence. Cette transformation contient deux parties : le réel, qui est représenté par le cosinus (DCT) et l'imaginaire, qui est représenté par le sinus (DST). De plus, nous pouvons utiliser ces deux parties pour extraire des informations sur l'amplitude (module) du signal et sa phase.

Les matrices de transformation 1D-DCT et 1D-DST pour une entrée carrée de taille $B \times B$ sont données par:

$$\varphi_{ij}^{\psi} = \begin{cases} \frac{1}{\sqrt{B}} & i = 0, 0 \leq j \leq B - 1 \\ \sqrt{\frac{2}{B}} \psi\left(\frac{\pi(2j+1)i}{2B}\right) & 1 \leq i \leq B - 1 \\ \sqrt{\frac{2}{B}} \psi\left(\frac{\pi(2j+1)i}{2B}\right) & 0 \leq j \leq B - 1 \end{cases} \quad (III.1)$$

Où $\psi(\cdot)$ désigne $\cos(\cdot)$ ou $\sin(\cdot)$. Le 2D-DCT n'est qu'un produit d'une base verticale et d'une base horizontale de 1D-DCT. Ainsi, en utilisant l'équation III.1, nous générons la matrice φ de taille $B \times B$ qui est utilisée pour créer les bases de DCT (ou DST) de taille $\rho \times \rho$ ($\rho = B \cdot B$):

$$\varphi = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1B} \\ a_{21} & a_{22} & \cdots & a_{2B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{B1} & a_{B2} & \cdots & a_{BB} \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_B \end{pmatrix} \in \mathbb{R}^{B \times B} \quad (III.2)$$

Afin de créer les différentes bases, les composantes de chaque ligne de la matrice φ sont utilisées comme poids pour toutes les lignes de la même matrice, comme le montre l'équation suivante:

$$M_k = v_i^T \cdot v_j \in \mathbb{R}^{B \times B}, \quad i, j = 1, 2, \dots, B, \quad k = 1, 2, \dots, \rho \quad (III.3)$$

Le calcul de $M_k|_{k=1}^{\rho}$ peut être facilement compris par l'exemple suivant:

$$M_1 = v_1^T \cdot v_1 = \begin{pmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1B} \end{pmatrix} \times (a_{11}, a_{12}, \dots, a_{1B}) = \begin{pmatrix} a_{11}(a_{11}, a_{12}, \dots, a_{1B}) \\ a_{12}(a_{11}, a_{12}, \dots, a_{1B}) \\ \vdots \\ a_{1B}(a_{11}, a_{12}, \dots, a_{1B}) \end{pmatrix} \in \mathbb{R}^{B \times B} \quad (III.4)$$

Après cela, chaque matrice résultante (M_k) est réorganisée en un 1D vecteur (V_k):

$$V_k = F_{\rho \times 1}(M_k) \in \mathbb{R}^{\rho \times 1} \quad (III.5)$$

Où $F_{\rho \times 1}(\cdot)$ est une fonction qui mappe la matrice $M_k \in \mathbb{R}^{B \times B}$ à un vecteur $V_k \in \mathbb{R}^{\rho \times 1}$. Les vecteurs obtenus ($V_k|_{k=1}^{\rho}$) sont alors concaténés en un seul vecteur (V):

$$V = [V_1, V_2, \dots, V_{\rho}] \in \mathbb{R}^{\rho \times \rho} \quad (III.6)$$

Nous utilisons maintenant la technique de balayage en *zig-zag* pour réorganiser les composants de ce vecteur:

$$\hat{V} = F_Z(V) = [\tilde{V}_1, \tilde{V}_2, \dots, \tilde{V}_{\rho}] \in \mathbb{R}^{\rho \times \rho} \quad (III.7)$$

La technique du *zig-zag* est utilisée pour réorganiser les composantes du vecteur (\hat{V}) en fonction de leur importance (basse fréquence à haute fréquence), dans laquelle les composantes importantes sont placées au début du vecteur. Enfin, il est important de noter que le premier vecteur (\tilde{V}_1) (première colonne de \hat{V}) représente la valeur moyenne qui est appelée la composante DC.

2. Formulation de filtres : Le système proposé (DFTNet) est similaire aux systèmes PCANet et DCTNet, et la principale différence entre eux réside dans la manière dont les filtres de convolution sont formés: les premier et second systèmes utilisent les techniques PCA et DCT, tandis que notre système utilise la technique DFT (DCT et DST). Par conséquent, dans cette sous-section, nous présenterons la méthode utilisée dans le système DFTNet pour sélectionner les filtres de convolution à partir des bases DCT et DST.

Tout d'abord, comme dans le système PCANet, l'image dans le système DCTNet (et donc DSTNet) est analysée par blocs. *Cong Jie Ng et al* dans (C. J. Nget al, 2016), démontre la forte similitude entre les vecteurs propres des blocs (filtres) et la base DCT (et donc DST). En effet, selon lui, si la corrélation entre les blocs est très élevée, les vecteurs propres de PCA de la matrice de covariance des blocs se rapprocheront de la base DCT (et donc DST). De plus, ils démontrent une grande valeur propre PCA correspondante à une basse fréquence dans le DCT (et donc DST) et vice versa. Cette propriété est vitale pour la sélection des bases DCT et DST pour DFTNet (sélection des filtres), qui suit le PCA en classant le vecteur propre selon la valeur propre respective.

Soit η le nombre de filtres de taille $k_1 \times k_2$ utilisés à un stage donné (en général, on utilise des filtres carrés et impairs, donc $k_1 = k_2$). Les filtres de convolution sont choisis parmi le vecteur 2D-DCT (\hat{V}) prouvé par l'équation III.7. Avant de sélectionner ces filtres, il faut noter que dans un système DCTNet ou DSTNet (et donc DFTNet), la composante DC (valeur moyenne) n'est pas considérée comme un filtre comme dans le système PCANet, dans lequel

la moyenne de chaque patch est supprimée pour de meilleures performances. La sélection des bases commence donc de 2 à $\eta + 1$.

$$V_F = \hat{V}_k(i) \Big|_{i=2}^{\eta+1} \in \mathbb{R}^{\rho \times \eta}, \quad \rho = k_1 \cdot k_2 \quad (III.8)$$

Enfin, l'ensemble des filtres est obtenu comme suit:

$$W_j = F_{k_1 \times k_2}[V_F(j)] \in \mathbb{R}^{k_1 \times k_2}, \quad j = 1, 2, \dots, \eta \quad (III.9)$$

Où W_j désigne filtre j , et $F_{k_1 \times k_2}$ est une fonction qui mappe le vecteur $V_F(j) \Big|_{j=1,2,\dots,\eta} \in \mathbb{R}^{\rho \times 1}$ à une matrice $W_j \in \mathbb{R}^{k_1 \times k_2}$.

III.2.2. Méthode proposée

Dans la Fig. III.2, nous présentons la structure proposée de notre méthode d'extraction de caractéristiques profondes (DFTNet deep learning) qui se compose de deux stages.

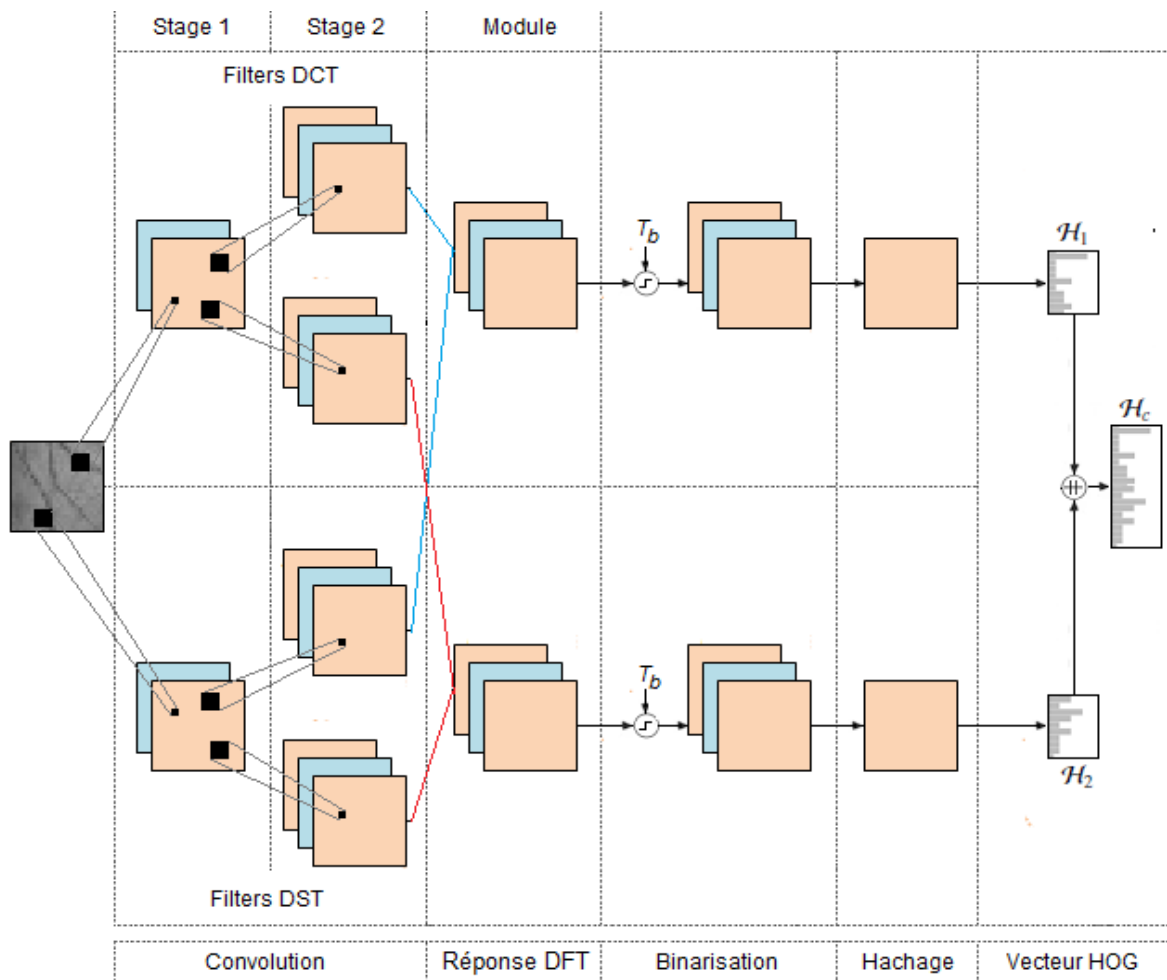


Fig.III.2 : Schéma fonctionnel de la méthode d'extraction de caractéristiques proposée (DFTNet). Un exemple de structure DFTNet à 2 stages avec 2 et 3 filtres de convolution dans le premier et le deuxième stage, respectivement.

En général, cette structure peut être divisée en cinq couches principales: couche de convolution (deux stages), couche pour calculer la réponse de DFT (la partie réelle (DCT), la partie imaginaire (DST), le module ou la phase), couche de binarisation, couche de hachage, et couche de vecteur de caractéristiques (HOG basée sur des blocs). Afin de décrire le cadre du système, nous supposons que les images d'entrée sont de la taille $H \times W$ et que la taille du patch, c'est-à-dire la taille du filtre de convolution (de DCT et DST) pour le stage ℓ est:

$$W_i^\ell = k_1^\ell \times k_2^\ell, \quad i \in [1, \dots, L_\ell], \quad \ell \in [1, \dots, S_\ell] \quad (III.10)$$

où L_ℓ désigne le nombre de filtres dans la couche de convolution ℓ et S_ℓ est le nombre de couches de convolution (stages). Il est important de noter que $k_j^\ell \big|_{j=1,2}$ est un entier impair satisfaisant les conditions $k_j^\ell \leq H$ et $k_j^\ell \leq W$.

1. Couche de convolution: La méthode d'extraction de caractéristiques profondes DFTNet est une technique incluse dans la catégorie des techniques d'apprentissage profond basées sur la convolution, telles que CNN. Ainsi, dans ces techniques, nous convoluons d'abord l'image d'entrée avec plusieurs filtres et après la couche de regroupement (*pooling*), le vecteur de caractéristiques peut être formé. La couche de convolution peut être réalisée à plusieurs stages différents en termes de taille de filtre et de nombre de filtres. Il est important de noter que contrairement à PCANet, DFTNet ne contiennent pas de phase d'apprentissage pour former les filtres de convolutions car il utilise directement les bases DCT et DST comme filtres. Il faut également noter que notre méthode (DFTNet) se compose de deux méthodes complètement similaires (DCTNet et DSTNet) et qui ne diffèrent que par le type de filtre. Par conséquent, pour être pratique, dans la couche de convolution, nous allons expliquer une seule méthode, qui est la DCTNet.

- **Stage 1 :** Dans ce stage, tout d'abord, nous utilisons l'équation III.7 pour former les bases de DCT (ou DST) (avec $B = k_1^1 = k_2^1$), après quoi le vecteur résultant \hat{V} est utilisé pour créer les L_1 filtres de convolution:

$$V_F = \hat{V}(j) \big|_{i=2}^{L_1+1} \in \mathbb{R}^{\rho \times L_1}, \quad \rho = k_1^1 \cdot k_2^1 \quad (III.11)$$

L'ensemble des filtres à ce stage est obtenu comme suit:

$$W_i^1 = F_{k_1^1 \times k_2^1}[V_F(i)] \in \mathbb{R}^{k_1^1 \times k_2^1}, \quad i = 1, 2, \dots, L_1 \quad (III.12)$$

Les sorties de ce stage sont obtenues en filtrant l'image d'entrée (J) par les filtres $W_i^1 \big|_{i=1}^{L_1}$, où les L_1 images filtrées peuvent être obtenues:

$$I_i^1 = I * W_i^1, \quad i = 1, 2, \dots, L_1 \quad (III.13)$$

où le symbole $*$ désigne le processus de convolution 2D et I_i^1 sont la sortie des images filtrées de premier stage. Il est important de noter que pour obtenir des images filtrées de la même taille que I ($H \times W$), un processus d'interpolation aux limites (remplissage de zéro) est appliqué.

- **Stage 2** : Comme le premier stage, les mêmes opérations sont effectuées au deuxième stage sur toutes les sorties du premier stage. De même, à ce stade, nous utilisons l'équation III.7 pour former les bases de DCT (ou DST) (avec $B = k_1^2 = k_2^2$), après quoi le vecteur résultant \hat{V} est utilisé pour créer les L_2 filtres de convolution:

$$V_F = \hat{V}(j) \Big|_{i=2}^{L_2+1} \in \mathbb{R}^{\rho \times L_2}, \quad \rho = k_1^2 \cdot k_2^2 \quad (III.14)$$

Les filtres de convolution à ce stade sont donnés comme suit:

$$W_i^2 = F_{k_1^2 \times k_2^2}[V_F(j)] \in \mathbb{R}^{k_1^2 \times k_2^2}, \quad j = 1, 2, \dots, L_2 \quad (III.15)$$

Les sorties du deuxième stage sont obtenues en filtrant les images d'entrée ($I_i^1 \Big|_{i=1}^{L_1}$) par les filtres $W_j^2 \Big|_{j=1}^{L_2}$:

$$I_{ij}^2 = I_i * W_j^2, \quad i = 1, 2, \dots, L_1, \quad j = 1, 2, \dots, L_2 \quad (III.16)$$

Enfin, en utilisant les L_2 filtres, on peut obtenir L_2 images filtrées pour chaque image filtrée issue du premier stage, donc pour chaque image d'entrée originale, on obtient $L_1 \cdot L_2$ images filtrées à la sortie du deuxième stage.

2. Couche de réponse de DFT : En fait, notre système contient quatre sous-systèmes, y compris le système original (DCTNet). Dans cette couche, nous pouvons configurer notre système pour qu'il fonctionne comme l'un des quatre systèmes suivants: DCTNet, DSTNet, DFT^{MD}Net ou DFT^{PH}Net. Ainsi, les résultats de filtrage du dernier stage $I_{ij}^2 \Big|_{i=1,2,\dots,L_1, j=1,2,\dots,L_2}^{\text{DCT}}$ (obtenus à partir des filtres DCT) et $I_{ij}^2 \Big|_{i=1,2,\dots,L_1, j=1,2,\dots,L_2}^{\text{DST}}$ (obtenus à partir des filtres DST) représentent respectivement les parties réelle et imaginaire du système DFTNet. Ces deux réponses sont combinées pour produire la phase et le module comme suit:

$$\Rightarrow \text{DCTNet} : I_{ij}^{\text{Re}} = I_{ij}^2 \Big|_{i=1,2,\dots,L_1, j=1,2,\dots,L_2}^{\text{DCT}} \quad (III.17)$$

$$\Rightarrow \text{DSTNet} : I_{ij}^{\text{Im}} = I_{ij}^2 \Big|_{i=1,2,\dots,L_1, j=1,2,\dots,L_2}^{\text{DST}} \quad (III.18)$$

$$\Rightarrow \text{DFT}^{\text{PH}}\text{Net} : I_{ij}^{\text{Ph}} = I_{ij}^2 \Big|_{i=1,2,\dots,L_1, j=1,2,\dots,L_2}^{\text{DCT-DST}}, \quad I_{ij}^{\text{Ph}} = \arctan\left(\frac{I_{ij}^{\text{DST}}}{I_{ij}^{\text{DCT}}}\right) \quad (III.19)$$

$$\Rightarrow \text{DFT}^{\text{MDNet}} : I_{ij}^{\text{Md}} = I_{ij}^2 \Big|_{i=1,2,\dots,L_1, j=1,2,\dots,L_2}^{\text{DCT-DST}}, \quad I_{ij}^{\text{Ph}} = \sqrt{(I_{ij}^{2\text{DCT}})^2 + ((I_{ij}^{2\text{DST}})^2)} \quad (\text{III.20})$$

3. Couche de binarisation : Le processus de binarisation (quantification binaire) transforme une valeur réelle en une valeur binaire. En fait, un principe de seuil est appliqué, comme suit:

$$I_{ij}^b(n, m) = \begin{cases} 1 & \text{si } I_{ij}^X(n, m) \geq \tau_b \\ 0 & \text{si } I_{ij}^X(n, m) < \tau_b \end{cases}, \quad X \equiv \{\text{Re, Im, Md, Ph}\} \quad (\text{III.21})$$

$i = 1, \dots, L_1, j = 1, \dots, L_2$

Où τ_b est le seuil de binarisation. En pratique, ce seuil est choisi égal à 0 ($\tau_b = 0$), pour la partie réelle et la partie imaginaire, car les valeurs de la réponse ont la même probabilité d'être négatifs ou positifs. Il est à noter que d'une part, comme le signe de phase suit toujours le signe de la partie imaginaire, un seuil égal à zéro produit les mêmes résultats de binarisation pour la partie imaginaire (DST) et la phase (PH). Deuxièmement, les composantes du module (MD) sont positives, donc un seuil de zéro donne un résultat dans lequel toutes les composantes sont égales à 1 (image blanche). Par conséquent, le seuil de binarisation pour le module (MD) et la phase (PH) est déterminé comme suit:

$$\tau_b = \kappa \cdot \varrho_X, \quad X \equiv \{\text{Md, Ph}\} \quad (\text{III.22})$$

Où ϱ_X est la valeur moyenne du module (ou de la phase) et κ est un facteur prédéfini varié dans l'intervalle $[0,05 \dots 1]$ avec un pas de 0,05.

4. Couche de hachage: Cette couche réduit la quantité de données dans laquelle les L_2 images binarisées sont converties en une image qui a des composantes entières. Ainsi, le code binaire (codé sur L_2 bits) autour de chaque pixel est converti à l'aide du polynôme de décodage suivant (processus de conversion binaire en décimal):

$$I_i^3 = \sum_{j=0}^{L_2-1} \hat{I}_{ij}^b \cdot 2^j, \quad i = 1, \dots, L_1 \quad (\text{III.23})$$

Comme dans PCANet, le nombre de sorties de cette couche dans DCTNet est égal au nombre de filtres utilisés dans le premier stage de convolution. Ainsi, après avoir décodé chaque L_2 groupe séparément, nous obtenons un ensemble d'images égal à L_1 .

5. Couche de vecteur de caractéristiques: Cette couche réduit également la taille de la caractéristique de chaque image. Ainsi, l'histogramme de chaque image parmi les L_1 images est calculé et tous ces histogrammes sont concaténés pour former le vecteur de caractéristique. Contrairement aux PCANet et DCTNet original, notre méthode proposée (DFTNet) utilise l'histogramme de gradient orienté (HOG) (Taha J. Alhindi et al, 2018).

Pour obtenir le vecteur de caractéristique de chaque image d'entrée, nous partitionnons d'abord chaque image $I_i^3 \Big|_{i=1}^{L_1}$ en plusieurs blocs (B). Ainsi, chaque image est partitionnée en N_b blocs comme suit:

$$N_b = \left\lfloor \frac{H-b_1}{o} + 1 \right\rfloor \cdot \left\lfloor \frac{W-b_2}{o} + 1 \right\rfloor \quad (III.24)$$

Où o désigne le chevauchement horizontal / vertical entre deux blocs adjacents, $b_1 \times b_2$ est la taille du bloc d'analyse et $\lfloor \cdot \rfloor$ est la partie entière de la valeur. Pour chaque image ($I_i^3 \Big|_{i=1}^{L_1}$), on obtient un ensemble de Φ_i blocs définis comme suit:

$$\Phi_i = \{B_1^i, B_2^i, \dots, B_{N_b}^i\} \in \mathbb{R}^{(b_1 \times b_2) \times N_b}, \quad B_j^i \Big|_{j=1}^{N_b} \in \mathbb{R}^{b_1 \times b_2}, \quad i = 1, \dots, L_1 \quad (III.25)$$

où B_j^i désigne le $j^{\text{ème}}$ bloc de l'image I_i^3 . Ensuite, un histogramme HOG pour chaque bloc (B_j^i) est calculé et tous les vecteurs résultants pour les L_1 images sont ensuite concaténés pour former le vecteur de caractéristiques final pour l'image examinée.

Dans la technique HOG, l'image d'entrée est analysée par fenêtre (W_{HOG}), dans laquelle chaque fenêtre (W_{HOG}) est divisée en cellules non superposées. Ensuite, l'orientation et la magnitude du gradient sont calculées pour chaque pixel. Un histogramme de ces orientations est formé pour chaque cellule. L'amplitude du gradient est utilisée comme poids de vote. Les histogrammes des cellules de chaque fenêtre sont concaténés pour former le descripteur HOG. Ainsi, en utilisant la technique HOG, nous pouvons extraire la caractéristique de chaque bloc comme:

$$H_j^i = F_{\text{HOG}}(B_j^i) \in \mathbb{R}^{\lambda \times 1}, \quad j = 1, 2, \dots, N_b, \quad i = 1, 2, \dots, L_1 \quad (III.26)$$

où F_{HOG} désigne le processus d'extraction de caractéristiques HOG et λ est la longueur de l'histogramme de bloc. Cette valeur (λ) est fonction du nombre de fenêtres de HOG (η_w) et du nombre de bins d'histogramme (η_b). Pour chaque image $I_i^3 \Big|_{i=1}^{L_1}$, les vecteurs de caractéristiques HOG extraits de tous les blocs sont concaténés en un seul vecteur (ϑ_i):

$$\vartheta_i = [H_1^i, H_2^i, \dots, H_{N_b}^i] \in \mathbb{R}^{(N_b \lambda) \times 1}, \quad i = 1, 2, \dots, L_1 \quad (III.27)$$

Enfin, le vecteur de caractéristiques de l'image d'entrée est obtenu comme suit:

$$V_T = [\vartheta_1, \vartheta_2, \dots, \vartheta_{L_1}] \in \mathbb{R}^{(N_b \lambda L_1) \times 1} \quad (III.28)$$

Il est important de noter que la longueur et la précision du vecteur (V_T), de chaque image d'entrée, changent en fonction de la taille du bloc ($b_1 \times b_2$) et du taux de chevauchement (o).

III.3. Résultats et discussions

L'objectif de cette section est d'évaluer les performances de la méthode proposée, nous avons donc implémenté notre méthode dans un système d'identification biométrique basé sur l'empreinte de la paume de la main dans le domaine multispectral. Dans nos expériences, nous avons utilisé une base de données de 100 personnes, ce qui est similaire à un nombre d'employés dans une petite et moyenne entreprise. Le système biométrique étant composé de deux phases: l'enrolement et l'identification, nous avons divisé la base de données en deux galeries. Dans la galerie d'enrolement, nous avons utilisé aléatoirement trois échantillons de la modalité biométrique (bande spectrale) pour chaque personne, soit $100 \times 3 = 300$ échantillons, tandis que les neuf autres échantillons ont été utilisés pour l'évaluation des performances du système, soit $100 \times 9 = 900$ échantillons. En utilisant toutes les images de test, 45450 scores correspondants ont été obtenus, dont 900 scores pour des expériences authentiques et 44550 scores pour des expériences d'imposteurs.

Dans ce travail, nous avons mené plusieurs expériences, que nous pouvons organiser en deux parties principales. Dans la première partie, nous présenterons des expériences pour évaluer les performances du système biométrique. Le but de cette partie est d'évaluer l'efficacité des systèmes proposés en sélectionnant leurs paramètres optimaux afin de choisir le meilleur système parmi les différents systèmes proposés (DCTNet, DSTNet, DFT^{MD}Net et DFT^{PH}Net). La deuxième partie se concentre sur l'efficacité de notre système par rapport aux systèmes basés sur des méthodes d'extraction de caractéristiques classiques (hand-crafted method) et apprises (learned hand-crafted method).

III.3.1. Performance des systèmes proposés

Cette partie vise à tester les systèmes biométriques proposés. Elle est également divisée en deux sous-parties: la première traite des expériences préliminaires pour sélectionner les paramètres optimaux pour chaque système, tandis que la seconde sous-partie se concentre sur la comparaison des systèmes proposés pour sélectionner les meilleurs.

1. Expériences préliminaires : En biométrie, il est difficile d'obtenir un résultat exact pour un système de reconnaissance des personnes. Cela est principalement dû au choix des caractéristiques qui doivent représenter exactement l'identité des personnes. Ce choix est important car il conditionne toute la méthodologie utilisée pour la reconnaissance. Ainsi, comme nous l'avons montré dans la section précédente, notre méthode (pour chaque sous-système) comprend deux paramètres (nombre de filtres dans chaque stage ($L_i|_{i=1}^2$) et taille des

filtres dans chaque stage ($k_i |_{i=1}^2$) pour contrôler la précision du vecteur de caractéristiques extrait. Il est donc impératif d'évaluer les performances du système d'identification pour sélectionner les paramètres appropriés donnant les meilleurs résultats. Mathématiquement, il n'y a pas de formule magique pour déterminer directement les paramètres optimaux qui fournissent les meilleures performances du système. En général, dans ces cas, les tests expérimentaux sont réalisés en faisant varier les différents paramètres dans des ensembles prédéfinis puis en sélectionnant la combinaison qui optimise une fonction objective (tests empiriques). Au début, il faut noter que nos tests sont réalisés sur un système d'identification biométrique basé sur la bande spectrale "WHITE" de l'empreinte palmaire, qui fonctionne en mode ensemble ouvert. En outre, le système est développé et implémenté en utilisant Matlab 2009 sur une plateforme Windows 7 et un PC embarqué avec un processeur Intel Pentium de 2,2 GH avec 2 Go de DRAM.

Les performances du système biométrique sont évaluées selon les quatre méthodes d'extraction de caractéristiques (DCTNet, DSTNet, DFT^{MD}Net et DFT^{PH}Net) en faisant varier un paramètre à chaque fois. Pour cela, sous le premier stage ($\ell = 1$), nous essayons de sélectionner le nombre de filtres (L_1) parmi cinq nombres prédéfinis (2, 4, 6, 8, 10). De plus, la taille des filtres de convolution ($k_1 \times k_2$) est choisie parmi cinq tailles prédéfinies (9×9, 11×11, 13×13, 15×15, 17×17). Il convient de noter que nos systèmes utilisent le classificateur SVM en raison de son efficacité dans les applications de classification. Il est important de noter que pour les systèmes DFT^{PH}Net et DFT^{MD}Net, le seuil de binarisation (κ) est présélectionné à 0,5. Pour examiner l'effet de ces paramètres sur la précision du système d'identification biométrique (mode ensemble ouvert), nous illustrons, pour les différentes méthodes d'extraction de caractéristiques, les performances des systèmes sous forme de taux d'identification (Genuine Acceptance Rate (GAR)) et les résultats expérimentaux obtenus, si les systèmes biométriques sont configurés en un seul stage de convolution, sont présentés sur la Fig. III.3. Ainsi, à partir des courbes des Fig. III.3.(a), Fig. III.3.(b), Fig. III.3.(c) et Fig. III.3.(d), on peut tirer quatre remarques importantes:

- Une performance très acceptable peut être obtenue avec toutes les combinaisons possibles de L_1 et ($k_1 \times k_2$) où un taux d'identification (ensemble ouvert) effectif (GAR) supérieur à 99,200% a déjà été obtenu.
- En général, plus le nombre de filtres est élevé, plus le taux d'identification est élevé, de sorte que les meilleurs résultats ont été obtenus avec 10 filtres, ce qui est meilleur que les performances obtenues avec 2 filtres.

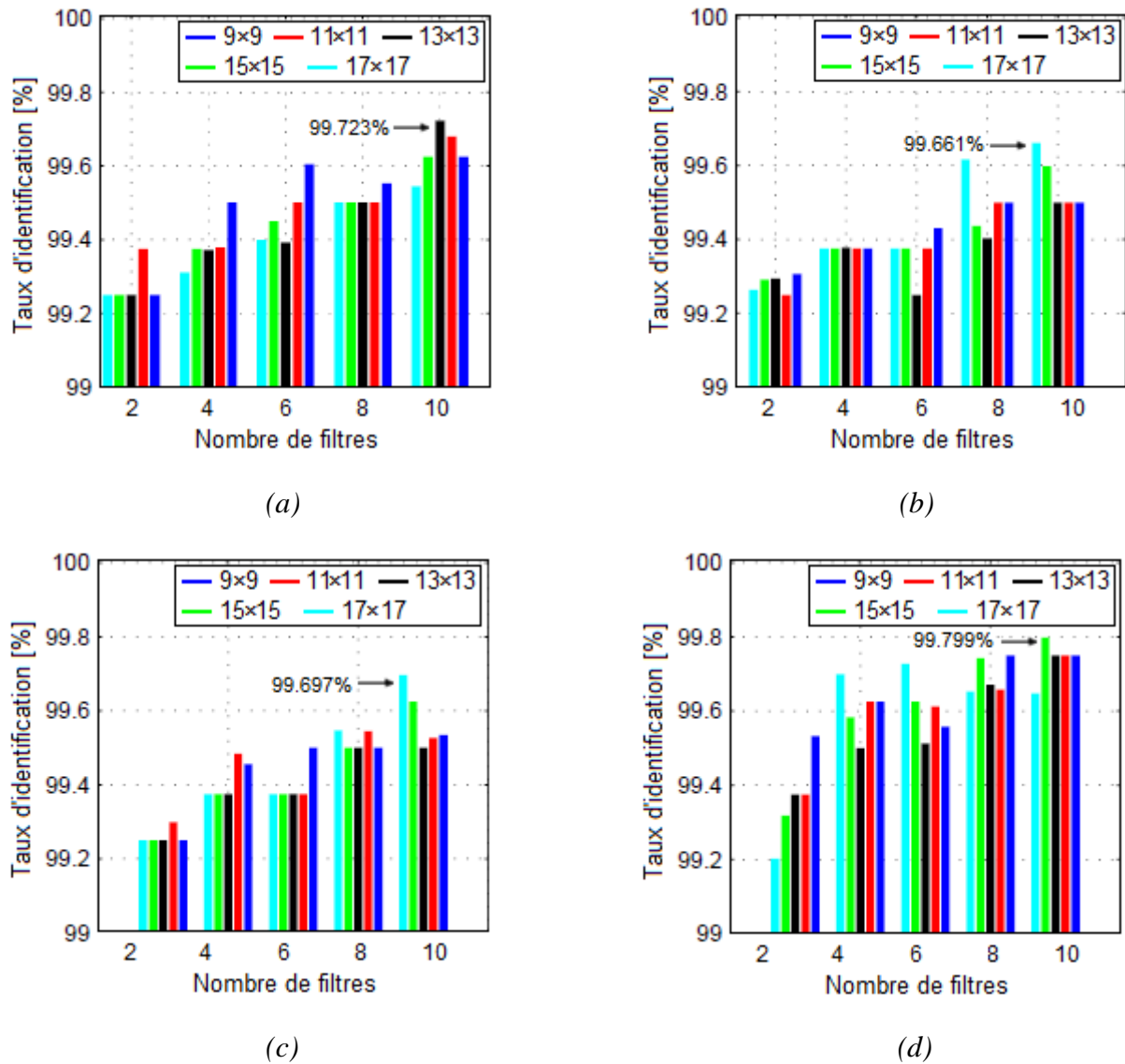


Fig.III.3 : Performances du système biométrique en un seul stage. (a) Système biométrique basé sur DCTNet, (b) Système biométrique basé sur DSTNet, (c) Système biométrique basé sur DFT^{PH}Net, et (d) Système biométrique basé sur DFT^{MD}Net.

- Par rapport aux filtres de petite taille, les performances du système peuvent être améliorées avec les filtres de grande taille, qui peuvent donner d'excellentes performances.
- Enfin, l'utilisation de la méthode d'extraction de caractéristiques DFT^{MD}Net peut améliorer efficacement les performances du système.

D'après la Fig. III.3, il est clair que 10 filtres de convolution offrent de meilleurs résultats en termes de GAR. Dans ce cas, le système d'identification peut atteindre un taux d'erreur (Equal Error Rate (EER)) de 0,2767% (à un seuil T_o égal à 0,6396), 0,3391 ($T_o = 0,6296$) et 0,3034% ($T_o = 0,6283$) avec des tailles de filtre 13x13, 17x17 et 17x17 pour les systèmes basés sur DCTNet, DSTNet et DFT^{PH}Net, respectivement. Des améliorations de 27,20%, 40,58% et

33,45% peuvent être obtenues en utilisant le système basé sur $DFT^{MD}Net$ avec une taille de filtre de 15×15 , dans ce cas, le système fonctionne avec un EER de 0,2015% ($T_o = 0,7068$).

Bien qu'un stage de convolution ait donné des résultats acceptables, elle n'exclut pas la possibilité d'accepter un utilisateur non autorisé ou de rejeter un client. Heureusement, notre système peut être configuré en plusieurs stages, ce qui peut réduire les erreurs de reconnaissance et améliorer l'efficacité du système. Pour cette raison, nous allons réexaminer les systèmes proposés sous deux stages de convolution afin d'améliorer leurs performances. Par conséquent, tous les systèmes proposés utilisent dans le premier stage 4 filtres de convolution de taille 15×15 , et nous effectuons la même méthodologie de test que dans la partie précédente. Ainsi, pour montrer l'efficacité de cette configuration (deux stages de convolution), les résultats des tests d'identification en mode ensemble ouvert, pour tous les systèmes biométriques, sont présentés dans la Fig. III.4.

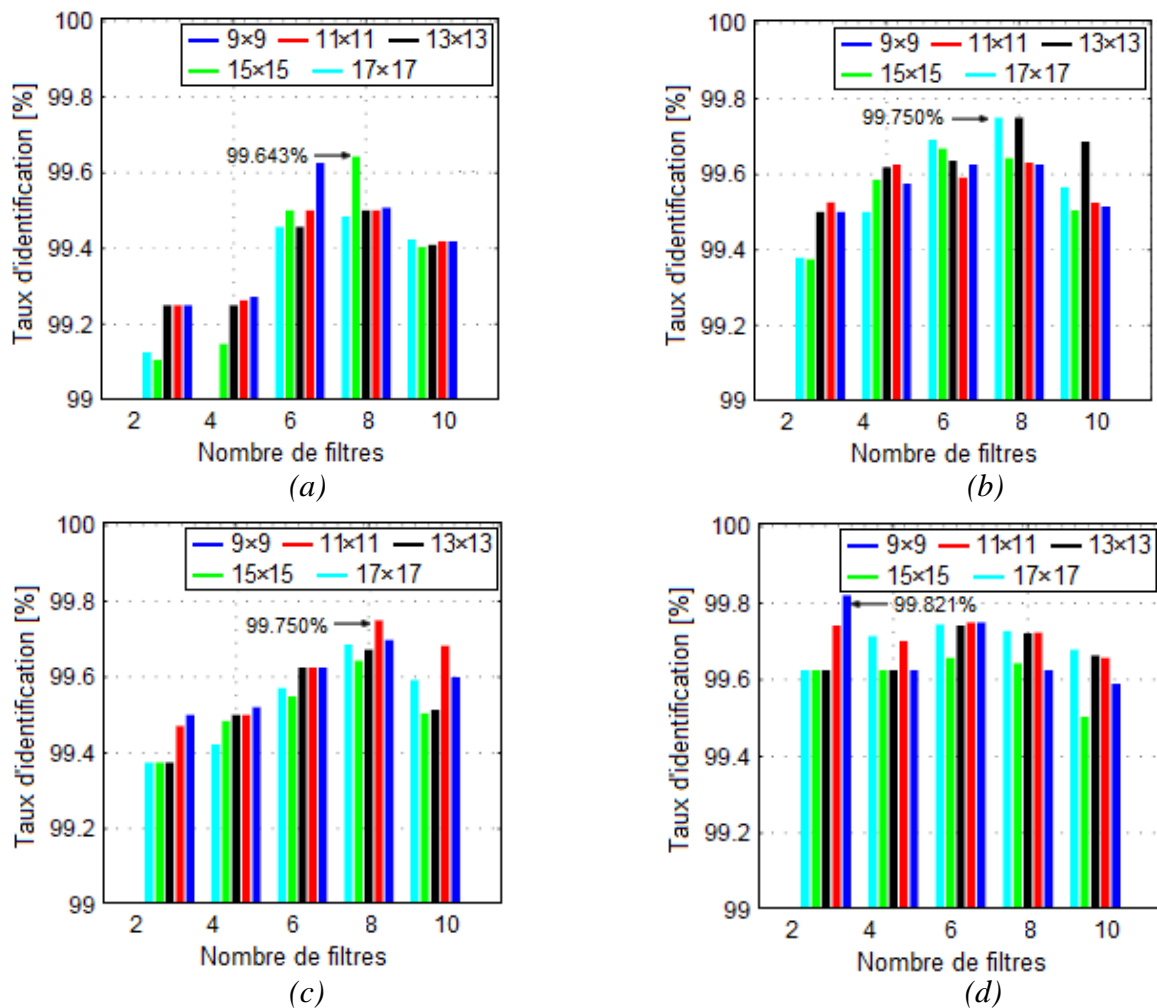


Fig.III.4 : Performances du système biométrique en deux stages. (a) Système biométrique basé sur DCTNet, (b) Système biométrique basé sur DSTNet, (c) Système biométrique basé sur $DFT^{PH}Net$, et (d) Système biométrique basé sur $DFT^{MD}Net$.

Dans ces tests également, nous pouvons tirer, de la Fig. III.4, deux remarques importantes:

- Généralement, l'utilisation de deux stages de convolution améliore les performances des systèmes biométriques proposés.
- Plus le nombre des stages de convolution est élevé, meilleurs sont les filtres de petite taille (ce qui donnera de meilleurs résultats), surtout si les tailles des filtres dans les premiers stages de convolution sont grandes.

Grâce à l'observation et à l'évaluation des Fig. III.4.(a), Fig. III.4.(b), Fig. III.4.(c) et Fig. III.4.(d), nous pouvons voir que des améliorations de performance de 26,250%, 22,120% et 10,950% peuvent être obtenues pour DSTNet, DFT^{PH}Net et DFT^{MD}Net, respectivement. Malheureusement, une détérioration des performances de 57,27% a été obtenue pour le système DCTNet. En général, ces améliorations sont très acceptables et leur importance augmente avec la croissance de la base de données. D'après la figure 3, les meilleures performances sont obtenues lorsque la méthode d'extraction de caractéristiques basée sur DFT^{MD}Net est utilisée. Dans ce cas, le système fonctionne avec un EER égal à 0,1790% à un seuil $T_o = 0,7003$. De plus, il est important de noter que de meilleures performances peuvent également être obtenues avec d'autres configurations. Par exemple, les systèmes biométriques proposés vont fonctionner avec un EER de 0,250% à $T_o = 0,6138$ et $T_o = 0,6170$ dans le cas de DSTNet et DFT^{PH}Net, respectivement. Le système basé sur DCTNet peut fonctionner avec un EER de 0,3570 à $T_o=0,5966$.

Puisque la méthode d'extraction de caractéristiques basée sur DFT^{MD}Net a donné les meilleurs résultats, dans les expériences restantes, cette méthode sera utilisée pour évaluer le système d'identification biométrique dans deux modes: ensemble ouvert et ensemble fermé. Mais avant cela, nous choisirons d'abord le meilleur seuil de binarisation (κ) parmi un ensemble de seuils appartenant à l'intervalle [0,05..1,00].

2. Performances du système biométrique : Pour évaluer judicieusement notre méthode, nous avons utilisé le KNN et le SVM comme classificateurs pour évaluer les performances du système biométrique basé sur DFT^{MD}Net. De plus, le système a été testé en deux stages de convolution. La variation du seuil de binarisation (κ) lui permet de représenter les caractéristiques de l'image avec le contraste le plus approprié. Ainsi, dans nos tests, il existe un large éventail de choix ($\kappa \in [0,05..1]$, par un pas de 0,05, ce qui donne 20 cas), nous pouvons donc choisir le nombre le mieux adapté à l'analyse de texture dans une application spécifique. Ainsi, pour montrer l'impact du seuil de binarisation sur les performances du système, les

meilleurs résultats des tests d'identification en ensemble ouvert au point EER sont présentés dans le tableau III.1.

Tableau III.1 : Performances du système biométrique (DFT^{MD}Net) sous les différentes bandes.

| STAGE | BANDES | KNN | | | SVM | | |
|-------|--------|----------|--------|---------|----------|--------|---------|
| | | κ | T_o | EER [%] | κ | T_o | EER [%] |
| 1 | WHITE | 0.60 | 0.2505 | 0.250 | 0.60 | 0.7695 | 0.125 |
| | 460 | 0.30 | 0.1935 | 0.125 | 0.40 | 0.6883 | 0.125 |
| | 630 | 0.25 | 0.2895 | 0.250 | 0.30 | 0.6275 | 0.250 |
| | 700 | 0.20 | 0.2662 | 0.410 | 0.45 | 0.7335 | 0.250 |
| | 850 | 0.20 | 0.3084 | 0.875 | 0.25 | 0.6687 | 0.875 |
| | 940 | 0.25 | 0.3033 | 1.000 | 0.20 | 0.6360 | 1.250 |
| 2 | WHITE | 0.30 | 0.2340 | 0.250 | 0.45 | 0.6973 | 0.178 |
| | 460 | 0.30 | 0.2171 | 0.250 | 0.45 | 0.7255 | 0.125 |
| | 630 | 0.30 | 0.2144 | 0.193 | 0.20 | 0.7080 | 0.250 |
| | 700 | 0.15 | 0.1487 | 0.180 | 0.30 | 0.7309 | 0.080 |
| | 850 | 0.15 | 0.2447 | 0.804 | 0.20 | 0.7185 | 0.689 |
| | 940 | 0.20 | 0.2709 | 0.963 | 0.25 | 0.6593 | 0.875 |

Ainsi, à l'exception des bandes spectrales de *WHITE* et *460nm*, qui donnent un EER égal à 0,125% aux seuils (T_o) égaux à 0,7695 et 0,6883 avec des seuils de binarisation de 0,6 et 0,4 en un seul stage de convolution, les bandes spectrales *630nm*, *700nm*, *850nm* et *940nm* fonctionnent mieux avec deux stages de convolution, dans lesquelles un excellent EER de 0,080% ($T_o = 0,7309$) peut être obtenu lorsque la bande spectrale de 700 nm est utilisée avec un seuil de binarisation égal à 0,30. Enfin, il ressort de ce tableau que le classificateur SVM est plus performant que le classificateur KNN en terme de taux d'identification, et bien sûr, sans oublier leur vitesse d'exécution qui est supérieure à celle de KNN.

En effet, le tableau III.2 récapitule les différentes performances du système basé sur DFT^{MD}Net utilisant les différentes bandes spectrales. À partir de ce tableau, nous remarquons que notre méthode proposée peut potentiellement offrir de meilleurs taux d'erreur. Cependant, dans l'identification en mode ensemble ouvert, cette méthode conduit à une erreur minimale égale à 0,080% ($T_o = 0,7309$) dans le cas de la bande spectrale de *700nm*. Ce résultat confirme la supériorité de notre méthode d'extraction de caractéristiques.

Tableau III.2 : Meilleur système biométrique (DFT^{MD}Net) pour chaque bande spectrale

| Bande spectrale | κ | Ensemble ouvert | | Ensemble fermé | | Classificateur | Nombre de stages |
|-----------------|----------|-----------------|---------|----------------|---------|----------------|------------------|
| | | T_o | EER [%] | RPR | ROR [%] | | |
| WHITE | 0.60 | 0.6973 | 0.125 | 13 | 98.375 | SVM | 1 |
| 460 | 0.40 | 0.6882 | 0.125 | 60 | 99.250 | SVM | 1 |
| 630 | 0.30 | 0.2144 | 0.193 | 71 | 98.875 | KNN | 2 |
| 700 | 0.30 | 0.7309 | 0.080 | 10 | 99.250 | SVM | 2 |
| 850 | 0.20 | 0.7185 | 0.689 | 45 | 96.625 | SVM | 2 |
| 940 | 0.25 | 0.6593 | 0.875 | 81 | 97.000 | SVM | 2 |

Pour valider l'efficacité de notre méthode, nous avons effectué des tests supplémentaires pour le mode d'identification en ensemble fermé, et les résultats du tableau 3 démontrent également l'efficacité de la bande spectrale de $700nm$. Dans ce cas, le système fonctionne bien (classificateur SVM) avec un taux d'identification (Rank-One Recognition (ROR)) égal à 99,250% et un minimum rang de reconnaissance parfaite (Rank of Perfect Recognition (RPR)) égal à 10.

III.3.2. Etude comparative

Pour prouver la haute performance de notre méthode, nous avons effectué deux comparaisons importantes. La première comparaison vise à démontrer l'efficacité des méthodes d'extraction de caractéristiques profonds par rapport à celles basées sur des méthodes classiques, tandis que la seconde vise à démontrer l'efficacité de notre méthode par rapport à d'autres méthodes (profondes) qui ont été utilisés dans la littérature. Il convient de noter que la comparaison est limitée aux systèmes d'identification biométriques basés sur certaines bandes spectrales uniquement ($850nm$, $940nm$ et *WHITE*) qui fonctionnent en mode d'identification ensemble ouvert et sont bien sûr suffisantes pour donner un aperçu de l'efficacité de notre méthode d'extraction de caractéristiques.

1. Comparaison par des méthodes classiques: Dans cette sous-partie, nous allons essayer d'examiner l'efficacité de la méthode d'extraction de caractéristiques profonds par rapport aux méthodes classiques. En effet, notre méthode est principalement destinée à l'analyse de la texture de l'image, nous avons également choisi deux catégories de méthodes classiques d'analyse de texture. Ces deux méthodes (l'une classique (*hand-crafted method*) et l'autre classique apprise (*learned hand-crafted method*)) sont bien connues dans ce domaine et sont basées sur le modèle binaire local (*Local Binary Pattern (LBP)*). Notez que pour une comparaison équitable, les conditions suivantes doivent être remplies:

- Toutes les méthodes doivent être effectuées avec la même méthodologie de test;
- Toutes les méthodes doivent être effectuées sur la même plateforme d'implémentation ;
- Toutes les méthodes doivent être implémenter sur la même machine.

Nous avons donc pris en compte ces conditions afin de démontrer l'efficacité de la méthode basée sur le deep learning, et ce sous l'aspect le plus important qui concerne la précision de l'identification. Pour limiter les tests, deux méthodes classiques basées sur l'analyse de texture, à savoir le LBP à trois et quatre patches (*Three-Patch LBP (TP-LBP)* et *Four-Patch LBP (FP-*

LBP)) et le descripteur binaire compact (*Compact Binary Descriptor (CBD)*) sont utilisés. Le détail de ces deux méthodes est présenté en **annexe 1**.

Parce que la méthode CBD présente deux paramètres très importants, qui sont la taille du voisinage des pixels (R) et la longueur de la matrice de projection (BIN) et pour améliorer les performances du système d'identification, nous avons testé ces deux paramètres afin d'adapter la méthode CBD à la base de données utilisée. Les résultats des tests obtenus pour les trois bandes spectrales sont illustrés dans le tableau III.3.

Tableau III.3 : sélection des paramètres CBD sous les trois bandes spectrales

| BD | Fenêtre = 15x15 (R = 7) | | | | | | Fenêtre = 19x19 (R = 9) | | | | | |
|------------|-------------------------|-------|----------|-------|--------------|--------------|-------------------------|-------|--------------|--------------|--------------|--------------|
| | BIN = 19 | | BIN = 23 | | BIN = 25 | | BIN = 19 | | BIN = 23 | | BIN = 25 | |
| | T_o | EER | T_o | EER | T_o | EER | T_o | EER | T_o | EER | T_o | EER |
| 850 | 0.614 | 1.298 | 0.658 | 1.196 | 0.684 | 1.018 | 0.619 | 1.307 | 0.617 | 1.375 | 0.654 | 0.994 |
| 940 | 0.705 | 0.485 | 0.655 | 0.678 | 0.674 | 0.750 | 0.659 | 0.500 | 0.700 | 0.409 | 0.653 | 0.625 |
| WHT | 0.712 | 1.157 | 0.698 | 0.875 | 0.726 | 0.250 | 0.688 | 1.000 | 0.736 | 0.921 | 0.711 | 0.875 |

Ainsi, à partir de ce tableau, nous pouvons clairement voir que les deux paramètres CBD peuvent donner de bons résultats dans toutes les bandes spectrales. L'EER le plus élevé ne dépasse pas 1 000% et ce, lors de l'utilisation de la bande spectrale de 850 nm (EER = 0,994% à un seuil, T_o , égal à 0,654). Il est à noter que dans tous ces tests, on modifie également le nombre de régions au sein de l'image (de 1 à 16) et la taille du codebook (de 25 à 500 avec un pas de 25) et les résultats obtenus montrent que 9 régions avec un codebook de taille 100 est suffisant pour permettre de meilleures performances.

Le but de la deuxième série d'expériences est d'évaluer les performances du système lors de l'utilisation d'une méthode d'extraction de caractéristiques classique au lieu des méthodes classiques apprises. Pour cela, nous comparons les performances de la méthode CBD (en utilisant les paramètres optimaux précédemment sélectionnés) avec deux variantes prometteuses de la méthode LBP (TP-LBP et FP-LBP) et les résultats, pour les trois bandes spectrales en mode d'ouverture d'identification, sont indiqués dans le tableau III.4.

Tableau III.4 : Comparaisons de performances de DFT^{MD}Net avec les méthodes classiques

| BD | Identification ensemble ouvert | | | | | | Identification ensemble fermé | | | | | |
|-------|--------------------------------|--------------|--------------|--------------|--------------|--------------|-------------------------------|-----------|---------------|-----------|---------------|-----------|
| | 850 nm | | 940 nm | | WHITE | | 850 nm | | 940 nm | | WHITE | |
| | T_o | EER | T_o | EER | T_o | EER | ROR | RPR | ROR | RPR | ROR | RPR |
| FPLBP | 0.257 | 6.798 | 0.296 | 4.750 | 0.245 | 1.625 | 71.505 | 92 | 81.124 | 89 | 93.500 | 65 |
| TPLBP | 0.265 | 3.250 | 0.248 | 2.936 | 0.206 | 1.125 | 86.875 | 90 | 87.375 | 85 | 94.875 | 69 |
| CBD | 0.654 | 0.994 | 0.700 | 0.406 | 0.726 | 0.250 | 97.000 | 72 | 98.500 | 93 | 98.025 | 44 |
| Notre | 0.719 | 0.689 | 0.659 | 0.875 | 0.697 | 0.125 | 96.625 | 45 | 97.000 | 81 | 98.375 | 13 |

D'après ce tableau, il est clair que le système fonctionne parfaitement en utilisant la méthode classique d'extraction de caractéristiques apprises. En effet, un EER de 0,994% ($T_o = 0,6538$),

0,409% ($T_o = 0,6998$) et 0,250% ($T_o = 0,7260$) peut être obtenu, en utilisant la méthode CBD (les bandes spectrales de 850 nm, 940 nm et *WHITE*). En général, une amélioration d'environ 77,78% et 84,62% a été obtenue en utilisant la méthode CBD au lieu de TP-LBP et FP-LBP, respectivement. Nous avons également examiné le mode d'identification en ensemble fermé en appliquant les paramètres optimaux de CBD obtenus précédemment, puis avons calculé les valeurs du ROR et du RPR. Le tableau 2 montre également que la méthode CBD fonctionne toujours mieux. Par exemple, dans le cas de la bande spectrale *WHITE*, un ROR de 97,000% (RPR = 72) au lieu d'un ROR de 98,500% (RPR = 93) et 98,025% (RPR = 44) pour TP-LBP et FP -LBP , respectivement. Le tableau III.4 compare également notre méthode aux méthodes classiques. En analysant les résultats de ce tableau, on peut dire que:

- Les méthodes basées sur l'apprentissage donnent de meilleurs résultats que les autres méthodes (sans apprentissage);
- En général, le deep learning est efficace par rapport aux autres méthodes, surtout si l'on utilise plusieurs niveaux d'analyse d'images (plusieurs stages de convolution).

Les résultats expérimentaux obtenus dans ce mémoire, à partir d'une base de données multispectrale de 100 personnes, démontrent clairement l'efficacité du principe d'apprentissage dans la tâche d'extraction de caractéristiques, rendant ces types de systèmes adaptés à de nombreuses applications de haute sécurité. Cet effet augmente si la taille de la base d'apprentissage augmente. Il est également important de noter que l'efficacité de l'apprentissage profond devient évidente dans les grandes bases de données.

2. Comparaison par des travaux récents : Dans cette partie et afin de montrer l'efficacité de notre méthode proposée via les autres méthodes (classique et avec apprentissage profond), une étude comparative avec certains travaux de la littérature est menée. Le tableau III.5 résume les principaux travaux utilisant la base de données multispectrale CASIA.

Tableau III.5 : Comparaisons de performances de DFT^{MD}Net avec l'état de l'art

| Méthodes | Année | acronyme | classificateur | Meilleur EER [%] |
|------------------------|---------------------------------|-----------------------|----------------------|------------------|
| Méthodes Classiques | 2010 (J. Chen et al, 2010) | SAX | KNN | 0.9000 |
| | 2011 (Z. Khan et al, 2011) | Contour-code | Hash Table | 0.3000 |
| | 2014 (W. Kang et al, 2014) | RootSIFT | Hierachique matching | 1.0000 |
| | 2016 (N. Charfi et al, 2016) | Sparse | SRC - SVM | 0.2800 |
| | 2016 (Q. Zheng et al, 2016) | DoN | KNN | 0.5300 |
| Méthodes Deep learning | 2016 (Y. Wenet al, 2016) | AlexNet | Hausdroff dist | 0.0803 |
| | 2016 (A. Meraoumia et al, 2016) | PCANet | SVM | 0.1200 |
| | 2016 (J. Svoboda et al, 2016) | Siamese Net | SVM | 1.8600 |
| | 2019 (A. Genovese et al, 2019) | PamNet | KNN | 0.7200 |
| | 2019 (L. Fei et al, 2018) | VGG | Softmax | 3.7800 |
| Notre | / | DFT ^{MD} Net | SVM | 0.0800 |

A partir de ce tableau, qui ne représente que l'identification en mode ensemble ouvert, nous observons que notre méthode proposée réalise les meilleures performances d'identification parmi les autres méthodes répertoriées dans le même tableau. En particulier, l'EER de notre méthode proposée est de 0,0800%, ce qui est en fait le meilleur résultat par rapport à d'autres travaux connexes. Ainsi, il y a vraiment peu de travail sur la base de données multispectrale CASIA, dans laquelle la plupart des techniques de la littérature sont testées uniquement sur la base de données multispectrale PolyU (The Hong Kong Polytechnic University (PolyU) Multispectral Palmprint Database,, s.d.).

Nous pensons que ce travail aboutit à une technique nouvelle et récente qui nous permet d'obtenir les meilleures performances d'identification dans toutes les bandes de la base de données CASIA qui est connue pour avoir un EER généralement supérieur à 0,200% soit par l'utilisation d'une méthode classique ou une méthode basée sur l'apprentissage profond. De plus, notre système d'identification de l'empreinte palmaire peut être amélioré en modifiant le processus d'extraction des caractéristiques, par exemple en augmentant le nombre des stages de DFTMDNet pour efficacement supprimer les caractéristiques inutiles.

Conclusion

Aujourd'hui, les techniques d'apprentissage profond sont largement utilisées dans de nombreuses applications pour fournir un système efficace de reconnaissance de formes. L'objectif principal de ce chapitre est de montrer les performances de la technique d'apprentissage profond dans le développement d'un système biométrique. Le système biométrique proposé est basé sur la modalité biométrique de la paume de la main. Les résultats expérimentaux obtenus, utilisant plusieurs systèmes basés sur le *deep learning*, montrent l'intérêt de ce type d'analyse approfondie pour améliorer la précision du système, notamment pour la grande taille de la base de données. Dans les travaux futurs, nous prévoyons de comparer les méthodes habituelles d'apprentissage profond (par exemple CNN, ICANet et PCANet) pour sélectionner la meilleure qui puisse être utilisée dans le système de reconnaissance de formes.

Conclusion Générale

Conclusion générale

Dans tout système de reconnaissance de formes, l'étape d'extraction de caractéristiques est l'une des étapes les plus importantes du système en raison de son impact significatif et direct sur la précision du système. En effet, les techniques d'extraction de caractéristiques sont des processus de transformation d'images brutes en informations plus significatives et utiles qui peuvent être utilisées dans une application de vision par ordinateur. Pour cette raison, ces processus ont été soigneusement conçus par des spécialistes (*hand-crafted methods*) utilisant leurs connaissances et leur expertise. Habituellement, ces techniques sont appliquées à l'image sans connaissance préalable de son contenu. Ainsi, pour améliorer leurs performances, ces techniques ont été dotées d'une connaissance préalable de leur contexte grâce au apprentissage machine (*learned hand-crafted methods*). Ces dernières années, avec l'avènement de l'analyse approfondie, les méthodes basées sur l'apprentissage profond convolutif font elles-mêmes automatiquement le travail de l'expert dans les couches convolutives.

L'objectif principal de cette étude est de montrer l'importance des principes d'apprentissage profond en développant une nouvelle méthode d'extraction de caractéristique basée sur le principe d'apprentissage profond convolutif. La méthode proposée est intégrée dans un système d'identification de l'empreinte palmaire multi-spectrale. Les résultats expérimentaux obtenus, utilisant plusieurs bandes spectrales de la base de données CASIA, montrent l'intérêt de ce principe pour améliorer la précision du système biométrique et ont également montré l'intérêt de l'analyse hiérarchique qui permet généralement d'obtenir de meilleures performances. Pour cela, nous prévoyons dans nos futurs travaux d'exploiter des caractéristiques plus hiérarchiques dans le cadre de l'apprentissage profond pour améliorer encore les performances de reconnaissance.

Glossaire

Les termes suivants, classés dans l'ordre alphabétique, sont utilisés dans le texte.

| | |
|-------------------------|---|
| ADN | : Acide Désoxyribo-Nucléique |
| ANN | : Réseau neuronal artificiel - Artificial Neural Network |
| CA-LBFL | : Apprentissage des fonctionnalités binaires locales contextuelles - Context-Aware Local Binary Feature Learning |
| CASIA | : Académie chinoise des sciences, Institut d'automatisation - Chinese Academy of Sciences Institute of Automation |
| CBD | : Descripteurs binaires compactés - Compact Binary Descriptor |
| CCD | : Dispositif à Transfert de Charges - Charge Coupled Device. |
| CNN | : Réseau Neuronal Convolutif – Convolutional Neural Network. |
| DCT | : Transformée en Cosinus Discrète - Discret Cosine Transform. |
| DCTNet | : Réseau DCT - DCT Network |
| DFD | : Descripteur de Visage Discriminant - Discriminant Face Descriptor |
| DFT | : Transformée de Fourier Discrète - Discrete Fourier Transform |
| DFT^{MD} | : Module de la Transformée de Fourier Discrète |
| DFT^{PH} | : Phase de la Transformée de Fourier Discrète |
| DST | : Transformée en Sinus Discrète - Discret Sinus Transform |
| DSTNet | : Réseau DST – DST Network |
| DWT | : Transformée en Ondelettes Discrète - Discrete Wavelet Transform |
| ECG | : Electro-Cardio-Gramme |
| EEG | : Electro-Encéphalo-Gramme |
| EER | : Taux d'Erreurs Egales - Equal Error Rate. |
| FAR | : Taux de Fausses Acceptations - False Acceptance Rate. |
| FP-LBP | : Motifs Locaux Binaires Quatre Patchs - Four-Patch Local Binary Pattern |
| FRR | : Taux de Faux Rejets - False Reject Rate. |
| GABOR | : Dennis Gabor |
| GAR | : Taux d'acceptation des clients - Genuine Acceptance Rate. |

| | |
|-----------------------|---|
| HAUSDROFF DIST | : Distance de Hausdroff - Hausdroff Distance |
| HOG | : Histogramme de Gradient Orienté - Histogram of Oriented Gradient |
| ICA | : Analyse en Composantes Indépendantes - Independent Component Analysis |
| ICANet | : Réseau ICA – ICA Network |
| IRM | : Imagerie par Résonance Magnétique - Magnetic Resonance Imaging |
| KNN | : K-plus Proches Voisins - K-Nearest Neighbours |
| LBP | : Motifs Locaux Binaires - Local Binary Pattern. |
| LDA | : Analyse Discriminante Linéaire – Linear Discrimination Analysis. |
| LQP | : Motifs Quantifiés Locaux - Local Quantized Pattern. |
| MD | : Module |
| PCA | : Analyse en Composantes Principales- Principal Components Analysis. |
| PCANet | : Réseau PCA – PCA Network. |
| PDV | : Pixel Difference Vector |
| PH | : Phase |
| Poly U | : Université Polytechnique Hong Kong – H.K Polytechnic University |
| RBF | : Fonctions de Base Radiale - Radial Basis Function |
| RdF | : Reconnaissance de Formes – Pattern Recognition |
| RFT | : Forêt Aléatoire - Random Forest Transform |
| ROI | : Région d'intérêt - Region Of Interest. |
| ROOT SIFT | : Racine de Transformation d'entité à échelle invariante - Root Scale Invariant Feature Transform |
| ROR | : Taux d'Identification - Rank-One Recognition |
| RPR | : Rang de Reconnaissance Parfaite - Rank of Perfect Recognition. |
| SAX | : Approche symbolique des agrégats -Symbolic Aggregate approxImation |
| Softmax | : Fonction Exponentielle Normalisée - Fonction Softmax |
| SRC | : Classificateur basé sur une Représentation Sparse - Sparse Representation based Classificateur |
| SVM | : Machines à Vecteurs de Support - Support Vector Machine. |
| TP-LBP | : Motifs Locaux Binaires Trois Patches - Three-Patch Local Binary Pattern |
| VGG | : Groupe de géométrie visuelle. Univ Oxford - Visual Geometry Group at Oxford University. |

Annexe 1

Méthodes classiques d'extraction de caractéristique

Comme mentionné précédemment, la précision des systèmes de reconnaissance de formes dépend fortement des méthodes d'extraction de caractéristiques, de sorte que toute complaisance dans le choix de la méthode appropriée conduira nécessairement à de très mauvais résultats. En fait, ces technologies présentent des avantages et des inconvénients. Par exemple, les performances des méthodes classiques sont faibles par rapport aux méthodes modernes qui dépendent du *deep learning*, mais ceci du point de vue de la précision de la classification. Du point de vue de la vitesse d'exécution, les méthodes modernes ont des performances médiocres par rapport aux méthodes classiques. Bien entendu, depuis plusieurs années, les chercheurs proposent de nombreuses approches, dont plusieurs entre eux utilisent le principe de l'apprentissage. Dans cette annexe, nous tenterons d'expliquer les deux méthodes classiques que nous avons utilisées dans le processus de comparaison dans le troisième chapitre.

A.1. Motifs locaux binaires

Le modèle binaire local (LBP) est la méthode la plus adoptée et la plus largement utilisée dans les systèmes de reconnaissance de formes. Certes, en raison de cet effet, de nombreux chercheurs ont tenté d'améliorer cette méthode, résultant en de nombreuses variantes de la LBP, dans lesquelles LBP à trois patchs (Three-Patch LBP (TP-LBP)) et LBP à quatre patchs (Four-Patch LBP (FP-LBP)) sont deux extensions efficaces de la LBP.

A.1.1. TP-LBP : Dans la méthode TP-LBP (voir Fig. A.1), le code pour le pixel de chaque image est obtenu en comparant les deux patchs avec le patch central de taille $w \times w$ autour du

pixel. Le code attribué à chaque pixel est de S bits. L'équation suivante donne l'expression de TP-LBP (Ameni Yengui Jammoussi et Dorra Sellami Masmoudi, 2011):

$$\text{TP-LBP}_{r,s,w,\alpha}(p) = \sum_i^S f(D_S)2^i \quad (\text{A.1})$$

$$D_S = d(C_i, C_p) - d(C_{i+\alpha \bmod S}, C_p) \quad (\text{A.2})$$

Où C_p est le patch central, C_i et $C_{i+\alpha \bmod S}$ sont deux patches le long de l'anneau de rayon r , $d(\cdot, \cdot)$ la distance entre deux patches et f est définie comme $f(x) = 1$ si $x > \tau$, sinon 0. La valeur de τ est fixée légèrement supérieure à zéro.

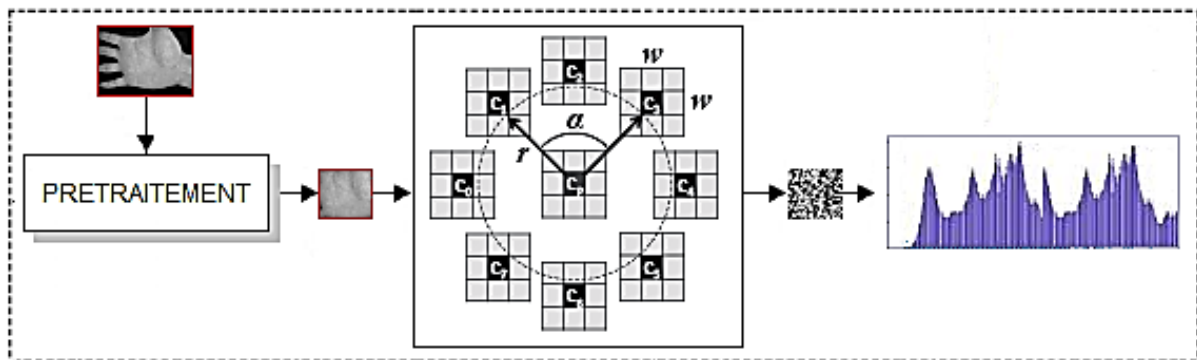


Fig. A.1 : Extraction de caractéristiques utilisant TP-LBP

Une fois codée, une signature d'image est produite, dans laquelle l'image est divisée en une grille de régions sans chevauchement et un histogramme mesurant la fréquence de chaque code binaire est calculé pour chacune de ces régions. Chacun de ces histogrammes est normalisé à la longueur unitaire. Une image est représentée par ces histogrammes concaténés en un seul vecteur.

A.1.2. FP-LBP : Dans le FP-LBP (voir Fig. A.2), deux anneaux de r_1 et r_2 sont considérés et centrés sur le pixel. Le code FP-LBP est obtenu en comparant deux patches symétriques centraux dans l'anneau intérieur avec deux patches symétriques centraux dans l'anneau extérieur positionnés α patches le long du cercle (Chouchane Ammar et al, 2016).

$$\text{FP-LBP}_{r_1,r_2,s,w,\alpha}(p) = \sum_i^{S/2} f(D_S)2^i \quad (\text{A.3})$$

$$D_S = d(C_{1,i}, C_{2,i+\alpha \bmod S}) - d(C_{1,i+S/2}, C_{2,i+S/2+\alpha \bmod S}) \quad (\text{A.4})$$

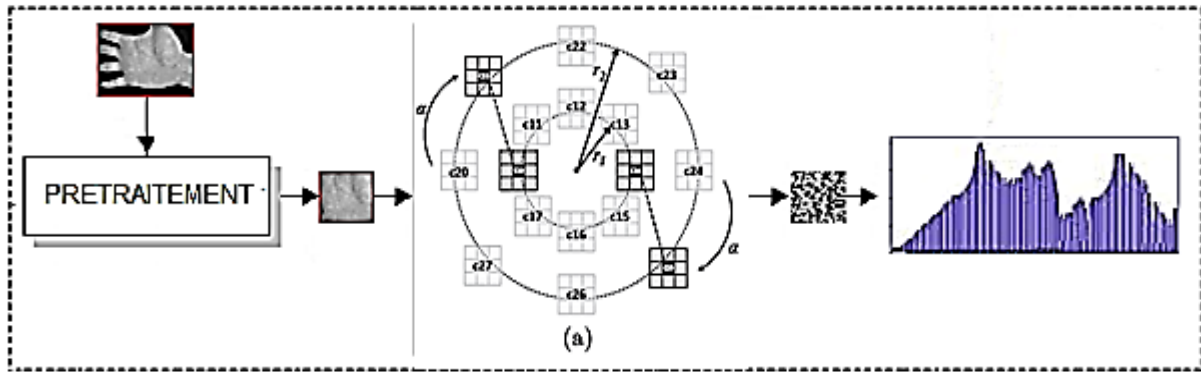


Fig. A.2 : Extraction de caractéristiques utilisant FP-LBP

Comme dans TP-LBP, la réponse résultante, dans FP-LBP, est ensuite divisée en régions non chevauchantes et un histogramme est calculé pour chaque région. Les histogrammes sont normalisés au rang unitaire et concaténés en un seul vecteur pour représenter les caractéristiques de l'image entière.

A.2. Descripteurs binaires compactés (CBD)

Comme déjà mentionné, compte tenu de l'intérêt croissant porté à l'utilisation des images dans les applications de vision par ordinateur, plusieurs méthodes sont utilisées pour analyser et donc comprendre le contenu des images. Ainsi, étant donné les deux avantages de la méthode d'analyse d'images basée sur la LBP (vitesse et faible occupation de la mémoire), *Jiwen Lu et al.* dans (*Jiwen Lu et al, 2015*), ils ont tenté d'améliorer les performances de cette technique en incluant une phase d'apprentissage pour éliminer ses faiblesses. Les phases d'apprentissage et d'extraction des caractéristiques sont résumées comme suit:

A.2.1. Phase d'apprentissage : Cette méthode repose sur la robustesse des codes binaires vis-à-vis des changements locaux de texture d'image. Les codes binaires compacts sont appris directement à partir des pixels bruts pour la représentation de l'image. En effet, pour garantir de meilleurs résultats de classification, les caractéristiques CBD doivent être construites à partir d'un ensemble d'images fournies à partir du même contexte. Pour mieux comprendre la phase d'apprentissage (formation de la matrice de projection (caractéristique Cbfd) et du dictionnaire (*codebook*)), la Fig. A.3 illustre le schéma de principe de la phase d'apprentissage.

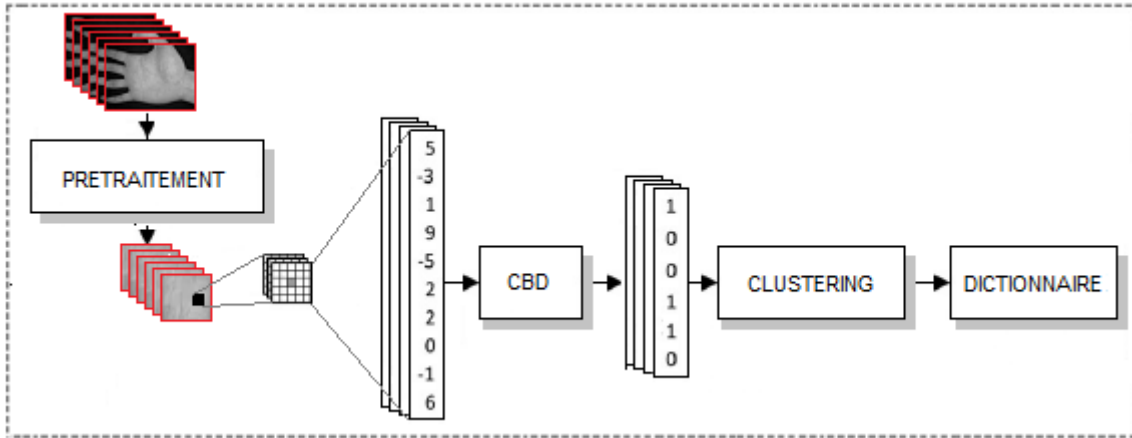


Fig. A.3 : Formation de la matrice de projection et du dictionnaire pour l'encodage des caractéristiques

1. Apprentissage de CBD : Les vecteurs d'apprentissage sont obtenus en utilisant la relation de chaque pixel avec leurs quartiers. En effet, l'image est analysée par une fenêtre rectangulaire de taille $(2R + 1) \times (2R + 1)$, où R est un entier positif, centré sur chaque pixel. Soit $X = [x_1, x_2, \dots, x_N]$ l'ensemble des vecteurs d'apprentissage (appelé Pixel Difference Vector (PDV)) obtenu en mesurant la différence entre le pixel central et les pixels voisins dans la fenêtre prédéfinie. La taille de chaque vecteur est égale à $(2R + 1) \cdot (2R + 1) - 1$ ou le PDV entre le pixel central et lui-même ($PDV_0 = 0$) est exclu. La caractéristique CBD (matrice de projection) vise à apprendre K fonctions de hachage $(\{wk\}_{k=1..K})$ afin de quantifier chaque vecteur $x_n|_{n=1..N}$ en un vecteur binaire $b_n = [b_{n1}, b_{n2}, \dots, b_{nk}]$:

$$b_n = 0.5 \times (\text{sgn}(w_k^T x_n) + 1) \quad (\text{A.5})$$

Où $\text{sgn}(v) = 1$ si $v \geq 0$ et -1 sinon. Nous devons donc d'abord construire la matrice de projection W qui comprend toutes les fonctions de hachage w_k :

$$W = [w_1, w_2, \dots, w_k] \quad (\text{A.6})$$

La matrice W est initialisée avec les K premiers vecteurs propres de la matrice de covariance ($C = XX^T$). Ensuite, une tâche d'optimisation est appliquée pour minimiser la fonction d'objectif $J(w_k)$, qui est définie comme suit:

$$\min J(w_k) = J_1(w_k) + \lambda_1 J_2(w_k) + \lambda_2 J_3(w_k) \quad (\text{A.7})$$

Les deux paramètres prédéfinis λ_1 et λ_2 sont choisis pour équilibrer les effets de différents termes. Les termes J_1 , J_2 et J_3 sont choisis pour garantir que: *i*) la variance des codes binaires appris est maximisée, *ii*) la perte de quantification entre la caractéristique originale et les codes binaires codés est minimisée et *iii*) les bins de caractéristiques dans les codes binaire appris se répartissent autant que possible. Après la convergence de l'algorithme d'optimisation, on obtient

une matrice de projection W (caractéristiques CBD) contenant toutes les fonctions de hachage et permettant de transformer un vecteur PDV en un vecteur binaire de taille K bits.

2. Apprentissage du dictionnaire : Le but du livre de codes est de réduire le nombre de vecteurs binaires trouvés pour chaque image. En effet, tous les vecteurs d'apprentissage (PDV) sont projetés dans la matrice W , puis l'algorithme de clustering *k-means* est appliqué à tous les vecteurs binaires obtenus. Enfin, le centroïde des classes obtenues représente le dictionnaire (*codebook*).

A.2.2. Phase d'extraction de caractéristiques :

La tâche d'extraction de caractéristiques (voir la Fig. A.4) est basée sur la matrice de projection W (caractéristique CBD) et le dictionnaire (*codebook*) obtenu pendant la phase d'apprentissage. Cependant, après la formulation de tous les vecteurs PDV ($X = \{x_n\}_{n=1..N}$) de l'image, leurs correspondances binaires sont réalisées par leur projection dans W :

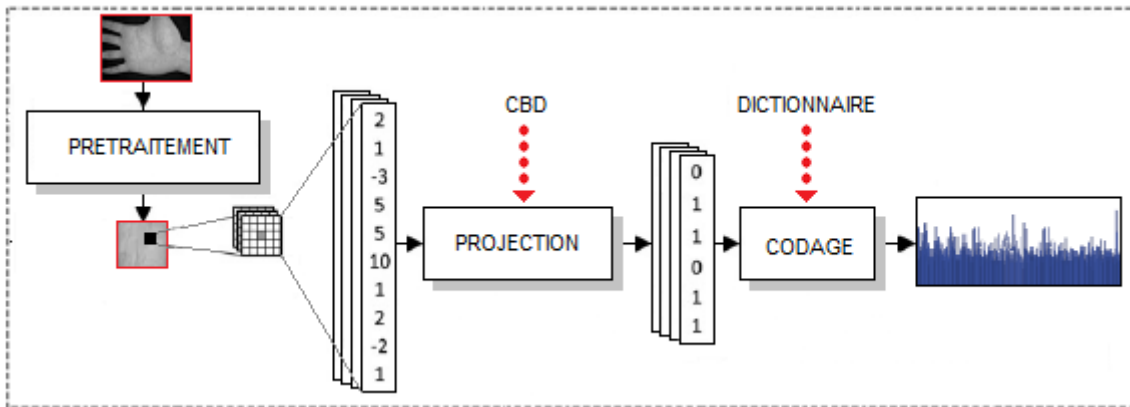


Fig. A.4 : Extraction de caractéristiques utilisant CBD

$$V_n = 0.5 \times (\text{sgn}(W^T X) + 1) \quad (\text{A.8})$$

Ensuite, chaque vecteur binaire est remplacé par la coordonnée vectorielle la plus proche dans le dictionnaire. Enfin, en utilisant les différentes coordonnées, un histogramme est formé pour représenter l'ensemble de l'image. Pour trouver des vecteurs de caractéristiques discriminantes, l'image brute est segmentée en plusieurs régions, et chaque région est analysée comme une image (chaque région a ses propres caractéristiques CBD (W) et son dictionnaire). Pour chaque région, un histogramme (\mathcal{H}_s) est créé, et enfin, la concaténation de tous les histogrammes permet d'obtenir un vecteur précis pour représenter l'image.

$$\mathcal{V} = [\mathcal{H}_1, \mathcal{H}_2, \dots, \mathcal{H}_M] \quad (\text{A.9})$$

Où M désigne le nombre de régions.

Bibliographie

1. (s.d.). Récupéré sur CASIA Palmprint Database: www.cbsr.ia.ac.cn/english/PalmprintDatabases.asp.
2. (2020, 06 15). Récupéré sur Geeks for geeks: <https://www.geeksforgeeks.org/pattern-recognition-introduction>
3. A. Genovese et al. (2019). Palmnet: Gaborpca convolutional networks for touchless palmprint recognition. *Transactions on Information Forensics and Security*. IEEE .
4. A. Meraoumia et al, ., 2.–6. (2016). Improving biometric identification performance using pcanet deep learning and multispectral palmprint. *Springer, Biometric Security and Privacy*, 51-69.
5. Abhilash Sharma et al. (2015). Biometric System- A Review. (*IJCSIT*) *International Journal of Computer Science and Information Technologies*, 4616-4619.
6. Alaa Tharwat. (2018). Independent Component Analysis: an Introduction. *Applied Computing and Informatics*, (p. doi.org/10.1016/j.aci.2018.08.006).
7. Alice Ahlem Othmani. (2015). *Tree species identification from T-LiDAR data*. An overview Chapter in *Computer Vision and Pattern Recognition in Environmental Informatics*.
8. Ameni Yengui Jammoussi et Dorra Sellami Masmoudi. (2011). Joint Integral Histogram based Adaboost for Face Detection System. *International Journal of Computer Applications*.
9. Belahcen, M. (2013). *Authentication et Identification en Biométrie*. Biskra: Doctoral thesis, Université Mohamed Khider .
10. Brian S. Everitt, G. D. (2013). Discrimination, Classification and Pattern Recognition. *Applied Multivariate Data Analysis*, 248-270.
11. C. J. Nget al. (2016). DCT based region log-tiedrank covariance matrices for face recognition, doi: 10.1109/ICASSP.2. *International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 2099-2103). Shanghai: IEEE .
12. C. Jia et al. (2019). Content-Aware Convolutional Neural Network for In-Loop Filtering in High Efficiency Video Coding. *Transactions on Image Processing*, vol. 28, no. 7, doi: 10.1109/TIP.2019.2896489 (pp. 3343-3356). IEEE .

13. C. Lai and C. Tsai. (2017). Digital Image Watermarking Using Discrete Wavelet Transform and Singular Value Decomposition. Dans *Transactions on Instrumentation and Measurement* (pp. 3060-3063). IEEE .
14. Chakladar D, C. S. (2019). Feature Extraction and Classification in Brain-Computer Interfacing. *Future Research Issues and Challenges*. In: Li X., Wong KC. (eds) *Natural Computing for Unsupervised Learning and Semi-Supervised Learn.*
15. Chelaoua R et Meraoumia A. (2019). Fusion of finger-knuckle-print and finger vein recognition using random forest tree. *International Conference of Mathematical Sciences (ICMS 2019)*, s. 156.
16. Chouchane Ammar et al. (2016). Evaluation of Histograms Local Features and Dimensionality Reduction for 3D Face Verification. *Journal of Information Processing Systems*.
17. Chuang S. (2018). Vein recognition based on minutiae features in the dorsal venous network of the hand. <https://doi.org/10.1007/s11760-017-1195-3> (pp. 573–581). SIViP 12.
18. D. Palma et al. (2019). Biometric Palmprint Verification: A Dynamical System Approach. *Transactions on Systems, Man, and Cybernetics Systems* (pp. 2676-2687). doi: 10.1109/TSMC.2017.277123; IEEE .
19. David Bařina. (2011). Gabor Wavelets in Image Processing. *Proceedings of the 17th Conference STUDENT*. EEICT .
20. Dvořák M et Draňanský M. (2019). 3D Hand Geometry Recognition, https://doi.org/10.1007/978-3-319-98734-7_4. Springer, Obaidat M., Traore I., Woungang I. (eds) *Biometric-Based Physical and Cybersecurity Systems*.
21. Edwin Hancock. (2020). Pattern recognition. *journal Elsevier*, volume 106, ISSN: 0031-3203.
22. Georgy Gimel'farb et al. (2012). Structural, Syntactic, and Statistical Pattern Recognition. Hiroshima, Japan: Joint IAPR International Workshop, SSPR & SPR.
23. Ghahramani Z. (2014). *Unsupervised Learning*. In: Bousquet O., von Luxburg U., Rätsch G. (eds) *Advanced Lectures on Machine Learning. ML. Lecture Notes in Computer Science, vol 3176*. https://doi.org/10.1007/978-3-540-28650-9_5. Berlin, Heidelberg: Springer.
24. Gilbert Strang, (. (1999). The Discrete Cosine Transform. *SIAM Rev*, 135-147.
25. Gudivada, B. I. (2020). Applied Computer Vision and Image Processing. *Proceedings of ICCET 2020, Volume 1, Springer*.

26. H. Setiawan et E. M. Yuniarno. (2017). Biometric Recognition Based on Palm Vein Image Using Learning Vector Quantization. *5th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering*. ICICI-BM.
27. Hamouda, A. (2019). *Extraction des caractéristiques pour la classification des images médicales par la DCT*. Mémoire de Master en Informatique, université de Guelma.
28. Henniger O et al. (2017). *Opportunities for Biometric Technologies in Smart Environments*. https://doi.org/10.1007/978-3-319-61110-9_13: Springer, Braun A., Wichert R., Maña A. (eds) Ambient Intelligence. AmI Lecture Notes in Computer Science.
29. J. Chen et al. (2010). Palmprint authentication using a symbolic representation of images. *Image and Vision Computing*, vol. 28, no. 3, 343–351.
30. J. Galbally et al. (2019). A Study of Age and Ageing in Fingerprint Biometrics. *IEEE Transactions on Information Forensics and Security*, [https://doi: 10.1109/TIFS.2018.2878160.](https://doi.org/10.1109/TIFS.2018.2878160), 1351-1365.
31. J. Svoboda et al. (2016). Palmprint recognition via discriminative index learning. *23rd International Conference on Pattern Recognition (ICPR)* (pp. 4232–4237). IEEE.
32. Jake Frankenfield. (2020, 06 13). Récupéré sur Artificial Neural Network (ANN): <https://www.investopedia.com/terms/a/artificial-neural-networks-ann.asp>
33. Jiwen Lu et al. (2015). Learning Compact Binary Face Descriptor for Face Recognition. *Transactions on Pattern Analysis and Machine Intelligence* (pp. 2041 – 2056). IEEE .
34. Joel Harp. (2009). A test of a pattern recognition system for identification of spiders. *Bulletin of entomological research*, 217-224.
35. Jonathon Shlens. (2014). A Tutorial on Principal Component Analysis. *arXiv:1404.1100v1*.
36. Jose G. Moreno-Torres et al. (2009). On the Homogenization of Data from Two Laboratories Using Genetic Programming. *International Workshop on Learning Classifier Systems*, 185-197.
37. Khodaskar, A., & Ladhake, S. A. (2014). Pattern recognition: Advanced development, techniques and application for image retrieval. *International Conference on Communication and Network Technologies*. Sivakasi, India: IEEE .
38. L. Fei et al. (2018). Feature extraction methods for palmprint recognition. *A survey and evaluation,*” *Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 2 (pp. 346–363). IEEE .
39. Li Deng, D. Y. (2019). Deep Learning: Methods and Applications. *Foundations and Trends in Signal Processing*, 197-387.

40. Loren Peter Lugosch. (2018). *Learning Algorithms for Error Correction*. University Montreal, Quebec, Canada.: Department of Electrical and Computer Engineering McGill.
41. Mehmet Umut Demircin et al. (2011). *Region Of Interest (Roi) Video encoding*. Pub. No.: US 2011/0235706 A1: Patent Application Publication Demircin et al.
42. N. Charfi et al. (2016). Local invariant representation for multi-instance toucheless palmprint identification. *International Conference on Systems, Man, and Cybernetics (SMC)*. . IEEE.
43. Olson T. (2017). The Discrete Fourier Transform. In: *Applied Fourier Analysis*. https://doi.org/10.1007/978-1-4939-7393-4_3. New York: Birkhäuser.
44. Onur C. Hamsici et al. (2016). Efficient Descriptor Extraction Over Multiple Levels Of An Image Scale Space. *United States Patent Hamsici et al.*, Patent No : USO0953 0073B2.
45. Pdraig Cunningham et Sarah Jane Delany. (2020). k-Nearest Neighbour Classificateurs: 2nd Edition (with Python examples. *arXiv:2004.04523*.
46. Philippe Bolon, e. a. (2017). Analyse d'images : Filtrage et segmentation. *Ouvrage collectif publié avec le soutien du Groupement de Recherche 134 : Traitement Du Signal et Images (CNRS)*, 277 – 283.
47. Pulkit Madaan et Abhishek Maiti. (2019). *Deep mean shift clustering*. New Delhi: BTP report submitted in partial fulfillment of the requirements for the Degree of B.Tech. in Computer Science & Applied Mathematics and Computer Science & Engineering respectively, Indraprastha Institute of Information Technology.
48. Q. Zheng et al. (2016). A 3d feature descriptor recovered from a single 2d palmprint image. *transactions on pattern analysis and machine intelligence, vol. 38, no. 6* (pp. 1272–1279). IEEE .
49. Ricco Rakotomalala. (2017). *Analyse de corrélation, Étude des dépendances-Variables quantitatives*. Lyon 2: Université Lumière.
50. Shifei Ding et al. (2011). A survey on feature extraction for pattern recognition. *DOI 10.1007/s10462-011-9225-y*. Springer Science+Business Media.
51. Shiny Chandra Bai et al. (2017). A biometric recognition system for human identification using finger vein patterns. *International Journal of Emerging Trends in Engineering and Development Issue 7*.
52. Srinidhi, C. L. (2019). Pattern Recognition and Machine Learning Framework for Automated Analysis of Retinal Images. *National Institute of Technology Karnataka, Surathkal*.

53. Taha J. Alhindi et al. (2018). Comparing LBP, HOG and Deep Features for Classification of Histopathology Images. *World Congress on Computational Intelligence (IEEE WCCI)*, Rio de Janeiro, Brazil: IEEE .
54. TAYEBI Abderrazeq et BOMEDIANA Touhami. (2018). *Utilisation des caractéristiques statistiques et les Moments de Zernike pour la Reconnaissance des Lettres Arabes Manuscripts*. ADRAR: Université Ahmed Draia.
55. *The Hong Kong Polytechnic University (PolyU) Multispectral Palmprint Database*,. (s.d.). Récupéré sur <https://www4.comp.polyu.edu.hk/~biometrics/MultispectralPalmprint/MSP.htm>.
56. Todd Rearick et al. (2019). Noise Suppression Methods And Apparatus. *10,488,482 B2*. USA: United States Patent Rearick et al.
57. U Raghavendra. (2018). Deep convolution neural network for accurate diagnosis of glaucoma using digital fundus images. *Information Sciences Volume 441*, doi.org/10.1016/j.ins.2018.01.051., 41-49.
58. V. Khryashchev et al. (2015). Face Recognition Using Local Quantized Patterns And Gabor Filters. *doi:10.5194/isprsarchives-XL-5-W6-59*.
59. V. Stimper et al. (2019). Multidimensional Contrast Limited Adaptive Histogram Equalization. (p. doi: 10.1109/ACCESS.2019.2952899). IEEE Access.
60. Vincent Dumoulin, F. V. (2018). *A guide to convolution arithmetic for deep learning*. arXiv:1603.07285v2, Version 2.
61. W. Kang et al. (2014). Contact-free palm-vein recognition based on local invariant features. *PloS one*, vol. 9, no. 5, p. e97548.
62. Wu Chou, W. (2010). Discriminant-function-based minimum recognition error rate pattern-recognition approach to speech recognition. *Proceedings of the IEEE*.
63. Y. Duan et al. (2018). Context-Aware Local Binary Feature Learning for Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 5, [doi: 10.1109/TPAMI.2017.2710183](https://doi.org/10.1109/TPAMI.2017.2710183)., 1139-1153.
64. Y. Wen et al. (2016). A discriminative feature learning approach for deep face recognition. *European conference on computer vision* (pp. 499–515). Springer.
65. Z. Guo et al. (2010). A Completed Modeling of Local Binary Pattern Operator for Texture Classification. *Transactions on Image Processing* (pp. 1657-1663). IEEE.
66. Z. Khan et al. (2011). Contour code: Robust and efficient multispectral palmprint encoding for human recognition. *International Conference on Computer Vision* (pp. 1935–1942). IEEE.

67. Z. Lei et al. (2014). Learning Discriminant Face Descriptor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 2, doi: 10.1109/TPAMI.2013.112, 289-302.
68. Zhang D et al. (2016). *Discriminative Learning in Biometrics*, in: *Discriminative Learning in Biometrics*. Singapore: Springer, https://doi.org/10.1007/978-981-10-2056-8_1.
69. Zhang D et al. (2018). *Advanced Biometrics*. Springer, Finger-Knuckle-Print Verification, https://doi.org/10.1007/978-3-319-61545-5_5.
70. Zhang XD. (2020). Support Vector Machines. *A Matrix Algebra Approach to Artificial Intelligence* (pp. https://doi.org/10.1007/978-981-15-2770-8_8). Singapore: Springer.
71. Zhi-Hua Zhou. (2018). A brief introduction to weakly supervised learning. *National Science Review*5: doi: 10.1093/nsr/nwx106, 44–53.
72. Zulfiqar Ali et al. (2017). Performance Analysis of Statistical Pattern Recognition Methods in KEEL. Marseille, France.: International Conference on Knowledge Based and Intelligent Information and Engineering Systems .
73. Zuzana Majdisova et Vaclav Skala. (2017). Radial basis function approximations: comparison and applications. *Applied Mathematical Modelling Volume 51*, doi.org/10.1016/j.apm.2017.07.033, 728-743.