République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université de Médéa
Faculté de Technologie
Département du Génie Electrique

THESE

En vue de l'obtention du diplôme de

Doctorat en Sciences
en Génie Électrique

THEME :

AUTHENTIFICATION PAR RECONNAISSANCE BIOMETRIQUE DES VEINES DES DOIGTS

Présenté par

BOUCHERIT ISMAIL

Le : 31/10/2020

Devant le jury composée de :

M. HOUASSINE Hamza	MCA. Université de Bouira	Président
M. GUESSOUM Abderrezak	Prof. Université de Blida	Examinateur
M. BOUHEDDA Mounir	MCA. Université de Médéa	Examinateur
M. OULD ZMIRLI Mohamed	Prof. Université de Médéa	Directeur de Thèse

Année Universitaire : 2019/2020

Remerciements

Tout d'abord, je remercie le bon dieu, tout puissant, de m'avoir donné la volonté d'entamer ce travail, la force et la patience pour le réaliser et le courage pour l'achever.

Je tiens à exprimer toute ma gratitude à mon directeur de thèse, le Professeur Mohamed Ould Zmirli pour l'aide compétente qu'il m'a apportée, pour sa patience et son encouragement. Sa disponibilité, son attention et sa compétence scientifique m'ont été précieux pour réaliser mon travail, le structurer et l'améliorer.

Je tiens également à remercier tous les membres du Jury : M.Hamza Houassine, M.Abderrezak Guessoum, M.Mounir Bouhedda, pour leur lecture et leurs commentaires très intéressants sur mon travail.

Je tiens à exprimer ma gratitude à l'ensemble des personnes qui ont contribué de près ou de loin à la réalisation de ce travail, surtout M.Hamza Hantabli.

Dédicace

Je remercie le dieu le tout puissant de m'avoir motivé à Réaliser ce modeste travail.

Je dédié ce travail à :

Mes très chers parents qui m'ont aidé durant les moments Les plus pénibles de ce long chemin, la Miséricorde d'Allah sur eux.

Merci mes parents

À ma famille

A mes collègues de l'université de MEDEA

ملخص

نماذج الوريد المستخدمة مؤخرًا هي تقنية بيومترية جذابة تعتمد على نماذج الوريد البشري . أظهرت العديد من الدراسات أن صور الأوردة في الأصابع تكون رديئة الجودة بسبب التباين في الأنسجة والإضاءة غير المتكافئة. لذلك . في هذه الدراسة ، اقترحنا تقنية جديدة تعمل على تحسين جودة صورة عروق الأصابع. تتضمن هذه الطريقة تحسين التباين واستخدام مرشحات Gabor

حاليًا ، حلت الأساليب الجديدة للتعرف على عروق الإصبع على أساس التعلم الآلي محل الأساليب التقليدية. في هذه الأطروحة ، استخدمنا بنية جديدة تعتمد على اندماج مخرجات شبكتي CNN تطابقتين ولكن مع مدخلين من نفس الصورة لأوردة الأصابع التي تمت معالجتها بتقنيتين مختلفتين

الكلمات المفتاحية الشبكة العصبية التلافيفية ، التعرف على الوريد القياسات الحيوية ، تحسين الصورة

Résumé

Les modèles de veines utilisées récemment est une technique biométrique attirante fondée sur les modèles de veines humaines. De nombreuses études ont montré que les images des veines des doigts étaient de mauvaise qualité en raison de la variation des tissus et de l'illumination inégale. Dans cette étude, nous avons proposé une nouvelle technique, qui améliore la qualité d'image des veines des doigts. Cette méthode comprend l'amélioration du contraste, et l'utilisation de filtres de Gabor.

Actuellement, de nouvelles méthodes de reconnaissance par les veines des doigts basées sur l'apprentissage automatique ont remplacé les méthodes conventionnelles. Dans cette thèse, nous avons utilisé une nouvelle architecture basée sur la fusion des sorties de deux CNN identiques mais avec deux entrées d’une même image des veines e doigt traité par deux techniques différentes.

Mots-clés : biométrie, amélioration de l'image, identification par veines des doigts, réseau neuronal convolutionnel.

Summary

The recently used vein models is an attractive biometric technique based on human vein models. Numerous studies have shown that the images of the veins of the fingers are of poor quality due to the variation in tissue and uneven illumination. In this study, we proposed a new technique, which improves the image quality of the finger veins. This method includes improving contrast, and use of Gabor filter.

Currently, new finger vein recognition methods based on machine learning have replaced conventional methods. In this thesis, we used a new architecture based on the fusion of the outputs of two identical CNNs but with two inputs of the same image vein of finger treated by two different techniques.

Keywords : biometrics, Image Enhancement, finger vein identification, convolutional neural network.

Table des matières

Introduction générale 1

1 Contexte 1

2 Vérification et identification 2

3 Objectifs principaux des travaux 2

4 Contributions 3

5 Organisation du manuscrit 4

Chapitre 1 5

La biométrie 5

1 Introduction 5

2 Description de quelques modalités biométriques 7

2.1 L'empreinte digitale 7
2.2 La biométrie visuelle 7
2.3 La reconnaissance vocale 9
2.4 L'ADN 9
2.5 La biométrie de la main 9
2.6 Biométrie des veines 10
2.7 La biométrie dentaire 10

3 Les systèmes biométriques 11

4 Evaluation des performances des Systèmes biométriques 13

5 Fiabilité des systèmes biométriques 13

5.1 Test de vérification 14
5.2 Test d'identification 15

6 Les bases d'apprentissage et les bases de test 17

7 Les avantages et les limites de la biométrie 17

8 Conclusion 18

Chapitre 2 19

la biométrie des veines des doigts 19

1 Introduction 19

2 Les limites des systèmes biométriques 19

3 La biométrie des veines 21

4 Acquisition de données 22

4.1 Capteurs et systèmes commerciaux 23
4.2 Nouvelle génération de scanner veineux 24

4.3 Les conditions d'acquisition des images des veines des doigts 24
4.4 Durabilité et anti-vandalisme 24

5 Les bases de données des veines de doigt 25

6 Les systèmes de reconnaissance classiques 27
6.1 Prétraitement 27
6.2 Extraction des caractéristiques 32
6.3 L'appariement des caractéristiques 39

7 Conclusion 42

Chapitre 3 43

Contribution à l'Amélioration de la qualité des images des veines de doigt par fusion DCT 43

1 Introduction 43

2 Etat de l'art des méthodes d'amélioration de l'image des veines de doigt 43

3 Méthode de Fusion par DCT 44
3.1 Égalisation d'histogramme 45
3.2 Filtrage par filtre de Gabor 47
3.3 Fusion d'images 51
3.4 Evaluation de la qualité 53
3.5 Résultats 55

4 Conclusion 59

Chapitre 4 60

L'apprentissage profond 60

1 Introduction 60

2 Intelligence artificielle, apprentissage automatique, et apprentissage profond 60
2.1 Intelligence artificielle 61
2.2 Apprentissage machine 61
2.3 Apprentissage des représentations à partir de données 62
2.4 Le «profond» dans l'apprentissage profond 65

3 Architecture des réseaux de neurones 66

4 Quelques architectures connues 67
4.1 Réseaux de neurones profonds DNNs 67
4.2 Les Réseaux de neurones récurrents (RNNs) 72

5 Réseau de neurones à convolution 73
5.1 Descente du gradient de l'erreur 74
5.2 Les blocs de construction 76

6 Sur-ajustement et régularisation (Overfitting and regularization) 85
6.1 Régularisation 87

7 Conclusion 89

Chapitre 5 90

Contribution à l'identification par les veines des doigts basé sur un réseau de cnn fusionnés *90*

1 Introduction **90**

2 Intérêt de l'apprentissage automatique pour les méthodes de reconnaissance **90**

3 Métriques **91**

3.1 Matrice de confusion 91

3.2 Techniques de validation croisée 92

4 Bases de données utilisées **93**

4.1 Base de données FV-USM 93

4.2 Base de données SDUMLA 95

4.3 Base de données THU-FVFDT2 96

4.4 Prétraitement des images des bases de données 97

5 Travaux expérimentaux **98**

5.1 Stratégies d'apprentissage 98

5.2 Optimisation d'une nouvelle architecture CNN 99

6 Comparaison avec des travaux précédents **106**

6.1 Comparaison des résultats en utilisant la base de données FV-USM 106

6.2 Comparaison des résultats en utilisant la base de données SDUMLA 107

6.3 Comparaison des résultats en utilisant la base de données THUFVDT2 108

7 Conclusion **108**

Conclusion générale *110*

Bibliographie *112*

Table des Figures

Figure 1. 1 : Quelques caractéristiques biométriques: a) ADN, b) Géométrie de la main, c) Visage, d)Empreintes digitales e)Iris, f) Signature, g) Veine de la main, h) Voix 6
Figure 1. 2: Anatomique de a) L'iris, b) La rétine. 8
Figure 1. 3 : Reconnaissance 3D. 8
Figure 1. 4 : (a) L'image des veines sur la paume et les doigts de la main (b) l'image des veines sur le dos de la main. 10
Figure 1. 5 : Exemple d'une radiographie panoramique dentaire [9]. 11
Figure 1. 6 : Principaux modules d'un système biométrique ainsi que les différentes modes. 12
Figure 1. 7 : Courbe du point d'équivalence des erreurs dans un système biométrique. 14
Figure 1. 8 : Courbe DET. 15
Figure 1. 9 : points de fonctionnement 16

Figure 2. 1 : Les traits biométrique de la main a) empreintes digitales, b) empreintes palmaires, c) géométrie de la main, d) veine de la paume e) les empreintes d'articulations des doigts 20
Figure 2. 2 : Imagerie des structures vasculaires par: a) MRA , b) Hyper-spectral Imaging [15] 21
Figure 2. 3: Dispositif de capture d'image IR 22
Figure 2. 4: Système d'authentification par veines de doigt 23
Figure 2. 5:Reconnaissance des veines des doigts dans le secteur bancaire : a) Barclays (Hitachi), b) Homebanking BPH Bank, c) Fingervein ATM 23
Figure 2. 6:Type d'éclairage :a) Aérien , b) Latéral 23
Figure 2. 7: Scanner veineux de deuxième génération de l'Université de Twente. 24
Figure 2. 8 : Dispositif de veine de doigt installé sur ATM 25
Figure 2. 9 : Enveloppe de veine de doigt pour les distributeurs automatiques de billets en plein air (prototype) 25
Figure 2. 10 : Quelques échantillons d'images de veines de doigts de différentes universités. (a) Images de veines de doigts de l'Université Polytechnique de Hong Kong; (a) Images de veines de doigts de l'Université de Shandong. (c) Images de veines de doigts de l'Université de Pékin. [19] 26
Figure 2. 11: Alignement de doigt par translation et rotation [21] 28
Figure 2. 12 :Extraction de la région du doigt 29
Figure 2. 13 :RI d'un échantillon de la base SDUMLA. 29
Figure 2. 14 : Exemple d'images de la Base « SDUMLA_HMT » 30
Figure 2. 15 : Fonction de recadrage de dynamique 31
Figure 2. 16: Extension linéaire de dynamique : a)Original, b) Recadrage : a = 30, b = 200 ... 31
Figure 2. 17 :(a) Image originale, (b) Histogramme de l'image originale, (c) Image égalisée(c) Histogramme de l'image égalisée. 32
Figure 2. 18: Section transversale a)Position sur l'image des veines. b) Profil de luminosité transversal d'une veine 32
Figure 2. 19 : Relation spatiale entre le point de suivi Pa et le profil transversal 33

Figure 2. 20: Masques circulaires pour détecter la largeur de ligne veine. 34
Figure 2. 21 : Caractéristiques des empreintes digitales..... 35
Figure 2. 22 : schéma bloc des étapes de la méthode MC..... 36
Figure 2. 23 :Profils transversaux d'une image de la veine 37
Figure 2. 24 :Quatre images binaires produites après avoir étiqueté les pixels. 38
Figure 2. 25 : application de la méthode MC, a) Image originale, b) Image des veines..... 38
Figure 2. 26 :Images des veines: a) Image des veines, b) Image post-traitée. 38
Figure 2. 27 :Superposition du motif veineux sur : a) Image MC, b)Image MC amélioré 39
Figure 2. 28 :Terminaisons de ligne et des bifurcations utilisées pour l'appariement 39
Figure 2. 29 :a) Image filtrée par le filtre de Gauss, b) Spectre de fréquence c) Spectre de phase 41
Figure 2. 30 : Résultat d'appariement Deux photos authentiques en utilisant : a) POC , b) BLPOC. 41
Figure 2. 31: Résultat d'appariement Deux photos imposteurs en utilisant : a) POC , b) BLPOC. 42

Figure 3. 7: Amélioration proposée de l'image des veines des doigts 45
Figure 3. 8: (a) image d'origine et son histogramme, (b) l'image prétraitée avec égalisation d'histogramme et son histogramme (c) l'image prétraitée avec CLAHE et son histogramme. 47
Figure 3. 9 : Exemple de filtre Gabor 2D 47
Figure 3. 10 : Une enveloppe gaussienne..... 49
Figure 3. 11 : Les parties réelles et imaginaires d'une fonction complexe de Gabor dans le domaine spatial 49
Figure 3. 12 : Ondelette Gabor 2D symétrique 50
Figure 3. 13 : Réponses spatiales 2D d'un dictionnaire de filtres de Gabor symétriques pairs. Ici, chaque image contient 65 x 65 pixels..... 51
Figure 3. 14: a)Images filtrées avec quatre orientations: 22,5º, 67,5 º, 112,5 º et 157,5 º, b) L'image reconstruite..... 51
Figure 3. 15: Fusion d'images à l'aide de la technique DCT 55
Figure 3. 16: Exemple de résultats obtenus de traitement d'une image de la base FV-USM: a) Image de veines, b) Image améliorée MCGF, c) Image améliorée CLAHE, d) Fusion d'images améliorées. 56
Figure 3. 17: Exemples de résultats obtenus à l'aide d'image SDUMLA: a) Image de veines, b) Image améliorée MCGF, c) Image améliorée CLAHE, d) Fusion d'images améliorées. 57
Figure 3. 18: Résultats de: a) valeur de gris moyenne, b) contraste de l'image, c) entropie de l'image 59

Figure 4. 1: Intelligence artificielle, apprentissage automatique et apprentissage en profondeur 61
Figure 4. 2 : Apprentissage automatique: un nouveau paradigme de programmation 62
Figure 4. 3 : Quelques exemples de données..... 63
Figure 4. 4 : Changement de coordonnées 64
Figure 4. 5 : Réseau de neurones profonds pour la classification des chiffres 66
Figure 4. 6 : Réseau de neurones à une seule couche cachée 66

Figure 4. 7 : Réseau de neurones à deux couches cachées 67
Figure 4. 8 :Illustration d'un perceptron multi-couches 68
Figure 4. 9 :Un schéma représentant la forme d'un réseau d'un RBM 69
Figure 4. 10 :Un schéma représentant la forme d'un DBN 70
Figure 4. 11 : compression et de décompression de données dans un auto-encodeur 71
Figure 4. 12 :Réseau récurrent : a) forme dépliée b) forme temporelle 73
Figure 4. 13 :Exemple d'unité LSTM 73
Figure 4. 14 : ajustement des poids des réseaux de neurones 74
Figure 4. 15 : Exemple de neurone formel 75
Figure 4. 16 : Exemple de modèles de CNN 76
Figure 4. 17 : Illustration de l'opération de convolution : a) Noyau 2 x 2, b) L'entrée et la sortie de convolution, c) La sortie de convolution. 76
Figure 4. 18 : Convolution sur tous les emplacements spatiaux 77
Figure 4. 19: L'image de Lenna et l'effet de différents noyaux de convolution[72]. 78
Figure 4. 20 : Le monde visuel forme une hiérarchie spatiale de modules visuels: les arêtes hyper locales se combinent pour former des objets locaux tels que les yeux ou les oreilles, qui se combinent pour former des concepts de haut niveau tels que «chat». 79
Figure 4. 21 : Représentation des cartes caractéristiques 3D 80
Figure 4. 22 : Le concept d'une carte de réponse: une carte en 2D de la présence d'un motif à différents endroits dans une entrée 80
Figure 4. 23 : Filtrage par un noyau de taille 3 x 3. 81
Figure 4. 24: La fonction ReLU, produit 0 comme sortie lorsque x <0, puis produit une linéaire avec une pente de 1 lorsque x> 0. 82
Figure 4. 25 : Exemples de fonctions de régulation 82
Figure 4. 26 : Exemple de la sortie d'une couche Max-pooling 83
Figure 4. 27 :Fonctionnement de la couche flatten 84
Figure 4. 28 : Les images peuvent être décomposées en motifs locaux tels que des bords, des textures, etc. 85
Figure 4. 29 : Variation de la précision dans la base de test en fonction du nombre d'époques 86
Figure 4. 30 : Variation de la précision dans la base d'apprentissage en fonction du nombre d'époques 86
Figure 4. 31 : Suppression de quelques neurones par Dropout. 88
Figure 4. 32 : Exemple d'augmentation de données par rotation de l'image. 89

Figure 5. 1: La validation croisée LOOCV 93
Figure 5. 2: Un aperçu sur les images de la base de données des veines des doigts FV-USM avec la RI, (a) index gauche, (b) majeur gauche, (c) index droit, et (d) majeur droit 95
Figure 5. 3 : Un aperçu de la base de données des veines des doigts SDUMLA-HMT, (a) (d) Index, (b) (e) majeur, (c) (f) annulaire 96
Figure 5. 4 : Un aperçu de la base de données des veines des doigts SDUMLA-HMT, (a) (d) Index, (b) (e) Majeur, (c) (f) Annulaire 96
Figure 5. 5 : Extraction de la RI de la base de données des veines des doigts SDUMLA-FV (a) Image des veines des doigts; (b) extraction de la RI; (c) Redimensionnement et amélioration de l'image. 97

Figure 5. 6 : Exemples de deux images originales et de deux images améliorées par CLAHE pour la base de données (a) SDUMLA; (b) FV-USM et (c) THU-FVFDT2 98
Figure 5. 7: Architecture du CNN proposé. 103
Figure 5. 8 : Architecture de fusion tardive. 104
Figure 5. 9 : Image utilisées pour la fusion : (a) l'original; (b) traitée par CLAHE (c); filtrée par Filtre de Gabor; (d) fusionnée par DCT 104

Liste des tableaux

Tableau 2. 1. Comparaison entre les bases de données typiques des veines du doigt [19]... 26

Tableau 3. 1.Valeur moyenne des images dans le jeu de données FV-USM..... 58
Tableau 3. 2.Valeur moyenne des images dans le jeu de données SDUMLA 58

Tableau 5. 1. Exemple de calcul par la matrice de confusion 91
Tableau 5. 2. Configurations adoptées du réseau CNN en variant la fonction d'activation. .. 99
Tableau 5. 3. Taux de reconnaissance obtenu pour les différentes fonctions d'activation adoptées..... 100
Tableau 5. 4. Configurations adoptées du réseau CNN an variant les caractéristiques de la taille de la carte et le nombre de CL..... 100
Tableau 5. 5. Taux de reconnaissance obtenu pour les différentes tailles du noyau et le nombre de CL 101
Tableau 5. 6.Configurations adoptées du réseau CNN en variant le nombre de couches de convolution et la couche max-pooling. 101
Tableau 5. 7. Taux de reconnaissance obtenu pour les différents nombres de couches de convolution et la couche max-pooling 102
Tableau 5. 8. Description des couches du CNN proposé..... 102
Tableau 5. 9. Tableau 8. Différentes combinaisons d'images d'entrée. 105
Tableau 5. 10. Taux d'identification obtenu des différentes configurations. 105
Tableau 5. 11.Taux d'identification obtenu des différentes bases de données pour différent nombre d'images d'apprentissage. 105
Tableau 5.12. Comparaisons du taux d'identification dans le cas de la base de données FV-USM en utilisant deux sessions. 106
Tableau 5. 13. Comparaisons du taux d'identification dans le cas de la base de données FV-USM en utilisant une seule session. 107
Tableau 5. 14. Comparaisons du un taux d'identification dans le cas de la base de données SDUMLA pour quatre images de d'entraînement et deux images de test. 107
Tableau 5. 15. Comparaisons du taux d'identification dans le cas de la base de données SDUMLA pour cinq images d'entraînement et une image de test..... 108
Tableau 5. 16. Comparaisons du taux d'identification dans le cas de la base de données THUFVDT2. 108

Liste des acronymes

AM : Ante-Mortem

BLPOC : Band Limited Phase Only Correlation

CCD : Charge-Coupled Device

CL : Convolution Layer

CLAHE : Contrast Limited Adaptive Histogram Equalization

CNN : Convolutional Neural Network

DBN : Deep Belief Network

DCT : Discrete Cosine Transform

DET : Detection Error trade-off

DNN : Deep Neural Network

EER : Equal Error Rate

FAR : False Acceptance Rate

FAR : False-Acceptance Rate

FN : false negative

FP: False Positive

FRR : False Rejection Rate

FRR : False-Rejection Rate

FV-USM : Finger Vein Sains Malaysia University

GLG : Gray-Level Grouping

HE : Histogram equalization

HSV (valeur de saturation de teinte)

IA : Intelligence Artificielle

IDCT : Inverse Discrete Cosine Transform

IR : Infra Rouge

LBP : Local Binary Pattern

LDC : Local Directional Code

LED : Light Emitting Diode

LLBP : Local Line Binary Pattern

LOOCV : Leave one out cross-validation

LSTM : Long Short-Term Memory

MC : Maximum Curvature

MCGF : Multi-Channel Gabor Filter

MGF : Modified Gaussian Filter

MRA : Magnetic Resonance Angiography

MSE : Mean Square Error

PBBM : Personalized Best BitMap

PM : Post-Mortem

POC : Phase Only Correlation

PSF : Point Spread Function

RBM : Restricted Boltzmann Machine

ReLU : Rectified Linear Unit

RI : Région d'Intérêt

RLT : Repeated Line Tracking

RNN : Récurrent Neuronal Network

RVB : Rouge-Vert-Bleu

SDUMLA -HMT : Group of Machine Learning and Applications - Homologous Multi-modal

SVM : Support Support Vector Machine

THU-FVFDT : Tsinghua University Finger Vein and Finger Dorsal Texture Database

TN : True Negative

TP : True Positive

WER : Weighted Error Rate

INTRODUCTION GÉNÉRALE

1.1 Contexte 1
1.2 Vérification et identification. 2
1.3 Objectifs principaux des travaux 2
1.4 Contributions 3
1.5 Organisation du manuscrit 4

1 Contexte

Avec les développements sociaux et économiques actuel, les exigences en matière de sécurité de l'information ont fortement augmenté. Cependant, les méthodes d'authentification traditionnelles, telles que les mots de passe et les clés, ne permettent pas de satisfaire ces attentes car elles risquent d'être oubliées, perdues ou volées. L'identification et l'authentification basées sur des caractéristiques biométriques ont émergé radicalement comme les techniques les plus sûres et les plus pratiques.

À l'origine, le mot « biométrie » désigne l'application au domaine des mesures mathématiques utilisées en biologie. Il renvoie maintenant à un éventail de techniques, d'appareils et de systèmes permettant aux machines de reconnaître des personnes ou d'authentifier leur identité. Ces systèmes mesurent et analysent les attributs physiques et comportementaux des personnes, comme les traits du visage, les inflexions de la voix, les empreintes digitales, les empreintes de la paume de la main, la forme des veines des doigts et de la main, la structure des yeux (iris ou rétine) ou la démarche.

Ces méthodes physiques et comportementales ont été considérablement appliquées pour la reconnaissance des malfaiteurs dans les contextes juridiques, pour rassurer l'accès aux endroits sensibles. Les modalités biométriques extrinsèques sont susceptibles de faire l'objet d'attaques par illégalité, car les fausses images de visage, les empreintes digitales et l'iris peuvent tromper le système de vérification. Par contre les modalités intrinsèques, comme la veine digitale, la veine de la main et la veine de la paume sont difficiles d'être attaquées. Pour dupliquer les veines de la main il faut soit avoir accès à la base de données, soit prendre une photo directement sur la personne vivante.

2 Vérification et identification

Il existe plusieurs types d'application qui nécessitent l'authentification des utilisateurs. Ces applications peuvent être séparées en deux catégories qui sont la vérification d'identité ou l'identification.

La vérification d'identité correspond à la vérification d'une identité proclamée. Le plus souvent elle est utilisée en association avec un moyen d'identification classique tel qu'une carte, un badge ou une clé, et sert de sécurité supplémentaire pour vérifier que la carte ou le badge n'a pas été volé ou n'est pas utilisé par une personne non autorisée. La vérification correspond à une décision du type OUI ou NON à la question : "l'individu est-il bien celui qu'il prétend être ?"

L'identification est utilisée dans des applications différentes souvent du type contrôle d'accès ou de passage à une frontière. Cela correspond à la recherche de l'identité de la personne qui se présente dans une base de données. L'identification peut servir à autoriser l'utilisation des services, comme par exemple contrôler l'accès à une zone très sécurisée pour laquelle seul un nombre restreint de personnes (enregistrées dans une base de données) a l'autorisation d'accès.

L'identification peut également servir à détecter un individu non autorisé dans des bases de données, par exemple pour les personnes ne possédant plus aucun papier d'identité (par exemple une personne en situation irrégulière qui a détruit ses papiers pour ne pas pouvoir se faire identifier), si elles sont enregistrées dans une base de données pour les visas, il est alors possible de les identifier dans la base par leurs données biométriques. La finalité n'est pas la même mais le procédé est le même, il consiste à chercher si l'individu appartient à une base de données.

3 Objectifs principaux des travaux

La biométrie basée sur le modèle vasculaire est un trait biométrique émergent en raison de ses divers avantages par rapport aux autres traits biométriques. La biométrie vasculaire basée sur les veines traite des modèles formés par les vaisseaux sanguins situés à l'intérieur du corps humain, dans lequel les modèles à l'intérieur des doigts et des mains sont les parties du corps les plus utilisées. Ceci est communément appelé reconnaissance des veines des doigts et des mains. Afin de mener des recherches sur n'importe quel trait biométrique, il y a deux conditions préalables importantes: la disponibilité d'ensembles de données d’apprentissage et de tests, et la disponibilité d'une chaîne d'outils de reconnaissance biométrique complète adaptée au trait biométrique spécifique pour pouvoir effectuer des évaluations des performances de reconnaissance.

Les méthodes classiques de reconnaissance par la veine passent par quatre étapes qui sont : l’acquisition de l’image, le prétraitement, l’extraction de caractéristiques et

l'appariement. En particulier, la capture d'image permet de visualiser les veines dans un doigt. Le prétraitement améliore principalement la qualité des images, puis une étape d'extraction de la Région d'Intérêt (RI) est nécessaire pour séparer la région où se trouve les veines du reste de l'image. L'extraction de caractéristiques détecte les caractéristiques du motif veineux pour la représentation et l'appariement mesure la similarité entre deux images de veines digitales pour la reconnaissance.

Récemment, l'apprentissage profond a remplacé toutes les étapes des méthodes classiques pour fournir de nouveaux systèmes de reconnaissance avec des performances remarquables. Peu d'études ont été menées sur la reconnaissance des veines des doigts en utilisant l'apprentissage profond. Notre travail vise à établir une méthode de reconnaissance basée sur l'utilisation d'une nouvelle architecture de réseaux de neurones profond permettant d'augmenter les performances de reconnaissance par rapport aux méthodes utilisées actuellement.

4 Contributions

Notre contribution comprend les quatre points suivants:

(1) L'amélioration de la qualité des images des veines de doigt qui présentent une qualité médiocre lors de la formation de la base de données, sachant que chaque base de donnée était obtenue par des systèmes d'acquisition différent.

(2) La conception d'un réseau de neurone profond par optimisation de plusieurs paramètres, tels que la fonction d'activation, la taille du noyau et le nombre de couches.

(3) Une amélioration de la précision dans le processus d'extraction de la région d'intérêt de l'image de la veine du doigt de la base de données (SDUMLA-HMT), en exploitant les techniques fournies dans la littérature. En utilisant les opérations de segmentation et de redimensionnement, avec plus de précision dans l'opération de séparation de la région veineuse du bas, ce qui contribue à l'amélioration des performances de reconnaissance.

(4) L'utilisation du réseau de neurones profond optimisé dans un nouveau mode d'architecture pour former un nouveau modèle qui prend en entrée différentes images, en termes de qualité, d'un même doigt. Contrairement à d'autres méthodes utilisant des CNN pour l'identification des veines des doigts où une seule voie est utilisée pour construire le modèle, notre modèle proposé utilise deux CNN en parallèle et fusionne leurs sorties en une seule couche, ce qui nous permet d'exploiter les informations contenues dans deux images au lieu d'une seule image.

5 Organisation du manuscrit

Nous avons choisi d'articuler le manuscrit autour de cinq chapitres principaux. Le premier chapitre est consacré à la présentation générale de la biométrie.

Dans le deuxième chapitre nous nous concentrons sur la biométrie des veines. Les capteurs d'images des veines dans les systèmes de reconnaissance seront présentés dans un premier temps. On parlera ici sur les conditions d'acquisition des images des veines des doigts puis nous présenterons quelques bases de données des veines des doigts existants. Dans un second temps, on présentera quelques méthodes classiques de reconnaissances. Ces méthodes ont des étapes communes qui sont le prétraitement de l'image, l'extraction des caractéristiques et l''appariement des caractéristiques. On passera en revue les principales méthodes proposées dans la littérature.

Comme la qualité de l'image des veines de doigt influe sur les performances des méthodes classiques. Le troisième chapitre sera consacré à l'amélioration de la qualité de l'image, nous proposons dans ce contexte une méthode hybride.

Le quatrième chapitre est consacré à l'éclaircissement du concept apprentissage en profondeur ou "Deep Learning". Plusieurs architectures seront présentées, après nous nous concentrerons sur un type particulier qui utilise des couches de convolution. Cette architecture a montré dans la littérature une très bonne performance notamment dans les applications liées aux traitement d'images.

Dans le dernier chapitre, nous donnons une description détaillée sur notre système d'identification proposée ainsi que les différents résultats obtenus. Ce système est basé sur une architecture de réseaux de neurones profond, optimisé puis testé sur trois bases de données publiques. Notre modèle sera comparé avec d'autres méthodes qui utilisent les mêmes bases de données.

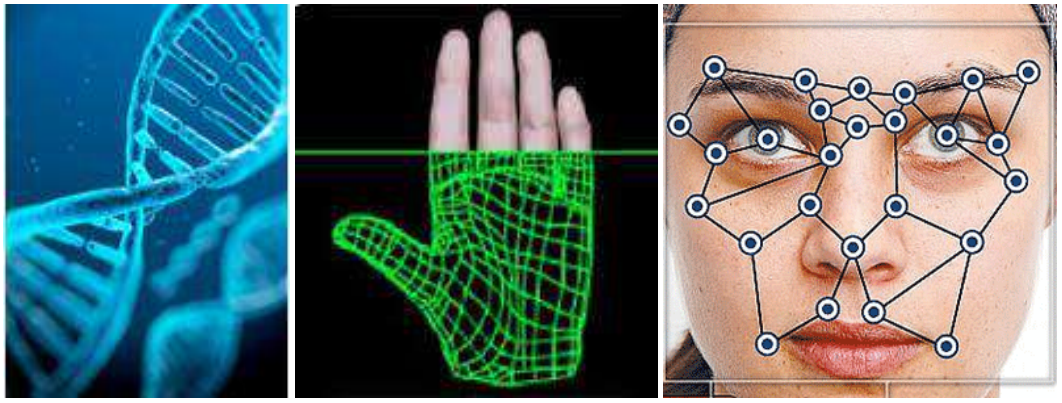
CHAPITRE 1
LA BIOMÉTRIE

1 Introduction

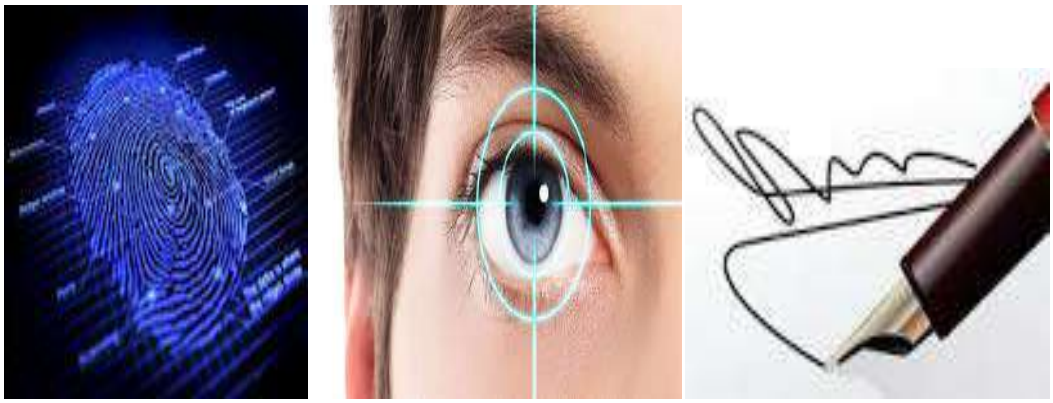
Le mot **biométrie** signifie "mesure du vivant", et désigne dans un sens très large l'étude quantitative des êtres vivants. L'usage de ce terme se rapporte de plus en plus à l'usage de ces techniques à des fins de reconnaissance, d'authentification. Nous introduirons dans ce chapitre quelques notions et définitions de base liées à la biométrie. Nous donnerons le principe de fonctionnement des systèmes biométriques ainsi que les outils utilisés pour mesurer leurs performances.

La biométrie consiste à identifier une personne à partir d'une ou de plusieurs caractéristiques physiologiques (empreintes digitales, visage, iris, contour de la main, etc.), ou comportementales (signature, démarche, etc.)
Les systèmes biométriques sont de plus en plus utilisés depuis quelques années. L'apparition de l'ordinateur et sa capacité à traiter et à stocker les données ont permis la création des systèmes biométriques informatisés. Il existe plusieurs caractéristiques physiques uniques pour un individu, ce qui explique la diversité des systèmes appliquant la biométrie, selon que l'on prend en compte [1] :

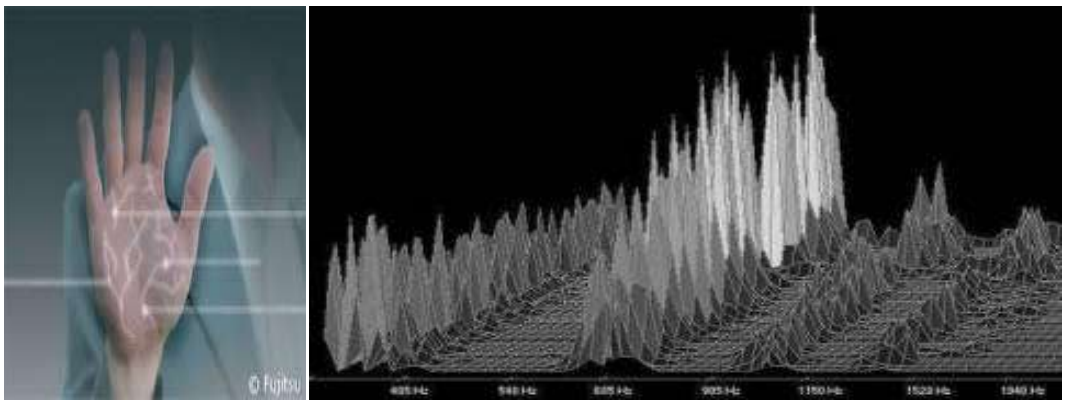
- ✓ L'empreinte digitale
- ✓ La géométrie de la main
- ✓ L'iris
- ✓ La rétine
- ✓ Le visage
- ✓ La dynamique de frappe au clavier
- ✓ La reconnaissance vocale
- ✓ La dynamique des signatures
- ✓ L'ADN



(a) (b) (c)



(d) (e) (f)



(g) (h)

Figure 1. 1 : Quelques caractéristiques biométriques: a) ADN, b) Géométrie de la main, c) Visage, d)Empreintes digitales e)Iris, f) Signature, g) Veine de la main, h) Voix

2 Description de quelques modalités biométriques

Dans ce qui suit, nous donnons attention aux avantages ainsi qu'aux inconvénients de certaines modalités biométriques, notamment celles qui utilisent les caractéristiques physiques et biologiques.

2.1 L'empreinte digitale

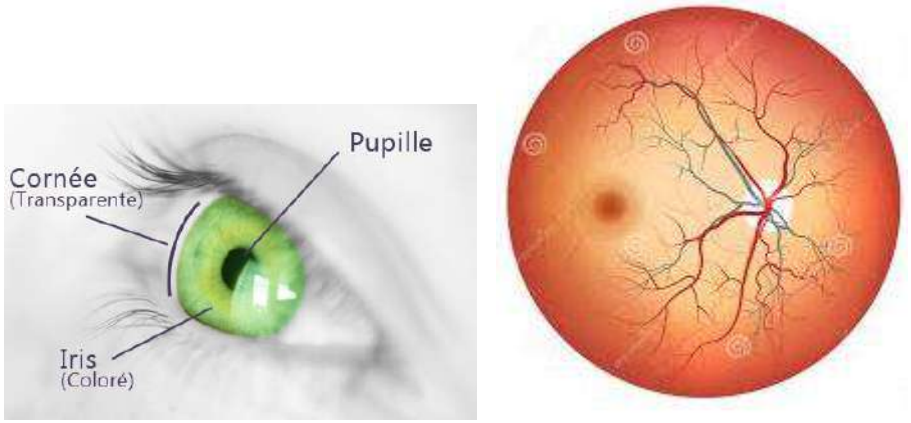
Dans le domaine judiciaire, l'utilisation de l'empreinte digitale est une pratique courante. Les systèmes biométriques utilisant cette modalité semblent être les plus développés et les plus usités. Sa généralisation a permis d'intégrer les systèmes intelligents à usage quotidien (ordinateurs portables, smart phones,…). Cependant, l'empreinte digitale présente des avantages et des inconvénients. En termes d'avantage, l'empreinte digitale est connue pour sa singularité, sa constance et la facilité de son acquisition. Par contre, cette modalité présente des inconvénients, notamment chez des personnes ayant des empreintes dégradées dues à l'âge, à certaines maladies, ou suite à la manipulation de certaines substances chimiques. Dans ce cas, les minuties (caractéristiques de l'empreinte), sont parfois difficilement détectables [2].

Un autre problème lié à cette modalité, est sa sensibilité face aux attaques potentielles « *Spoofing* ». Plus précisément, les systèmes biométriques utilisant un capteur d'empreintes digitales peuvent présenter des failles de vérification. Ceci est possible en utilisant de fausses empreintes que l'imposteur reproduit sur une matière spécifique. L'attaque peut être détectée, soit au niveau de l'acquisition même, soit a posteriori par analyse de la texture de l'image acquise. Cela suppose que le système biométrique soit équipé par un système d'anti-Spoofing [3].

2.2 La biométrie visuelle

On peut distinguer ici, la biométrie de l'iris ou la biométrie de la rétine. L'iris est l'anneau coloré autour de la pupille de l'œil. La reconnaissance des individus par l'iris est considérée comme une biométrie très fiable. Cette modalité commence à devenir de plus en plus utilisée pour le contrôle d'accès, notamment pour les applications gouvernementales. Les systèmes de reconnaissance par iris utilisent souvent des caméras fonctionnant en mode infrarouge. Ceci est considéré comme une technique non intrusive puisqu'il est possible de capturer une image de l'iris à une distance allant de quelques centimètres à quelques mètres. Comme ceci est le cas dans d'autres systèmes biométriques, la biométrie de l'iris n'échappe aux risques d'attaques. Plusieurs applications montrent qu'il est possible d'attaquer ce type de systèmes en utilisant des objets reproduisant les caractéristiques de l'œil.

Considérée également comme une biométrie oculaire [4], l'extraction des caractéristiques de la rétine s'effectue en envoyant un faisceau de lumière à l'intérieur de l'œil afin de capturer une image de la rétine. Étant donné que la rétine est située à l'arrière de l'œil (figure 1.2.b), la lecture d'empreintes rétiniennes n'est pas totalement acceptée en raison du processus intrusif pour capturer une image.



(a) (b)

Figure 1. 2: Anatomique de a) L'iris, b) La rétine.

3.3 La reconnaissance faciale

La reconnaissance faciale est une modalité biométrique facile à mettre en œuvre. Elle peut être fondée sur l'extraction des caractéristiques locales ou globales du visage à partir d'images capturées par des caméras en 2D ou en 3D (figure 1.3). Les systèmes de reconnaissance du visage sont pratiques et ne nécessitent aucun contact avec les utilisateurs. Cependant, dans de tels systèmes, il est bien connu que le taux d'erreur peut accroitre en raison de la sensibilité de cette modalité aux conditions de capture, à savoir, l'illumination, la pose, ou à d'autres facteurs liés au vieillissement, à la prise ou la perte de poids [5]. De plus, les systèmes de reconnaissance faciale peuvent être sujettes à des attaques de différents types, d'où la nécessité d'intégrer des modules « *anti-Spoofing* » pour accroître les performances du système.

Les systèmes de reconnaissance de visage infrarouges utilisent des ondes infrarouges pour mesurer les radiations thermiques émises dans la gamme du spectre infrarouge. Ces systèmes ne sont pas sensibles à la couleur de la peau, bronzage, usage de produits de beauté, les lentilles de l'œil coloré, la chirurgie plastique et les photos peuvent être prises dans l'obscurité. Néanmoins, le coût extrêmement élevé de cette technologie rend son utilisation prohibitive pour les applications.

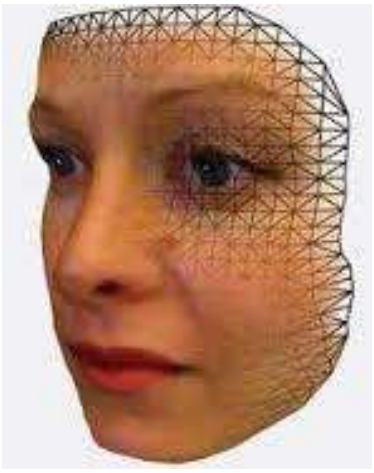


Figure 1. 3 : Reconnaissance 3D.

2.3 La reconnaissance vocale

La reconnaissance de la voix est une technique biométrique ayant une variété d'applications. Elle peut être utilisée pour contrôler certains systèmes tels que l'accès physique ou logique (à des services). Bien qu'elle soit facile à utiliser et souvent acceptée par les utilisateurs, cette modalité s'avère sensible au changement de la voix dû à l'âge, au rhume, etc. Par ailleurs, le système à reconnaissance vocale peut être défaillant dans des milieux bruyants (faible rapport signal sur bruit). Il peut également être sensible aux attaques (enregistrement et réinjection du signal vocal au système) [6].

2.4 L'ADN

Chaque personne possède 23 paires de chromosomes qui contiennent le plan génétique de l'ADN. Chaque paire contient un élément maternel ainsi qu'un élément paternel. Considérée comme une biométrie biologique, l'ADN d'un individu peut être obtenu grâce à une mèche de cheveux, un échantillon de salive, ou de sang, etc. Cinq pour cent de l'ADN est constitué de la partie codante (gènes). Les gènes sont analysés afin d'identifier une personne. Parmi les avantages de cette modalité, est l'unicité du génome à chaque individu. Cela rend évidemment l'identification très précise. En revanche, l'inconvénient majeur de cette modalité est le temps nécessaire à l'identification. De récents progrès montrent qu'il est maintenant possible d'extraire des caractéristiques ADN en seulement quelques heures [7].

2.5 La biométrie de la main

Utilisé fréquemment pour l'accès aux zones réglementées, le système biométrique de la main utilise soit les caractéristiques géométriques soit les caractéristiques de la texture de la main. Lorsque l'on considère la géométrie de la main, cette modalité requiert une mesure et une analyse descriptive méticuleuse des paramètres géométriques, y compris ceux des doigts (leur longueur, leur largeur, leur finesse, etc). Cette modalité peut être réalisée, soit en 2D, soit en 3D. Elle est considérée moins intrusive et fortement acceptable. De plus, elle est simple à mettre en place puisqu'elle ne nécessite pas un équipement spécifique et onéreux. Une simple caméra peut être déployée dans ce type d'applications. Cependant, les caractéristiques géométriques de la main ne sont pas uniques, rendant ainsi son utilisation restreinte à l'authentification. Cette modalité ne peut être fiable qu'à l'âge adulte en raison des modifications morphologiques rapides à l'âge enfant. En termes de robustesse, les dispositifs utilisant la géométrie de la main peuvent être attaqués en utilisant une fausse main ayant la même géométrie que celle d'un individu donné. En revanche, certaines modalités de cette catégorie analysent la texture de la main, notamment au niveau de la paume « *Palm-Print* ». Dans certaines approches, les caractéristiques extraites de la texture peuvent être fusionnées avec les caractéristiques géométriques [3].

2.6 Biométrie des veines

L'analyse et la caractérisation des veines, notamment de la main permettent l'identification ainsi que l'authentification unique des individus [8][85] (figure 1.4). Cette biométrie nécessite l'acquisition d'images en proche infrarouge. Les vaisseaux de la veine paraissent noirs lorsque le rayon infrarouge est absorbé par l'hémoglobine désoxydée dans les veines. Une fois que la forme des veines caractérisées, elle est comparée à une référence.

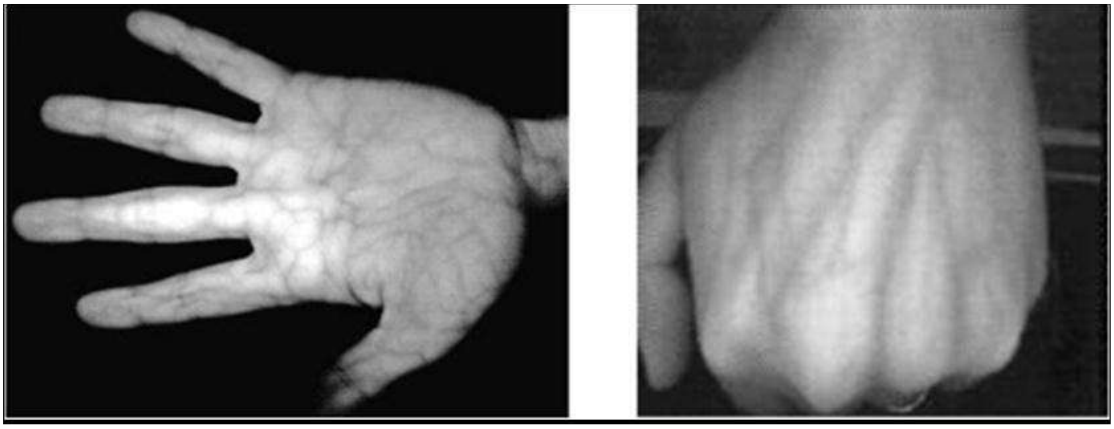


Figure 1. 4 : (a) L'image des veines sur la paume et les doigts de la main (b) l'image des veines sur le dos de la main.

La biométrie des veines est considérée comme une modalité fiable et robuste, car la structure globale des veines ne change pas au cours du temps (à l'exception de certaines modifications dues à la dilatation des veines ou à des cas pathologiques). Par ailleurs, le « *Spoofing* » des systèmes biométriques utilisant les veines peut paraître difficile, mais des expériences récentes ont montré que la modalité peut présenter des failles, notamment lorsque l'identification et l'authentification sont effectuées en mode non-supervisé (en l'absence d'un contrôle humain).

2.7 La biométrie dentaire

Souvent utilisée en Odontologie (médico-légale), les dentistes légistes sont impliqués dans des enquêtes visant l'identification des corps humains. Dans ce contexte, les dentistes légistes sont parfois amenés à déterminer l'âge, l'origine, l'historique dentaire, voire le statut socioéconomique des êtres humains. L'identification se fait par la comparaison des données *Ante-Mortem* (AM)et *Post-Mortem* (PM) à partir de radiographies dentaires (imagerie par rayon X) (figure 1.5). Les radiographies PM, sont évidemment acquises après le décès d'une personne alors que les radiographies AM sont, quant à elles, acquises durant sa période de vie. Ainsi, la biométrie dentaire requiert l'extraction des caractéristiques dentaires à partir des images radiographiques et leur mise en correspondance. Bien que l'avantage de cette modalité réside dans sa possibilité de son utilisation en PM et en AM, son automatisation n'est pas évidente, notamment lorsque les variations AM/ PM sont importantes [9].



Figure 1. 5 : Exemple d'une radiographie panoramique dentaire [9].

3 Les systèmes biométriques

Un système biométrique est essentiellement un système de reconnaissance de formes qui utilise les données biométriques d'un individu. Selon le contexte de l'application, un système biométrique peut fonctionner en mode d'enrôlement ou en mode de vérification ou bien en mode d'identification [1] :

- Le mode d'**enrôlement** est une phase d'apprentissage qui a pour but de recueillir des informations biométriques sur les personnes à identifier. Plusieurs campagnes d'acquisitions de données peuvent être réalisées afin d'assurer une certaine robustesse au système de reconnaissance aux variations temporelles des données. Pendant cette phase, les caractéristiques biométriques des individus sont saisies par un capteur biométrique, puis représentées sous forme numérique (signatures), et enfin stockées dans la base de données. Le traitement lié à l'enrôlement n'a pas de contrainte de temps, puisqu'il s'effectue « hors-ligne ».

- Le mode de ***vérification*** est une comparaison "un à un", dans lequel le système valide l'identité d'une personne en comparant les données biométriques saisie avec le modèle biométrique de cette personne stockée dans la base de données du système. Dans un tel mode, le système doit alors répondre à la question suivante: *«Suis-je réellement la personne que je suis en train de proclamer ?»*. Actuellement la vérification est réalisée via un numéro d'identification personnel, un nom d'utilisateur, ou bien une carte à puce.

- Le mode d'***identification*** est une comparaison "Un à N", dans lequel le système reconnaît un individu en l'appariant avec un des modèles de la base de données. La personne peut ne pas être dans la base de données. Ce mode consiste à associer une identité à une personne. En d'autres termes, il répond à des questions du type: *« Qui suis-je ? »*

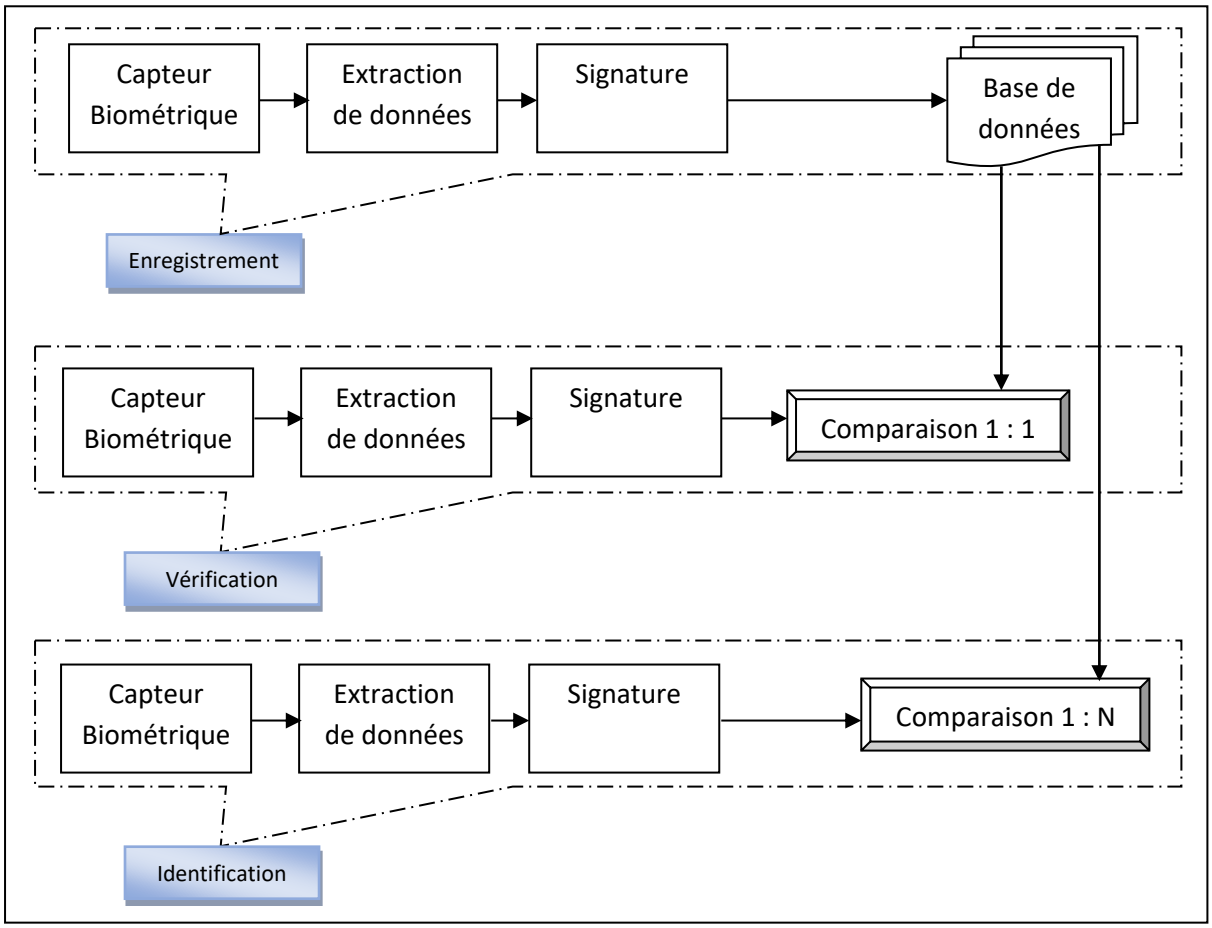


Figure 1. 6 : Principaux modules d'un système biométrique ainsi que les différentes modes.

Les différents modules qui composent un système biométrique sont représentés sur la figure 1.6 ; leur fonctionnement peut être résumé comme suit [1]:

1. *Module capteur biométrique* : correspond à la lecture de certaines caractéristiques physiologiques, comportementales ou biologiques d'une personne, au moyen d'un terminal de capture biométrique (ou capteur biométrique).

2. *Module extraction des données* : extrait les informations pertinentes à partir des données biométriques brutes, par exemple des images de visage ou des régions caractéristiques de visage.

3. *Module création d'une signature* : crée un modèle numérique afin de représenter la donnée biométrique acquise. Ce modèle, appelé aussi signature, sera conservé sur un support portable (puce ou autre) ou dans une base de données.

4. *Module comparaison :* compare les caractéristiques biométriques d'une personne soumise à contrôle avec les « signatures » mémorisées. Ce module fonctionne soit en mode vérification (pour une identité proclamée) ou bien en mode identification (pour une identité recherchée).

5. *Module base de données :* stocke les modèles biométriques des utilisateurs enrôlés.

4 Evaluation des performances des Systèmes biométriques

Chaque caractéristique (ou modalité) biométrique a ses forces et ses faiblesses, et le choix dépend de l'application visée. On ne s'attend à ce qu'aucune modalité biométrique ne réponde efficacement aux exigences de toutes les applications. En d'autres termes, aucun système biométrique n'est "optimal". Faire correspondre un système biométrique spécifique à une application dépend du mode opérationnel de l'application et des caractéristiques biométriques choisies. Plusieurs études ont été menées afin d'évaluer les performances des systèmes biométriques. La société américaine –*'International Biometric Group* (IBG) – à par exemple effectué une étude basée sur quatre critères d'évaluation :

- ✓ *Intrusivité :* ce critère permet de classifier les systèmes biométriques en fonction de l'existence d'un contact direct entre le capteur utilisé et l'individu à reconnaître. La reconnaissance faciale est une technique « non intrusive », car il n'existe aucun contact entre le capteur (la caméra) et le sujet, elle est bien acceptée par les utilisateurs à l'inverse d'autres techniques « intrusives » comme l'iris où un contact direct est nécessaire entre le capteur et l'œil.

- ✓ *Fiabilité :* dépend de la qualité de l'environnement (éclairage par exemple) dans lequel l'utilisateur se trouve. Ce critère influe sur la reconnaissance de l'utilisateur par le système.

- ✓ *Coût :* doit être modéré. À cet égard nous pouvons dire que la reconnaissance faciale ne nécessite pas une technologie coûteuse. En effet, la plupart des systèmes fonctionnent en utilisant un appareil à photo numérique de qualité standard.

- ✓ *Effort :* requis par l'utilisateur lors de la saisie de mesures biométriques, et qui doit être réduit le plus possible. La reconnaissance faciale est la technique biométrique la plus facile à utiliser car elle est non contraignante.

Les techniques les moins "intrusives" aux plus "intrusives" : la voix, la frappe sur le clavier, la signature, la main, le visage, l'empreinte digitale, l'iris et enfin la rétine. Les techniques les plus fiables aux moins fiables: l'iris, la rétine, l'empreinte digitale, le visage, la main, la voix, et enfin à un niveau équivalent, la frappe sur le clavier et la signature. Les techniques les moins coûteuses aux plus coûteuses: la frappe sur le clavier, la voix, la signature, l'empreinte digitale, le visage, la main, la rétine et enfin l'iris. Les techniques les plus faciles d'utilisation aux plus difficiles: le visage, la signature, L'iris, la frappe sur le clavier, la voix, l'empreinte digitale, la main et enfin la rétine.

5 Fiabilité des systèmes biométriques

Afin de mesurer la fiabilité d'un système biométrique en modes de vérification et/ou d'identification, deux principaux tests sont utilisés :

5.1 Test de vérification

Dans la tâche de vérification, un utilisateur doit faire une demande d'authentification de son identité, alors le système biométrique doit déterminer si l'identité proclamée par l'utilisateur est acceptée ou rejetée. Deux taux sont alors calculés :
Le Taux de Faux Rejets ou False Rejection Rate (FRR), il exprime le pourcentage d'utilisateurs rejetés alors qu'ils devraient être acceptés par le système.

Le Taux de Fausses Acceptations ou False Acceptance Rate (FAR), il exprime le pourcentage d'utilisateurs acceptés par le système alors qu'ils devraient être rejetés.

On peut formuler le test de vérification de la façon suivante [10] :
Soient *XQ* le vecteur de caractéristiques de la personne proclamée *I*, *Xt* le vecteur de caractéristiques de la personne *I* stockée dans la base de données, *S(XQ , Xt)* la fonction de similarité entre le vecteur *XQ* et *Xt* . La fonction *S* donne le score de similarité entre les mesures biométriques de la personne de la base de données et la personne proclamée. Le test de vérification est alors défini par la fonction *(I, XQ)*, telle que :

$$(I, X_Q) = \begin{cases} w_1 & Si \quad S(X_Q, X_t) \geq \theta \\ w_2 & Sinon \end{cases} \tag{1.1}$$

Où w_1 indique que la personne proclamée est vraie et w_2 qu'elle est un imposteur. Le choix du seuil de similarité θ est important car il influe directement sur les performances du système. Un seuil θ trop grand entraîne l'apparition d'un grand nombre de faux rejets, tandis qu'un seuil θ trop petit engendre un taux important de fausses acceptations. La statistique la plus simple pour mesurer la performance d'un algorithme dans le contexte de la vérification est de calculer le point d'équivalence des erreurs (Equal Error Rate - EER) (figure 1.7)**.** Le point d'équivalence des erreurs, ou taux d'exactitude croisée, est déterminé par le point d'intersection entre la courbe du taux de fausses acceptations et la courbe du taux de faux rejets.

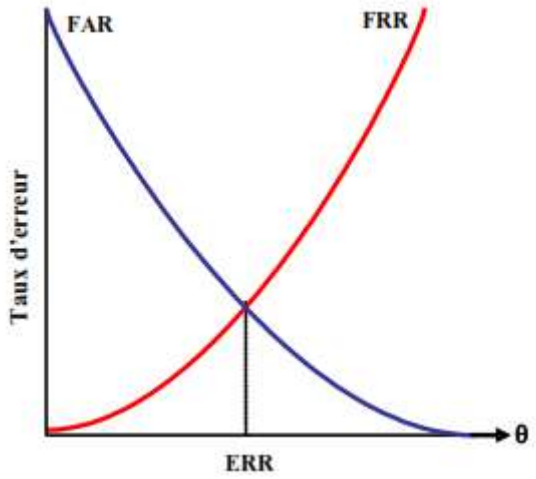


Figure 1. 7 : Courbe du point d'équivalence des erreurs dans un système biométrique.

L'évaluation passe également par le tracé de statistiques complexes, comme la courbe

« Detection Error trade-off (DET) » (figure 1.8). Cette courbe donne le FRR en fonction du FAR. Elle est tracée de manière paramétrique en fonction des valeurs du seuil.

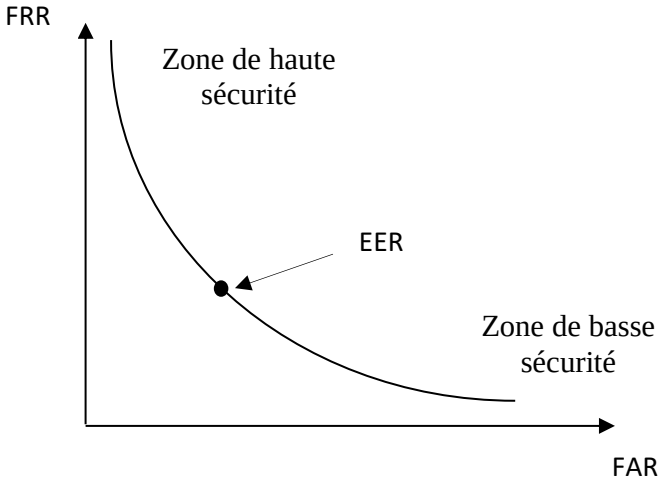


Figure 1. 8 : Courbe DET.

5.2 Test d'identification

On peut formuler le test d'identification de la façon suivante:
Soient XQ le vecteur de caractéristiques, pour déterminer l'identité I_k , $K \in \{1,2,3, \dots, N, N+1\}$
on calcule la fonction *(I,XQ)* définie par :

$$(I, X_Q) = \begin{cases} I_k Si \max\left(S(X_Q, X_{I_k})\right) \geq \theta, k = 0,1,2, \dots N \\ I_{N+1} Sinon \end{cases} \quad (1.2)$$

où I_1,..,I_N sont les identités enrôlées, I_{N+1} une identité rejetée, X_{I_k} la signature biométrique qui correspond à l'identité I_k , et θ le seuil.

En résumé, il existe deux façons de **mesurer la performance** en termes de Taux d'acceptation Faux **FAR et** le taux de faux rejet **FRR**, défini comme suit:

$$\boldsymbol{FRR} = \frac{\boldsymbol{Nombre\ des\ clients\ rejetés\ (FR)}}{\boldsymbol{Nombre\ total\ d'accès\ clients}} \quad (1.3)$$

$$\boldsymbol{FAR} = \frac{\boldsymbol{Nombre\ des\ imposteurs\ acceptés\ (FA)}}{\boldsymbol{Nombre\ total\ d'accès\ imposteurs}} \quad (1.4)$$

Les 4 types de points de fonctionnement les plus utilisés sont [11]:

- EER : "Equal Error Rate" ou taux d'erreurs égales. Ce point de fonctionnement correspond au seuil qui donne des taux FAR et FRR égaux.

- WER : "Weighted Error Rate" ou taux d'erreur pondéré. Ce point de fonctionnement correspond au seuil tel que le FRR est proportionnel au FAR avec un coefficient qui dépend de l'application. Le seuil du WER est égal au seuil de l'EER lorsque ce coefficient est égal à 1.

- FAR fixé. Ce point de fonctionnement correspond au seuil tel que le taux FAR est égal à un taux fixé par l'application (par exemple 1%). La performance du système est donnée par le taux FRR pour cette valeur de FAR fixée.

- FRR fixé. Ce point de fonctionnement correspond au seuil tel que le taux FRR est égal à un taux fixé par l'application (par exemple 1%). La performance du système est donnée par le taux FAR pour le FRR fixé.

Sur la Figure 1.9 sont représentés des exemples des deux types de points de fonctionnement présentés ci-dessus.

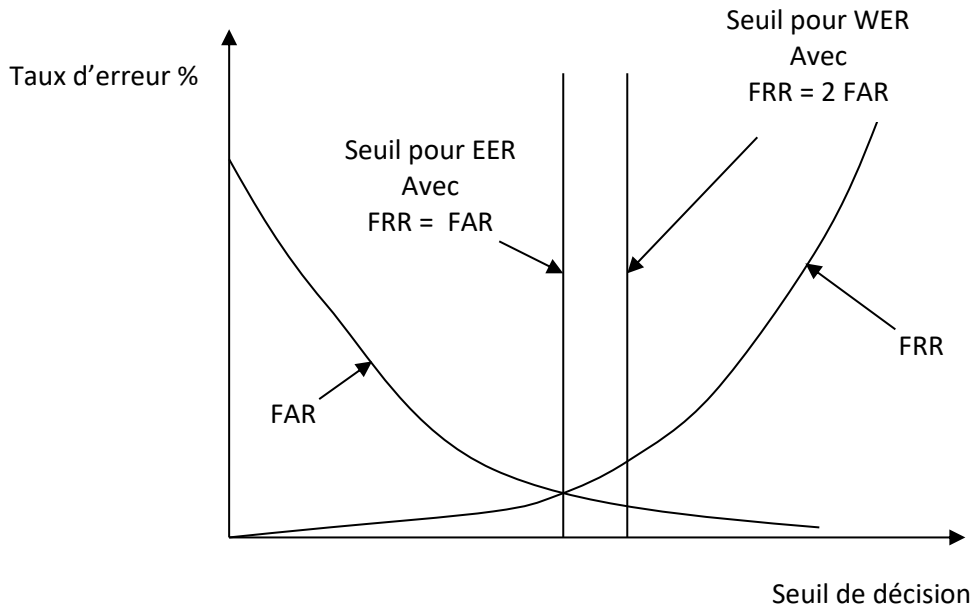


Figure 1. 9 : points de fonctionnement

Le point de fonctionnement qui définit le choix du seuil dans le module de décision dépend de l'application visée. En général lorsqu'il n'y a pas d'application définie mais qu'il s'agit d'un test de performance sur une base de données pré-enregistrée, on utilise le plus souvent l'EER (c'est-à-dire les deux taux d'erreurs égaux) car c'est un point de fonctionnement assez neutre qui ne favorise aucun des deux types d'erreurs. En revanche lorsqu'une application est définie ou lorsque l'on connaît les objectifs de performance, on peut utiliser les autres points de fonctionnement et le plus

souvent les points de fonctionnement correspondant à des niveaux fixés pour l'un des deux types d'erreurs.

Le compromis à faire pour le réglage du seuil de décision est le compromis entre confort et sécurité. Le confort correspond à un taux de Faux Rejets bas et la sécurité à un taux de Fausses Acceptations bas.
En général pour un besoin de sécurité on fixe FAR = 1% selon le niveau de sécurité souhaité. Pour un besoin de confort on fixe FRR = 1% en fonction du degré de confort souhaité.

Cependant, il est important de noter que le seuil de décision associé au point de fonctionnement choisi va être estimé sur une base de données, dite de développement, avec laquelle on règle les paramètres qui seront ensuite utilisés pour l'application en condition réelle. Le choix de cette base de données est primordial pour le bon réglage d'un système biométrique. Tout d'abord la base de données doit être représentative de l'application visée mais surtout elle doit être de taille suffisante. En particulier, pour estimer des seuils de décision pour des taux de FAR ou de FRR fixés et faibles (par exemple FAR = 0,1%), cela nécessite un grand nombre de données pour avoir de la précision dans des zones où il y a peu d'erreurs.

6 Les bases d'apprentissage et les bases de test

Un système monomodal doit être optimisé sur des bases de donnée. En général, on dispose d'une base de données de scores (si l'on veut évaluer un système de fusion de scores) que l'on va devoir utiliser à la fois pour l'apprentissage et pour le test. Il va donc falloir découper cette base en deux sous-ensembles, un pour l'apprentissage et un pour le test.
La méthode la plus simple est donc de séparer la base de données en deux sous-ensembles disjoints (de tailles égales ou non), l'un servant à l'apprentissage et l'autre au test. Ce protocole d'évaluation pose un problème de généralisation de performances. En effet, les performances obtenues sont très dépendantes du découpage apprentissage/test et ne sont pas forcément représentatives de l'ensemble des données disponibles. Pour cela on utilise un protocole d'évaluation qui consiste en un certain nombre de découpages de la base, réalisés de manière aléatoire. En effet on reproduit un certain nombre de fois le découpage en deux sous-ensembles (apprentissage et test), on réalise ensuite pour chaque tirage l'apprentissage des paramètres de fusion et le test de performances. Cela permet d'avoir des estimations statistiques des erreurs de classification en calculant la moyenne et l'écart-type sur les différents tirages et de voir si les performances varient beaucoup d'un tirage à l'autre [11].

7 Les avantages et les limites de la biométrie

La biométrie est une technologie récente et commence à être adoptée par de grands constructeurs de matériel informatique. L'usage de la biométrie est un complément de l'utilisation des méthodes d'authentification comme des mots de passe, des badges, des cartes à puce. Malheureusement la biométrie présente un certain nombre d'inconvénients parmi eux le problème de la qualité de l'authentification. Ces méthodes ne sont en effet pas toujours fiables à 100%, ce qui empêche des utilisateurs de bonne foi d'accéder à leur système. Car il s'agit bien là d'une des caractéristiques majeures de tout organisme vivant : on s'adapte à l'environnement, on vieillit, on subit des traumatismes plus ou moins importants, bref on évolue et les mesures changent.

8 Conclusion

La technologie biométrique sert généralement à établir l'identité des personnes ou à vérifier si celles-ci détiennent certains privilèges, comme conduire une voiture ou avoir accès à une zone sécurisée ou réservée.

Quoi qu'il en soit, la biométrie peut apporter beaucoup dans la vie quotidienne en termes de sécurité mais également de confort. Comme pour d'autres nouvelles technologies son essor devra s'accompagner d'une réflexion approfondie sur le respect des libertés individuelles.

CHAPITRE 2

LA BIOMÉTRIE DES VEINES DES DOIGTS

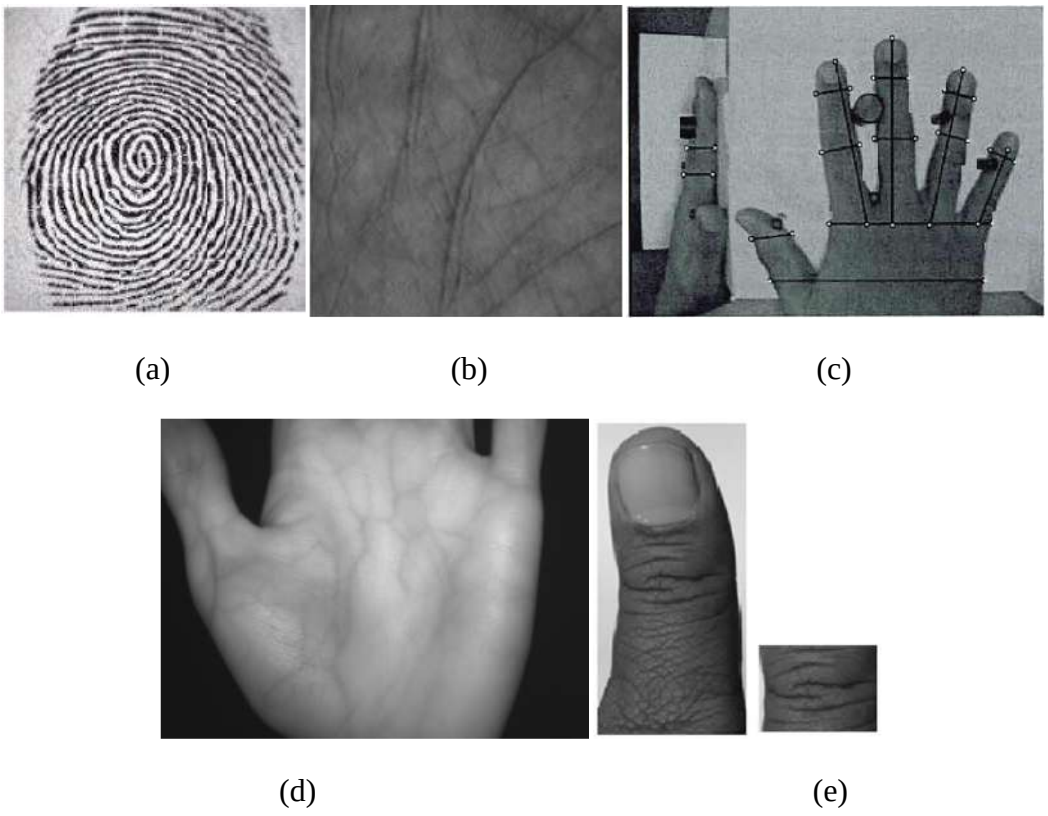
1 Introduction

La biométrie des veines s'est rapidement distinguée comme la plus pertinente pour identifier et authentifier les personnes de manière fiable, en fonction de cette caractéristique biologique consubstantiel. Ce chapitre commence par introduire la biométrie des veines, ensuite nous présentons la technologie utilisé pour acquérir les images des veine de doigt, en présentant d'abord les systèmes d'acquisitions, les conditions qui affectent la vérification par ces systèmes, après nous citons quelques bases de données existantes. La dernière section sera consacrée à la présentation des méthodes classique les plus utilisées dans chaque étape du système biométrique qui sont le prétraitement, l'extraction des caractéristiques, et enfin l'appariement.

2 Les limites des systèmes biométriques

Dans le domaine des sciences de l'information, le terme biométrie désigne la technologie d'un système d'identification qui utilise des caractéristiques physiques humaines telles que le visage, les empreintes digitales, la géométrie de la main ou des caractéristiques comportementales comme la voix et la démarche. De nos jours, avec la demande accrue de sécurité de l'information, la biométrie est de plus en plus répandue dans l'authentification des utilisateurs pour l'octroi ou le refus des droits de contrôle d'accès aux données et aux ressources système [12].

Les applications biométriques les plus connues incluent le contrôle d'accès aux portes, le contrôle des passages frontaliers, la sécurité financière et les systèmes de fréquentation. Parmi les biométries disponibles, les systèmes à main (empreintes digitales, empreintes palmaires, géométrie de la main, veine de la paume et les empreintes d'articulations des doigts) sont les plus populaires en raison de leur grande acceptation par les utilisateurs (figure 2.1).



(a) (b) (c)

(d) (e)

Figure 2. 1 : Les traits biométrique de la main a) empreintes digitales, b) empreintes palmaires, c) géométrie de la main, d) veine de la paume e) les empreintes d'articulations des doigts

Les systèmes d'empreintes digitales sont la technologie biométrique la plus mature. Ils ont été utilisés dans de nombreuses applications depuis des années. Bien que l'utilisation des empreintes digitales soit l'approche la plus courante, cette méthode est vulnérable à la falsification.

Cela peut se produire lorsque des informations cachées sont volées par une personne non autorisée à partir des capteurs d'empreintes digitales. En outre, il nécessite également un contact physique, ce qui n'est pas souhaitable d'un point de vue hygiénique et peut causer des problèmes avec le dispositif. De plus, l'état de la surface d'un doigt tel que la sueur et la sécheresse peut empêcher l'obtention d'un motif d'empreinte digitale clair. Cela peut dégrader la précision de reconnaissance.

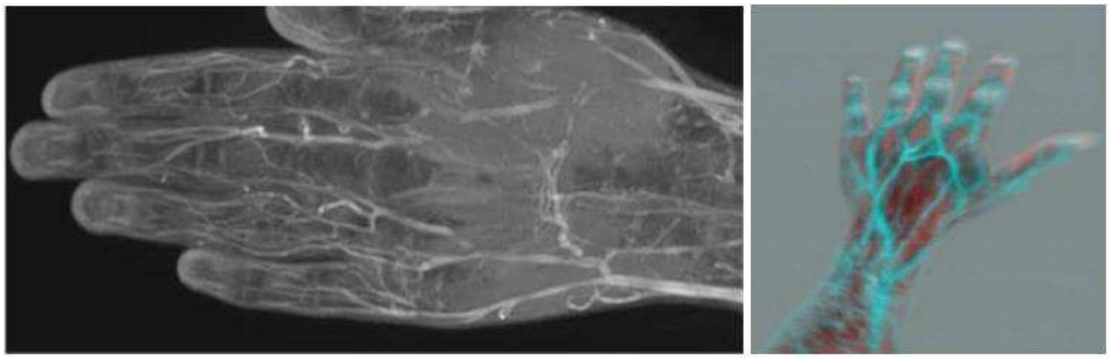
En ce qui concerne les systèmes d'empreintes palmaires et les empreintes d'articulations des doigts, les traits biométriques sont faciles à reproduire, car les traits sont extraits de l'extérieur du corps humain. Pour surmonter les limites des systèmes biométriques actuels basés sur les mains, une nouvelle technologie biométrique basée sur la forme des veines de doigts humains a été développé. Aujourd'hui, elle reçoit beaucoup plus d'attention des chercheurs et est devenu une direction de recherche attrayante. Par rapport à d'autres biométries à base de main, les systèmes de veine de doigt présentent certains excellents avantages dans leur application.

Tout d'abord, les veines du doigt sont difficiles à contrefaire. Cela est attribué à la veine cachée sous la surface de la peau, ce qui rend la duplication des motifs des veines pratiquement impossible. Deuxièmement, la qualité de la veine capturée n'est pas facilement influencée par les conditions de la peau. En outre, par rapport aux systèmes de vérification à base de veine de doigt, la taille du dispositif peut être beaucoup plus petite.

De plus, les motifs veineux sont stables et clairement définis, permettant ainsi à un utilisateur d'utiliser un dispositif d'acquisition d'images peu onéreux et peu coûteux. Enfin, en termes d'hygiène, la reconnaissance des veines utilise parfois une imagerie sans contact, ce qui garantit à la fois la commodité et la propreté. En raison des propriétés uniques susmentionnées, l'utilisation de la reconnaissance des veines devient un candidat biométrique approprié pour l'identification personnelle dans les systèmes de sécurité de haut niveau. De plus, il peut être combiné avec d'autres modalités biométriques telles que les empreintes digitales et la géométrie des doigts pour former des biométriques multimodales basées sur les doigts. Par conséquent, cela garantira que les systèmes de sécurité sont très robustes contre la contrefaçon.

3 La biométrie des veines

En principe, l'imagerie de haute précision des structures vasculaires humaines, y compris celles à l'intérieur de la main humaine, est un problème résolu. La figure 2.2a montre les vaisseaux correspondants à l'aide d'un appareil d'imagerie médicale par résonance magnétique (Magnetic Resonance Angiography (MRA)), tandis que la figure 2.2b montre le résultat de l'application de l'imagerie hyperspectrale à l'aide d'un STEMMER Dispositif d'imagerie utilisant leur logiciel Perception Studio pour visualiser les données capturées dans la plage 900-1700 nm. Cependant, les capteurs biométriques ont une limitation en termes de coûts. Pour un déploiement pratique pour des solutions d'authentification réelles, les technologies utilisées pour produire les images de la figure 2.2 ne sont pas une option pour cette raison. La solution est beaucoup plus simple et donc plus rentable par utilisation de l'imagerie infrarouge (IR).



(a) (b)

Figure 2. 2 : Imagerie des structures vasculaires par: a) MRA , b) Hyper-spectral Imaging [15]

JoeRice a breveté son système «Veincheck» basé sur l'imagerie IR au début des années 1980, qui est souvent considéré comme la naissance de la biométrie vasculaire de la main. Après l'expiration de ce brevet, Hitachi, Fujitsu et Techsphere ont lancé des produits de sécurité reposant sur la biométrie veineuse (tous détenant divers brevets dans ce domaine maintenant)[12].

4 Acquisition de données

Kono et al [86] ont été les premiers qui ont proposé les veines comme une nouvelle modalité biométrique. L'utilisation de la reconnaissance par les veines présente de nombreux avantages, tels que la difficulté de produire des faux, une immunité relative à l'état de surface du doigt, une définition stable et claire des motifs de la veine et la capacité d'utiliser des dispositifs de capture compacts [13].

Un dispositif d'imagerie spécial est utilisé pour obtenir l'image infrarouge du doigt. Une lumière infrarouge irradie l'arrière de la main et la lumière passe dans le doigt (figure 2.4). Une caméra située à l'autre côté de la main capture cette lumière. L'intensité de la lumière provenant de la LED est réglée en fonction de la luminosité de l'image. Comme l'hémoglobine dans le sang absorbe la lumière infrarouge, le motif des veines est capturé sous forme d'ombres [14]. La détection du motif veineux est basée sur la lumière IR. Lors de l'acquisition de l'image des veines, le temps d'exposition de la lumière infrarouge est très petit pour qu'il soit assuré pour le corps humain. Le niveau de pénétration et le temps d'exposition de la lumière infrarouge sont différents des rayons X.

Il est important de capturer les échantillons réels des veines du doigt qui ne devraient pas être manipulés ou modifiés. L'imagerie IR est plus sûre car elle pénètre dans la seule surface superficielle du doigt. Lorsque la lumière est rayonnée dans le doigt, la température de la peau augmente, mais elle est contrôlée en ajustant le paramètre opérationnel de la LED pour garantir une température sûre. [15]
Il existe plusieurs schémas pour la capture des images comme celle proposée dans [16], l'auteur a développé une méthode pour capturer l'image de la veine de doigt en utilisant une lumière infrarouge à intensité non uniforme.

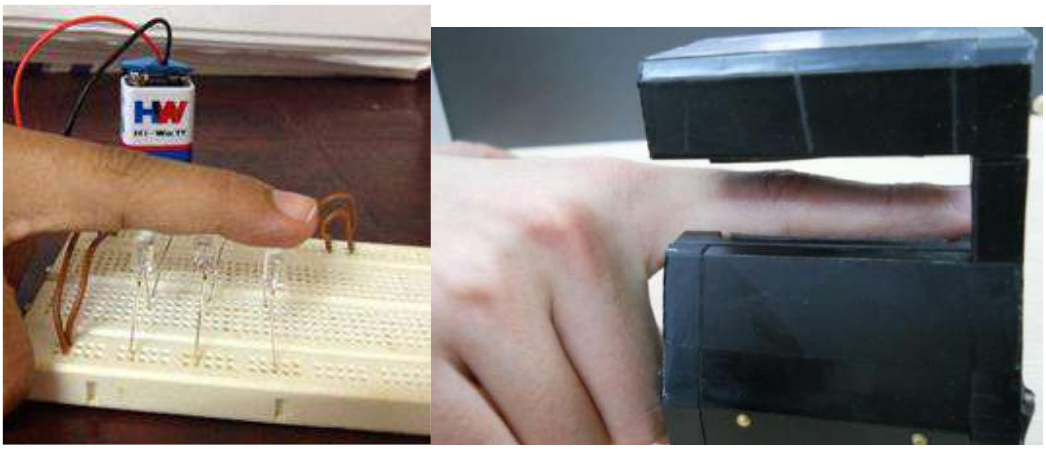


Figure 2. 3: Dispositif de capture d'image IR

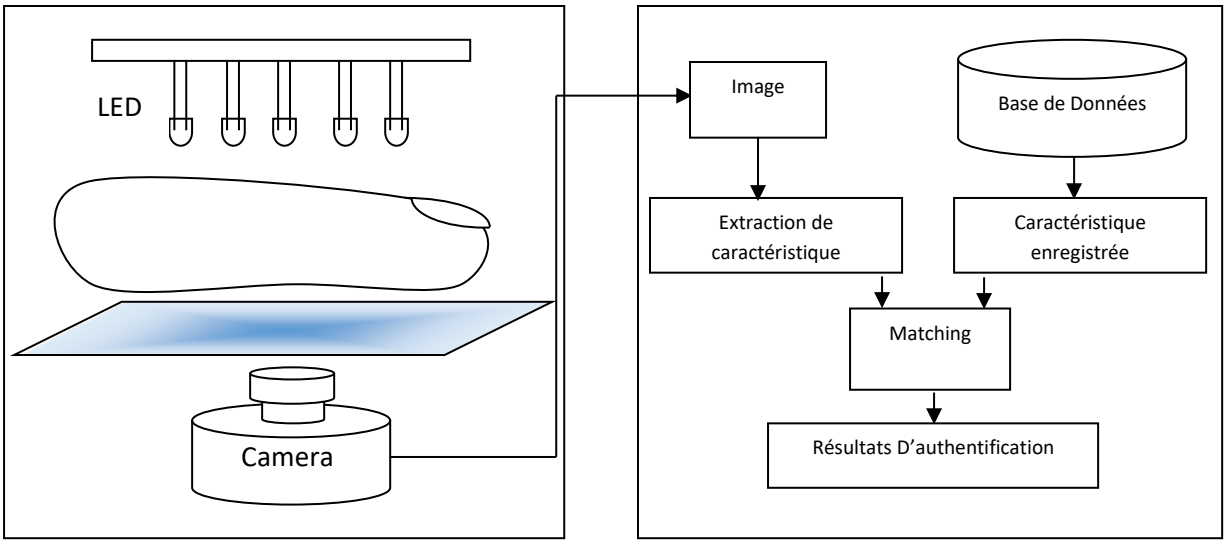


Figure 2. 4: Système d'authentification par veines de doigt

4.1 Capteurs et systèmes commerciaux

La figure 2.5 montre les trois capteurs de la veine du doigt actuellement disponibles. Le capteur Hitachi est basé sur un principe de trans-illumination pure, les deux autres capteurs éclairent le doigt sur le côté tandis que la capture est effectuée par le bas (tous les capteurs capturent le côté palmaire du doigt). Au cours des dernières années, d'autres concurrents sont entrés sur le marché [12].

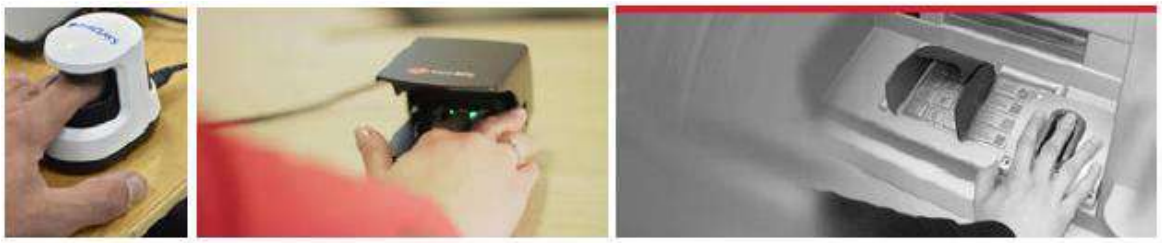
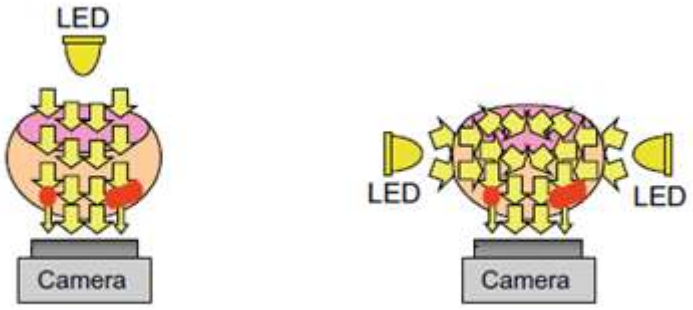


Figure 2. 5:Reconnaissance des veines des doigts dans le secteur bancaire : a) Barclays (Hitachi), b) Homebanking BPH Bank, c) Fingervein ATM [12]



(a) (b)

Figure 2. 6:Type d'éclairage :a) Aérien , b) Latéral [12]

4.2 Nouvelle génération de scanner veineux

Une nouvelle conception du scanner de la veine du doigt est développée avec de nouvelles capacités, (Figure 2.7) [12]. Le scanner est beaucoup plus compact et est construit à l'aide de composants moins chers: il utilise des cartes de traitement et des caméras Raspberry Pi. Ce nouveau scanner a été conçu de manière à soutenir la recherche de diverses manières. Il prend en charge plusieurs bandes LED IR qui peuvent être positionnées dans un demi-cercle de 0◦ à 180◦. Il prend également en charge trois caméras, permettant ainsi la reconstruction 3D des veines des doigts.

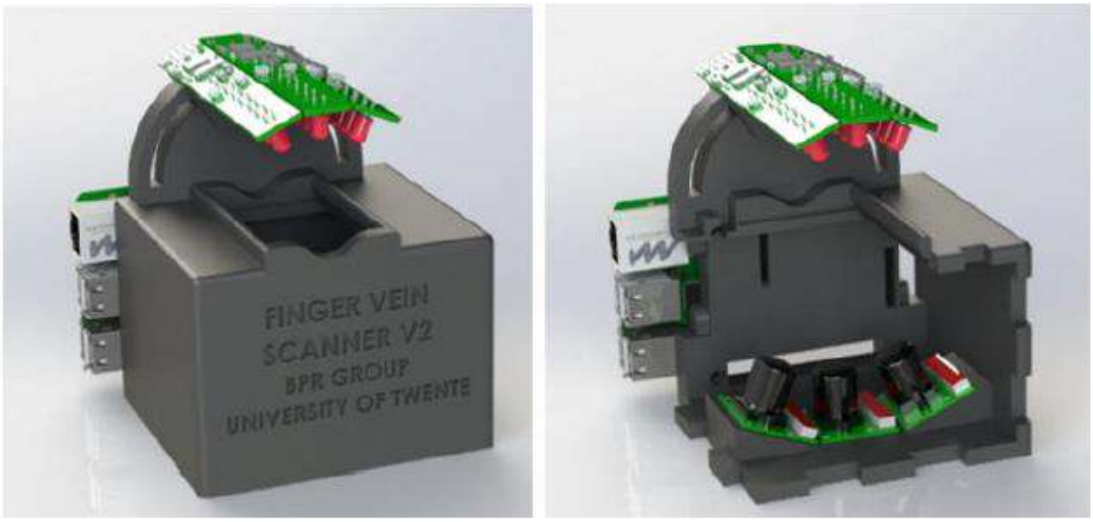


Figure 2. 7: Scanner veineux de deuxième génération de l'Université de Twente [12].

4.3 Les conditions d'acquisition des images des veines des doigts

Actuellement, il n'existe aucune base de données des veines de doigt accessible au public qui présente diverses conditions d'acquisition environnementales et non environnementales.

En général, chaque système d'authentification biométrique comprend une étape d'enregistrement et une étape de vérification. Pendant la phase d'enregistrement obligatoire, le modèle biométrique considéré de l'utilisateur est acquis dans des conditions contrôlées et supervisées. Cette situation idéale ne sera probablement pas présente lors de la deuxième étape du processus d'authentification. Chaque fois que l'utilisateur souhaite s'authentifier, il doit présenter son trait biométrique qui est ensuite acquis à nouveau. Cette acquisition de données biométriques lors de l'authentification peut être effectuée dans différentes conditions environnementales.

4.4 Durabilité et anti-vandalisme

Dans de nombreux cas, les guichets automatiques sont situés à l'extérieur pour fournir aux utilisateurs des services financiers 24/7. Les conditions environnementales des cas d'utilisation en extérieur sont beaucoup plus difficiles que les cas d'utilisation

en intérieur. Les utilisateurs de l'ATM étant grand public, l'équilibre entre la convivialité et la durabilité est un facteur clé.



Figure 2. 8 : Dispositif de veine de doigt installé sur ATM [12]

Afin d'intégrer l'appareil sur des ATM extérieur, l'appareil devait être robuste contre les conditions ambiantes difficiles. Les exigences suivantes devaient être satisfaites:

- L'appareil doit pouvoir fonctionner sous des conditions météorologiques extrêmes telles que la pluie, la neige ou la lumière du soleil.
- L'enceinte de l'appareil doit résister au vandalisme. L'enceinte du dispositif veineux a été repensée pour faire face aux scénarios de cas d'utilisation de l'ATM d'extérieur (figure 2.9). Un capot pour protéger la platine de numérisation a été ajouté pour augmenter la durabilité. Le boîtier rond illustré sur la Figure 2.9 est conçu pour résister au vandalisme et peut supporter le poids d'un homme adulte.



Figure 2. 9 : Enveloppe de veine de doigt pour les distributeurs automatiques de billets en plein air (prototype)[12]

5 Les bases de données des veines de doigt

Les bases d'images disponibles sur les veines ne sont pas parfaites dans certains aspects. Par exemple, seules les images de la veine du doigt ou les images de texture des doigts sont disponibles dans ces bases de données, sans autres informations

personnelles telles que les nationalités, les groupes sanguins, le genre et les couleurs de la peau [17]. Presque toutes les recherches sur la reconnaissance des doigts utilisaient leur propre base de données pour vérifier leurs algorithmes proposés. Certaines des bases de données n'ont pas assez de quantité d'images. Par conséquent, il est difficile à vérifier les performances des méthodes existantes, puisqu'elles sont effectuées sur de petites bases de données. En outre, peu de bases de données ont été disponibles en ligne [18].

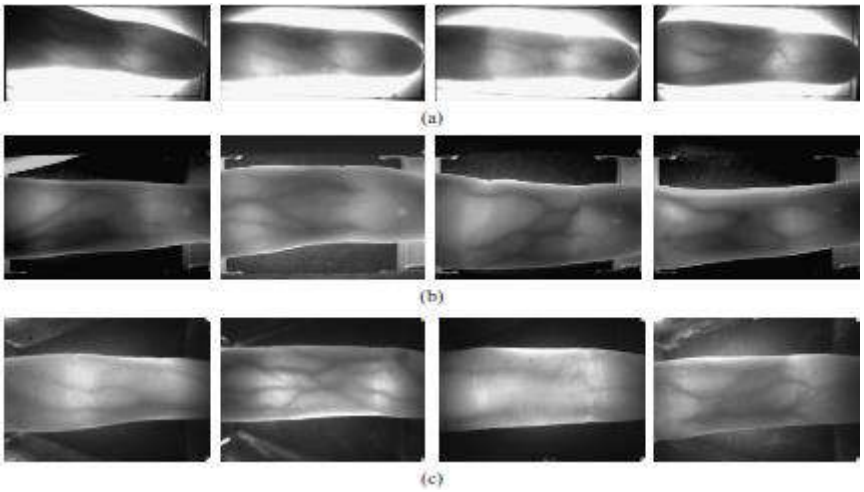


Figure 2. 10 : Quelques échantillons d'images de veines de doigts de différentes universités. (a) Images de veines de doigts de l'Université Polytechnique de Hong Kong; (a) Images de veines de doigts de l'Université de Shandong. (c) Images de veines de doigts de l'Université de Pékin. [19]

Le tableau suivant présente quelques Bases de données disponibles.

Tableau 2. 1. Comparaison entre les bases de données typiques des veines du doigt [19]

Base de données	N ° de sujet	Nombre de doigts par sujet	Nombre d'image par doigt	Taille de l'image (pixels)	Format
SDUMLA-FV	106	6 (Index, anneau, milieu, des deux mains)	6	320×240	.BMP
HKPU-FV	156	2	12/6	513×256	.BMP
UTFV	60	6	4	672×380	8 bit niveau

		(Index, anneau, milieu, des deux mains)			de gris .PNG
THU-FV	610	1	2	200×100	.BMP
MMCBNU_6000	100	6 (Index, anneau, milieu, des deux mains)	10	640×480	.BMP
FV-USM	123	4 (index gauche, milieu gauche, et index droit, milieu droit)	6	640 x 480	JPEG

6 Les systèmes de reconnaissance classiques

Dans le processus d'acquisition de l'image de la veine du doigt, il existe une condition aussi commune: la structure de la topologie de la veine du doigt est trop simple et l'information de reconnaissance utile est trop petite, cette condition est causée par une acquisition incorrecte d'image ou une différence individuelle. Dans ce cas, la méthode d'extraction des caractéristiques conventionnelles tend à avoir une mauvaise performance et des nouvelles approches sont maintenant nécessaires. Un système d'authentification classique basé sur la biométrie veineuse se compose principalement de trois étapes [14]:

1. Prétraitement de l'image IR des veines de doigt.
2. Extraction des caractéristiques.
3. L'appariement des caractéristiques

6.1 Prétraitement

Il y a trois sous-étapes principales dans l'étape de prétraitement [20]:

1. Alignement de la rotation et de la translation.
2. Localisation de la région d'intérêt (RI).
3. Amélioration de l'image pour extraire des modèles vasculaires stables / fiables.

6.1.1 Alignement de la rotation et de la translation

Si le dispositif de capture ne comporte pas de chevilles ou de barres de guidage, l'alignement des doigts doit être accompli, comme le montre la Figure 2.11. Cette situation entraînera que l'image du doigt sera affectée par des problèmes de rotation qui peuvent entraver une reconnaissance précise. Pour réduire ce problème, l'alignement en rotation est effectué. En générale l'alignement est achevé selon la base de donnée en question.

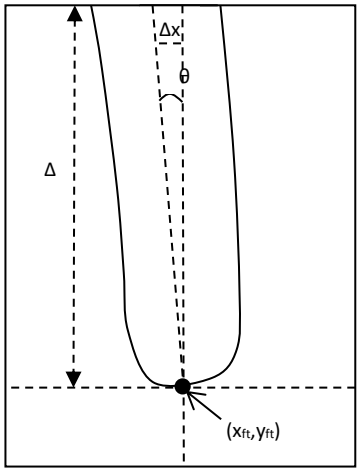


Figure 2. 11: Alignement de doigt par translation et rotation [21]

Le dispositif de capture utilisé dans la formation de la base SDUMLA ne comporte pas de chevilles ou de barres de guidage, l'alignement des doigts ne varie pas beaucoup dans les collections. Cette situation entraînera que l'image du doigt ne sera pas trop affectée par des problèmes de rotation qui peuvent entraver une reconnaissance précise. Les problèmes de déplacement ont été résolus en évaluant les paramètres de translation entre deux images améliorées en utilisant la fonction de Corrélation de Phase Seule (POC) [21]. Ensuite, les régions communes pour les deux images sont recadrées sur la base des paramètres de translation évalués comme le montré dans la Figure 2.11.

6.1.2 Extraction de la RI

Dans l'image des veines de doigt, les bords de l'image contiennent du bruit et ne fournissent pas beaucoup d'informations sur les veines en raison d'une exposition moins élevée à la lumière infrarouge. Ainsi, les bords de l'image peuvent être découpés. La RI peut-être extrait par certaines opérations morphologiques ainsi que par la connaissance du fond du dispositif d'acquisition.

Pour ce faire, les images sont recadrées de sorte que les informations proviennent principalement de la région du doigt. Le contraste de l'image est ajusté,

afin d'augmenter la différence entre les veines et leur arrière-plan. De plus, le bruit existant et indésirable est supprimé en filtrant l'image avec les masques appropriés[22].

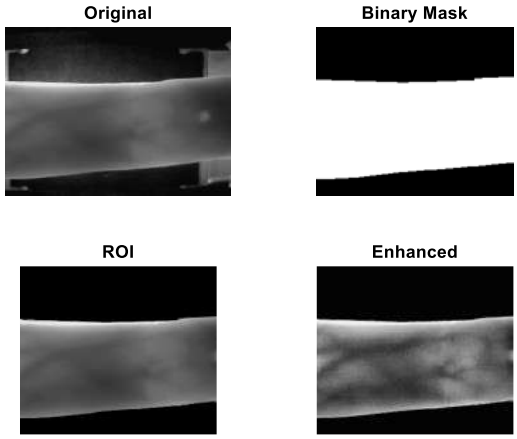


Figure 2. 12 :Extraction de la région du doigt

Yang et al. ont proposé une méthode utilisant une fenêtre prédéfinie avec un hauteur fixe et une largeur réglable dans le plan d'imagerie des doigts [23]. Mahri et al. ont proposé une méthode basée sur les bords de doigts pour obtenir la RI [24]. Malheureusement, ces méthodes ne sont pas fiables dans la localisation des RI des veines de la même personne à différentes sessions en application réelle [25].

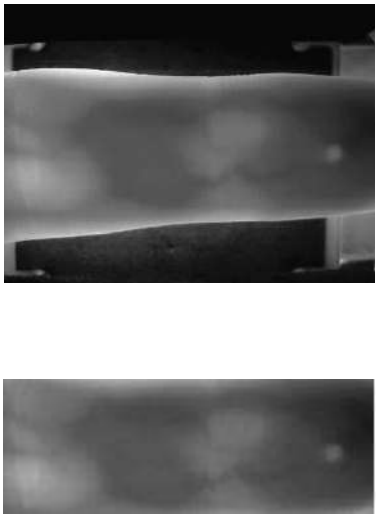


Figure 2. 13 :RI d'un échantillon de la base SDUMLA.

Dans l'étape d'extraction de la RI utilisé dans [21], la largeur de l'image est fixée empiriquement à 300 pixels. La figure 2.13 montre le sous-échantillonnage de la région veineuse, où la largeur N est définie de manière adaptative en mesurant la largeur aux extrémités inférieures et supérieurs de la géométrie de doigt. Dans ce cas, les dimensions peuvent différer entre les différents doigts. Afin de standardiser la taille de la région veineuse, la taille de l'image est étirée à une valeur normalisée.

6.1.3 Amélioration de l'image

Une image des veines acquise à partir de scanners à rayons X ou ultrasons (en imagerie médicale) a une qualité élevée. Mais il est très gênant pour les utilisateurs en raison d'une acquisition lente et donc à éviter en biométrie. On observe que les zones contenant le schéma des veines ont des intensités sombres par rapport aux tissus environnants lorsqu'elles sont observées sous la lumière infrarouge (IR) car la lumière infrarouge est absorbée par le sang qui s'écoule dans la veine. Ainsi, l'imagerie infrarouge est utilisée dans la biométrie pour l'acquisition sans contact et non invasif d'images veineuses [26]. Les détails de la veine du doigt dans les images acquises, en particulier les minces, ne sont pas très clairs. Cela peut être attribué à l'illumination inégale et / ou au placement imparfait des doigts pendant l'imagerie. Par conséquent, les images de veines avec un faible contraste et un éclairage irrégulier sont soumises à une amélioration de l'image non linéaire [20].

L'application des algorithmes d'amélioration des images est nécessaire pour les images des veines de faible qualité. Un exemple d'une image des veines peut être vu dans la figure 2.14. L'image de faible qualité est caractérisée par un faible contraste.

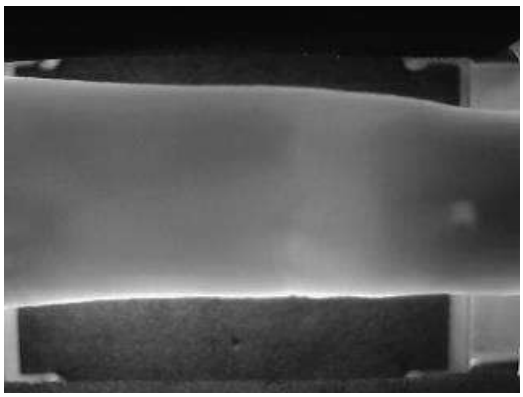


Figure 2. 14 : Exemple d'images de la Base « SDUMLA_HMT »

Le prétraitement d'image implique l'amélioration de l'image, la réduction du bruit, la segmentation et la normalisation de l'image.

6.1.3.1 Transformation de recadrage (Normalisation)

On suppose une image de départ présentant un histogramme concentré dans l'intervalle [a,b]. Les valeurs a, b correspondent aux niveaux de gris extrêmes présents dans cette image. Le recadrage de dynamique consiste à étendre la dynamique de l'image transformée à l'étendue totale [0, 255]. La transformation de recadrage est donc une application affine qui s'écrit:

$$t(f) = \begin{cases} 255 * \frac{f-a}{b-a} & pour \ \ a < f \ < b \\ 0 & Si \quad f < a \\ 255 & Si \ \ f > b \end{cases} \tag{2.1}$$

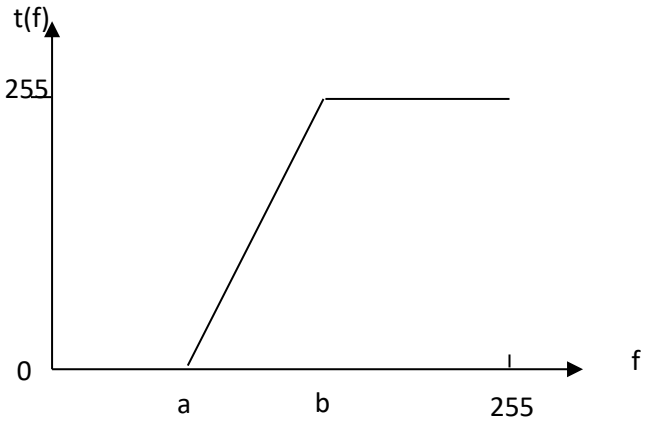
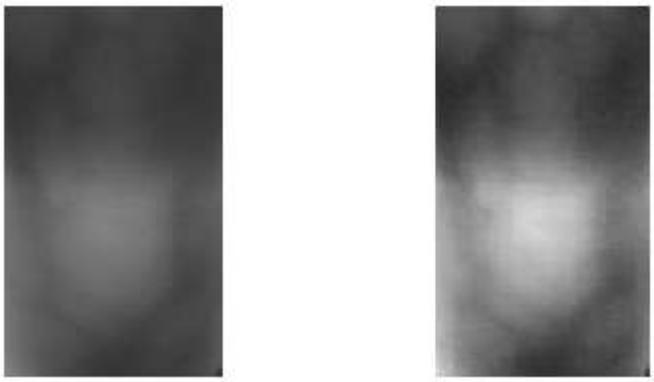


Figure 2. 15 : Fonction de recadrage de dynamique

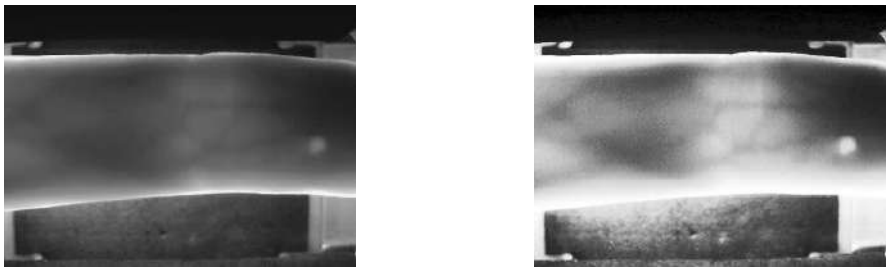


(a) (b)

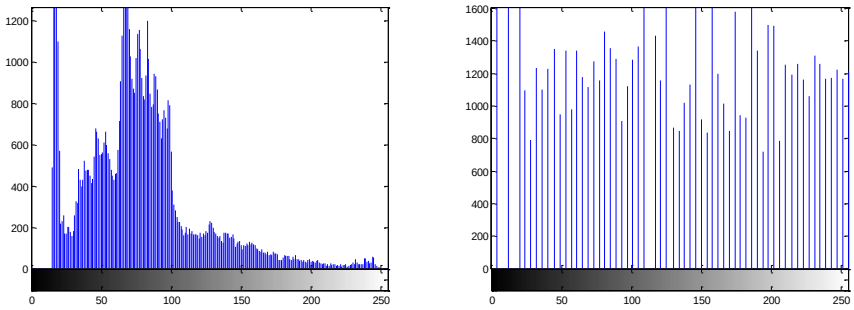
Figure 2. 16: Extension linéaire de dynamique : a)Original, b) Recadrage : a = 30, b = 150

6.1.3.2 Egalisation d'histogramme

Cette étape consiste à rendre l'histogramme de l'image aussi uniforme que possible afin de donner une chance d'apparition équiprobable pour tous les niveaux de gris des pixels. Cette technique améliore le contraste et permet d'augmenter artificiellement la clarté d'une image grâce à une meilleure répartition des intensités relatives.



(a) (c)

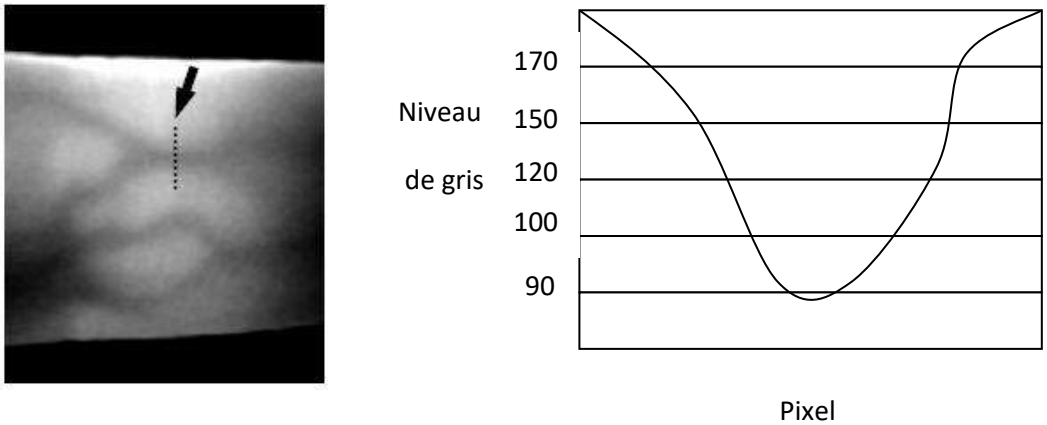


(b) (d)

Figure 2. 17 :(a) Image originale, (b) Histogramme de l'image originale, (c) Image égalisée(c) Histogramme de l'image égalisée.

6.2 Extraction des caractéristiques

Pour l'exploitation de l'information des doigts, Miura et al. ont proposé une méthode de suivi de ligne répétée ou Repeated Line Tracking (RLT) [27]. Pour l'extraction robuste des modèles de veine du doigt à partir des images non uniformes, dans la méthode RLT, l'extraction des motifs est basée sur le nombre de fois que les lignes de suivi passent à travers les points. Le profil transversal d'une veine apparaît comme une vallée, comme indiqué dans la Figure 2.18. La profondeur de la vallée varie avec l'ombrage de l'image. Cependant, la vallée reste détectable. Par conséquent, les profils nous donnent une méthode robuste de détection des veines. L'opération de suivi de ligne commence à n'importe quel pixel de l'image capturée. La position des pixels est appelée "point de suivi actuel" et est déplacée pixel par pixel le long de la ligne sombre. La profondeur du profil transversal est vérifiée autour du point de suivi actuel.



(a) (b)

Figure 2. 18: Section transversale a)Position sur l'image des veines. b) Profil de luminosité transversal d'une veine

La figure 2.19 montre un exemple de la relation spatiale entre le point de suivi actuel (Pa) et le profil transversal. Le pixel P est un voisin du point de suivi actuel Pa dans la direction supérieure droite. Le profil Transversal s-p-t ressemble à une vallée. Par conséquent, le point de suivi actuel est sur une ligne sombre. La direction de cette ligne sombre peut être détectée en vérifiant la profondeur de la vallée avec θ variable. Cela nous donne le θ auquel la vallée est la plus profonde. Après cela, le point de suivi actuel se déplace vers le pixel le plus proche de cette direction, pixel p. Si la vallée n'est détectable pour aucune direction θ, le point de suivi actuel n'est pas sur une ligne sombre et une nouvelle opération de poursuite commence à une autre position.

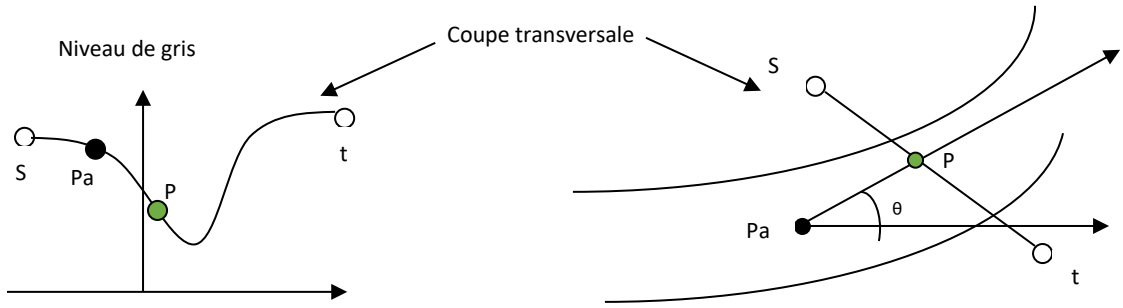


Figure 2. 19 : Relation spatiale entre le point de suivi Pa et le profil transversal

Si une opération de suivi de ligne est effectuée, seulement une partie des veines de l'image sera suivie. Pour résoudre ce problème, des séquences de suivi de veines sont démarrées à différentes positions qui sont déterminées de sorte que les essais de suivi de ligne sont menés uniformément à travers l'image. Le point de suivi actuel peut suivre une région de bruit par hasard, cependant, les lignes sombres sont de plus en plus suivies plus souvent avec des opérations répétées. Cela rend une extraction robuste des modèles de veines du doigt. Le nombre de fois que chaque pixel est devenu le point de suivi actuel est enregistré dans une matrice appelée "espace de lieu". La taille de l'espace du lieu a le même nombre de pixels que l'image capturée. Le nombre total d'essais sur lesquels chaque pixel est devenu le point de suivi actuel est enregistré dans l'élément de matrice.

En toute insuffisance, cette méthode n'est pas adaptée à la description de la veine mince et sensible au bruit. Compte tenu du faible contraste entre les régions vasculaires et non vasculaires, une autre approche utilise un seuil adaptatif local pour la segmentation des doigts [28]. Sur la base de cette méthode, une amélioration a été proposée dans [29]. Cependant, il est difficile d'obtenir un compromis entre la réduction du bruit et la préservation du réseau veineux en utilisant les méthodes basées sur le seuil.

Les caractéristiques de la veine du doigt peuvent être extraites à l'aide d'un détecteur de largeur de ligne (wide line detector) [30]. Les caractéristiques de la veine ont été présentées par des lignes avec une largeur différente, et tous les points sur les lignes de veine de l'image sont extraits. Ce détecteur de largeur de ligne utilise un filtre non linéaire pour extraire complètement une ligne. Les réponses isotropes sont obtenues en utilisant des masques circulaires qui contiennent soit une pondération constante normalisée, soit un profil gaussien. Cette méthode comporte deux paramètres, la taille du masque circulaire et du seuil de contraste de luminosité qui peut être automatiquement sélectionné. Le détecteur largeur de ligne est robuste contre le bruit car la détection n'est pas basée sur des dérivés d'images. Par conséquent, il est efficace pour les applications pratiques dans lesquelles le bruit est inévitable.

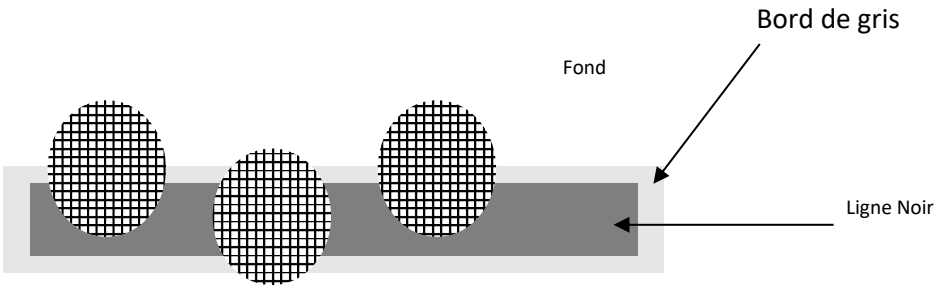


Figure 2. 20: Masques circulaires pour détecter la largeur de ligne veine.

Basé sur la considération que les veines du doigt apparaissent toujours sous forme de crêtes noires, des méthodes basées sur la courbure ont été utilisées pour détecter les centres des veines.

Yu, Qin, Cui et Hu [33] ont extrait des modèles de veines en fonction des caractéristiques de minuties. Ces points caractéristiques sont utilisés comme représentation géométrique de la forme des motifs veineux. Les caractéristiques minuties sont extraites des motifs de la veine pour la reconnaissance, qui incluent les points de bifurcation et les points terminaisons. Ces éléments sont utilisés comme représentation géométrique de la forme des motifs veineux.

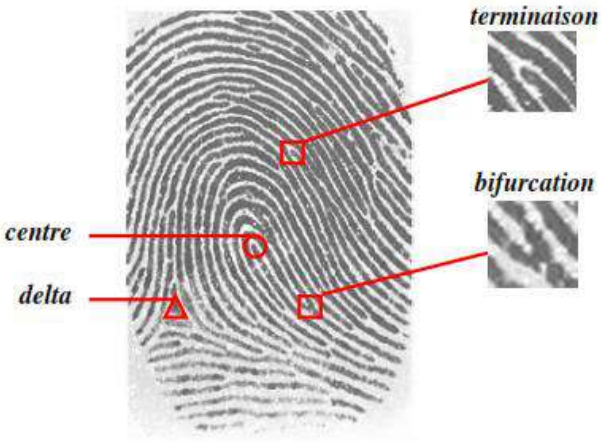


Figure 2. 21 : Caractéristiques des empreintes digitales.

Les méthodes basées sur les minuties sont reconnues comme une branche importante, qui tente de découvrir des modèles de minutie à partir d'images de veines de doigts pour l'appariement et la reconnaissance. Cependant, l'exactitude de ces méthodes n'est généralement pas satisfaisante. L'un des problèmes les plus difficiles est que les correspondances de deux ensembles de minuties sont difficiles à obtenir résultant de la rotation, et de la déformation des images de la veine du doigt[30]. L'application de la méthode de minuties pour la reconnaissance de la veine digitale n'a pas donné un bon résultat malgré que cette méthode est très reconnue, l'image de la veine ne comporte pas beaucoup de descripteurs de caractéristiques, donc un nombre faible de minuties qui pourra être obtenu (figure 2.28) [31].

Song et al. (2011) [32] ont proposé un nouveau système de vérification de la veine du doigt en utilisant la courbure moyenne. Une approche alternative pour l'extraction des fonctions de la veine du doigt utilisant des filtres Gabor a été proposée dans [20]. La forme analytique des filtres Gabor 2D dans le domaine spatial peut s'exprimer comme suite:

$$h(p)_\theta = \frac{1}{2\pi|C|^{\frac{1}{2}}} cos\omega_m^T (p_n - p_0) \exp\left[-\frac{1}{2}(p_n - p_0)^{\mathrm{T}} \mathrm{C}^{-1}(p_n - p_0)\right]^{\mathrm{T}} \quad (2.2)$$

Il existe d'autres méthodes comme les méthodes basées sur la zone locale, où les caractéristiques extraites sont en forme binaire, comme le modèle binaire local (LBP), le modèle binaire linéaire local (LLBP) [33], les cartes bitmap personnalisées (Personalized Best Bit Map PBBM), et le code directionnel local (LDC) [87].

Naoto Miura (2007) [31] a développé un algorithme de calcul des courbures maximales locales (MC) dans les profils transversaux d'une image de veine. Cet algorithme se compose de trois étapes (figure 2.22). Tout d'abord, une extraction des positions centrales des veines est effectuée en calculant la courbure du profil de la section transversale de l'image. Deuxièmement, la connexion des positions centrales

par une opération de filtrage spécifique et enfin l'étiquetage de l'image, les pixels avec des valeurs inférieures au seuil sont étiquetés comme des parties du fond, et ceux qui ont des valeurs supérieures ou égales au seuil sont étiquetés comme parties de région veineuse.

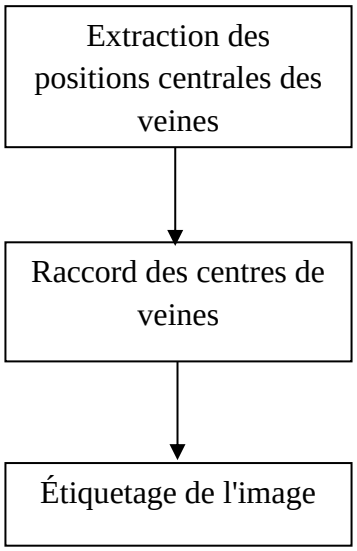


Figure 2. 22 : schéma bloc des étapes de la méthode MC

Pour extraire les lignes des veines avec de nombreuses largeurs et de luminosités, cette méthode vérifie les profils transversaux d'une image de la veine. Le profil en coupe transversale d'une veine ressemble à une dent parce que la veine est plus foncée que le reste de la région des veines, comme le montre la Figure 2.18. Ces courbes concaves présentent de grandes courbures. Même si les veines étroites ou larges, lumineuses ou sombres sont représentées dans une image (positions A, B et C de la figure 2.23) et la position centrale des veines n'a pas de luminosité minimale locale (position C), les courbures des profils des veines sont importants. Par conséquent, la position centrale des veines peut être obtenue en calculant les courbures maximales locales dans les profils transversaux.

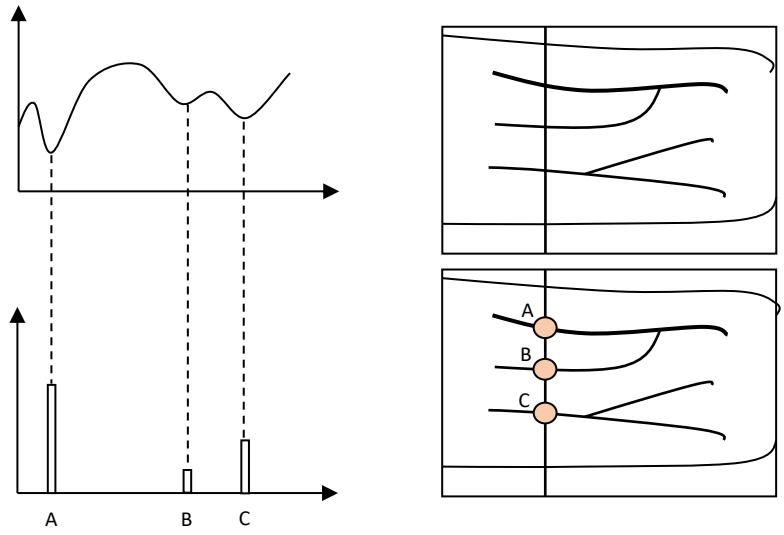


Figure 2. 23 :Profils transversaux d'une image de la veine

Pour obtenir l'extension de la veine dans une image entière, tous les profils d'une direction sont analysés. Les directions utilisées sont horizontales, verticales et les deux directions obliques coupant l'horizontale et la verticale à 45°. Ainsi, toutes les positions centrales des veines sont détectées en calculant les courbures maximales locales.

Une opération de filtrage est appliquée pour brancher les centres de veines, elle est basée sur la vérification des valeurs des pixels voisins. On considère par exemple un pixel p, on vérifie les pixels des deux côtés, s'ils ont de grandes valeurs, on dessine une ligne horizontale. Dans le cas contraire ce pixel est considéré comme un bruit qui doit être éliminé. Ce calcul est fait à tous les pixels et pour toutes les directions.

Enfin, un seuillage est appliqué pour distinguer deux régions, la région veineuse formée par étiquetage des pixels dont les valeurs sont supérieures ou égales au seuil, les reste sont étiquetés comme des parties de l'arrière-plan.

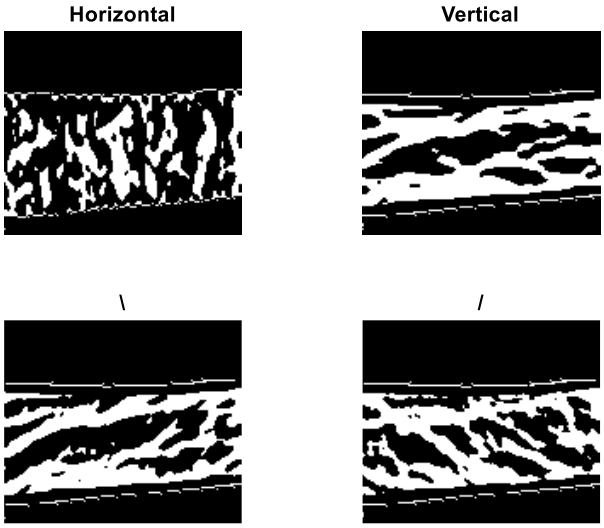


Figure 2. 24 :Quatre images binaires produites après avoir étiqueté les pixels.

La figure 2.24 illustre l'image de la veine du doigt tirée de la base de données SDUMLA-HMT [36] et la veine extraite par MC.



(a) (b)

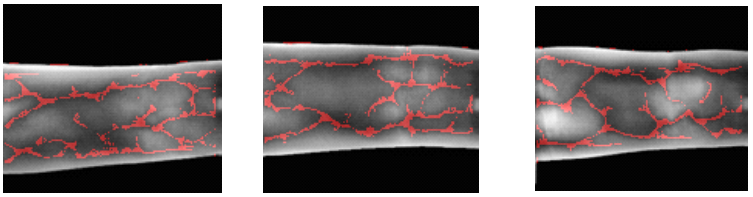
Figure 2. 25 : application de la méthode MC, a) Image originale, b) Image des veines

La qualité de l'image de sortie de l'algorithme MC n'est pas bonne et doit être améliorée. Pour cela, des opérations morphologiques sont utiles, telles que les opérations d'érosion et de dilatation. L'opération de dilatation morphologique élargis les lignes de veines. La forme de la veine a un impact sur le résultat final de reconnaissance. La figure 2.26 montre un exemple d'image de sortie de l'étape de post-traitement.

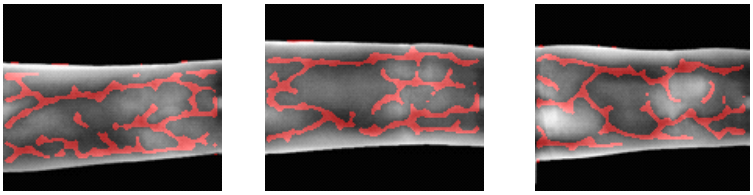


Figure 2. 26 :Images des veines: a) Image des veines, b) Image post-traitée.

La qualité de la structure de la veine des doigts obtenue peut être évaluée en utilisant uniquement la visualisation humaine, comme le montre la figure 2.27. Les résultats montrent que les images contenant les lignes de veine après post-traitement de l'image extraite ont une meilleure qualité d'image. En fait, l'observation visuelle confirme que l'image veineuse améliorée contient moins de vacillation par rapport à l'image obtenu par MC [35].



(a)



(b)

Figure 2. 27 :Superposition du motif veineux sur : a) Image MC, b)Image MC amélioré

6.3 L'appariement des caractéristiques

Généralement, l'appariement des veines s'effectue en utilisant les caractéristiques extraites telles que les minuties, le modèle binaire local, SIFT, les bio-métriques souples, les mesures statistiques, l'apprentissage par machine, les méthodes de corrélation (ou modèle) et algorithmes hybrides [25]. L'utilisation de minuties pour l'appariement des veines requiert une extraction supplémentaire de points caractéristiques tels que les terminaisons de ligne et les bifurcations. Étant donné qu'un modèle de veine de doigt présente peu de ces points, l'appariement des modèles basé sur la comparaison des valeurs de pixels est plus approprié pour l'appariement des motifs de la veine.

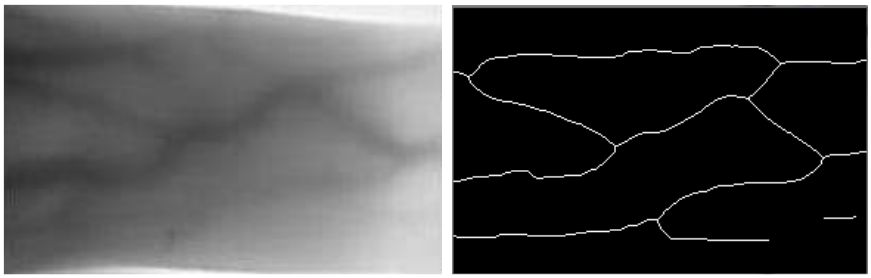


Figure 2. 28 :Terminaisons de ligne et des bifurcations utilisées pour l'appariement

La corrélation ou l'appariement basé sur un modèle donne des résultats précis même lorsque des veines authentiques sont manquantes ou que des veines parasites sont introduites. Il élimine une déformation rigide à travers l'enregistrement d'image et peut donc être désigné comme correspondant en fonction d'un enregistrement rigide.

Dans [21], les scores de correspondances ont été mesurés en appliquant une corrélation de phase (POC). La corrélation entre deux images est traitée dans le domaine de la fréquence en considérant uniquement les informations de phase. Le score correspondant est déterminé en calculant la mesure de similarité en fonction du pic de corrélation de phase limitée en bande.

Dans la tâche d'appariement d'image, la propriété remarquable de la méthode POC réside dans sa capacité à fournir une grande capacité de discrimination basée sur sa valeur de crête de corrélation. La figure 2.30.a montre un exemple d'image veineuse correspondante en utilisant la fonction POC d'origine pour deux images identique. Lorsque deux images de veine identiques sont associées, leur fonction de corrélation de phase donne un pic pointu distinct. Mais, lorsque les deux images sont différentes, le pic de corrélation de phase baisse énormément comme il est montré dans la figure 2.31.a.

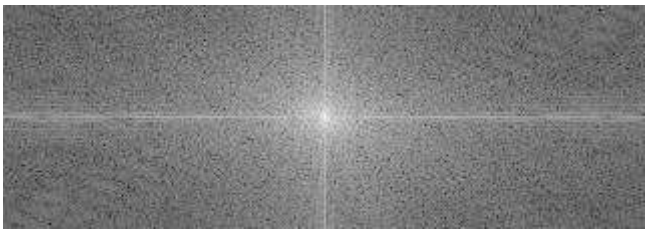
Ito et al. [34] ont suggéré que la performance de correspondance pourrait être améliorée si les composantes de phase dans le spectre de fréquence non dominant sont exclus du calcul du POC. Généralement, dans un cas idéal, les informations de l'image de la veine sont principalement concentrées dans le domaine de basse fréquence, alors que l'information moins significative est généralement concentrée à la plus haute fréquence. Par exemple, comme le montre la Figure 2.29.b, le spectre à basse fréquence met en évidence les informations détaillées de l'image de la veine, ce qui est indiqué par une région beaucoup plus brillante et répartie autour du centre du spectre d'image.

L'idée de base consiste à éliminer ces composantes à haute fréquence sans importance est d'imposer une limite de bande sur la gamme de fréquences d'origine lors du calcul du spectre cross-phase. Cette méthode est également connue sous le nom Bande limitée de corrélation en phase seule « Band-Limited Phase-Only Correlation » (BLPOC).

Les figure 2.30 et 2.31 montrent des exemples d'appariement authentique et imposteur utilisant les deux fonctions POC et BLPOC respectivement.



(a)

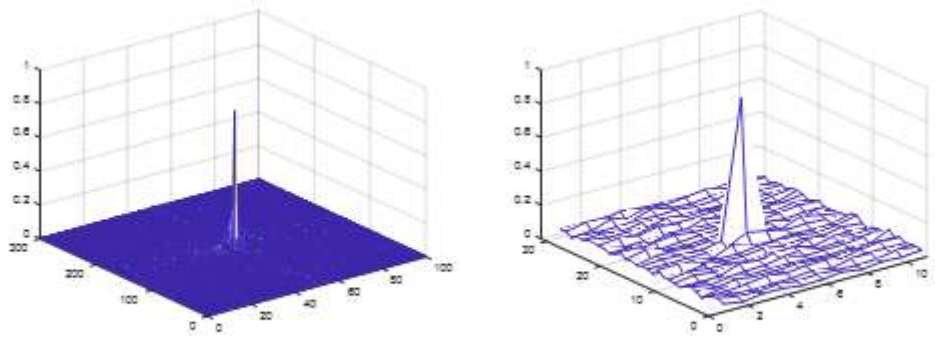


(b)



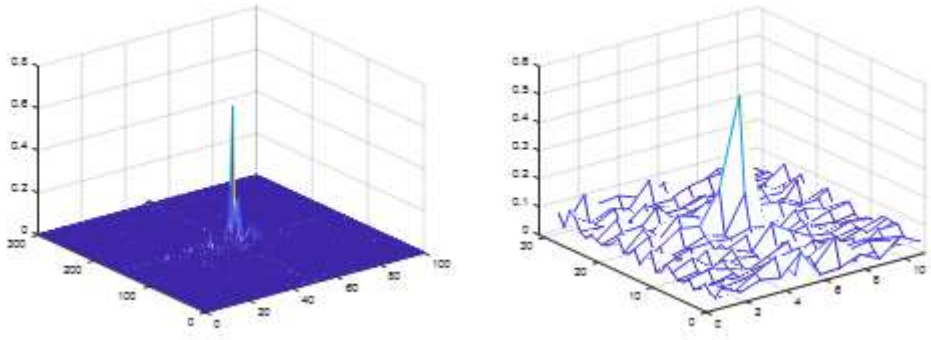
(c)

Figure 2. 29 :a) Image filtrée par le filtre de Gauss, b) Spectre de fréquence c) Spectre de phase



(a) (b)

Figure 2. 30 : Résultat d'appariement Deux photos authentiques en utilisant : a) POC , b) BLPOC.



(a) (b)

Figure 2. 31: Résultat d'appariement Deux photos imposteurs en utilisant : a) POC , b) BLPOC.

7 Conclusion

Dans ce chapitre, nous avons décrit les différentes étapes constituant le système de reconnaissance par les veines de doigt. Nous avons cité quelques exemples des premiers travaux établis sur cette modalité. Le système d'authentification des veines de doigts combine une facilité d'utilisation et un haut niveau de sécurité et de précision. Ce système a l'avantage d'être résistant aux contrefaçons car le motif se trouve dans le doigt. Nous avons exposé quelques bases de données publiques existantes dont la plupart sont accessibles uniquement à des fins de recherche. Les technologies existantes possèdent diverses techniques d'extraction des caractéristiques des veines du doigt. Une nouvelle méthode d'identification des veines devrait être proposée pour surmonter les faiblesses des méthodes classiques cités dans ce chapitre.

CHAPITRE 3 CONTRIBUTION À L'AMÉLIORATION DE LA QUALITÉ DES IMAGES DES VEINES DE DOIGT PAR FUSION DCT

1 Introduction

L'amélioration de l'image est un vaste domaine d'étude dédié à l'amélioration de la qualité d'une image d'entrée, renvoyant une image de sortie qui est plus appropriée que l'image originale à des fins de visualisation simple ou pour une application spécifique.

De nombreuses études ont montré que les images de la veine de doigt étaient de mauvaise qualité en raison de la variation dans les tissus et de la luminosité inégale. Par conséquent, il existe un besoin de techniques efficaces pouvant améliorer la qualité des images.

Dans ce chapitre, un état de l'art des méthodes d'amélioration existantes sont dévoilé, ensuite, nous proposons une nouvelle approche d'amélioration de la qualité d'image des veines de doigt basée sur la fusion DCT (Discrete Cosine Transform).

2 Etat de l'art des méthodes d'amélioration de l'image des veines de doigt

Pour améliorer le contraste dans les images de la veine des doigts, les chercheurs ont utilisé la fonction PSF (Point Spread Function) pour résoudre les problèmes de flou à l'aide d'un modèle de diffusion spécifique basé sur les propriétés optiques de la peau des doigts. Cependant, il y avait un phénomène de dispersion de la lumière provenant de la peau et des tissus internes des doigts comme les muscles, les vaisseaux sanguins et même les os. En raison de ces problèmes, la PSF n'est pas suffisante pour améliorer efficacement la visibilité du motif veineux au doigt [25].

Pour améliorer l'image IR. Zang et al. [40] ont proposé un algorithme flou à plusieurs seuils. Premièrement, un seuil amélioré a été obtenu en calculant la moyenne

de tous les points adjacents, suivi du comptage de la moyenne pour affilier chaque point. Ensuite, une transformation inverse a été appliquée pour obtenir les images.

Cong-li et al. [41] ont proposé une méthode pour extraire la caractéristique de la veine du doigt en détectant le pic et la vallée en utilisant une opération morphologique. Cette méthode est faible dans l'amélioration de la veine du doigt car l'image IR a un faible contraste. Pour améliorer le contraste, Zhao et al. [42] ont proposé un algorithme qui applique un filtrage d'accentuation haute fréquence suivi d'une égalisation d'histogramme.

Yang [29] a utilisé un filtre Gabor à canaux multiples (Multi-Channel Gabor Filter MCGF) pour extraire des informations d'une image de veine de doigt de faible qualité. Cette méthode a pu améliorer l'image après avoir appliqué la règle de reconstruction sur les informations de la veine. Une autre amélioration de l'image a été proposée par Zhang et al. [43] qui ont utilisé un filtre circulaire de Gabor avec un algorithme de groupement de niveaux de gris (Gray-Level Grouping GLG). Le filtre circulaire de Gabor a été utilisé après traitement d'image par GLG pour mettre en évidence les crêtes des veines. Une méthode pour extraire des informations géométriques de l'image a été proposée par Jianping et al. [44]. Elle est basée sur la transformée de contourlet non sous-échantillonnée. Cette technique permet de distinguer les bruits des bords faibles. Un banc de filtres d'amélioration qui génère de meilleures performances a été proposée par Truce et al. [45]. Il a été utilisé pour l'amélioration d'un banc de filtres directionnels. Une succession d'opérations de filtrage a été proposée par Jinfeng [46] pour réduire le bruit et améliorer la qualité de l'image veineuse. Tout d'abord, une opération d'extraction de la région d'intérêt a été réalisée pour effacer les lumières présentes en arrière-plan de l'image, puis l'image a été filtrée par un banc de filtres directionnels. Ensuite, l'image directionnelle a été filtrée par un filtre adaptatif de Wiener et suivie par le filtre de Frangi. Enfin, une règle de reconstruction a été appliquée. Dans leur étude [47], Park et al. ont proposé et développé une technique d'amélioration de la qualité d'image, qui a pris en compte l'épaisseur et la direction des lignes veineuses à l'aide d'un filtre de Gabor optimal. Dans une autre étude [48], les chercheurs ont proposé une technique d'évaluation de la qualité d'image pour les images des veines des doigts; cependant, ils n'ont pas pris en compte l'amélioration de la qualité. Ils ont utilisé un filtre gaussien modifié (MGF) [22, 21] pour améliorer les images des veines des doigts.

3 Méthode de Fusion par DCT

Dans cette section, nous proposons une nouvelle technique qui améliore la qualité d'image des veines des doigts. Ce procédé comprend l'amélioration du contraste, l'utilisation du filtre de Gabor et la fusion d'images qui génère une image ayant des motifs connectifs. Cette méthode d'amélioration est présentée dans la figure 3.7. Deux opérations sont appliquées à l'image d'entrée, la première opération consiste à appliquer le filtre de Gabor et la seconde est l'égalisation d'histogrammes adaptatifs

à contraste limité CLAHE (Contrast Limited Adaptive Histogram Equalization) [28], suivie d'un filtrage à l'aide d'un filtre médian. Les deux images sont ensuite fusionnées par la méthode DCT (Discrete Cosinus Transform). Enfin, un ajustement du contraste est effectué sur l'image résultante. Les détails de chaque étape sont décrits ci-dessous.

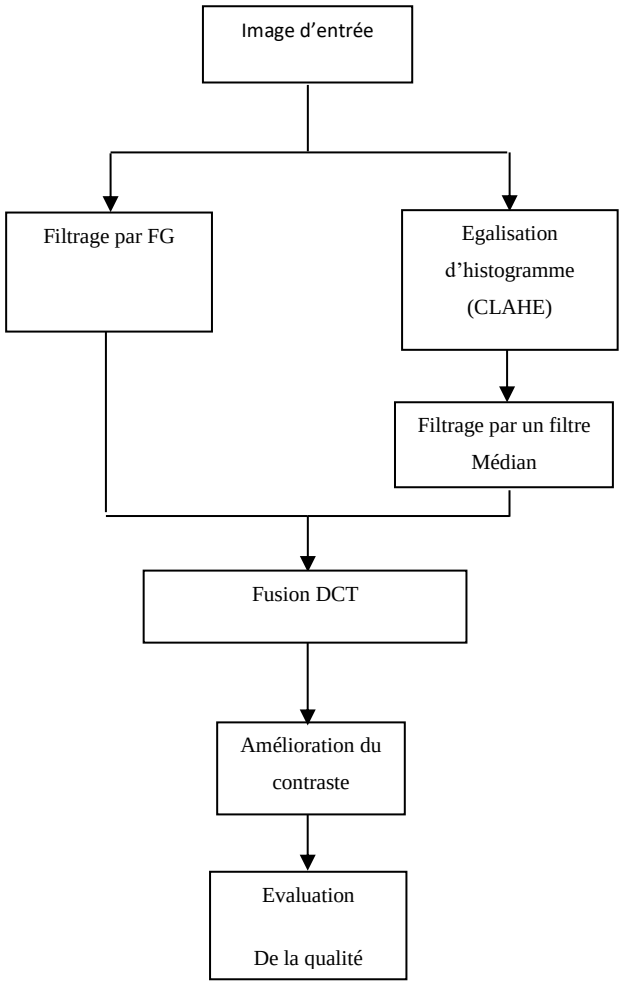


Figure 3. 1: Amélioration proposée de l'image des veines des doigts

3.1 Égalisation d'histogramme

Cette technique est parmi les plus populaires pour le rehaussement de contraste des images, proposée par Zimmerman et Pizer. Elle s'appuie sur l'histogramme des niveaux d'intensité des pixels de l'image. En effet, la forme de l'histogramme donne une idée sur la qualité perceptible globale de l'image, elle est sombre ou lumineuse suivant la répartition des niveaux d'intensité. L'objectif de l'amélioration des images en utilisant l'égalisation d'histogramme est d'obtenir une répartition uniforme et un ajustement de l'intensité [49].

Tout d'abord, certains chercheurs ont appliqué la technique d'égalisation d'histogramme d'image pour ajuster la dispersion des niveaux de gris dans les images.

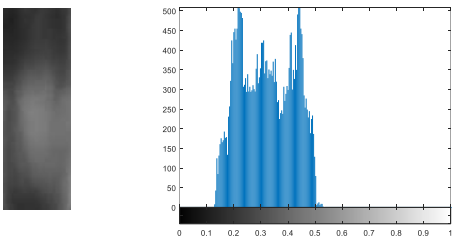
Cependant, cette technique n'a pas pu être appliquée aux images des veines du doigt [32]. Une nouvelle technique, à savoir la méthode d'égalisation d'histogrammes adaptatifs CLAHE, restreint la taille des histogrammes locaux, elle consiste à manipuler localement l'histogramme de l'image. Cette méthode limitait l'amplification du bruit et augmentait le contraste local dans les images.

Cependant, si l'image est constituée de certaines régions nettement plus sombres ou plus claires que l'image restante, le contraste dans ces régions n'est pas amélioré. Cette technique fonctionne sur les plus petites régions de données (tuiles), au lieu de l'image complète. Le contraste de chaque mosaïque est amélioré, ce qui fait correspondre l'histogramme de la région résultante à l'histogramme indiqué. Ensuite, les tuiles voisines sont combinées en fonction de l'interpolation bilinéaire afin d'éliminer les limites induites artificiellement dans l'image [50].

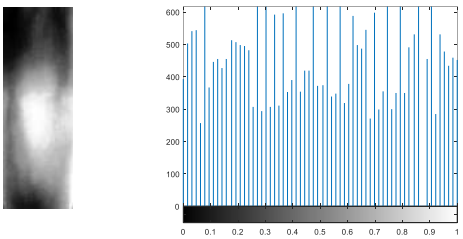
Ici, la CLAHE a été calculée comme suit:

$$S = Hist(i) \times \frac{255}{M \times M} \tag{3.1}$$

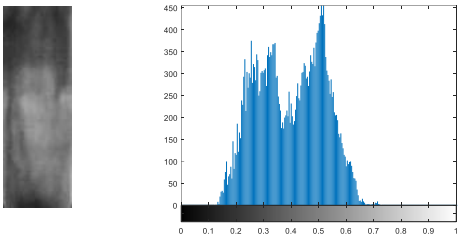
Où; Hist fait référence à la distribution cumulative de l'histogramme. Cela reflète également le nombre des niveaux de gris qui apparaissent dans une région spécifique dans l'image i. S fait référence aux nouvelles valeurs de pixel, tandis que M indique la taille de la fenêtre locale. Cette étape d'égalisation est suivie par un filtrage par un filtre médian qui diminue le bruit de l'image et les autres irrégularités de l'image afin d'améliorer la qualité visuelle de celle-ci. Ce résultat est montré dans la figure 3.8 [51].



(a)



(b)



(c)

Figure 3. 2: (a) image d'origine et son histogramme, (b) l'image prétraitée avec égalisation d'histogramme et son histogramme (c) l'image prétraitée avec CLAHE et son histogramme.

3.2 Filtrage par filtre de Gabor

Les chercheurs ont appliqué les filtres de Gabor dans de nombreuses applications d'analyse d'images, ces filtres étant réglables en orientation et en échelle. Le filtre de Gabor à deux dimensions apparaît comme une fonction de forme gaussienne associées à une onde plane complexe. Ceci est estimé comme suit [52]:

$$g(x,y) = s(x,y).w_r(x,y) \tag{3.2}$$

où s (x, y) est une sinusoïde complexe, connue sous le nom de porteuse, et w_r (x, y) est une fonction gaussienne 2D, connue sous le nom d'enveloppe.

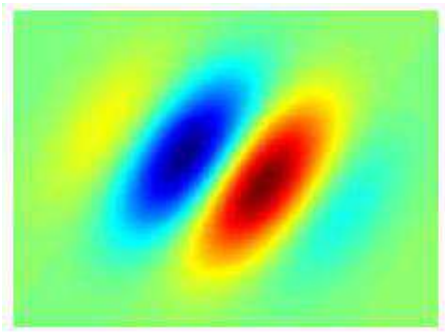


Figure 3. 3 : Exemple de filtre Gabor 2D

3.2.1 La porteuse sinusoïdale complexe

La sinusoïde complexe est définie comme suit,

$$s(x,y) = \exp\big(i(2\pi(u_0x + v_0y) + P)\big) \tag{3.3}$$

Où $(u_0x + v_0y)$ et P définissent respectivement la fréquence spatiale et la phase de la sinusoïde.

Cette sinusoïde peut être représentée comme deux fonctions réelles distinctes, classiquement allouées dans la partie réelle et imaginaire d'une fonction complexe. La partie réelle et la partie imaginaire de cette sinusoïde sont

$$Re\left(s(x,y)\right) = cos(2\pi(u_0x + v_0y) + P) \quad (3.4)$$

$$Im\left(s(x,y)\right) = sin(2\pi(u_0x + v_0y) + P) \quad (3.5)$$

Les paramètres u_0 et v_0 définissent la fréquence spatiale de la sinusoïde en coordonnées cartésiennes. Cette fréquence spatiale peut également être exprimée en coordonnées polaires comme l'amplitude F_0 et la direction ω_0 :

$$F_0 = \sqrt{u_0^2 + v_0^2} \quad (3.6)$$

$$\omega_0 = tan^{-1}\left(\frac{v_0}{u_0}\right) \quad (3.7)$$

$$u_0 = F_0\, cos\omega_0 \quad (3.8)$$

$$v_0 = F_0\, sin\omega_0 \quad (3.9)$$

En utilisant cette représentation, la sinusoïde complexe est

$$s(x,y) = \exp\big(i(2\pi F_0(xcos\omega_0 + ysin\omega_0) + P)\big) \quad (3.10)$$

3.2.2 L'enveloppe gaussienne

L'enveloppe gaussienne se présente comme suit:

$$\omega_r(x,y) = K\ exp\big(-\pi(a^2(x - x_0)_r^2 + b^2(y - y_0)_r^2)\big) \quad (3.11)$$

où $(x - x_0)$est le pic de la fonction, a et b sont des paramètres de mise à l'échelle de la gaussienne, et l'indice r représente une opération de rotation telle que

$$(x - x_0)_r\ = (x - x_0)cos\theta + (y - y_0)sin\theta \quad (3.12)$$

$$(y - y_0)_r\ = -(x - x_0)sin\theta + (y - y_0)cos\theta \quad (3.13)$$

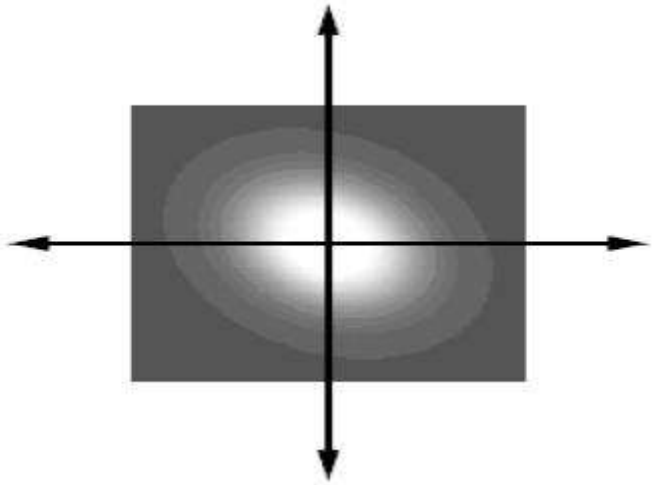


Figure 3. 4 : Une enveloppe gaussienne

3.2.3 La fonction de Gabor complexe

La fonction complexe de Gabor est définie par les 9 paramètres suivants;

K: met à l'échelle l'amplitude de l'enveloppe gaussienne.
(a, b): Met à l'échelle les deux axes de l'enveloppe gaussienne.
θ: Angle de rotation de l'enveloppe gaussienne.
(x_0, y_0): Emplacement du pic de l'enveloppe gaussienne.
(u_0, v_0): Fréquences spatiales de la porteuse sinusoïdale en coordonnées cartésiennes, ça peut être également exprimé en coordonnées polaires comme(F_0, ω_0):
P: Phase de la porteuse sinusoïdale.

Chaque fonction de Gabor complexe se compose de deux fonctions en quadrature (déphasées de 90 degrés), idéalement situées dans les parties réelles et imaginaires d'une fonction complexe.

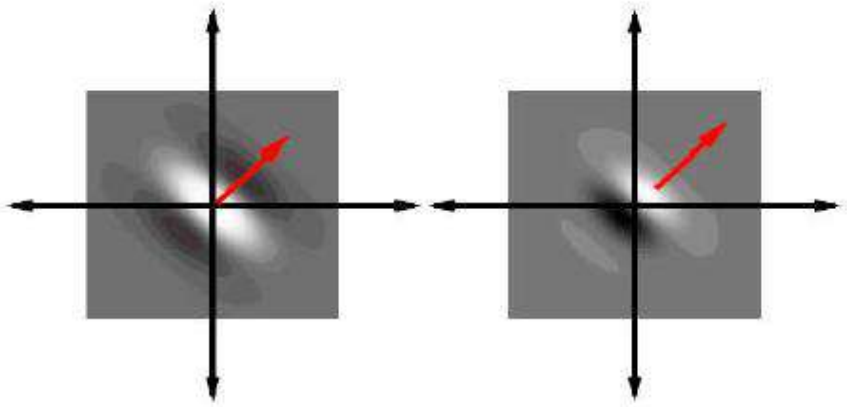


Figure 3. 5 : Les parties réelles et imaginaires d'une fonction complexe de Gabor dans le domaine spatial

Nous avons maintenant la fonction complexe de Gabor dans le domaine spatial:

$$g(x,y) = K\ exp\big(-\pi(a^2(x-x_0)_r^2 + b^2(y-y_0)_r^2)\big).\exp\big(i(2\pi(u_0x + v_0y) + P)\big) \quad (3.14)$$

Ou en coordonnées polaires

$$g(x,y) = Kexp\big(-\pi(a^2(x-x_0)_r^2 + b^2(y-y_0)_r^2)\big)exp\big(i(2\pi F_0(xcos\omega_0 + ysin\omega_0) + P)\big) \quad (3.15)$$

La transformée de Fourier 2D de ce Gabor est la suivante:

$$\hat{g}(u,v) = \frac{K}{ab}exp\left(i\big(-2\pi\big(x_0(u-u_0) + y_0(v-v_0)\big) + P\big)\right)exp\left(-\pi\left(\frac{(u-u_0)_r^2}{a^2} + \frac{(v-v_0)_r^2}{b^2}\right)\right) \quad (3.16)$$

Ou en coordonnées polaires, l'amplitude :

$$Amplitude\big(\hat{g}(u,v)\big) = \frac{K}{ab}\ exp\left(-\pi\left(\frac{(u-u_0)_r^2}{a^2} + \frac{(v-v_0)_r^2}{b^2}\right)\right) \quad (3.17)$$

$$Phase\big(\hat{g}(u,v)\big) = -2\pi\big(x_0(u-u_0) + y_0(v-v_0)\big) + P \quad (3.18)$$

La composante réelle, connu sous le nom de filtre de Gabor à symétrie paire, est utilisée pour la détection de crête, tandis que la composante imaginaire, appelé filtre de Gabor à symétrie impaire, est utilisée pour la détection de bord. L'orientation et la fréquence de ces filtres de Gabor sont efficaces pour améliorer les images. Cependant, aucune méthode ne peut remplacer l'intuition basée sur l'expérience pour la sélection des paramètres.

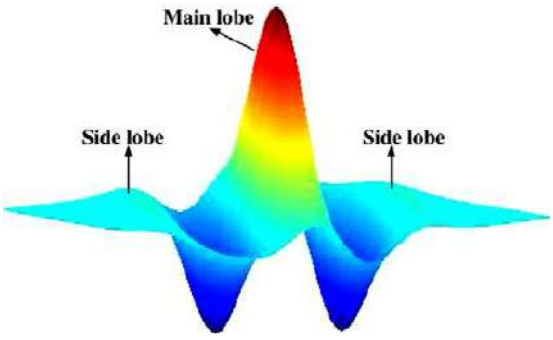


Figure 3. 6 : Ondelette Gabor 2D symétrique

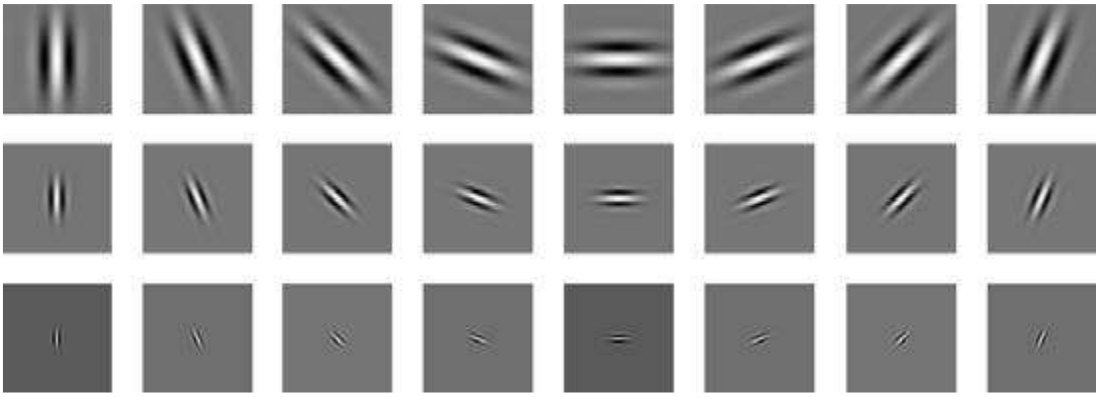


Figure 3. 7 : Réponses spatiales 2D d'un dictionnaire de filtres de Gabor symétriques pairs. Ici, chaque image contient 65 x 65 pixels.

L'image est filtrée en utilisant les parties réelles du filtre Gabor avec quatre orientations 22,5º, 67,5 º, 112,5 º et 157,5 º. La règle de reconstruction est définie comme:

$$RI(x_i, y_i) = argmin([I_1(x_i, y_i), I_2(x_i, y_i), I_3(x_i, y_i), I_4(x_i, y_i)]) \quad (3.19)$$

Où RI (x, y) représente l'image reconstruite et I (x, y) représente l'image filtrée. Un exemple d'image obtenue par MCGF est illustré dans la figure 3.8.



(a) (b)

Figure 3. 8: a)Images filtrées avec quatre orientations: 22,5º, 67,5 º, 112,5 º et 157,5 º, b) L'image reconstruite

3.3 Fusion d'images

La fusion d'images combine toutes les informations pertinentes obtenues à partir de plusieurs images en une seule image. Les techniques de fusion d'images basées sur la transformation en cosinus discrète (DCT) sont très efficaces et permettent de gagner beaucoup de temps dans les systèmes temps réel qui utilisent des normes basées sur la DCT pour les images. La méthode de transformation DCT 2D pour un bloc N × N dans l'image, x (m, n), peut être définie comme suit [53]:

$$d(k,l) = \frac{2\alpha(k)\alpha(l)}{N} . \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} x(m,n). \cos\left(\frac{(2m+1)\pi k}{2N}\right). \cos\left(\frac{(2n+1)\pi k}{2N}\right) \quad (3.20)$$

Avec k = 0, 1, . . . , N - 1 et

$$\alpha(k) = \begin{cases} \frac{1}{\sqrt{2}} & Si\ \ k = 0 \\ 1 & Sinon \end{cases} \quad (3.21)$$

La transformation DCT inverse (IDCT) peut être définie comme suit:

$$x(m,n) = \sum_{k=0}^{N-1}\sum_{l=0}^{N-1} \frac{2\alpha(k)\alpha(l)}{N} . d(k,l). \cos\left(\frac{(2m+1)\pi k}{2N}\right). \cos\left(\frac{(2n+1)\pi k}{2N}\right) \quad (3.22)$$

Où m, n = 0, 1, . . . , N - 1.

Dans l'équation (3.20), d (0, 0) fait référence au coefficient DC, tandis que d (k, l) est le coefficient AC du bloc. De plus, les coefficients de transformation normalisés peuvent être calculés comme suit:

$$\hat{d}(k,l) = \frac{d(k,l)}{N} \quad (3.23)$$

La valeur moyenne, μ, et la variance, σ^2, pour le bloc N × N dans le domaine spatial sont estimées comme suit:

$$\mu = \frac{1}{N^2}\sum_{m=0}^{N-1}\sum_{n=0}^{N-1} x(m,n) \quad (3.24)$$

$$\sigma^2 = \frac{1}{N^2}\sum_{m=0}^{N-1}\sum_{n=0}^{N-1} x^2(m,n) - \mu^2 \quad (3.25)$$

De plus, la variance dans le bloc de pixels N × N est calculée à l'aide de ses coefficients DCT comme suite :

$$\sigma^2 = \sum_{k=0}^{N-1}\sum_{l=0}^{N-1} \frac{d^2(k,l)}{N^2} - \hat{d}^2(0,0) \quad (3.26)$$

La valeur de la variance est supposée être une mesure de contraste dans les applications de traitement d'image. De plus, la technique de fusion, qui utilise la DCT, est basée sur le maximum d'informations disponibles pour chaque bloc de pixels pour les 2 images DCT. Toutes les informations utilisées ici dépendaient des valeurs de la variance pour chaque bloc [54].

La méthode de fusion DCT est basée sur le principe de prendre le bloc qui contient le maximum d'informations. Nous faisons donc une comparaison entre les blocs des deux

images (égalisation d'histogramme (HE) et filtrage de Gabor (GF)). Le bloc contenant la valeur de variance la plus élevée sera choisi et l'autre sera rejeté. À partir des deux images DCT traitées, une image DCT finale a été obtenue en utilisant l'équation (3.27) [50].

$$d_{fus}(i,j) = \begin{cases} d_{HE}(i,j) & Si \quad var_{d_{HE}} > var_{d_{GF}} \\ d_{GF}(i,j) & Si \quad var_{d_{HE}} < var_{d_{GF}} \end{cases} \tag{3.27}$$

3.4 Evaluation de la qualité

Au cours des dernières années, plusieurs méthodes ont été proposées pour l'évaluation de la qualité des images des veines des doigts, qui peuvent être classées en deux catégories: les méthodes basées sur la classification et les méthodes basées sur les scores.

Les méthodes basées sur la classification classent généralement les images des veines des doigts en plusieurs classes. Yang et al. [55] ont utilisé le modèle SVM (Support Vector Machine) pour classer les images des veines des doigts en deux types: basse qualité et haute qualité. Wang et al. [56] ont extrait cinq caractéristiques de chaque image en tant que paramètres d'évaluation de la qualité de l'image de la veine du doigt, et les images sont classés en trois types, à savoir, de faible qualité, de meilleure qualité et de haute qualité.

En revanche, les méthodes basées sur le score calculent toujours un score de qualité pour chaque image de la veine du doigt. Xie et al. [57] ont présenté une méthode d'évaluation de la qualité des images des veines des doigts basée sur la caractéristique hiérarchique des veines et des scores de qualité calculés à partir de la qualité du motif veineux majeur et de la qualité veineuse mineure. Ma et al. [58] ont présenté un rapport signal / bruit basé sur le système visuel humain comme indice d'évaluation de la qualité, et l'ont intégré à quatre autres indices d'évaluation de la qualité pour obtenir le score de qualité d'image total de l'image de la veine du doigt. Qin et al. [59] ont divisé une image de la veine du doigt en un ensemble de blocs qui ne se chevauchent pas et ont évalué un score de qualité local pour chaque bloc en fonction de la courbure dans l'espace Radon correspondant, sur la base duquel un score de qualité global de l'image de la veine du doigt est calculé.

Dans de nombreux cas, les méthodes basées sur le score sont supérieures aux méthodes basées sur la classification car les méthodes basées sur la classification sont trop arbitraires. Par exemple, il existe deux images des veines des doigts, la première image avec un score de qualité 59 peut être classée en faible qualité mesurée par la méthode de classification, tandis que la seconde avec un score de qualité 60 peut être classée en haute qualité. Dans ce cas, ils seront confrontés à un traitement très différent, même leur différence de qualité est très faible. Cependant, il existe encore des limites aux méthodes basées sur les scores existants. Certaines des méthodes précédentes

pondèrent linéairement les fonctionnalités d'évaluation pour obtenir un score de qualité d'image, d'autres utilisent une seule fonctionnalité pour obtenir le score. Mais la pesée est un modèle linéaire qui se limitait à résoudre des problèmes non linéaires et une seule caractéristique ne peut pas évaluer de manière globale la qualité des images des veines des doigts [60].

Pour comparer la qualité des images améliorées avec les bases de données des veines des doigts disponibles, nous avons évalué les valeurs de niveau de gris moyennes, l'entropie des images et le contraste de l'image. Le contraste de l'image et l'entropie ont été considérés comme les critères représentatifs pour évaluer la qualité de l'image [84]. Les valeurs de niveau de gris moyennes reflètent la volatilité des valeurs de gris dans l'image. En supposant que I (x, y) est l'image, les valeurs de niveau de gris moyennes ont été déterminées comme suit:

$$M = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} I(i,j)}{m \times n} \tag{3.28}$$

Le contraste de l'image représente la différence de niveau de gris dans l'image [54]. Pour déterminer le contraste de l'image, nous avons partitionné l'image en t sous-images non chevauchées de résolution 8 × 8. Le contraste de l'image peut être décrit comme le contraste moyen des valeurs de variance pour chaque bloc, déterminé comme suit:

$$C_t = \sqrt{\frac{\sum_{k=1}^{64}(I_k - M_t)^2}{64}} \tag{3.29}$$

$$Contraste = \frac{1}{t}\sum_{i=1}^{t} C_t \tag{3.30}$$

où; M_t et C_t se réfèrent aux valeurs de gris moyennes et à la variance dans la $t^{ème}$ image secondaire, respectivement. En outre, I_k fait référence à l'une des valeurs de niveau de gris dans les 64 pixels. L'entropie indique une mesure statistique du caractère aléatoire qui caractérise la texture de l'image. L'entropie de l'image est calculée comme suit:

$$Entropy = -\sum_{i=1}^{m}\sum_{j=1}^{n} I(i,j)\log(I(i,j)) \tag{3.31}$$

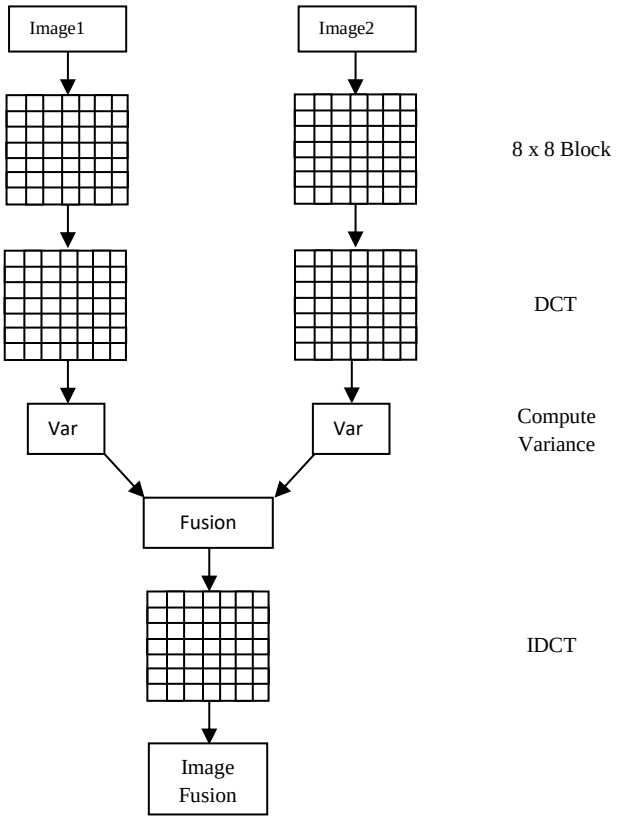
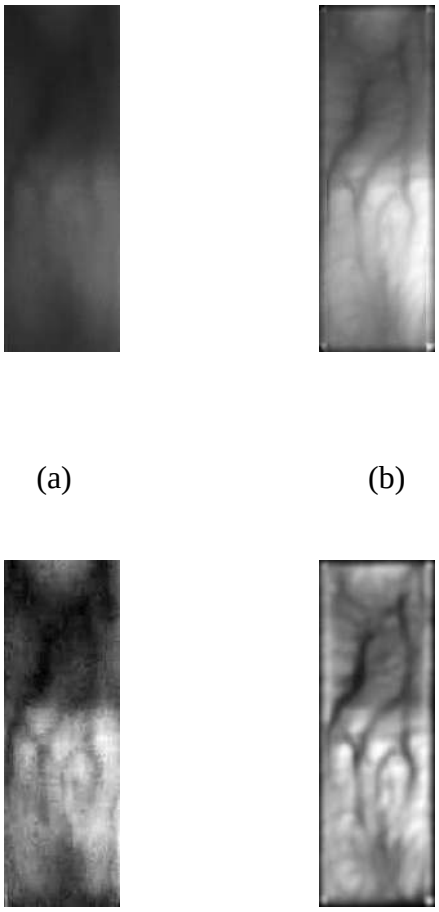


Figure 3. 9: Fusion d'images à l'aide de la technique DCT

3.5 Résultats

Nous avons utilisé deux bases de données, la base de données FV-USM et la base de données SDUMLA. Les images de la base de données FV-USM ont été composées à partir de 123 personnes. Chaque sujet a fourni quatre doigts, soit un total de 492 classes de doigts obtenues avec 6 images dans chaque classe, soit un total de 2952 images de veines de doigt. Nous utilisons uniquement les images extraites du fond des images fournies dans cette base de données. Dans le jeu de données SDUMLA, il était demandé à chaque personne de soumettre des images de son majeur, de son index et de son annulaire pour les deux mains. Les images pour les 6 doigts ont été collectées 6 fois, pour générer un total de 36 images dans chaque classe. Le jeu de données de la veine du doigt était constitué de 3816 images. Les figures 3.16 et 3.17 montrent deux exemples d'images de sortie en appliquant notre méthode proposée.



(a) (b)

(c) (d)

Figure 3. 10: Exemple de résultats obtenus de traitement d'une image de la base FV-USM: a) Image de veines, b) Image améliorée MCGF, c) Image améliorée CLAHE, d) Fusion d'images améliorées.

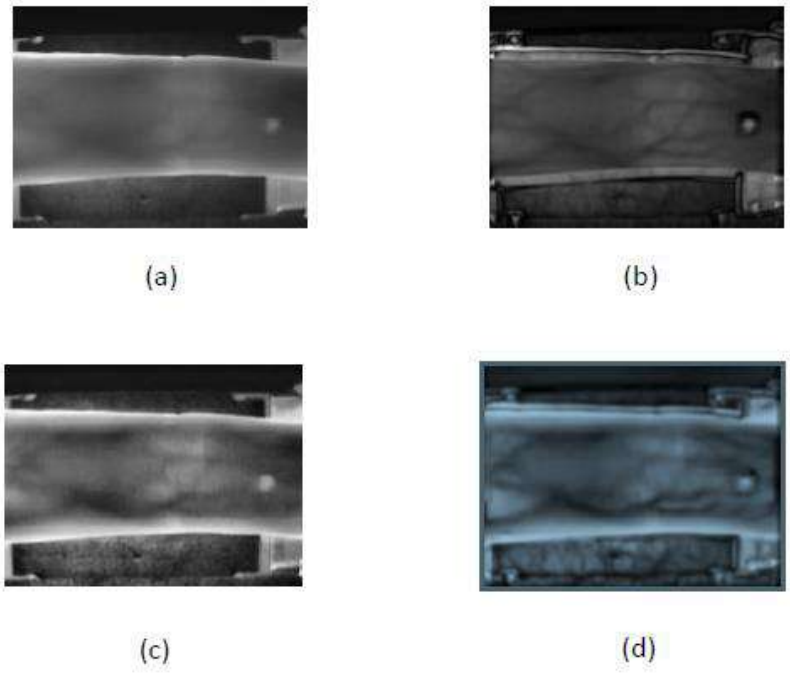


Figure 3. 11: Exemples de résultats obtenus à l'aide d'image SDUMLA: a) Image de veines, b) Image améliorée MCGF, c) Image améliorée CLAHE, d) Fusion d'images améliorées.

Nous avons comparé les valeurs moyennes de niveau de gris, le contraste et l'entropie de toutes les images. Les images de bonne qualité affichent des valeurs plus élevées par rapport aux images de faible qualité. Les résultats obtenus sont répartis dans les tableaux 3.1 et 3.2.

Les résultats dans le cas de deux images des deux bases de données sont présentés dans la figure 3.18. Nous avons remarqué que la valeur de la moyenne de niveau de gris, le contraste et l'entropie deviennent plus élevés lorsque nous avons appliqué des algorithmes d'amélioration aux images originales dans les deux ensembles de données. Pour la base de données SDUMLA, Nous avons remarqué que la valeur moyenne prend sa valeur maximale pour les images traitées par notre approche, alors qu'elle prend la valeur la plus faible pour CLAHE et par la Fusion. Mais pour la base de données FV-USM, la valeur moyenne est égale à 0,4 pour la méthode CLAHE et dépasse 0,5 pour les deux méthodes MCGF et la Fusion.

La valeur de contraste dans la base de données SDUMLA est la même pour MCGF et CLAHE. Cependant, dans la base FV-USM, le contraste atteint sa valeur minimale. Une valeur de contraste plus élevée est obtenue pour la méthode de Fusion pour les deux bases de données. Dans la figure 3.18.c, nous avons remarqué que l'entropie prend des valeurs proches de 7 pour toutes les méthodes dans les deux bases de données, une légère supériorité a été notée pour la méthode de fusion dans la base de données SDUMLA et une légère supériorité ont ainsi été notées pour la méthode MCGF dans la base de données FC-USM. À partir de cette analyse, nous avons constaté

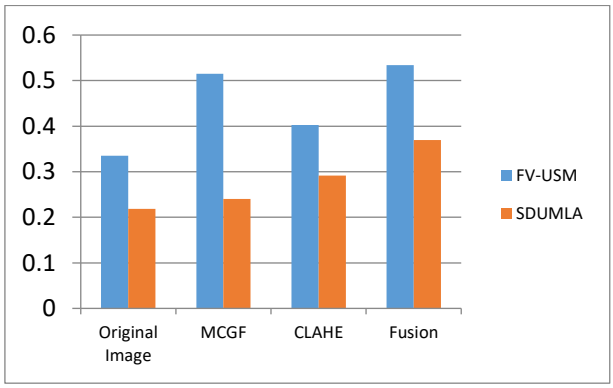
que la méthode de Fusion par DCT et la méthode MCGF prennent les valeurs maximales de contraste et d'entropie dans les deux bases de données, mais la méthode MCGF semble plus pratique pour la base de données FC-USM.

Tableau 3. 1.Valeur moyenne des images dans le jeu de données FV-USM

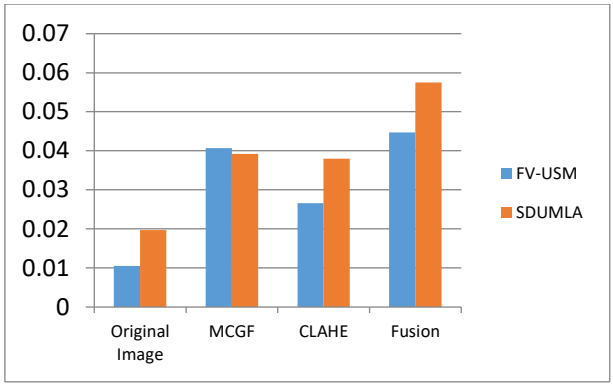
	Average gray value	Image contrast	Image entropy
Original Image	0,3353	0,0105	6,202
MCGF	0,5152	0,0407	7,1738
CLAHE	0,4028	0,0266	6,712
DCT Fusion	0,5338	0,0447	7,0874

Tableau 3. 2.Valeur moyenne des images dans le jeu de données SDUMLA

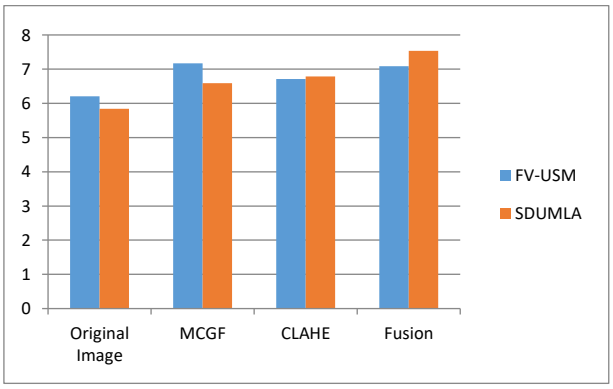
	Average gray value	Image contrast	Image entropy
Original Image	0,2184	0,0197	5,844
MCGF	0,2402	0,0392	6,5933
CLAHE	0,2914	0,038	6,7865
DCT Fusion	0,3698	0,0575	7,5373



(a)



(b)



(c)

Figure 3. 12: Résultats de: a) valeur de gris moyenne, b) contraste de l'image, c) entropie de l'image

4 Conclusion

Dans ce chapitre, nous avons proposé une nouvelle méthode d'amélioration de l'image des vaisseaux sanguins en appliquant diverses techniques d'amélioration sur deux bases de données d'images de veines des doigts disponibles dans le domaine public. La technique de fusion utilisant DCT a été appliquée sur les résultats obtenus de l'égalisation d'histogramme et du filtrage de Gabor. L'utilisation du filtre de Gabor et de l'égalisation de l'histogramme adaptatif permet de mettre en évidence et de visualiser ainsi que d'améliorer le motif veineux de l'image d'origine.

Notre approche montre de bonnes performances lorsqu'elle est utilisée dans la base de données SDUMLA. Cependant, une performance similaire peut être obtenue par la méthode MCGF lorsque les deux méthodes ont été testées avec la base de données FV-USM. Un autre choix de paramètres tels que les orientations et la taille du masque devrait permettre de meilleurs résultats. Cette méthode proposée pourrait produire des images de veines claires, améliorant ainsi les performances de reconnaissance des veines des doigts.

CHAPITRE 4
L'APPRENTISSAGE PROFOND

1 Introduction

Depuis quelques années, le domaine de l'Intelligence Artificielle (IA) a connu une nouvelle tendance avec les techniques de l'apprentissage profond. Un réseau de neurones artificiels à apprentissage profond acquiert sans cesse de l'expertise à partir de nouvelles données. Dans ce chapitre, nous définissons tout d'abord le concept Deep learning ou bien apprentissage profond, nous présentons quelques architectures particulières puis nous nous concentrons sur le réseau de neurones convolutif, une description détaillée sera donnée pour chaque couche constitutive.

2 Intelligence artificielle, apprentissage automatique, et apprentissage profond

Au cours des dernières années, l'intelligence artificielle a fait l'objet d'un battage médiatique intense. L'apprentissage automatique ou "Maching learning", l'apprentissage en profondeur ou apprentissage profond et l'IA apparaissent dans de nombreux articles, souvent en dehors des publications axées sur la technologie [61].

Premièrement, nous devons définir clairement le concept de l'IA. Qu'est-ce que l'IA, l'apprentissage machine et l'apprentissage profond (voir figure 4.1)? Comment se rapportent-ils les uns aux autres?

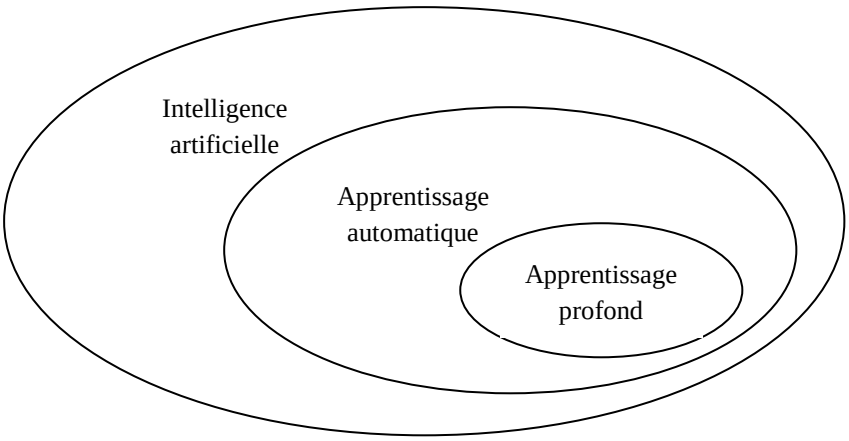


Figure 4. 1: Intelligence artificielle, apprentissage automatique et apprentissage en profondeur

2.1 Intelligence artificielle

L'IA est née dans les années 50, quand un ensemble de pionniers de l'informatique ont commencé à demander si les ordinateurs pouvaient « penser ». Une définition concise du champ serait la suivante: « c'est l'effort d'automatisation des tâches intellectuelles effectuées normalement par les humains ». L'IA est donc un domaine général qui englobe l'apprentissage par machine et l'apprentissage en profondeur, mais qui comprend également de nombreuses autres approches qui n'impliquent aucun apprentissage. Les premiers programmes d'échecs, par exemple, n'impliquaient que des règles codées en dur, conçues par des programmeurs, et n'étaient pas qualifiés d'apprentissage automatique. Pendant assez longtemps, de nombreux experts ont estimé que l'IA au niveau humain pouvait être obtenue en faisant en sorte que les programmeurs fabriquent un ensemble suffisamment grand de règles explicites pour manipuler les connaissances. Cette approche est connue sous le nom d'intelligence artificielle et c'était le paradigme dominant [61].

2.2 Apprentissage machine

Dans l'Angleterre victorienne, Lady Ada Lovelace était une amie et une collaboratrice de Charles Babbage, l'inventeur du moteur d'analyse: le premier ordinateur mécanique polyvalent connu. Le moteur d'analyse n'était pas conçu comme un ordinateur polyvalent lorsqu'il a été conçu dans les années 1830 et 1840, car le concept de calcul polyvalent n'était pas encore inventé. C'était simplement un moyen d'utiliser des opérations mécaniques pour automatiser certains calculs du domaine de l'analyse mathématique, d'où le nom Analytical Engine. En 1843, Ada Lovelace a commenté l'invention: « Le moteur d'analyse n'a aucune prétention de créer quoi que ce soit. Il peut faire tout ce que nous savons lui ordonner de jouer. Sa province est là pour nous aider à rendre disponible ce que nous connaissons déjà » [61].

L'apprentissage automatique découle de cette question: un ordinateur pourrait-il aller au-delà de «ce que nous savons comment l'ordonner» et apprendre par lui-même à exécuter une tâche donnée? Un ordinateur pourrait-il nous surprendre? Plutôt que des programmeurs élaborant des règles de traitement de données à la main, un ordinateur pourrait-il apprendre automatiquement ces règles en consultant des données?

Cette question ouvre la porte à un nouveau paradigme de programmation. En programmation classique, le paradigme de l'IA symbolique, les règles de saisie de l'homme (un programme) et les données à traiter conformément à ces règles, et les réponses obtenues (voir figure 4.2). Avec l'apprentissage automatique, les humains saisissent les données ainsi que les réponses attendues à partir des données, et sortent les règles. Ces règles peuvent ensuite être appliquées à de nouvelles données pour produire des réponses originales.

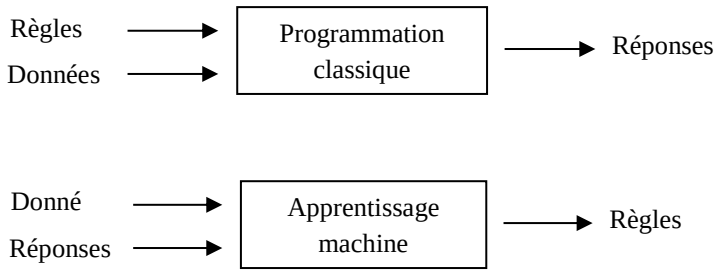


Figure 4. 2 : Apprentissage automatique: un nouveau paradigme de programmation

Un système d'apprentissage automatique est formé plutôt qu'explicitement programmé. Il présente de nombreux exemples relatifs à une tâche et trouve dans ces exemples une structure statistique qui permet au système de définir des règles pour l'automatisation de la tâche. Par exemple, si vous souhaitez automatiser le marquage des images de vacances, vous pouvez présenter un système d'apprentissage automatique avec de nombreux exemples d'images déjà étiquetées par des humains, et le système apprendra des règles statistiques pour associer des images spécifiques à des étiquettes spécifiques.

2.3 Apprentissage des représentations à partir de données

Pour définir l'apprentissage en profondeur et comprendre la différence entre l'apprentissage en profondeur et les autres approches d'apprentissage automatique, nous devons d'abord avoir une idée de ce que font les algorithmes d'apprentissage automatique. L'apprentissage automatique découvre les règles permettant d'exécuter une tâche informatique, à partir d'exemples de ce à quoi on s'attend. Donc, pour apprendre la machine, nous avons besoin de trois choses:

1. Points de données d'entrée — Par exemple, si la tâche est la reconnaissance vocale, ces points de données peuvent être des fichiers sonores des personnes qui parlent. Si la tâche consiste à baliser des images, il peut s'agir d'images.

2. Exemples de sortie attendue: dans une tâche de reconnaissance vocale, il peut s'agir de transcriptions de fichiers de son générées par l'homme. Dans une tâche d'image, les résultats attendus peuvent être des balises telles que «chien», «chat», etc.

3. Un moyen de mesurer si l'algorithme fait du bon travail - Cela est nécessaire pour déterminer la distance entre la sortie actuelle de l'algorithme et sa sortie attendue. La mesure est utilisée comme signal de retour pour ajuster le fonctionnement de l'algorithme. Cette étape d'ajustement est ce que nous appelons l'apprentissage.

Un modèle d'apprentissage automatique transforme ses données d'entrée en sorties significatives, un processus «appris» à partir de l'exposition à des exemples connus d'entrées et de sorties. Par conséquent, le problème central de l'apprentissage machine et de l'apprentissage en profondeur est de transformer de manière significative les données: autrement dit, d'apprendre des représentations utiles des données d'entrée disponibles, représentations qui nous rapprochent du résultat attendu. Avant d'aller plus loin: qu'est-ce qu'une représentation? À la base, c'est une façon différente d'analyser les données - de représenter ou d'encoder des données. Par exemple, une image couleur peut être encodée au format RVB (rouge-vert-bleu) ou au format HSV (valeur de saturation de teinte): il s'agit de deux types différents de représentations des mêmes données. Certaines tâches qui peuvent être difficiles avec une représentation peuvent devenir faciles avec une autre. Par exemple, la tâche "sélectionner tous les pixels rouges de l'image" est plus simple au format RVB, tandis que "rendre l'image moins saturée" est plus simple au format HSV. Les modèles d'apprentissage automatique consistent essentiellement à trouver les représentations appropriées pour leurs données d'entrée, c'est-à-dire des transformations des données qui la rendent plus adaptée à la tâche à accomplir, telle qu'une tâche de classification.

Considérons un axe des x, un des y et certains points représentés par leurs coordonnées dans le système (x, y), comme illustré dans la figure 4.3.

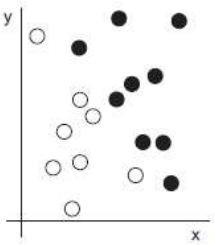


Figure 4. 3 : Quelques exemples de données

Comme vous pouvez le constater, nous avons quelques points blancs et quelques points noirs. Supposons que nous voulions développer un algorithme pouvant prendre les coordonnées (x, y) d'un point et indiquer si ce point est susceptible d'être noir ou blanc. Dans ce cas :

- Les entrées sont les coordonnées de nos points.
- Les résultats attendus sont les couleurs de nos points.
- Un moyen de mesurer si notre algorithme fait du bon travail pourrait être, par exemple, le pourcentage de points correctement classés.

Nous avons besoin ici d'une nouvelle représentation de nos données qui sépare clairement les points blancs des points noirs. Une transformation que nous pourrions utiliser, parmi beaucoup d'autres, serait un changement de coordonnées, illustré à la figure 4.4.

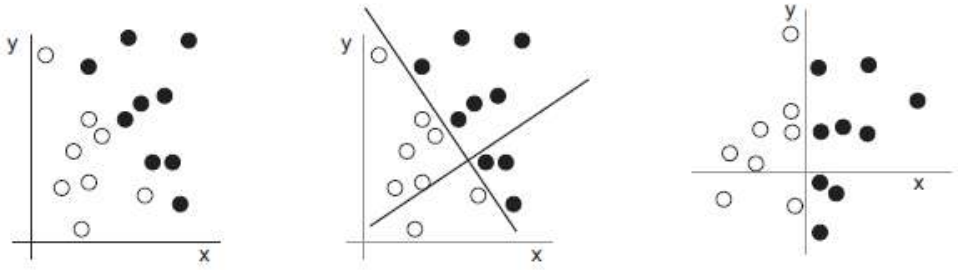


Figure 4. 4 : Changement de coordonnées

Dans ce nouveau système de coordonnées, les coordonnées de nos points peuvent être considérées comme une nouvelle représentation de nos données. Et c'est un bon exemple. Avec cette représentation, le problème de classification noir / blanc peut être exprimé sous la forme d'une règle simple: « Les points noirs sont tels que x > 0 » ou « Les points blancs sont tels que x <0 ». Cette nouvelle représentation résout fondamentalement le problème de classification.

Dans ce cas, nous avons défini le changement de coordonnées à la main. Mais si au lieu de cela nous essayions systématiquement de rechercher différentes modifications possibles des coordonnées et utilisions comme retour le pourcentage de points correctement classés, nous ferions un apprentissage automatique. Learning, dans le contexte de l'apprentissage automatique, décrit un processus de recherche automatique pour de meilleures représentations.

Tous les algorithmes d'apprentissage automatique consistent à rechercher automatiquement les transformations permettant de transformer les données en représentations plus utiles pour une tâche donnée. Ces opérations peuvent être des modifications de coordonnées, comme nous venons de le voir, ou des projections linéaires (pouvant détruire des informations), des traductions, des opérations non linéaires telles que « sélectionner tous les points tels que x> 0), etc. Les algorithmes

d'apprentissage automatique ne sont généralement pas créatifs pour trouver ces transformations; ils ne font que rechercher dans un ensemble prédéfini d'opérations, appelé espace d'hypothèses.

Techniquement, l'apprentissage automatique c'est rechercher des représentations utiles de certaines données d'entrée, dans un espace de possibilités prédéfini, en s'appuyant sur un signal de retour. Cette idée simple permet de résoudre un très large éventail de tâches intellectuelles, de la reconnaissance de la parole à la conduite autonome.

2.4 Le «profond» dans l'apprentissage profond

L'apprentissage en profondeur est un sous-domaine spécifique de l'apprentissage automatique: une nouvelle approche de l'apprentissage des représentations à partir de données qui met l'accent sur l'apprentissage de couches successives de représentations de plus en plus significatives. L'apprentissage en profondeur n'est pas une référence à une forme de compréhension plus profonde obtenue par l'approche; il représente plutôt cette idée de couches successives de représentations. Le nombre de couches contribuant à un modèle de données est appelé profondeur du modèle. D'autres noms appropriés pour le champ auraient pu être apprentissage des représentations en couches et apprentissage des représentations hiérarchiques. L'apprentissage en profondeur moderne implique souvent des dizaines, voire des centaines, de couches successives de représentations, et elles s'apprennent toutes automatiquement à partir de l'exposition aux données d'entraînement. En même temps, d'autres approches de l'apprentissage automatique ont tendance à privilégier l'apprentissage d'une ou deux couches de représentations des données; par conséquent, ils sont parfois appelés apprentissage superficiel.

En apprentissage en profondeur, ces représentations en couches sont apprises (presque toujours) au moyen de modèles appelés réseaux de neurones, structurés en couches littérales superposées. Le terme réseau de neurones est une référence à la neurobiologie, bien que certains des concepts centraux de l'apprentissage en profondeur aient été développés en partie en s'inspirant de notre compréhension du cerveau, les modèles d'apprentissage en profondeur ne sont pas des modèles du cerveau. Rien ne prouve que le cerveau mette en œuvre des mécanismes d'apprentissage utilisés dans les modèles modernes d'apprentissage en profondeur. On peut rencontrer des articles affirmant que l'apprentissage en profondeur fonctionne comme le cerveau ou a été modelé sur le cerveau, mais ce n'est pas le cas. Ce serait déroutant et contre-productif pour les nouveaux arrivants dans ce domaine de penser que l'apprentissage en profondeur est en aucune manière lié à la neurobiologie; vous n'avez pas besoin de ce voile de mystique et de mystère « comme notre esprit », et vous pouvez aussi oublier tout ce que vous avez pu lire sur les liens hypothétiques entre l'apprentissage en profondeur et la biologie. Pour nos besoins, l'apprentissage en profondeur est un cadre mathématique pour l'apprentissage des représentations à partir de données [61].

À quoi ressemblent les représentations apprises par un algorithme d'apprentissage en profondeur? Voyons maintenant comment un réseau de plusieurs couches (voir figure 4 .5) transforme l'image d'un chiffre afin de reconnaître son chiffre.

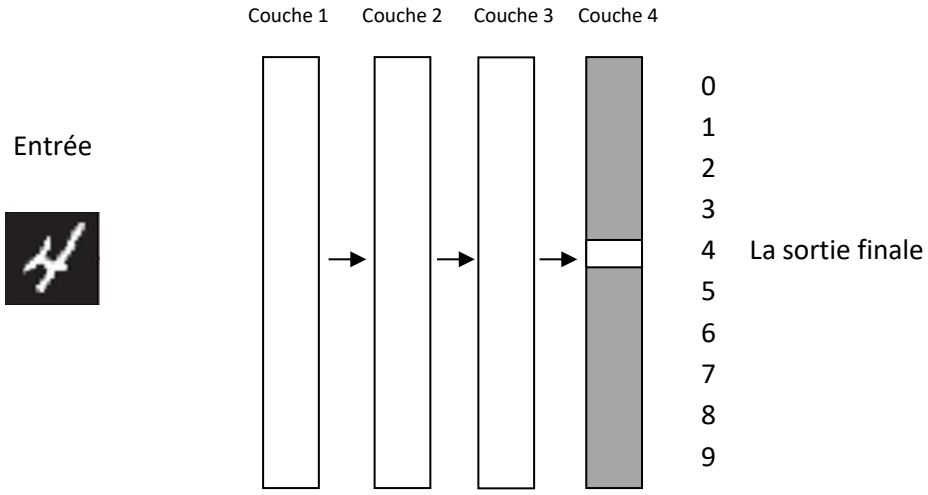


Figure 4. 5 : Réseau de neurones profonds pour la classification des chiffres

Le réseau transforme l'image numérique en représentations qui sont de plus en plus différentes de l'image originale et de plus en plus informatives sur le résultat final. Vous pouvez considérer un réseau en profondeur comme une opération de distillation d'informations en plusieurs étapes, dans laquelle l'information passe par des filtres successifs et ressort de plus en plus épurée (c'est-à-dire utile pour certaines tâches).

3 Architecture des réseaux de neurones

Un exemple d'architecture d'un réseau de neurones avec une seule couche cachée est représenté sur la figure 4.6.

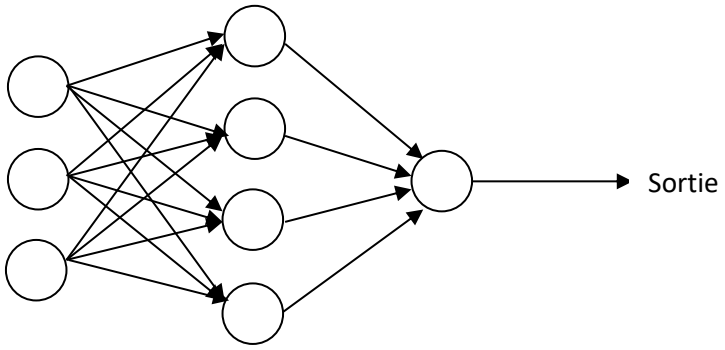


Figure 4. 6 : Réseau de neurones à une seule couche cachée

La couche à gauche de ce réseau s'appelle la couche d'entrée et les neurones de la couche s'appellent des neurones d'entrée. La couche à droite ou la couche de sortie

contient les neurones de sortie ou, comme dans ce cas, un seul neurone de sortie. La couche intermédiaire est appelée couche cachée, car les neurones de cette couche ne sont ni des entrées ni des sorties. Le terme « caché » semble peut-être un peu mystérieux, on pense qu'il devait revêtir une signification philosophique ou mathématique profonde, mais il ne voulait en réalité rien de plus que « pas une entrée ou une sortie ». Le réseau ci-dessus n'a qu'une seule couche cachée, mais certains réseaux ont plusieurs couches cachées [62]. Par exemple, le réseau à quatre couches suivant comporte deux couches cachées:

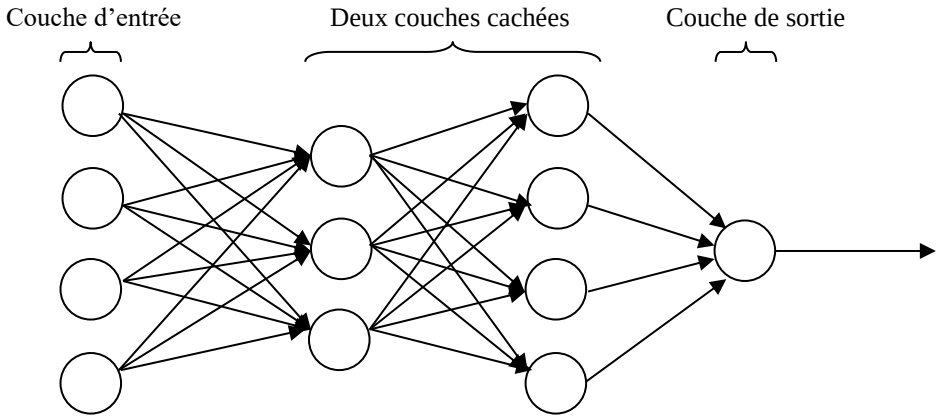


Figure 4. 7 : Réseau de neurones à deux couches cachées

4 Quelques architectures connues

4.1 Réseaux de neurones profonds DNNs

Les réseaux de neurones profonds ou DNNs (Deep Neural Network) possèdent une structure basique : une couche d'entrée, plusieurs couches cachées et une couche de sortie. Quand les données d'entrée sont fournies aux DNN, les valeurs de sorties sont calculées séquentiellement tout au long des couches du réseau. A chaque couche, le vecteur d'entrée comprenant les valeurs de sortie de chaque unité dans la couche inférieure est multiplié par le vecteur de pondération pour chaque unité de la couche courante afin de produire la somme pondérée, Ensuite, une fonction non linéaire, telle qu'une tangente sigmoïde, hyperbolique ou linéaire rectifiée, est appliquée à la somme pondérée pour calculer les valeurs de sortie de la couche.

En fonction des types de couches utilisées dans les DNN et de la méthode d'apprentissage correspondante, les DNN peuvent être classés en :

- Perceptron multicouche.
- Réseau de croyance profond.
- Auto-encodeur.

4.1.1 Perceptron Multicouche

Le perceptron multicouche ou MLP (Multi Layer Perceptron) possède une structure similaire aux réseaux de neurones habituels tout en contenant plus de couches empilées. Il est utilisé en cas de présence de grands nombre de données étiquetées (apprentissage supervisé).

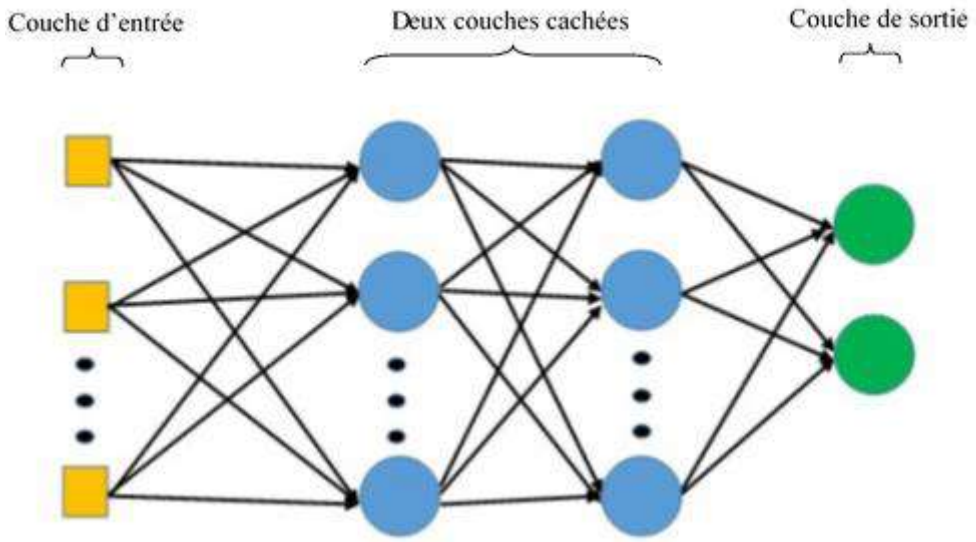


Figure 4. 8 :Illustration d'un perceptron multi-couches

4.1.2 Réseau de croyance profond

Un DBN (Deep Belief Network) est un réseau de neurones profond qui se comporte comme un modèle génératif dans lequel les couches de neurones sont entraînées les unes après les autres et chacune d'entre elles essaye de modéliser la distribution de ses entrées. Pendant l'entrainement d'une couche, il est considéré qu'elle forme une machine de Boltzmann restreinte avec la couche des unités visibles c'est-à-dire la couche qui précède la couche courante celle des unités cachées. Chaque machine de Boltzmann donne des poids par divergence contrastive d'une façon non-supervisée, la couche entraînée est maintenant considérée comme une fonction déterministe qui peut calculer l'espérance d'activation de ses propres unités ainsi que celles de la couche précédente [63].

4.1.2.1 Machine de Boltzmann restreinte

Les machines de Boltzmann restreintes connues en anglais sous l'abréviation RBM (Restricted Boltzmann Machine) sont des extracteurs de fonctionnalités utiles, ils sont principalement utilisés pour initialiser les réseaux de neurones profonds à action directe. Les RBMs offrent une flexibilité pour la modélisation des données avec des variables lentes [64].

Les RBMs sont des réseaux neuronaux à deux couches qui constituent les éléments constitutifs des DBNs. La première couche de la RBM est appelée la couche

visible ou d'entrée et la seconde est la couche cachée. Les nœuds sont connectés les uns aux autres à travers les couches, mais deux nœuds de la même couche ne sont pas liés. Autrement dit, il n'y a pas de communication intra-couche - c'est la restriction dans une RBM. Chaque nœud est un point de calcul qui traite les entrées et commence par prendre des décisions stochastiques sur la transmission ou non de ces entrées. Chaque nœud visible prend une entité de bas niveau d'un élément du jeu de données à apprendre. Par exemple, à partir d'un ensemble de données d'images en niveaux de gris, chaque nœud visible recevrait une valeur de pixel pour chaque pixel d'une image. (Les images MNIST ont 784 pixels, donc les réseaux neuronaux les traitant doivent avoir 784 nœuds d'entrée sur la couche visible).

Suivons maintenant cette valeur de pixel unique, x, à travers le réseau à deux couches. Au nœud 1 de la couche cachée, x est multiplié par un poids et ajouté à un soi-disant biais. Le résultat de ces deux opérations est introduit dans une fonction d'activation, qui produit la sortie du nœud, ou la force du signal qui le traverse, étant donné l'entrée x. Chaque x est multiplié par un poids distinct, les produits sont additionnés, ajoutés à un biais, et encore une fois, le résultat est transmis via une fonction d'activation pour produire la sortie du nœud.

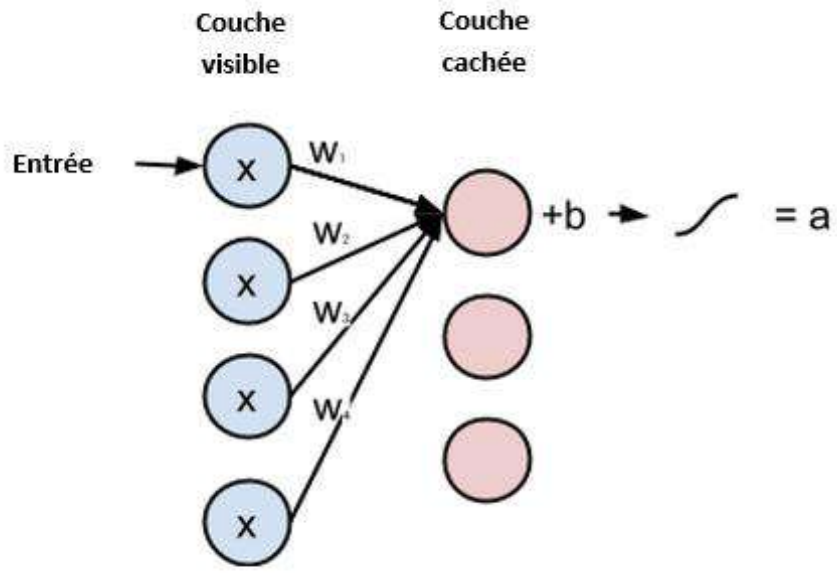


Figure 4. 9 :Un schéma représentant la forme d'un réseau d'un RBM

Étant donné que les entrées de tous les nœuds visibles sont transmises à tous les nœuds cachés, une RBM peut être définie comme un graphique bipartite symétrique. Symétrique signifie que chaque nœud visible est connecté à chaque nœud caché. Bipartite signifie qu'il a deux parties, ou couches, et le graphique est un terme mathématique pour un réseau de nœuds.

Chaque nœud caché reçoit les quatre entrées multipliées par leurs poids respectifs. La somme de ces produits est à nouveau ajoutée à un biais (ce qui oblige au

moins certaines activations à se produire), et le résultat est transmis par l'algorithme d'activation produisant une sortie pour chaque nœud caché.

4.1.2.2 Structure d'un DBN

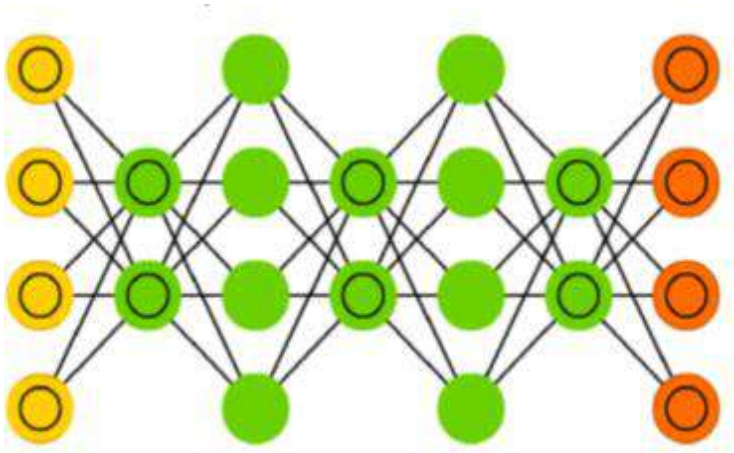


Figure 4. 10 :Un schéma représentant la forme d'un DBN

On remarque qu'un DBN possède le même nombre de nœuds dans la couche d'entrée et celle de sortie.

Contrairement à d'autres modèles, chaque couche d'un DBN apprend la totalité de l'entrée. Dans les CNNs, les premières couches filtrent uniquement les entrées pour les caractéristiques de base, telles que les bords, et les couches ultérieures recombinent tous les motifs simples trouvés par les couches précédentes. D'autre part, les réseaux de croyances profondes, fonctionnent globalement.

Le réseau est comme une pile de machines Boltzmann restreintes (RBM), où les nœuds de chaque couche sont connectés à tous les nœuds de la couche précédente et suivante. Cependant, contrairement aux RBM, les nœuds d'un DBN ne communiquent pas latéralement au sein de leur couche. Un réseau de poids symétriques relie différentes couches. Les nœuds de la couche cachée remplissent deux rôles - ils agissent comme une couche cachée pour les nœuds qui la précèdent et comme des couches visibles pour les nœuds qui lui succèdent. Ces nœuds identifient les corrélations dans les données.

Des algorithmes d'apprentissage gourmands (Greedy learning algorithms) sont utilisés pour pré-former les DBNs. Il s'agit d'une approche de résolution de problèmes qui implique de faire le choix optimal à chaque couche de la séquence, pour finalement trouver un optimum global. Les algorithmes d'apprentissage gourmands commencent à partir de la couche inférieure et montent, affinant les poids génératifs. L'apprentissage se déroule couche par couche, ce qui signifie que les couches des DBNs sont formées une par une. Par conséquent, chaque couche reçoit également une version différente des données et chaque couche utilise la sortie de la couche précédente comme entrée. Les algorithmes d'apprentissage gourmands sont utilisés pour former les DBNs car ils sont rapides et efficaces. De plus, ils aident à optimiser les poids à chaque couche.

4.1.3 Auto-encodeur

Un auto-encodeur est un modèle de réseau neuronal qui cherche à apprendre une représentation compressée d'une entrée. Il s'agit d'une méthode d'apprentissage non supervisée, bien que techniquement, ils soient formés à l'aide de méthodes d'apprentissage supervisées, appelées auto-supervisées. Ils sont généralement formés dans le cadre d'un modèle plus large qui tente de recréer l'entrée.

La conception du modèle auto-encodeur rend délibérément cela difficile en limitant l'architecture à un goulot d'étranglement au milieu du modèle, à partir duquel la reconstruction des données d'entrée est effectuée.

Il existe de nombreux types de codeurs automatiques, et leur utilisation varie, mais peut-être que l'utilisation la plus courante est un modèle d'extraction de fonctionnalités appris ou automatique. Dans ce cas, une fois le modèle ajusté, l'aspect de reconstruction du modèle peut être ignoré et le modèle jusqu'au point de goulot d'étranglement peut être utilisé. La sortie du modèle au goulot d'étranglement est un vecteur de longueur fixe qui fournit une représentation compressée des données d'entrée. Les données d'entrée du domaine peuvent ensuite être fournies au modèle et la sortie du modèle au goulot d'étranglement peut être utilisée comme vecteur d'entité dans un modèle d'apprentissage supervisé, pour la visualisation, ou plus généralement pour la réduction de la dimensionnalité.

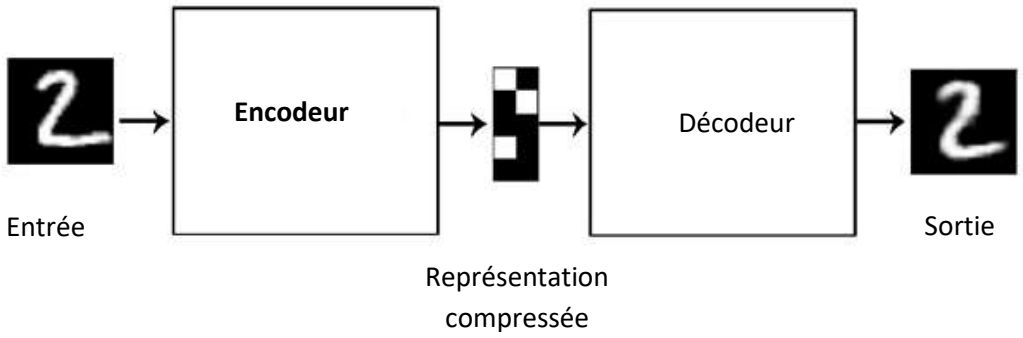


Figure 4. 11 : compression et de décompression de données dans un auto-encodeur

Un auto-encodeur est un algorithme de compression de données où les fonctions de compression et de décompression sont 1) spécifiques aux données, 2) avec perte, et 3) apprises automatiquement à partir d'exemples plutôt que conçu par un humain. De plus, dans presque tous les contextes où le terme "auto-encodeur" est utilisé, les fonctions de compression et de décompression sont implémentées avec des réseaux de neurones [65].

1) Les encodeurs automatiques sont spécifiques aux données, ce qui signifie qu'ils ne pourront compresser que des données similaires à celles sur lesquelles ils ont été formés.

C'est différent, disons, de l'algorithme de compression MPEG-2 Audio Layer III (MP3), qui ne contient que des hypothèses sur le "son" en général, mais pas sur des types de sons spécifiques. Un auto-encodeur formé sur les images de visages ferait un assez mauvais travail de compression des images d'arbres, car les fonctionnalités qu'il apprendrait seraient spécifiques aux visages.

2) Les encodeurs automatiques sont avec perte, ce qui signifie que les sorties décompressées seront dégradées par rapport aux entrées d'origine.

3) Les auto-encodeurs sont appris automatiquement à partir d'exemples de données, ce qui est une propriété utile: cela signifie qu'il est facile de former des instances spécialisées de l'algorithme qui fonctionneront bien sur un type spécifique d'entrée. Il ne nécessite aucune nouvelle ingénierie, juste des données d'apprentissage appropriées. Pour construire un auto-encodeur, on est besoin de trois choses: une fonction d'encodage, une fonction de décodage et une fonction de distance entre la quantité de perte d'informations entre la représentation compressée des données et la représentation décompressée (c'est-à-dire une fonction "perte"). L'encodeur et le décodeur seront choisis pour être des fonctions paramétriques (généralement des réseaux de neurones) et pour être différentiables par rapport à la fonction de distance, afin que les paramètres des fonctions d'encodage / décodage puissent être optimisés pour minimiser la perte de reconstruction, en utilisant la descente de gradient stochastique [65].

4.2 Les Réseaux de neurones récurrents (RNNs)

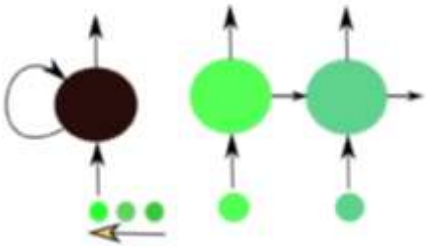
Toutes les architectures proposées jusqu'à maintenant traitent l'entrée comme une « image » : il n'y a pas de notion d'ordre entre les différents éléments de l'entrée. Une telle représentation est peu satisfaisante si on veut traiter des suites d'entrées comme une phrase. Il faut une architecture capable de tenir compte des entrées déjà lues. On appelle réseau récurrent une unité particulière qui prend en entrée une suite d'éléments, et à chaque étape de calcul renvoi une sortie et un état caché qu'il se renvoie dans une boucle de rétroaction. Ainsi, le traitement de l'élément suivant de l'entrée dépend de celui de l'élément courant. Si on « déplie » l'exécution du problème, on trouve que la séquence passe en fait par un nombre de couches cachées égal au nombre d'éléments dans la suite.

Il est possible de faire fonctionner cette unité cachée comme un neurone classique avec une sortie et une entrée supplémentaire représentant l'état caché, mais cela pose des problèmes d'évanouissement du gradient lors de la rétro-propagation. Une unité cachée qui résout ce problème et est largement utilisée dans la littérature est l'unité LSTM (Long Short-Term Memory) introduite dans [66].

L'idée est qu'on a un état de cellule, un état caché et une entrée. A chaque passage dans la couche, on calcule quel degré d'information il faut retenir de l'état de la cellule en fonction de l'état caché et de l'entrée (porte d'oubli). Ensuite on calcule s'il faut

ajouter de l'information à l'état de la cellule (porte de mise à jour). Enfin, on fait le calcul de la sortie et de l'état caché à partir des trois données précédentes [67]

Un LSTM élimine le défaut de disparition ou le problème de gradient explosif du RRN et cela rend LSTM adapté à apprendre de l'histoire à classer, traiter et prévoir les séries chronologiques quand il y a beaucoup de temps long et inconnu entre les événements importants [68].



(a) (b)

Figure 4. 12 :Réseau récurrent : a) forme dépliée b) forme temporelle

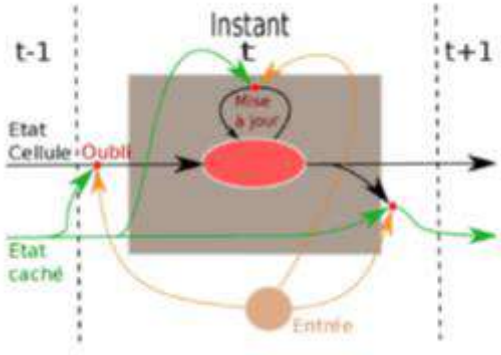


Figure 4. 13 :Exemple d'unité LSTM

Les RNNs ne semblent pas être profonds comme DNN ou CNN en termes de nombre de couches mais ils peuvent être considérés comme une structure encore plus profonde s'ils sont déroulés dans le temps car ils ont été utilisés avec succès dans de nombreux domaines, y compris le traitement du langage naturel et la traduction de la langue. Même si les RNNs ont été explorés moins que les DNN et les CNN, ils fournissent encore des méthodes d'analyse très puissantes pour l'information séquentielle.

5 Réseau de neurones à convolution

Les réseaux de neurones à convolution (CNN), ou réseaux de neurones convolutionnel ont démontré des performances exceptionnelles dans le domaine de compréhension et de la reconnaissance d'images, améliorant considérablement l'état de

l'art dans de nombreuses applications. Un CNN est l'un des structures de réseau les plus représentatives pour les technologies d'apprentissage profond et a connu un grand succès dans le domaine du traitement de l'image, ce qui nous a poussés à appliquer le CNN dans le domaine de la reconnaissance par image des veines de doigt. Cependant, la reconnaissance de la veine du doigt reste un domaine relativement peu exploré par l'apprentissage profond, à l'exception de quelques travaux [69] [70]. Les techniques d'apprentissage en profondeur nécessitent un très grand nombre d'échantillons d'apprentissage, mais les bases de données de veines de doits existantes ont des tailles relativement petites. Par conséquent, les applications de l'apprentissage en profondeur pour la reconnaissance de la veine de doigt sont encore au stade des tentatives. Le CNN est de type MLP constitué de plusieurs couches. Ces couches extraient des caractéristiques invariantes et discriminantes à partir d'informations d'entrée telles que des images [70].

5.1 Descente du gradient de l'erreur

Les CNNs sont constitués de neurones qui ont des poids (W) ou des paramètres d'apprentissage et des biais (b). Les poids des neurones sont déterminés de manière itérative en utilisant des données étiquetées à l'entrée du modèle. Ces poids sont ajustés en calculant l'erreur entre les entrées représentatives et leur réponse attendue en sortie. Une évaluation de la caractéristique de la décision par rapport à la décision prévisible est faite, ceci afin de réduire l'espace entre les deux (figure 4.14) [67].

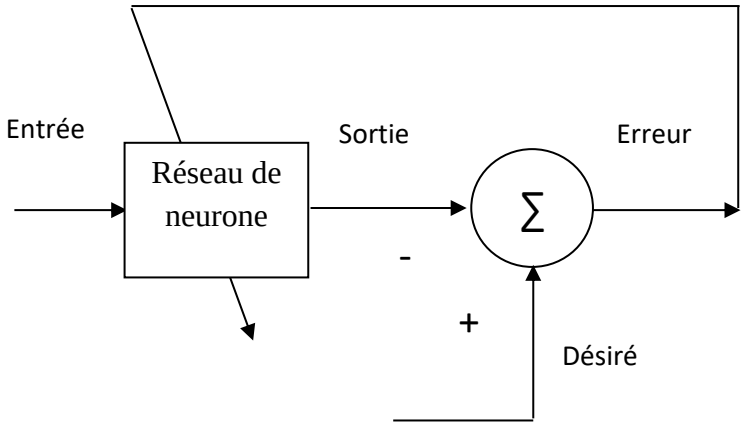


Figure 4. 14 : ajustement des poids des réseaux de neurones

Le terme de neurone (ou unité cachée) désigne une unité de calcul (figure 4.15). Cette unité prend une entrée, c'est à dire un vecteur A dont la taille est celle de l'entrée et lui applique une transformation linéaire WA+b, où W et b sont des paramètres du neurone. Après cette transformation linéaire, on applique généralement une fonction dite « d'activation » non linéaire f. Ainsi la transformation effectuée sur l'entrée A retourne la sortie f (WA + b). L'entraînement du neurone se fait par rétro-propagation à partir de la sortie finale du réseau de neurones : une évaluation de la qualité de la décision finale par rapport à la décision attendue est effectuée, et il s'agit alors de

minimiser la distance entre les deux. Pour ce faire, on peut par exemple procéder par descente de gradient. En effet, pour chaque paramètre, on calcule la dérivée partielle du score obtenu et on ajuste en conséquence les paramètres du réseau jusqu'à l'obtention d'un résultat satisfaisant [67].

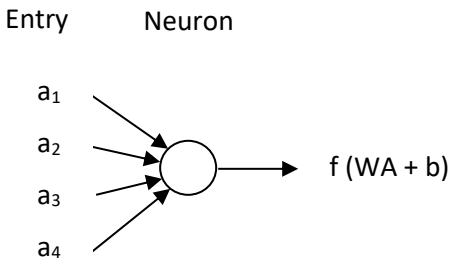


Figure 4. 15 : Exemple de neurone formel

L'objectif de descente du gradient est de trouver les valeurs des poids qui minimisent la fonction d'erreur E comme suit :

$$w_{t+1} = w_t - \eta \frac{\partial E}{\partial w_t} \tag{4.1}$$

Elle suit la direction inverse de la dérivée en chaque point de la courbe de l'erreur qui doit être une fonction dérivable, par exemple : L'Erreur Quadratique Moyenne (EQM) qui est définie par :

$$E = \frac{1}{2}\sum_{i=1}^{n}(Y_i - x_i)^2 \tag{4.2}$$

Le gradient stochastique est une approximation de la descente de gradient(Annexe)

Le réseau profond peut être entrainer suivant trois stratégies :

1. Descente de Gradient classique (batch gradient descent) : On fait passer la totalité des observations, le gradient est calculé et les poids sont corrigés.
2. Online : le gradient est calculé pour chaque observation.
3. Mini-batch (traitement par lots) : On fait passer un groupe d'observations avant le calcul du gradient.

• Le traitement par lots permet d'améliorer la convergence en réduisant le nombre de passage sur la base entière.

• Il permet également de se contenter de charger partiellement les données en mémoire au fur et à mesure.

5.2 Les blocs de construction

Un CNN contient une ou plusieurs couches de convolution, des couches de sous-échantillonnage et souvent des couches entièrement connectées. Un CNN typique est illustré à la figure 4.16.

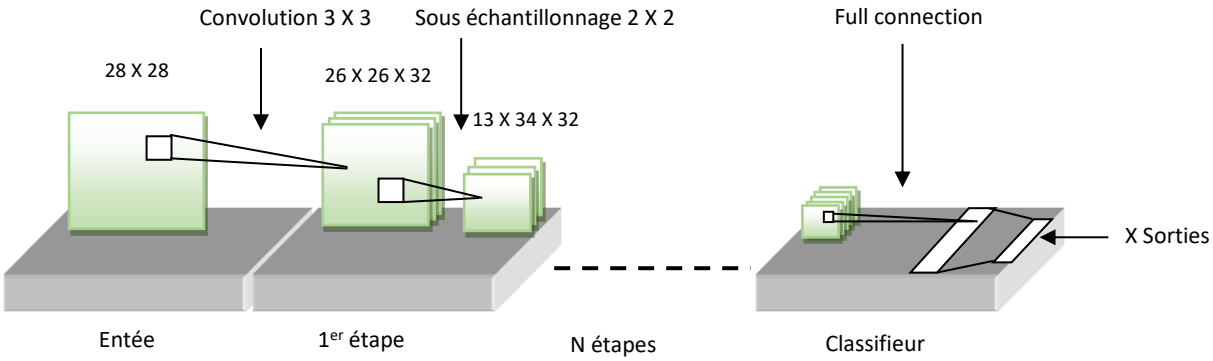


Figure 4. 16 : Exemple de modèles de CNN

Maintenant que l'architecture CNN est claire, nous allons discuter en détail les différents types de couches et fonctions.

5.2.1 La couche convolution

C'est la couche la plus impliqué parmi ceux que nous discutons dans cette section. On commence par octroyer la définition de la convolution[83]. Supposons que l'image d'entrée a une de taille 3 x 4 pixels et que la taille du noyau de convolution soit 2 x 2, comme illustré dans la figure 4.17.

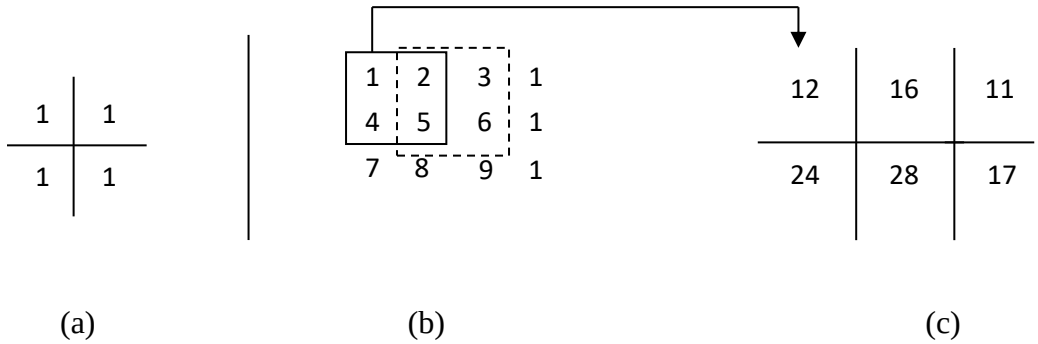


Figure 4. 17 : Illustration de l'opération de convolution : a) Noyau 2 x 2, b) L'entrée et la sortie de convolution, c) La sortie de convolution.

Si nous chevauchons le noyau de convolution au-dessus de l'image d'entrée, nous pouvons calculer le produit entre les nombres situés au même emplacement dans le noyau et l'entrée, et nous obtenons un nombre unique en additionnant ces produits. Par exemple, si nous chevauchons le noyau avec la région en haut à gauche dans l'entrée, le résultat de la convolution à cet emplacement spatial est le suivant:

1x1+1x4+1x2+1x5 = 12. Nous déplaçons ensuite le noyau d'un pixel vers la droite et obtenons le résultat de convolution suivant : 1x2+1x5+1x3+1x6 = 16. Nous continuons à déplacer le noyau jusqu'à ce qu'il atteigne la bordure de la matrice d'entrée (image). Nous répétons la convolution pour chaque emplacement de pixel possible jusqu'à ce que nous ayons déplacé le noyau dans le coin inférieur droit de l'image d'entrée, comme illustré dans la figure 4.18. Dans une couche de convolution, plusieurs noyaux de convolution sont généralement utilisés.

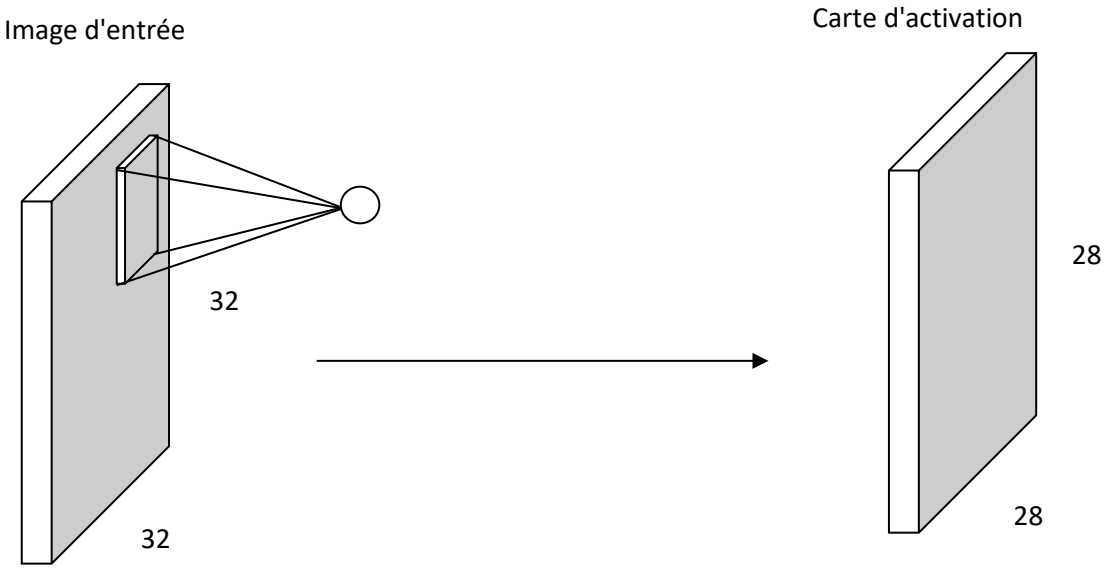
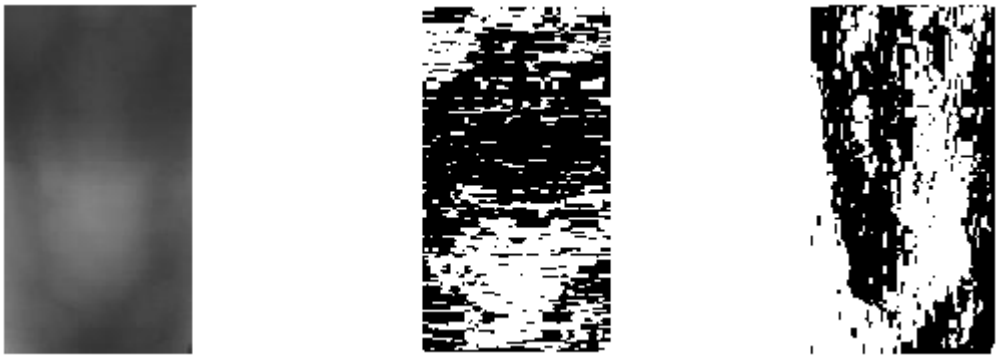


Figure 4. 18 : Convolution sur tous les emplacements spatiaux

La figure 4.19.a montre une image d'entrée et ses résultats de convolution à l'aide de deux noyaux différents (figure 4.19.b et figure 4.19.c). Un filtre de convolution de taille 3 x 3 , $k = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$ est employée.

Le noyau de convolution doit être de taille 3 x 3 x 3, dans lequel chaque canal est défini sur K. Lorsqu'il existe un bord horizontal à l'emplacement (x,y) (c'est-à-dire lorsque les pixels à l'emplacement spatial (x + 1; y) et (x- 1; y) diffère d'une grande quantité), nous nous attendons à ce que le résultat de convolution ait une magnitude élevée. Comme le montre la figure 4.20.b, les résultats de convolution mettent bien en évidence les bords horizontaux. Lorsque nous définissons chaque canal du noyau de convolution sur K^T (la transposée de K), le résultat de la convolution amplifie les arêtes verticales, comme est illustré dans la figure 4.19.c. La matrice (ou filtre) K et K^T s'appellent les opérateurs Sobel.



(a) (b) (c)

Figure 4. 19: Effet de différents noyaux de convolution [72].

Si nous ajoutons un terme de biais à l'opération de convolution, nous pouvons rendre le résultat de la convolution positif aux bords horizontaux (verticaux) dans une certaine direction (par exemple, un bord horizontal avec les pixels au-dessus de celui-ci plus lumineux que les pixels au-dessous de celui-ci) et négatif en-dessous. D'autres endroits. Si la couche suivante est une couche ReLU, la sortie de la couche suivante définit en fait de nombreuses "fonctions de détection de bord", qui s'activent uniquement aux bords horizontaux ou verticaux dans certaines directions. Si nous remplaçons le noyau Sobel par d'autres noyaux, nous pouvons apprendre des fonctionnalités qui s'activent pour des arêtes ayant des angles différents. Lorsque nous descendons plus bas dans le réseau profond, les couches suivantes peuvent apprendre à s'activer uniquement pour des motifs spécifiques (mais plus complexes), par exemple des groupes d'arêtes qui forment une forme particulière. Ces motifs plus complexes seront ensuite assemblés par des couches plus profondes afin d'être activés pour des parties d'objet significatives sur le plan sémantique ou même pour un type d'objet particulier, par exemple un chien, un chat, un arbre, une plage, etc.

Un autre avantage de la couche de convolution est que toutes les localisations spatiales partagent le même noyau de convolution, ce qui réduit considérablement le nombre de paramètres nécessaires à une couche de convolution. Par exemple, si plusieurs chiens apparaissent dans une image d'entrée, la même fonctionnalité "motif (patern)" ressemblant à une tête de chien sera activée à plusieurs endroits, correspondant aux têtes de chiens différents.

Dans une configuration de réseau neuronal profond, la convolution encourage également le partage de paramètres. Par exemple, supposons que les motifs en forme de tête de chien et en forme de tête de chat, soient deux caractéristiques apprises par un réseau de convolution profonde. Le CNN n'a pas besoin de lui consacrer deux jeux de paramètres disjoints (par exemple, des noyaux de convolution dans plusieurs couches). Les couches inférieures de CNN peuvent apprendre un "motif en forme d'œil" et un "motif de fourrure", partagés par ces deux caractéristiques plus abstraites. En bref, la combinaison de noyaux de convolution et de structures profondes et hiérarchiques est

très efficace pour apprendre de bonnes représentations (caractéristiques) à partir d'images pour des tâches de reconnaissance visuelle (figure 4.20).

La couche convolution à deux propriétés intéressantes:

Les modèles qu'ils apprennent sont des invariants de translation : Après avoir appris un certain motif dans le coin inférieur droit d'une image, un CNN peut le reconnaître n'importe où: par exemple, dans le coin supérieur gauche. Cela permet d'utiliser efficacement les données lors du traitement des images, car le monde visuel est fondamentalement complexe (Translation invariante).

Ils peuvent apprendre les hiérarchies spatiales des modèles : Une première couche de convolution apprendra de petits motifs locaux tels que des bords, une seconde couche de convolution apprendra des motifs plus grands constitués des caractéristiques des premières couches, etc. Cela permet aux couches de convolution d'apprendre efficacement des concepts visuels abstraits de plus en plus complexes (parce que le monde visuel est fondamentalement hiérarchisé spatialement).

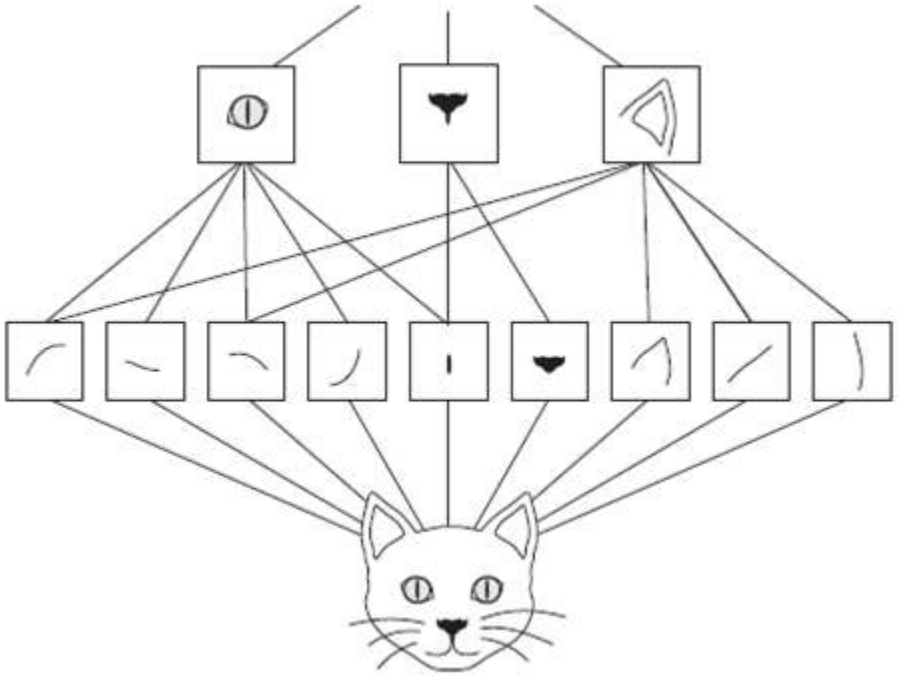


Figure 4. 20 : Le monde visuel forme une hiérarchie spatiale de modules visuels: les arêtes hyper locales se combinent pour former des objets locaux tels que les yeux ou les oreilles, qui se combinent pour former des concepts de haut niveau tels que «chat».

Les convolutions fonctionnent sur des tenseurs 3D (figure 4.21), appelés cartes de caractéristiques (*feature maps)*, avec deux axes spatiaux (hauteur et largeur) ainsi qu'un axe de profondeur (également appelé axe des canaux). Pour une image RGB, la

dimension de l'axe de profondeur est 3, car l'image comporte trois canaux de couleur: rouge, vert et bleu. Pour une image en noir et blanc, comme pour les chiffres du MNIST, la profondeur est de 1 (niveaux de gris). L'opération de convolution extrait les patchs de cartes de caractéristiques d'entée et applique la même transformation à tous ces patchs, produisant des cartes de caractéristiques de sortie. Cette carte d'entités en sortie est toujours un tenseur 3D: elle a une largeur et une hauteur. Sa profondeur peut être arbitraire, car la profondeur de sortie est un paramètre de la couche, et les différents canaux dans cet axe de profondeur ne correspondent plus à des couleurs spécifiques comme en entrée RVB; ils représentent plutôt des filtres. Les filtres codent des aspects spécifiques des données d'entrée: à un niveau élevé, un seul filtre pourrait coder le concept par exemple de «présence d'un visage dans l'entrée».

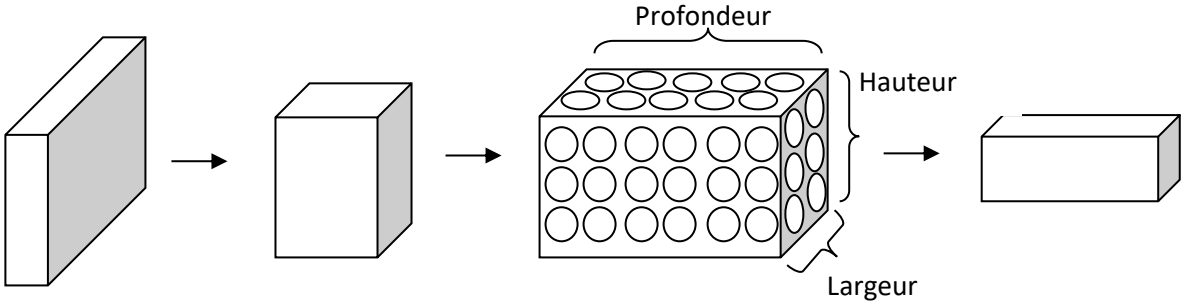


Figure 4. 21 : Représentation des cartes caractéristiques 3D

Dans l'exemple MNIST, la première couche de convolution prend une carte de caractéristiques de taille (28, 28, 1) et génère une carte de caractéristiques de taille (26, 26, 32): elle calcule 32 filtres sur son entrée. Chacune de ces 32 voies de sortie contient une grille de valeurs 26 × 26, qui est une carte de réponse du filtre sur l'entrée, indiquant la réponse de ce motif de filtre à différents emplacements de l'entrée (voir la figure 4.16). C'est ce que signifie le terme carte d'entités: chaque dimension de l'axe de profondeur est une entité (ou un filtre), et la sortie du tenseur 2D [: , : , n] est la carte spatiale 2D de la réponse de ce filtre sur l'entrée.

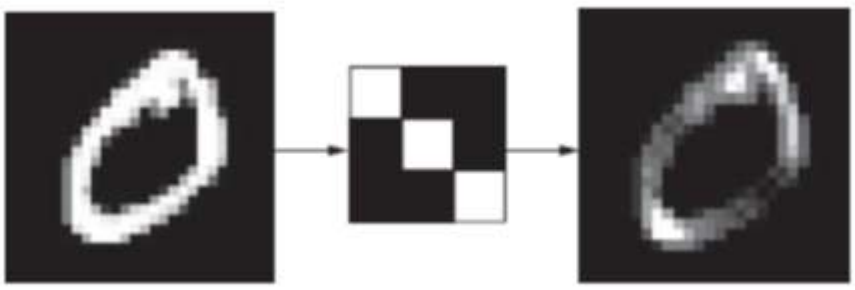


Figure 4. 22 : Le concept d'une carte de réponse: une carte en 2D de la présence d'un motif à différents endroits dans une entrée

Les convolutions sont définies par deux paramètres clés:

Taille des patchs extraits des entrées : Il s'agit généralement de 3 × 3 ou de 5 × 5. Dans l'exemple, il s'agissait de 3 × 3, (figure 4.23).

Profondeur de la carte des caractéristiques en sortie: nombre de filtres calculés par la convolution. L'exemple a commencé avec une profondeur de 32 et s'est terminé avec une profondeur de 64.

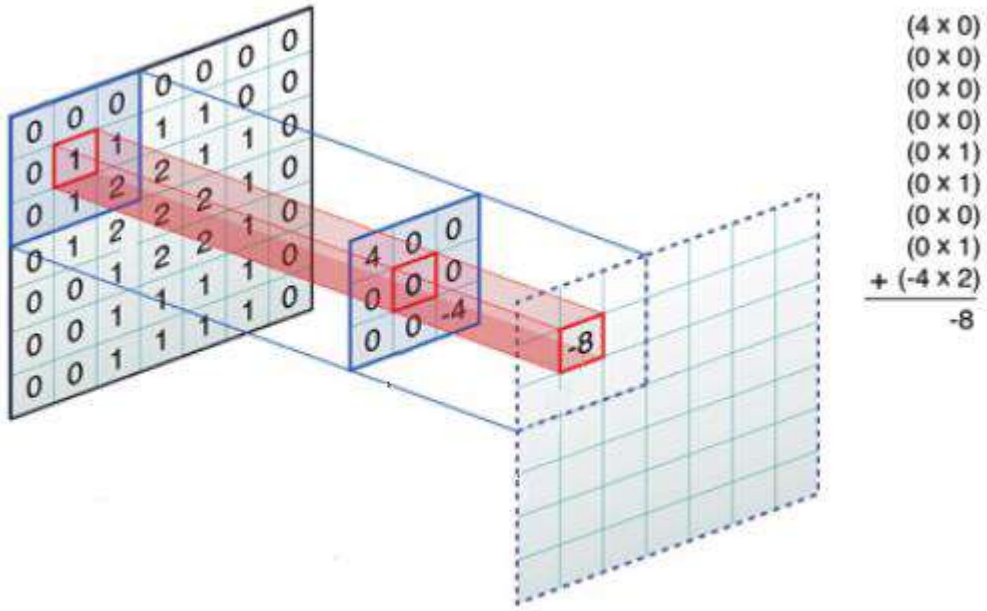


Figure 4. 23 : Filtrage par un noyau de taille 3 x 3.

5.2.2 La fonction ReLU (Rectified Linear Unit)

La fonction d'activation, ou fonction de transfert, est un seuil de stimulation qui, une fois atteint entraine une réponse du neurone. La couche ReLU ne change pas la taille de l'entrée, en fait, l'unité linéaire rectifiée (d'où le nom ReLU) peut être considéré comme une troncature effectuée individuellement pour chaque élément de l'entrée. ReLU a pour but d'augmenter la non-linéarité de CNN. Étant donné que les informations sémantiques dans une image (par exemple, une personne et un chien assis côte à côte sur un banc dans un jardin) constituent de toute évidence un mappage hautement non linéaire des valeurs de pixels en entrée, nous voulons que le mappage de l'entrée CNN jusqu'à sa sortie soit également hautement non linéaire. La fonction ReLU, bien que simple, est une fonction non linéaire, comme illustré à la figure 4.24.

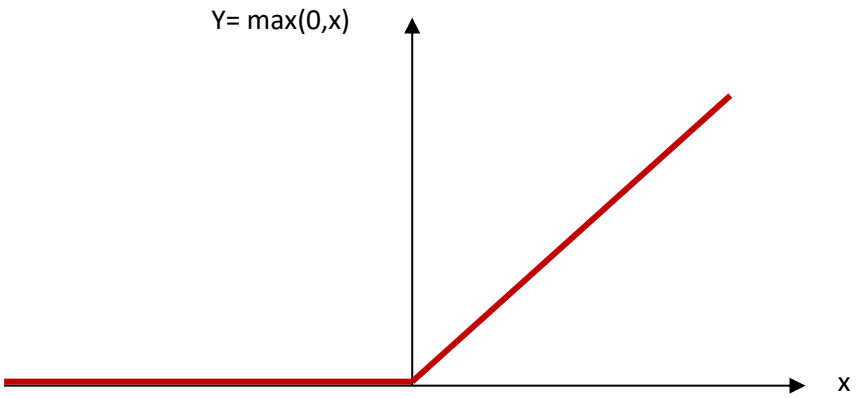


Figure 4. 24: La fonction ReLU, produit 0 comme sortie lorsque x <0, puis produit une linéaire avec une pente de 1 lorsque x> 0.

D'autres transformations non linéaires ont été utilisées dans la recherche sur les réseaux de neurones pour produire une non-linéarité, par exemple la fonction logistique sigmoïde $y = \sigma = \frac{1}{1+\exp(-x)}$. La fonction tangente hyperbolique *f(x)=tanh(x).* La fonction tangente hyperbolique : *f(x)=|tanh(x)|* [72].

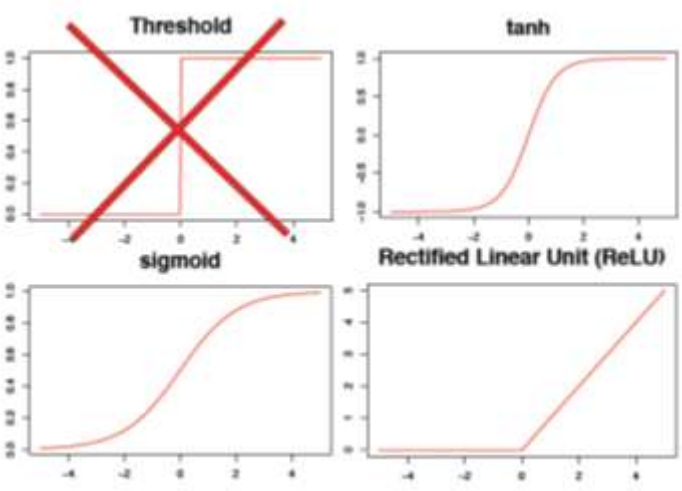


Figure 4. 25 : Exemples de fonctions de régulation

5.2.3 La couche Max-pooling

Un autre concept important des CNN est le max-pooling, qui est une forme de sous-échantillonnage non linéaire. Max-pooling partitionne l'image d'entrée en un ensemble de rectangles non chevauchés et, pour chacune de ces sous-régions, affiche la valeur

maximale (figure 4.26). Le Max-pooling élimine les valeurs non maximales, ce qui réduit le calcul pour les couches supérieures [73].

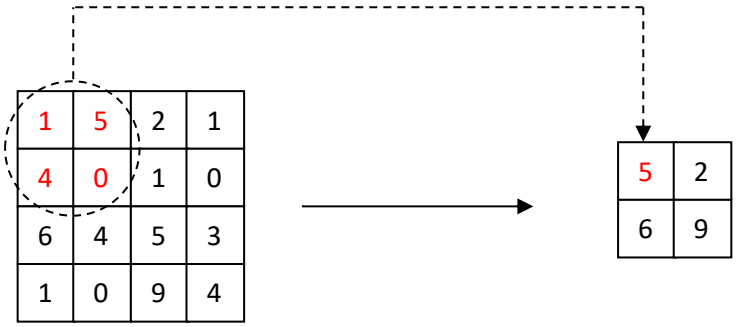


Figure 4. 26 : Exemple de la sortie d'une couche Max-pooling

Par exemple, avant les premières couches MaxPooling2D, la carte de caractéristiques est 26 × 26, mais l'opération de regroupement maximal elle est divisée par deux en 13×13. C'est le rôle du Max-pooling: sous-échantillonner de manière agressive les cartes de caractéristiques.

Le Max-pooling consiste à extraire des fenêtres des cartes d'entités en entrée et à afficher la valeur maximale de chaque canal. Il est conceptuellement similaire à la convolution, sauf qu'au lieu de transformer des patchs locaux via une transformation linéaire apprise (la convolution par un noyau), ils sont transformés via une opération de tenseur max codé en dur. Une grande différence par rapport à la convolution est que la mise en commun maximale est généralement effectuée avec 2 x 2 fenêtres et décalage de 2, afin de sous-échantillonner les cartes d'entités par un facteur de 2. Par ailleurs, la convolution est généralement réalisée avec 3 x 3 fenêtres et un décalage égale à 1.

5.2.4 flattening (ou mise à plat)

Dans cette couche les pixels sont récupérés ligne par ligne et ajoutés au vecteur final [83].

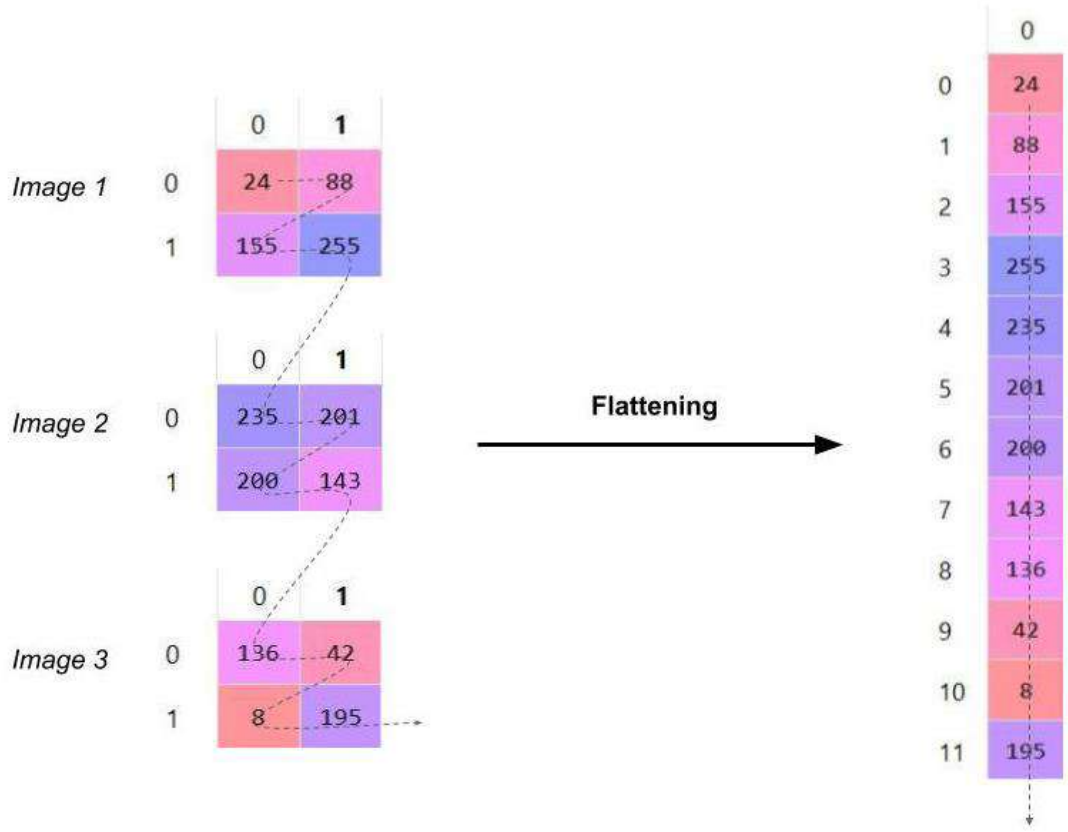


Figure 4. 27 :Fonctionnement de la couche flatten

5.2.5 La couche dense ou Fully connected

C'est un réseau densément connecté, (une pile de couches denses). Ces classificateurs traitent les vecteurs, qui sont 1D, alors que la sortie précédente est un tenseur 3D. Nous devons d'abord aplatir (flatten) les sorties 3D à 1D, puis ajouter quelques couches denses en haut. L'application de la ***fonction softmax*** aux sorties de la couche fully connection, nous donne la probabilité d'une classe.

La différence fondamentale entre une couche densément connectée et une couche de convolution est la suivante: les couches denses apprennent des motifs globaux dans leur espace de fonctions en entrée (par exemple, pour un chiffre MNIST, des motifs impliquant tous les pixels), tandis que les couches de convolution apprennent des motifs locaux (voir figure 4.28) : dans le cas des images, motifs trouvés dans de petites fenêtres 2D des entrées. Dans l'exemple précédent, ces fenêtres étaient toutes en 3 × 3 [61].

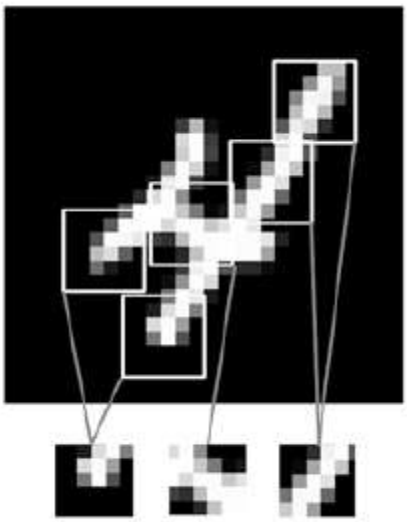


Figure 4. 28 : Les images peuvent être décomposées en motifs locaux tels que des bords, des textures, etc.

5.2.6 La fonction Softmax

Les solutions d'apprentissage approfondi aux problèmes de classification utilisent généralement la fonction softmax comme fonction de classification (dernière couche). La fonction softmax spécifie une distribution de probabilité discrète pour K classes, notée $\sum_{k=1}^{K} p_k$[64].

Si nous prenons x comme activation au niveau de l'avant-dernière couche d'un réseau de neurones et que θ en tant que paramètres de poids au niveau de la couche softmax, nous avons o en tant qu'entrée de la couche softmax [75],

$$o = \sum_{i}^{n-1} \theta_i x_i \tag{4.3}$$

Par conséquent, nous avons

$$p_k = \frac{\exp(o_k)}{\sum_{k=0}^{n-1} \exp(o_k)} \tag{4.4}$$

Par conséquent, la classe prédite serait $\hat{y}$

$$\hat{y} = \arg\max p_i \quad , \quad i \in 1, \dots, N \tag{4.5}$$

6 Sur-ajustement et régularisation (Overfitting and regularization)

La figure 4.29 montre un exemple de la variation de la précision pour une classification des chiffres du MNIST (des données de test) évolue au fil du temps [62].

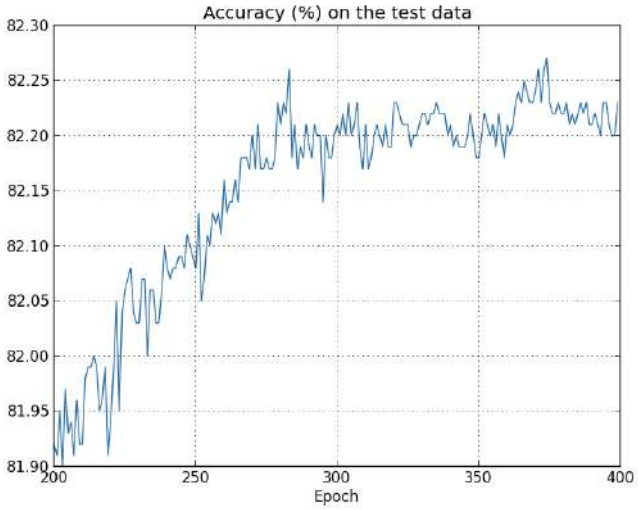


Figure 4. 29 : Variation de la précision dans la base de test en fonction du nombre d'époques

Au cours des 200 premières époques, la précision atteint un peu moins de 82%. L'apprentissage ralentit ensuite progressivement. Enfin, vers l'époque 280, la précision de la classification cesse de s'améliorer. Les époques ultérieures ne voient que de petites fluctuations stochastiques proches de la valeur de la précision à l'époque 280, ce que le réseau apprend après l'époque 280 ne se généralise plus aux données de test. Et donc ce n'est pas un apprentissage utile. Nous disons que le réseau est sur-adapté ou surentraîné au-delà de l'époque 280.

Un autre signe de sur-ajustement peut être vu dans la précision de la classification des données d'entraînement:

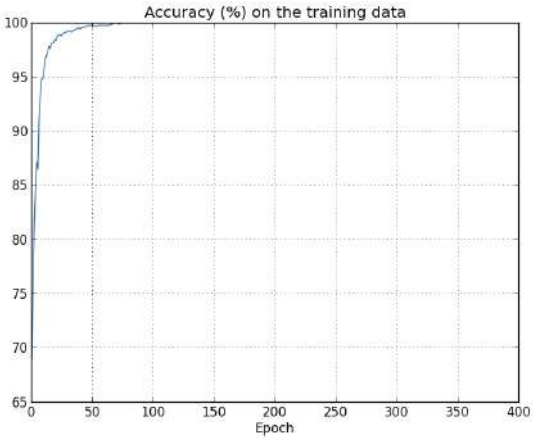


Figure 4. 30 : Variation de la précision dans la base d'apprentissage en fonction du nombre d'époques

La précision augmente jusqu'à 100%. Autrement dit, le réseau classe correctement les images d'entraînement. Pendant ce temps, notre précision de test ne dépasse que 82,27%. Donc, notre réseau apprend vraiment les particularités de l'ensemble d'apprentissage, pas seulement la reconnaissance des chiffres en général.

C'est presque comme si notre réseau mémorisait simplement l'ensemble d'entraînement, sans comprendre suffisamment les chiffres pour généraliser à l'ensemble de teste.

La façon évidente de détecter le sur-ajustement est d'utiliser l'approche ci-dessus, en gardant une trace de la précision des données de test pendant que notre réseau s'entraîne. Si nous constatons que la précision des données de test ne s'améliore plus, alors nous devons arrêter l'apprentissage. Bien sûr, à proprement parler, ce n'est pas nécessairement un signe de sur-ajustement. Il se peut que la précision des données de test et des données d'entraînement cesse de s'améliorer en même temps. Pourtant, l'adoption de cette stratégie empêchera le sur-ajustement.

L'un des meilleurs moyens de réduire le sur-apprentissage est d'augmenter la taille des données d'apprentissage. Avec suffisamment de données d'entraînement, même un très grand réseau est difficile à équiper. Malheureusement, les données de formation peuvent être coûteuses ou difficiles à acquérir, ce n'est donc pas toujours une option pratique.

6.1 Régularisation

Il existe beaucoup de techniques de régularisation. Nous décrivons brièvement deux autres approches pour réduire le sur-apprentissage: Dropout et l'augmentation artificielle de la taille de l'ensemble d'entraînement.

6.1.1 Dropout:

Dans cette technique on modifie le réseau lui-même. En particulier, supposons que nous ayons une entrée d'apprentissage x et une sortie souhaitée correspondante y. Normalement, nous nous entraînions en propageant vers l'avant x à travers le réseau, puis en rétropropagant pour déterminer la contribution au gradient. Avec l'abandon, ce processus est modifié. Nous commençons par supprimer aléatoirement (et temporairement) la moitié des neurones cachés du réseau, tout en laissant les neurones d'entrée et de sortie intacts. Après cela, nous nous retrouverons avec un réseau selon les lignes suivantes. Notez que les neurones de décrochage, c'est-à-dire les neurones qui ont été temporairement supprimés, sont toujours fantômes dans le réseau.

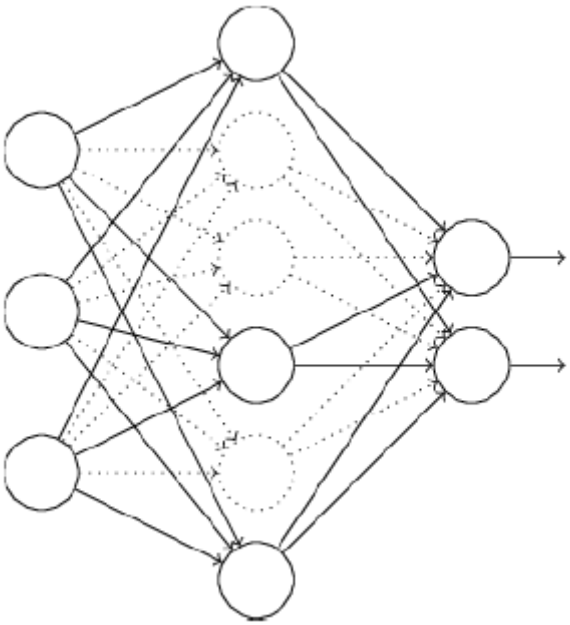


Figure 4. 31 : Suppression de quelques neurones par Dropout.

Nous propageons vers l'avant l'entrée x à travers le réseau modifié, puis propageons le résultat en retour, également à travers le réseau modifié. Après avoir fait cela sur un mini-lot d'exemples, nous mettons à jour les poids et les biais appropriés. Nous répétons ensuite le processus, en restaurant d'abord les neurones de décrochage, puis en choisissant un nouveau sous-ensemble aléatoire de neurones cachés à supprimer, en estimant le gradient pour un mini-lot différent et en mettant à jour les poids et les biais dans le réseau. En répétant ce processus encore et encore, notre réseau apprendra un ensemble de poids et de biais. Bien sûr, ces poids et biais auront été appris dans des conditions dans lesquelles la moitié des neurones cachés ont été abandonnés. Lorsque nous exécutons le réseau complet, cela signifie que deux fois plus de neurones cachés seront actifs. Pour compenser cela, nous divisons par deux les poids sortant des neurones cachés.

Cette procédure d'abandon peut sembler étrange et ponctuelle. Pour expliquer ce qui se passe, imaginant plutôt entraîner les réseaux de neurones de manière standard (pas de Dropout). Imaginez en particulier que nous formions plusieurs réseaux de neurones différents, tous utilisant les mêmes données d'entraînement. Bien sûr, les réseaux peuvent ne pas démarrer de manière identique et, par conséquent, après la formation, ils peuvent parfois donner des résultats différents. Lorsque cela se produit, nous pourrions utiliser une sorte de calcul de moyenne ou de vote pour décider quel résultat accepter. Par exemple, si nous avons formé cinq réseaux et que trois d'entre eux classent un chiffre en «3», il s'agit probablement vraiment d'un «3». Les deux autres réseaux font probablement une erreur. Ce type de schéma de moyenne est souvent considéré comme un moyen puissant (bien que coûteux) de réduire le sur-ajustement.

La raison en est que les différents réseaux peuvent s'ajuster de différentes manières, et la moyenne peut aider à éliminer ce type de sur-ajustement.

Lorsque nous abandonnons différents ensembles de neurones, c'est un peu comme si nous formions différents réseaux de neurones. Et donc la procédure de décrochage est comme la moyenne des effets d'un très grand nombre de réseaux différents. Les différents réseaux se surchargeront de différentes manières, et donc, espérons-le, l'effet net du décrochage sera de réduire le sur-ajustement.

6.1.2 Extension artificielle des données de d'apprentissage:

Obtenir plus de données d'apprentissage est une excellente idée. Malheureusement, cela peut être coûteux et n'est donc pas toujours possible dans la pratique. Cependant, il y a une autre idée qui peut fonctionner presque aussi bien, qui consiste à étendre artificiellement les données de formation. Cette idée est très puissante et a été largement utilisée. Le principe général est d'élargir les données d'entraînement en appliquant des opérations qui reflètent les variations du monde réel. Par exemple, faites tourner une image légèrement, disons 15 degrés.



Figure 4. 32 : Exemple d'augmentation de données par rotation de l'image.

7 Conclusion

En quelques années seulement, le Deep learning a permis de réaliser des percées considérables dans un large éventail de tâches qui étaient historiquement perçues comme extrêmement difficiles pour les ordinateurs, en particulier dans le domaine de la perception des machines: extraire des informations utiles à partir d'images, de vidéos, de sons, etc. Avec des données d'apprentissages suffisantes, il est possible d'extraire des données perceptuelles presque tout ce qu'un humain pourrait extraire. Par conséquent, on dit parfois que le Deep learning a résolu la perception, bien que ce soit vrai seulement pour une définition assez étroite de la perception.

CHAPITRE 5 CONTRIBUTION À L'IDENTIFICATION PAR LES VEINES DES DOIGTS BASÉ SUR UN RÉSEAU DE CNN FUSIONNÉS

1 Introduction

Actuellement, les systèmes basés sur l'apprentissage en profondeur ont remplacé les systèmes conventionnels en utilisant un mécanisme de bout en bout. En plus, le CNN se caractérise par une précision et une vitesse remarquable par rapport aux algorithmes traditionnels. À notre connaissance, aucun travail n'a été mené sur le CNN fusionné pour la reconnaissance des veines des doigts. Un problème lors de l'utilisation de CNN se produit dans la recherche de paramètres appropriés qui produisent le taux de reconnaissance le plus élevé. Pour cela, nous avons commencé une succession de tests où nous avons estimé un seul paramètre à chaque fois car la charge de calcul était très lourde. Nous avons choisi trois paramètres importants : la fonction d'activation, la taille de filtre et finalement le nombre approprié de couches CNN. L'architecture CNN obtenue a été utilisée pour construire un nouveau modèle pour la reconnaissance. Ce chapitre commence par l'intérêt des méthodes de reconnaissances basées sur l'apprentissage automatique. Puis nous exposons les métriques et les bases de données utilisées dans ce travail, avec plus de description sur notre prétraitement mené sur la base SDUMLA. La section 5 détaille le travail expérimental. Les résultats et la discussion sont présentés dans la section 6. Enfin, nous présentons nos conclusions dans la section 7.

2 Intérêt de l'apprentissage automatique pour les méthodes de reconnaissance

La correspondance biométrique est une « comparaison floue », parce que les traits biométriques capturés une seconde fois ne sont jamais exactement les mêmes que ceux de la première fois. Cette caractéristique d'appariement biométrique a conduit à l'utilisation de techniques d'apprentissage automatique, telles que les réseaux de neurones, la logique floue, l'apprentissage profond, etc., dans les algorithmes

biométriques. L'apprentissage automatique possède les propriétés essentielles d'être robuste au bruit et peut résoudre efficacement des problèmes complexes de reconnaissance de formes. En outre, l'apprentissage automatique est adaptatif, ordonné et possède généralement une architecture de calcul parallèle. Il modélise les caractéristiques biométriques complexes de manière très adaptative, sans faire beaucoup d'hypothèses à l'aide d'un modèle mathématique précis. Grâce à ces propriétés, l'apprentissage automatique s'est révélé une méthode efficace pour l'extraction de caractéristiques biométriques et le processus d'appariement. Il existe peu de recherches appliquant une approche d'apprentissage automatique dans le domaine de reconnaissance biométrique basée sur les veines de doigt. Cependant, la majorité tous des travaux antérieurs sur la reconnaissance de la veine des doigts reposent sur leur propre base de données interne, établie à cet effet [79].

3 Métriques

3.1 Matrice de confusion

Une matrice de confusion de taille n x n associée à un classifieur indique la classification prédite et réelle, où n est le nombre de différentes classes. Le tableau 5.1 présente une matrice de confusion pour n = 2, dont les entrées ont les significations suivantes:
vrai positif (true positive TP) est le nombre de prédictions positives correctes
vrai négatif (true negative TN) est le nombre de prédictions négatives correctes
faux positif (false positive (FP) FP) est le nombre de prédictions positives incorrectes
faux négatif (false negative FN) est le nombre de prédictions négatives incorrectes

Tableau 5. 1. Exemple de calcul par la matrice de confusion

	Condition Vrai	
Population total	Condition positive	Condition négative
Condition prévue positive	TP = 20	FP = 180
Condition prédite négative	FN = 10	TN= 1820

L'**accuracy** (Taux d'identification) et l'erreur de classification peuvent être obtenues à partir de cette matrice comme suit:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{5.1}$$

$$\text{Erreur} = \frac{\text{FP+FN}}{\text{TP+TN+FP+FN}} \tag{5.2}$$

3.2 Techniques de validation croisée

La validation croisée a deux intérêts par rapport aux tirages simples :

- Elle permet d'obtenir des résultats sur la base entière sans utiliser les mêmes données en apprentissage et en test.
- Elle permet d'augmenter la taille de la base d'apprentissage sans réduire la taille de la base de test.

Cela est possible par une idée simple qui consiste à poser que lorsque l'on fait un découpage en deux sous-bases, on va utiliser à tour de rôle chaque base en apprentissage et en test. En effet on fera une première étape pour laquelle un sous-ensemble servira pour l'apprentissage et l'autre pour le test et dans la deuxième étape on échangera les rôles. Cela permet ainsi d'avoir des résultats sur la base complète car à chaque étape on a des résultats sur la moitié de la base, mais toujours suite à des apprentissages sur des données différentes de celles de test.
On peut donc par cette méthode de validation croisée obtenir des résultats sur l'ensemble de la base de données et donc ne plus se réduire à seulement un sous-ensemble plus petit. Cela est très utile surtout car la taille réduite des bases de données est un problème récurrent en reconnaissance des formes et particulièrement en fusion de données biométriques.

Dans ce contexte de base de taille réduite, il est également intéressant de pouvoir augmenter la taille de la base d'apprentissage en ne se limitant pas à la moitié des données, sans pour autant réduire le nombre de données en test car on teste toujours sur l'ensemble de la base par le procédé de "croisement" évoqué précédemment. Pour augmenter la taille de l'ensemble d'apprentissage on utilise de la validation croisée dites "à k blocs" (k-fold cross validation). La validation croisée à k blocs consiste à découper des données en k sous-ensembles de même taille. (k - 1) sont alors utilisés en apprentissage et le dernier en test. Cette opération est reproduite k fois pour que chaque sous-ensemble soit utilisé une fois en test. On obtient donc comme dans le cas précédent des résultats pour l'ensemble des données mais pas toujours appris sur le même sous-ensemble d'apprentissage.

Le cas extrême de la validation croisée à k blocs est appelé "Leave-One-Out-Cross-Validation". Dans LOOCV, nous divisons le jeu de données en deux parties. Dans une partie, nous avons une seule observation, qui est nos données de test, et dans l'autre partie, nous avons toutes les autres observations de l'ensemble de données formant nos données d'apprentissage. Si nous avons un ensemble de données avec n observations, les données d'apprentissage contiennent n-1 observation et les données de test contiennent 1 observation. Ce processus est itéré pour chaque point de données, comme indiqué ci-dessous. La répétition de ce processus n fois génère n fois l'erreur quadratique moyenne (MSE) [11].

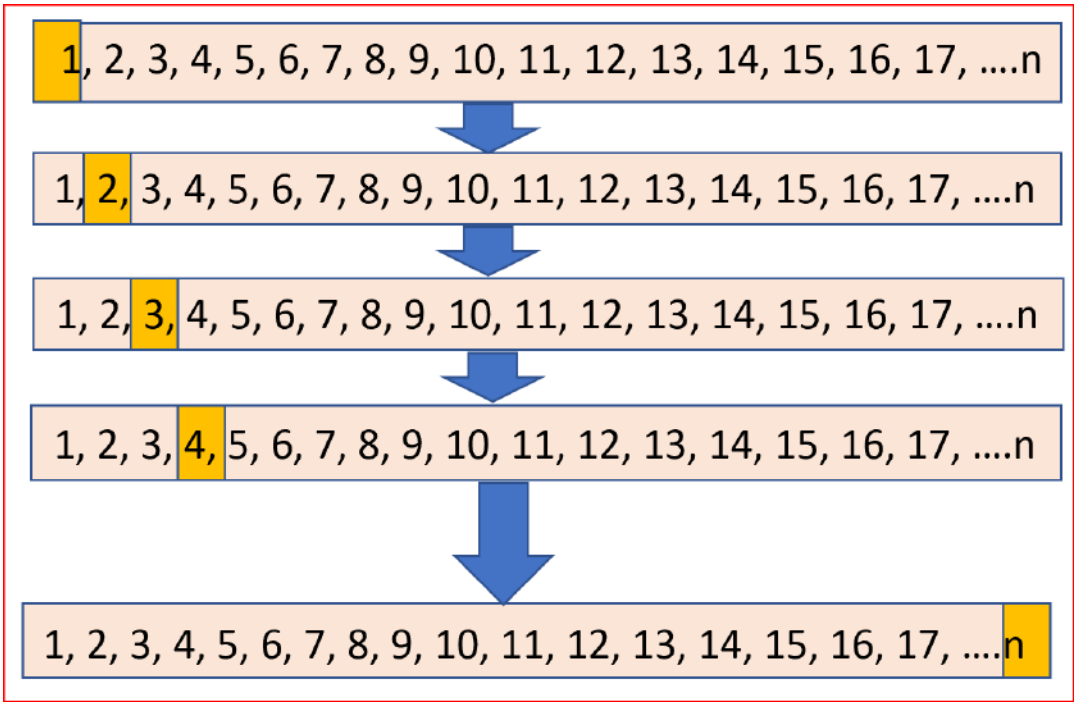


Figure 5. 1: La validation croisée LOOCV

4 Bases de données utilisées

Nous avons utilisé trois bases de données des veines des doigts: la base de données FV-USM de l'Université Sains Malaysia [21], la base de données SDUMLA-HMT de l'Université du Shandong [36] et la base de données THU-FVFDT2 de l'Université Tsinghua [80].

4.1 Base de données FV-USM

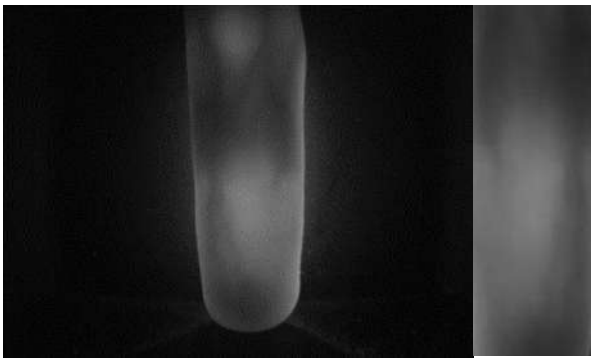
La base de données contient les informations sur la veine du doigt et de la géométrie du doigt. Elle peut être utilisé pour vérifier les systèmes biométriques uni-modaux (la veine du doigt ou la géométrie du doigt) ou bimodal (fusion de la veine et de la géométrie).

Compléter avec la RI extrait pour la reconnaissance des veines de doigts, cette base de données est disponible sur le Web afin d'aider d'autres chercheurs dans les domaines connexes à tester et à évaluer leurs algorithmes en fonction de la base de données de référence standard.

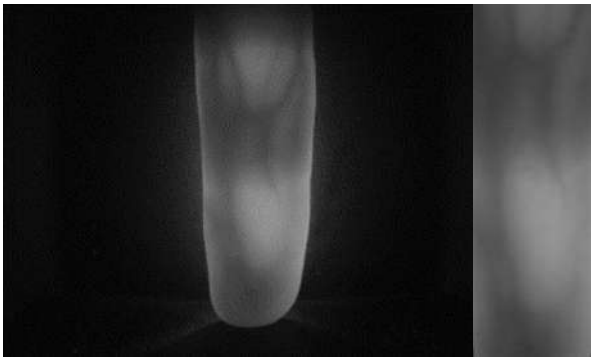
Les images de la base de données ont été recueillies auprès de 123 bénévoles, soit 83 hommes et 40 femmes, membres du personnel et étudiants de l'université Sains Malaisie. L'âge du sujet variait de 20 à 52 ans. Chaque sujet a fourni quatre doigts: index gauche, majeur gauche, et index droit, majeur droit, ce qui a donné au total 492 doigts obtenus. Les images de doigts capturées ont fourni deux caractéristiques importantes: la géométrie et la forme de la veine. Chaque doigt a été capturé six fois en une session et chaque individu a participé à deux sessions, séparées par plus de deux semaines. Dans la première session, un total de 2952 (123 x 4 x 6) images ont été collectées. Par conséquent, à partir de deux sessions, un total de 5904 images provenant

de 492 classes de doigts a été obtenu. La résolution spatiale des images des doigts capturés était de 640 x 480 et de 256 en niveaux de gris.

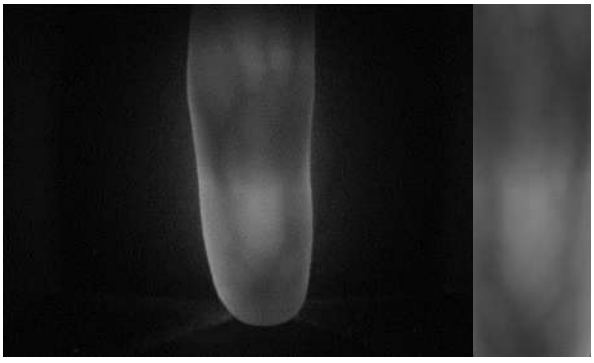
La RI de l'image de la veine est de 300 × 100 pixels. Les images de la première session sont généralement utilisées pour l'apprentissages et les images de la deuxième session sont utilisées comme images de test.



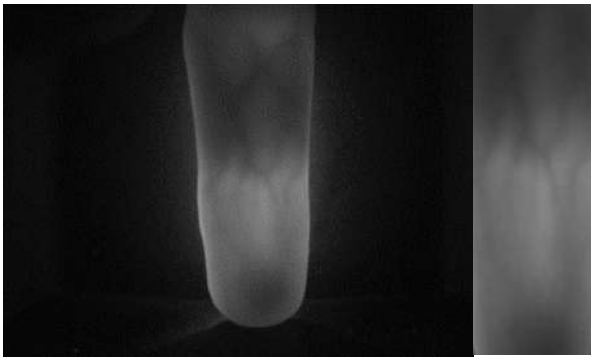
(a)



(b)



(c)



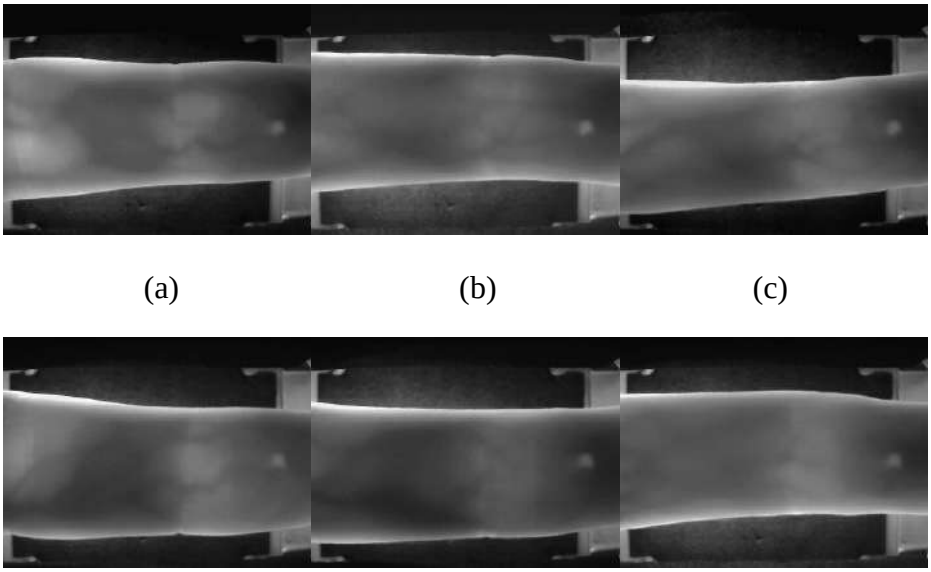
(d)

Figure 5. 2: Un aperçu sur les images de la base de données des veines des doigts FV-USM avec la RI, (a) index gauche, (b) majeur gauche, (c) index droit, et (d) majeur droit

4.2 Base de données SDUMLA

En 2010, le laboratoire d'apprentissage automatique et applications de l'Université du Shandong (SDUMLA) a mis en place la base de données sur les traits multimodaux homologues, appelée SDUMLA-HMT Database. Cette base comprend 5 traits biométriques, à savoir le visage, la veine du doigt, la démarche, l'empreinte digitale et l'iris.

Pour la reconnaissance par les veines des doigts, l'appareil utilisé pour capturer les images des veines est conçu par Joint Lab for Intelligent Computing and Intelligent Systems de l'Université de Wuhan. Dans le processus de capture, chaque sujet a été invité à fournir des images de son index, de son majeur et de son annulaire, gauche et droit, des deux mains, et la collecte pour chacun des 6 doigts est répétée 6 fois pour obtenir 6 images de la veine du doigt. Par conséquent, Cette base contient 3816 images de 320 × 240 pixels au format de fichier BMP.



(a) (b) (c)

(d) (e) (f)

Figure 5. 3 : Un aperçu de la base de données des veines des doigts SDUMLA-HMT, (a) (d) Index, (b) (e) majeur, (c) (f) annulaire

4.3 Base de données THU-FVFDT2

Cette base de données offre 610 images, dont les 220 premiers RI sont extraits de THU-FVFDT1, les 390 derniers RI sont capturés en deux sessions différentes avec un intervalle allant d'environ trois jours à une semaine, et chaque image correspond à un doigt d'un seul sujet. La majorité des sujets sont des étudiants et du personnel bénévole de la Graduate School de Shenzhen, Université Tsinghua. Un prétraitement des images a été appliqué en extrayant la RI et en normalisant l'image à 200 x 100 pixels.

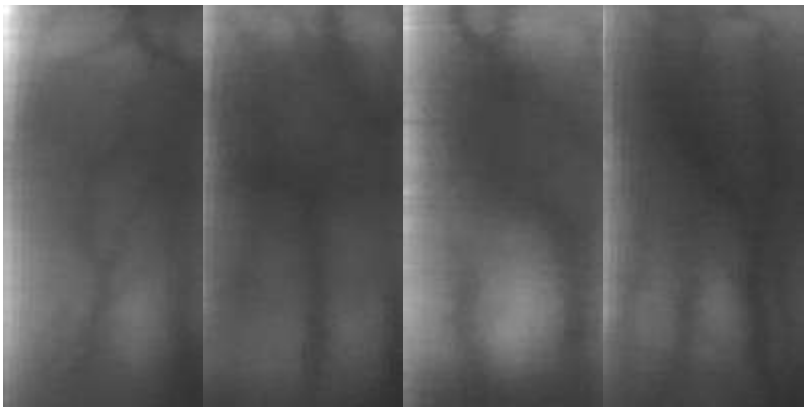


Figure 5. 4 : Un aperçu de la base de données des veines des doigts SDUMLA-HMT, (a) (d) Index, (b) (e) Majeur, (c) (f) Annulaire

4.4 Prétraitement des images des bases de données

Les images des veines des bases de données ont été acquises par différents systèmes d'acquisition. Malheureusement, les images infrarouges étaient sensibles à l'éclairage et contiennent de grands arrière-plans, ce qui rend le prétraitement particulièrement important.

Dans tous les modes de reconnaissance biométrique, le prétraitement est la première étape pour adapter les images d'entrée aux étapes suivantes. Il regroupe toutes les opérations de segmentation, de redimensionnement et de filtrage pour rendre l'image plus claire et donc permettre l'extraction de caractéristiques [82].

L'extraction de la RI est la première étape pour séparer le doigt de l'arrière-plan et supprimer les zones indésirables. Les bases de données FV-USM et THU-FVFDT2 contiennent des images brutes et des images avec RI, cependant, la base de données SDUMLA nécessite une opération d'extraction supplémentaire de la région du doigt pour obtenir ces informations. Pour cette dernière base de données, nous avons adopté la procédure d'extraction de la RI décrite dans Banerjee et al [77].

Pour améliorer la qualité des images, nous avons utilisé la méthode d'égalisation de l'histogramme CLAHE, qui est la méthode d'amélioration la plus utilisée. Cette technique normalise la luminosité de l'image en augmentant le contraste de petites régions. Toutes les images des trois bases de données ont été redimensionnées à une résolution de 300 × 100 pixels afin qu'elles soient soumises au même traitement. Un exemple de prétraitement de l'image des doigts est illustré dans la figure 5.5. La figure 5.6 montre des échantillons d'images originales et améliorées provenant des trois bases de données.

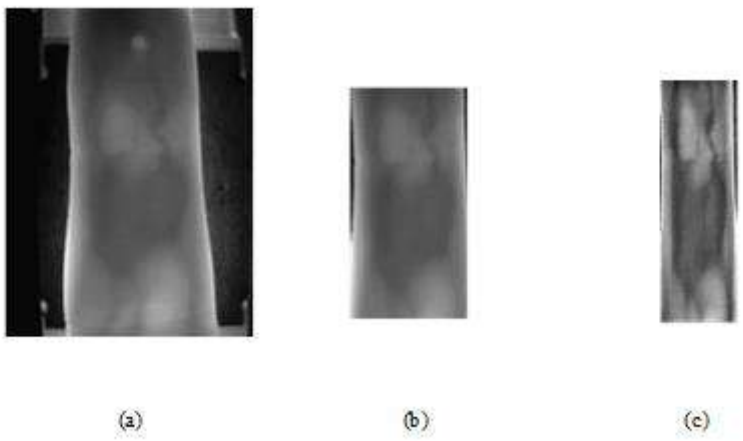


Figure 5. 5 : Extraction de la RI de la base de données des veines des doigts SDUMLA-FV (a) Image des veines des doigts; (b) extraction de la RI; (c) Redimensionnement et amélioration de l'image.

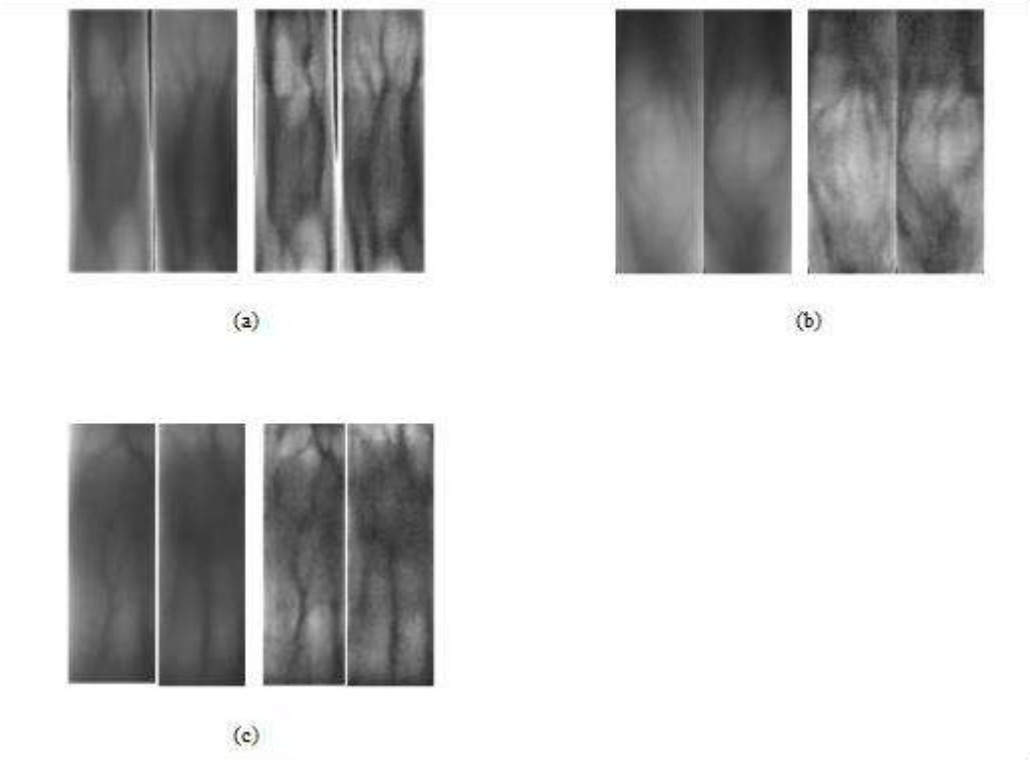


Figure 5. 6 : Exemples de deux images originales et de deux images améliorées par CLAHE pour la base de données (a) SDUMLA; (b) FV-USM et (c) THU-FVFDT2

5 Travaux expérimentaux

5.1 Stratégies d'apprentissage

Dans toutes les expériences, chaque doigt de chaque personne était considéré comme une classe distincte. En général, l'apprentissage a été effectué avec les images de la première session, et le test a été effectué avec les images de la deuxième session. Selon les bases disponibles, certaines bases comprennent deux sessions comme dans le cas des bases de données FV-USM et THU-FVFDT2, tandis que la base de données SDUMLA ne présente qu'une seule session, ce qui nous a conduit à choisir un certain nombre d'images sur les 6 images de chaque classe comme images d'apprentissage et le reste sera utilisé pour le test. Par exemple, quatre images ont été utilisées pour l'apprentissage et deux images ont été utilisées pour le test. Cette stratégie de formation d'une session a également été utilisée pour la base de données FV-USM dans Das et al [70], les auteurs ont montré que l'utilisation de toutes les classes en une session produisait de meilleurs taux de reconnaissance. Par conséquent, dans la base de données FV-USM, nous avons considéré toutes les classes pour l'analyse. Cela signifiait que 984 classes étaient considérées au lieu de 492 classes et assuraient le même usage que celui de la base de données SDUMLA. Nous avons appliqué cette stratégie d'apprentissage pour rechercher le CNN optimale. Le modèle final a été ensuite testé pour les deux sessions. La base de données THU-FVFDT2 contient une image dans la

première session et une image dans la deuxième session, nous avons donc utilisé cette base de données contenant 610 classes.

En revue, nous avons utilisé deux stratégies d'apprentissage: une session pour les bases de données FV-USM et SDUMLA et deux sessions pour les bases de données FV-USM et THU-FVFDT2. Les expériences ont été implémentées en utilisant python (3.5) avec une configuration système de 8 Go de RAM; Carte graphique NVIDIA GeForce GT 710, processeur i5-7500 3,40 GHz et le système d'exploitation Windows® 10.

5.2 Optimisation d'une nouvelle architecture CNN

Dans cette section, nous avons testé plusieurs topologies de réseau pour trouver celle qui donne le meilleur taux de reconnaissance. De nombreuses configurations ont été utilisées et le nombre de couches de convolution (CL) a été varié de 1 à 6 avec une ou plusieurs couches de max-pooling dans chaque CNN. Pour évaluer la performance du modèle, nous avons utilisé la technique LOOCV (Leave-One-Out-Cross-Validation). La base de données FV-USM contient N = 6 images pour chaque individu, donc les images d'entraînement contiennent N-1 images et les données de test contiennent une seule image. Ce processus a été répété N fois dans chaque classe, et la moyenne des N cycles de validation croisée a été calculée, ce qui donne une meilleure estimation des performances du modèle.

5.2.1 Estimation des paramètres du CNN

Dans cette étape, nous avons utilisé différents CNN et modifié plusieurs paramètres, tels que la fonction d'activation, la taille du noyau, le nombre des CL et le nombre de couche Max-pooling, comme il est décrit dans les tableaux 5.1, 5.3 et 5.5.

La taille de la fenêtre de Max-pooling était deux, et tous les CNN avaient les mêmes couches terminales, qui sont Flatten, Dropout et Softmax. La taille de l'image d'entrée était de 120 x 120 pixels.

Dans la première expérience, nous avons testé cinq fonctions d'activation: linéaire, tanh, relu, softsign et softplus. Nous avons utilisé un CNN simple contenant deux couches de convolution, et les résultats sont présentés dans le tableau 5.3. La fonction relu a donné la précision maximale, nous avons donc utilisé cette fonction dans tous les CNN ultérieurement.

Tableau 5. 2. Configurations adoptées du réseau CNN en variant la fonction d'activation.

	Configuration CNN				
CNN label	A	B	C	D	E
function	Linear	Tanh	Relu	Softsign	Softplus
Number of layers	2 CL	2 CL	2 CL	2 CL	2 CL

Max-Pooling					
Flatten					
Dense					
Dropout (0.5)					
Dense					
Softmax					

Tableau 5. 3. Taux de reconnaissance obtenu pour les différentes fonctions d'activation adoptées

CNN	A	B	C	D	E
Accuracy	0.929878	0.903625	0.946477	0.936484	0.624153

La deuxième expérience a été porté sur la taille du filtre. Nous avons choisi trois valeurs: 3, 5 et 7. Nous avons aussi utilisé neuf CNN, trois CNN étiquetés H, K et N, qui n'utilisaient qu'une seule couche avec les trois tailles présentées, trois CNN étiquetés I, L et O, qui utilisaient deux couches et trois CNN étiquetés J, M et P, qui utilisaient trois couches. Les résultats sont présentés dans le tableau 5.5. Le taux de reconnaissance le plus élevé a été obtenu lorsque nous avons utilisé une taille de noyau égale à 3.

Tableau 5. 4. Configurations adoptées du réseau CNN an variant les caractéristiques de la taille de la carte et le nombre de CL.

Configuration CNN								
H	I	J	K	L	M	N	O	P
1 Weight Layers	2 Weight Layers	3 Weight Layers	1 Weight Layers	2 Weight Layers	3 Weight Layers	1 Weight Layers	2 Weight Layers	3 Weight Layers
CL (3X3)	CL (3X3) CL (3X3)	CL (3X3) CL (3X3) CL (3X3)	CL (5X5)	CL (5X5) CL (5X5)	CL (5X5) CL (5X5) CL (5X5)	CL (7X7)	CL (7X7) CL (7X7)	CL (7X7) CL (7X7) CL (7X7)
Max-pooling								
Flatten								

Dense
Dropout (0.5)
Dense
Softmax

Tableau 5. 5. Taux de reconnaissance obtenu pour les différentes tailles du noyau et le nombre de CL

Taille du noyau	3	5	7
CNN	H	K	N
Accuracy of 1 Weight Layers	0.913279	0.870088	0.892107
CNN	I	L	O
Accuracy of 2 Weight Layers	0.907774	0.896426	0.30708
CNN	J	M	P
Accuracy of 3 Weight Layers	0.923272	0.893631	0.417514

Dans la troisième expérience, nous avons testé la topologie du réseau pour obtenir une estimation de l'identification en fixant les deux paramètres les plus optimaux précédemment obtenus, qui étaient la fonction relu et une taille de noyau égale à 3. Plusieurs configurations ont été utilisées (voir Tableau 5.6), le nombre de CL a été varié de 1 à 6 avec une ou plusieurs couches max-pooling dans chaque CNN. Le tableau 5.6 montre les résultats pour les différentes configurations choisies.

Tableau 5. 6.Configurations adoptées du réseau CNN en variant le nombre de couches de convolution et la couche max-pooling.

CNN Configuration					
Q	R	S	T	U	V
2 Weight Layers	4 Weight Layers	6 Weight Layers	3 Weight Layers	6 Weight Layers	9 Weight Layers
CL	CL	CL	CL	CL	CL
	CL	CL		CL	CL
		CL			CL
Max-pooling					
CL	CL	CL	CL	CL	CL
	CL	CL		CL	CL
		CL			CL
Max-pooling					

			CL	CL	CL
				CL	CL
					CL
			Max-pooling		
Flatten					
Dense					
Dropout (0.5)					
Dense					
Softmax					

Tableau 5. 7. Taux de reconnaissance obtenu pour les différents nombres de couches de convolution et la couche max-pooling

CNN Label	Q	R	S	T	U	V
Accuracy	0.912093	0.921917	0.901762	0.928692	0.606707	0.123137

Le résultat optimal a été obtenu en utilisant le CNN étiqueté T, qui se composait de trois couches de convolution avec la fonction d'activation relu pour calculer la sortie, suivie des couches max-pooling, pour le sous-échantillonnage. Dans la couche max-pooling, la valeur maximale des régions carrées avec une taille de 2 × 2 a été prise. Enfin, les dernières couches étaient Flatten, denses, Dropout et softmax comme il est décrit dans le tableau 5.8 et présentées dans la figure 5.7.

Tableau 5. 8. Description des couches du CNN proposé.

Index	Layer type	Filter size	Number of filters	Output shape
1	Convolution	3	16	(118, 118, 16)
2	Max-pooling	2		(59, 59, 16)
3	Convolution	3	16	(57, 57, 16)
4	Max-pooling	2	-	(28, 28, 16)
5	Convolution	3	16	(26, 26, 16)
6	Max-pooling	2	-	(13, 13, 16)
7	Dropout	-	-	(13, 13, 16)
8	Flatten	-	-	2704
9	Dense	-	-	1500
10	Dropout	-	-	1500

11	Dense	-	-	984
12	Softmax	-	-	

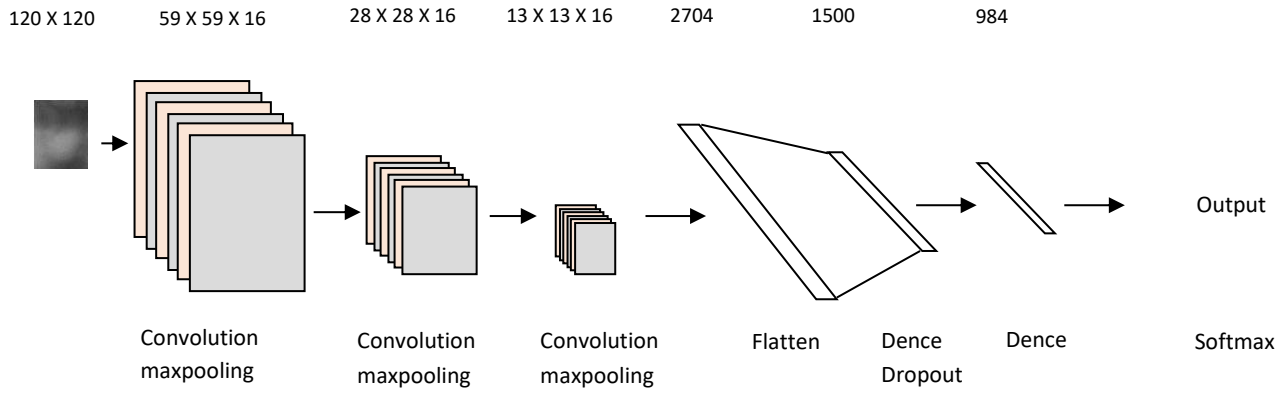


Figure 5. 7: Architecture du CNN proposé.

5.2.2 Fusion d'informations dans les CNN

Dans cette partie, nous proposons un modèle basé sur la fusion des informations à travers plusieurs CNN ayant la configuration optimale obtenue précédemment. Cette architecture augmente en largeur en rassemblant les CNN en parallèle et fusionne l'information en entrée, ou en milieu, ou en sortie.

Il existe trois façons pour fusionner les informations dans les CNN. Cela peut se faire dans la première couche, 'fusion précoce', dans la couche intermédiaire, 'fusion moyenne', ou à la fin du réseau, 'fusion tardive'. Un exemple de fusion tardive utilisant trois CNN est illustré sur la figure 5.8. Les résultats des trois réseaux sont fusionnés à la fin du processus [83]. L'idée de la fusion est que les résultats des réseaux séparés peuvent faire progresser la robustesse du modèle [81].

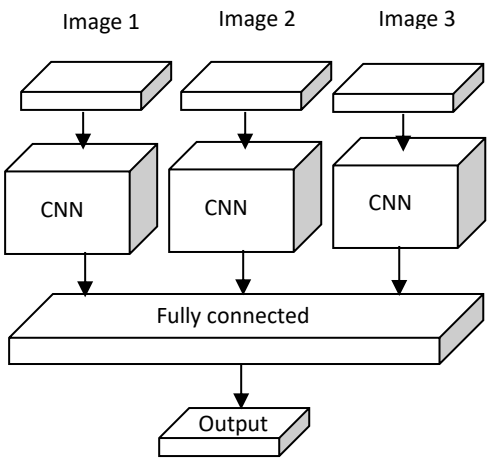


Figure 5. 8 : Architecture de fusion tardive.

Dans ce travail, nous avons changé les informations dans l'image d'entrée et utilisé la même configuration des réseaux CNN. Nous avons formé les images d'entrée à partir de l'image originale. Cette dernière a été traitée par trois méthodes: la méthode CLAHE, le filtrage par les filtres Gabor [82] et la fusion DCT [84] (figure 5.9). Ultérieurement, chaqu'une de ces images sera l'entrée d'un CNN afin de mieux présenter leurs informations

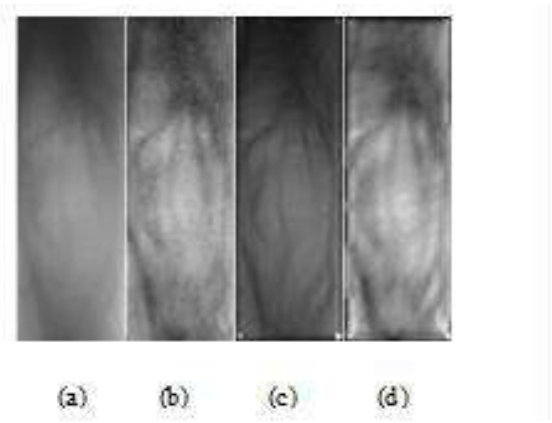


Figure 5. 9 : Image utilisées pour la fusion : (a) l'original; (b) traitée par CLAHE (c); filtrée par Filtre de Gabor; (d) fusionnée par DCT

Plusieurs configurations ont été envisagées à ce stade, en fonction du nombre d'images. Le tableau 5.9 montre les différentes configurations utilisées pour tester et rechercher le modèle optimal. Tous les tests ont été effectués à l'aide de la base de données FV-USM. Enfin, nous avons testé notre modèle sur les bases de données SDUMLA et THU-FVFDT2.

Tableau 5. 9. Tableau 8. Différentes combinaisons d'images d'entrée.

étiquettes	Combinaison	Original	CLAHE	Gabor	Fusion
A	**2**	√	√		
B	**2**	√		√	
C	**2**	√			√
D	**2**		√	√	
E	**2**		√		√
F	**2**			√	√
G	**3**	√	√	√	
H	**3**	√	√		√
I	**3**	√		√	√
J	**3**		√	√	√
K	**4**	√	√	√	√

Tableau 5. 10. Taux d'identification obtenu des différentes configurations.

CNN label	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**
Accuracy	0,9695	0,9676	0,9524	0,9663	0,9596	0,9603	0,9665	0,9545	0,9471	0,9412	0,9551

Toutes les combinaisons ont une précision supérieure à 94% (tableau 5.10). Par conséquent, la fusion utilisant plusieurs CNN améliore les performances du système avec une amélioration signification statistique de 5% de taux d'identification par rapport au résultat obtenu en utilisant le CNN étiqueté T. En effet, le réseau qui donne le taux d'identification le plus élevé est la combinaison étiquetée A, qui combine l'image originale et l'image améliorée par la méthode CLAHE. Par conséquent, nous choisissons cette combinaison comme modèle optimal. Le tableau 5.11 présente les résultats complets obtenus lorsque nous avons utilisé le modèle proposé pour différent nombre d'images d'apprentissage et de test pour toutes les bases de données. Le taux d'identification augmente avec l'augmentation du nombre d'images d'apprentissage.

Tableau 5. 11.Taux d'identification obtenu des différentes bases de données pour différent nombre d'images d'apprentissage.

Database	Training images				
	1	2	3	4	5
FV-USM	89.91%	91.13%	94.21%	96.15%	96.75%
One session					
SDUMLA	65%	66%	84.64%	89.88 %	99.48%
THUFVDT2	99.51%	-	-	-	-

6 Comparaison avec des travaux précédents

Il existe dans la littérature des méthodes qui ont donné des résultats atteignant 98% de taux d'identification. Cependant, ces méthodes avaient un ensemble d'apprentissage supérieur à l'ensemble de test et un petit groupe de sujets, c'est-à-dire un nombre de classe réduit. Dans cette section, nous choisissons de comparer notre méthode avec les méthodes qui utilisent les mêmes bases de données et stratégies d'apprentissage.

Les performances du système de reconnaissance des veines des doigts dépendent des stratégies de test dans une base de données qui présente différents problèmes, tels que celui du faible contraste et du bruit. Pour garantir que le résultat expérimental illustre la robustesse, les trois bases de données publiques seront utilisées avec différents nombres d'images d'apprentissage pour évaluer les performances du système proposé. Les résultats d'analyse comparative de notre approche avec les méthodes précédentes sont présentés dans les tableaux 5.12, 5.13, 5.14, 5.15 et 5.16 pour les trois bases de données FV-USM, SDUMLA et THUFVDT2.

6.1 Comparaison des résultats en utilisant la base de données FV-USM

Dans le cas de la base de données FV-USM, nous avons effectué deux comparaisons en fonction de la stratégie d'apprentissage utilisée.

Avec la première stratégie utilisant deux sessions (tableau 5.12), notre approche a donné un bon résultat car elle a dépassé toutes les autres méthodes envisagées à l'exception de la méthode MC lorsque six images ont été utilisées pour l'apprentissage et six images ont été utilisées pour le test. Un score minimal a été noté lors de l'utilisation d'une seule image pour l'apprentissage.

Tableau 5.12. Comparaisons du taux d'identification dans le cas de la base de données FV-USM en utilisant deux sessions.

FV-USM Tow Session	**MC [31]**	**RLT[27]**	**Das et al [70]**	**Notre approche**
Training 1 Testing6	90.34 %	78.28 %	72.97 %	70.10 %
Training 6 Testing6	85.24 %	78.28 %	71.11 %	81.71 %

Pour la deuxième stratégie (tableau 5.13), lorsque nous avons utilisé une session d'apprentissage, notre approche n'a été comparée qu'avec la méthode de Das et al. [70] puisque c'était la seule recherche qui proposait l'utilisation d'une session sur la base de données FV-USM.

Les résultats obtenus à partir de la base de données FV-USM étaient légèrement insuffisants lorsque nous avons appliqué notre modèle en utilisant une seule session.

Notre approche a été dépassée par Das et al. [70] avec une différence d'environ 2 à 3% pour différents nombres d'images d'apprentissage.

Tableau 5. 13. Comparaisons du taux d'identification dans le cas de la base de données FV-USM en utilisant une seule session.

FV –USM One Session	Training 1 Testing 5	Training 2 Testing 4	Training 3 Testing 3	Training 4 Testing 2
[70]	91.75 %	94.82 %	97.83 %	98.58 %
Notre approche	89.91 %	91.13 %	94.21 %	96.15 %

Il est clair que la valeur de la précision augmente avec l'augmentation du nombre d'images d'apprentissage, notre approche a donné un taux d'identification de 81,71% lors de l'utilisation de deux sessions (tableau 5.12) par rapport à la méthode Das et al. (71.11%) en utilisant six images pour la formation, ce qui vaut 10% de plus.

6.2 Comparaison des résultats en utilisant la base de données SDUMLA

La base de données SDUMLA contient des images obtenues avec une seule session, nous présentons ici deux comparaisons selon les résultats disponibles dans la littérature pour cette base de données.

Le tableau 5.14 montre les résultats d'un premier test utilisant quatre images d'entraînement. Comme indiqué dans ce tableau, notre système de reconnaissance ne peut garantir un taux d'identification supérieur aux méthodes [91] and [70], bien qu'il dépasse les résultats obtenus par [31] and [27].

Tableau 5. 14. Comparaisons du un taux d'identification dans le cas de la base de données SDUMLA pour quatre images de d'entraînement et deux images de test.

SDUMLA	Van et al**[91]**	Van et al **[91]**	MC **[31]**	RLT **[27]**	Das et al **[70]**	Notre approche
Training 4 Testing2	92.50%	95.67%	86.01%	87.11%	97.48%	89.88%

Le tableau 5.15 montre les résultats d'un deuxième test utilisant toujours la base de données SDUMLA mais avec cinq images d'apprentissage. En comparaison avec la base de données FV-USM, une situation différente se produit lors de la comparaison du système proposé avec cinq travaux précédents, comme il est exposé dans ce tableau. Dans l'ensemble, notre approche a donné la plus haute précision d'identification d'environ 99,48% par rapport à toutes les méthodes invoquées lors de l'utilisation de

cinq images pour l'apprentissage. Par exemple, notre approche dépasse la méthode Das et al. [96] de 0,58% et la méthode Banerjee et al [88] de 8,76%.

En comparaison, en utilisant la base de données FV-USM, le modèle proposé a présenté une amélioration significative du taux d'identification.

Tableau 5. 15. Comparaisons du taux d'identification dans le cas de la base de données SDUMLA pour cinq images d'entraînement et une image de test.

SDUMLA	[101]	[88]	MC [33]	RLT [29]	Das et al [70]	Notre approche
Training 5 Testing1	97.76%	90.72%	97.95%	96.06%	98.90%	99.48%

6.3 Comparaison des résultats en utilisant la base de données THUFVDT2

Notre modèle de reconnaissance a été comparé aux méthodes typiques utilisant la même base de données (THUFVDT2), à savoir [89], [90] and [80] comme le montre le tableau 5.16. Cette base de données contient 610 images et chaque classe ne contient qu'une image de test et une image d'apprentissage. Cela diminue l'erreur de classification et donne un taux de précision plus élevé par rapport aux résultats obtenus lors de l'utilisation de la base de données FV-USM, qui contient 5 904 images, et de la base de données SDUMLA, qui contient 3 816 images. Nous avons obtenu un taux d'identification atteignant 99,56%, ce qui dépasse les autres méthodes utilisant cette base de données.

Tableau 5. 16. Comparaisons du taux d'identification dans le cas de la base de données THUFVDT2.

THUFVDT2	Shazeedaet al.	Qui et al.	Wang et al.	Notre approche
Training 1Testing1	**[89]**	**[90]**	**[80]**	
Accuracy	99.51%	96.52%	90 .33%	99.56%

7 Conclusion

Dans ce travail, nous avons présenté une nouvelle approche basée sur le modèle d'apprentissage en profondeur pour réaliser l'identification personnelle à travers des modèles de veines des doigts. La supériorité du CNN est liée à la distinction des données d'entraînement. Dans le cas de la reconnaissance des veines des doigts, les bases de données sont petites et les structures des veines sont simples et influencent le processus d'apprentissage.

Le CNN fusionné est une architecture récemment envisagée qui utilise de nombreux réseaux qui détiennent des tâches différentes. Pour cette étude, au lieu d'utiliser une seule image dans un modèle multi-réseaux, nous avons utilisé la même

architecture pour différentes images des veines des doigts puis nous avons fusionné leurs sorties.

Il n'y a pas de règles absolues pour former un modèle CNN. Nous avons fait des choix qui semblent absolutistes, comme la fonction d'activation, la taille du filtre et le nombre de couches, et nous avons examiné régulièrement les opportunités. Nous avons étudié les performances de généralisation de notre architecture et observé des améliorations significatives.

Le modèle conçu a été testé pour trois bases de données publiques. Comparé à un CNN unique, le CNN fusionné apporte des informations complémentaires, contribuant à une identification plus précise.

CONCLUSION GÉNÉRALE

En raison de l'évolution rapide de la situation économique et sociale, la sécurité de l'information s'est accrue. Cependant, de nombreuses techniques d'authentification classiques, telles que l'utilisation de clés et de mots de passe, ne peuvent répondre à toutes les attentes, car elles peuvent être volées, perdues ou même être oubliées.

Etant donné que les nouvelles techniques basées sur l'authentification et l'identification avec des caractéristiques biométriques ne peuvent pas être oubliées, perdues ou copiées; ces techniques sont apparues comme des méthodes pratiques, fiables et sécurisées.

Les techniques de reconnaissance basées sur les images des veines sont non invasives, pratiques, précises et ne peuvent être forgées facilement. Elles ont donc suscité beaucoup d'attention au cours des dernières années. Comparées aux techniques biométriques classiques.

En premier lieu, nous avons utilisé les méthodes d'amélioration de la qualité de l'image comme CLAHE et le filtrage par le filtre de Gabor, puis nous avons proposé une technique hybride basée sur la fusion DCT.

Ensuite, nous avons présenté dans cette thèse l'identification des veines de doigts basée sur l'utilisation de l'apprentissage profond. Comme toutes les techniques d'apprentissage machine, l'apprentissage profond vise à entraîner un système pour qu'il résolve des situations sans que tous les paramètres nécessaires à la résolution du problème n'aient été calculés par l'implémentateur. L'objectif est d'entraîner un algorithme à paramètres variables (une « boîte noire ») à prendre une décision correcte concernant une tâche donnée. L'apprentissage se fait en optimisant les paramètres variables pour améliorer les décisions prises.

Nous avons conçu une nouvelle architecture pour améliorer le taux de reconnaissance. Il a été constaté que sa performance est comparable à celles des meilleures méthodes existantes dans l'état de l'art.

Les recherches futures peuvent inclure la collecte de plus grands ensembles de données, y compris des données 3D, ainsi que des données démographiques des personnes concernées.

Ces ensembles de données plus volumineux nous permettront de communiquer des chiffres de performance avec une plus grande confiance. Il permettra également la recherche de facteurs tels que l'âge, le sexe et l'origine ethnique sur la performance. La communauté de la recherche bénéficierait également grandement de méthodes d'essai et d'ensembles de données normalisés. Les performances biométriques peuvent encore

être améliorées en fusionnant d'autres traits de doigt tels que les empreintes digitales, le motif de pli du doigt et la forme du doigt.

BIBLIOGRAPHIE

[1]Ababsa, Souhila Guerfi. "Authentification d'individus par reconnaissance de caractéristiques biométriques liées aux visages 2D/3D." *Evry-Val d'Essonne* (2008).

[2]Blue, James L., et al. "Evaluation of pattern classifiers for fingerprint and OCR applications." *Pattern Recognition* 27.4 (1994): 485-501.

[3] Kabbara, Yeihya. *Caractérisation des images à Rayon-X de la main par des modèles mathématiques: application à la biométrie*. Diss. 2015.

[4]Latha, L., M. Pabitha, and S. Thangasamy. "A novel method for person authentication using retinal images." *2010 International Conference on Innovative Computing Technologies (ICICT)*. IEEE, 2010.

[5]De Marsico, Maria, Michele Nappi, and Daniel Riccio. "Face authentication with undercontrolled pose and illumination." *Signal, Image and Video Processing* 5.4 (2011): 401.

[6] Dai, Peng, and Yann Soon. "An improved model of masking effects for robust speech recognition system." *Speech Communication* 55.3 (2013): 387-396.

[7] Michael, Katina. "The legal, social and ethical controversy of the collection and storage of fingerprint profiles and DNA samples in forensic science." *2010 IEEE International Symposium on Technology and Society*. IEEE, 2010

[8] Michael, Goh Kah Ong, et al. "A contactless biometric system using palm print and palm vein features." *Advanced Biometric Technologies* (2011): 155-177.

[9] Abdel-Mottaleb, Mohamed, et al. "Challenges of developing an automated dental identification system." *2003 46th Midwest Symposium on Circuits and Systems*. Vol. 1. IEEE, 2003

[10] Jain, Anil K., Arun Ross, and Salil Prabhakar. "An Introduction to Biometric Recognition IEEE Transactions on Circuits and Systems for Video Technology." *Special Issue on Image-and Video-Based Biometrics* 14.1 (2004)

[11] Allano, Lorène. *La Biométrie multimodale: stratégies de fusion de scores et mesures de dépendance appliquées aux bases de personnes virtuelles*. Diss. Evry, Institut national des télécommunications, 2009.

[12] Uhl, Andreas, et al., eds. *Handbook of vascular biometrics*. Springer Open, 2020.

[13] Nguyen, Dat Tien, et al. "Fake finger-vein image detection based on fourier and wavelet transforms." *Digital Signal Processing* 23.5 (2013): 1401-1413.

[14] Yang, Xiaofei, Chunhua Yang, and Zhijun Yao. "The finger vein recognition based on shearlet." *2016 8th International Conference on Wireless Communications & Signal Processing (WCSP)*. IEEE, 2016.

[15] Miura, Naoto, Akio Nagasaka, and Takafumi Miyatake. "Feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification." *Machine vision and applications* 15.4 (2004): 194-203.

[16] Kulkarni, Sujata, R. D. Raut, and P. K. Dakhole. "A novel authentication system based on biometric trait." *Procedia Computer Science*. Vol. 85. Elsevier, 2016.

[17] Dai, Yanggang, et al. "A method for capturing the finger-vein image using nonuniform intensity infrared light." *2008 Congress on Image and Signal Processing*. Vol. 4. IEEE, 2008.

[18] Lu, Yu, et al. "An available database for the research of finger vein recognition." *2013 6th International congress on image and signal processing (CISP)*. Vol. 1. IEEE, 2013

[19] Yang, Lu, et al. "A survey of finger vein recognition." *Chinese Conference on Biometric Recognition*. Springer, Cham, 2014.

[20] Kumar, Ajay, and Yingbo Zhou. "Human identification using finger images." *IEEE Transactions on image processing* 21.4 (2011): 2228-2244

[21] Asaari, Mohd Shahrimie Mohd, Shahrel A. Suandi, and Bakhtiar Affendi Rosdi. "Fusion of band limited phase only correlation and width centroid contour distance for finger based biometrics." *Expert Systems with Applications* 41.7 (2014): 3367-3382.

[22] Lee, Eui Chul, Hyunwoo Jung, and Daeyeoul Kim. "New finger biometric method using near infrared imaging." *Sensors* 11.3 (2011): 2319-2333.

[23] Yang, Jinfeng, Yihua Shi, and Renbiao Wu. "Finger-vein recognition based on gabor features." *Biometric Systems, Design and Applications* (2011): 17-32.

[24] Mahri, Nurhafizah, Shahrel Azmin Sundi Suandi, and Bakhtiar Affendi Rosdi. "Finger vein recognition algorithm using phase only correlation." *2010 International Workshop on Emerging Techniques and Challenges for Hand-Based Biometrics*. IEEE, 2010.

[25] Gupta, Puneet, and Phalguni Gupta. "An accurate finger vein based verification system." *Digital Signal Processing* 38 (2015): 43-52.

[26] Gupta, Puneet, and Phalguni Gupta. "Multi-modal fusion of palm-dorsa vein pattern for accurate personal authentication." *Knowledge-Based Systems* 81 (2015): 117-130.

[27]Miura, Naoto, Akio Nagasaka, and Takafumi Miyatake. "Feature extraction of finger vein patterns based on iterative line tracking and its application to personal identification." *Systems and Computers in Japan* 35.7 (2004): 61-71.

[28] Zuiderveld, Karel. "Contrast limited adaptive histogram equalization." Graphics gems (1994): 474-485.

[29] Yang, Jinfeng, Jinli Yang, and Yihua Shi. "Finger-vein segmentation based on multi-channel even-symmetric Gabor filters." *2009 IEEE International Conference on Intelligent Computing and Intelligent Systems*. Vol. 4. IEEE, 2009.

[30] Liu, Laura, David Zhang, and Jane You. "Detecting wide lines using isotropic nonlinear filtering." *IEEE Transactions on image processing* 16.6 (2007): 1584-1595.

[31] Miura, Naoto, Akio Nagasaka, and Takafumi Miyatake. "Extraction of finger-vein patterns using maximum curvature points in image profiles." *IEICE TRANSACTIONS on Information and Systems* 90.8 (2007): 1185-1194.

[32]Caixia, Liu. "The Research on Finger Vein Image Preprocessing Based on Mathematical Morphology." *Information Engineering and Applications*. Springer, London, 2012. 465-471.

[33] Rosdi, Bakhtiar Affendi, Chai Wuh Shing, and Shahrel Azmin Suandi. "Finger vein recognition using local line binary pattern." *Sensors* 11.12 (2011): 11357-11371.

[34] Ito, Koichi, et al. "A fingerprint matching algorithm using phase-only correlation." *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* 87.3 (2004): 682-691.

[35] Ismail, Boucherit, and Mohamed Ould Zmirli. "Enhancement of Finger Vein Patterns Extracted by Maximum Curvature Method." *2018 International Conference on Applied Smart Systems (ICASS)*. IEEE, 2018.

[36] Yin, Yilong, Lili Liu, and Xiwei Sun. "SDUMLA-HMT: a multimodal biometric database." *Chinese Conference on Biometric Recognition*. Springer, Berlin, Heidelberg, 2011.

[37] Lu, Yu, et al. "Efficient descriptor of histogram of salient edge orientation map for finger vein recognition." *Applied optics* 53.20 (2014): 4585-4593.

[38] Petrovic, Igor. *Biométrie basée sur la reconnaissance des veines de la main*. Diss. Haute Ecole d'Ingénierie, 2013.

[39]Hashimoto, Junichi. "Finger vein authentication technology and its future." *2006 Symposium on VLSI Circuits, 2006. Digest of Technical Papers.*. IEEE, 2006.

[40] Yu, Cheng-Bo, et al. "Finger-vein image enhancement based on muti-threshold fuzzy algorithm." *2009 2nd International Congress on Image and Signal Processing*. IEEE, 2009.

[41]Mei, Cong-li, et al. "Feature extraction of finger-vein image based on morphologic algorithm." *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery*. Vol. 3. IEEE, 2009.

[42]Zhao, Jianjun, et al. "A new approach to hand vein image enhancement." *2009 Second International Conference on Intelligent Computation Technology and Automation*. Vol. 1. IEEE, 2009.

[43]Zhang, Jing, and Jinfeng Yang. "Finger-vein image enhancement based on combination of gray-level grouping and circular gabor filter." *2009 International Conference on Information Engineering and Computer Science*. IEEE, 2009.

[44]Zhou, Jianping, Arthur L. Cunha, and Minh N. Do. "Nonsubsampled contourlet transform: construction and application in enhancement." *IEEE International Conference on Image Processing 2005*. Vol. 1. IEEE, 2005.

[45] Truc, Phan TH, et al. "Vessel enhancement filter using directional filter bank." *Computer Vision and Image Understanding* 113.1 (2009): 101-112.

[46]Yang, Jinfeng, and Minfu Yan. "An improved method for finger-vein image enhancement." *IEEE 10th International Conference On Signal Processing Proceedings*. IEEE, 2010.

[47] Park, Young Ho, and Kang Ryoung Park. "Image quality enhancement using the direction and thickness of vein lines for finger-vein recognition." *International Journal of Advanced Robotic Systems* 9.4 (2012): 154.

[48] Nguyen, Dat Tien, et al. "New Finger-vein Recognition Method Based on Image Quality Assessment." *Ksii Transactions on Internet & Information Systems* 7.2 (2013).

[49]Chihaoui, Takwa. *Système d'identification de personnes basé sur la rétine*. Diss. 2018.

[50] Zuiderveld, Karel. "Contrast limited adaptive histogram equalization." *Graphics gems IV*. Academic Press Professional, Inc., 1994.

[51] Khellat-kihel, Souad, et al. "Finger vein recognition using Gabor filter and Support Vector Machine." *International image processing, applications and systems conference*. IEEE, 2014.

[52]Krishan, Abhay, and M. D. Singh. *Evaluation of Gabor filter parameters for image enhancement and segmentation*. Diss. 2009.

[53]Haghighat, Mohammad Bagher Akbari, Ali Aghagolzadeh, and Hadi Seyedarabi. "Multi-focus image fusion for visual sensor networks in DCT domain." *Computers & Electrical Engineering* 37.5 (2011): 789-797.

[54] Abbood, Alaa Ahmed, et al. "New Combined Technique for Fingerprint Image Enhancement." *Modern Applied Science* 11.1 (2016): 222-234.

[55] Yang, Lu, et al. "Finger vein image quality evaluation using support vector machines." *Optical engineering* 52.2 (2013): 027003.

[56] Wang, K. J., et al. "A finger vein image quality assessment method." *CAAI Trans. Intell. Syst* 6.4 (2011): 324-327.

[57]Xie, Shan Juan, et al. "Novel hierarchical structure based finger vein image quality assessment." *Chinese Conference on Biometric Recognition*. Springer, Cham, 2013.

[58]Ma, Hui, et al. "A finger vein image quality assessment method using object and human visual system index." *International Conference on Intelligent Science and Intelligent Data Engineering*. Springer, Berlin, Heidelberg, 2012.

[59]Qin, Huafeng, et al. "Quality assessment of finger-vein image." *Proceedings of the 2012 Asia Pacific signal and information processing association annual summit and conference*. IEEE, 2012.

[60] Zhou, Lizhen, et al. "Finger vein image quality evaluation based on support vector regression." *International Journal of Signal Processing, Image Processing and Pattern Recognition* 8.8 (2015): 211-222.

[61] Chollet, Francois. *Deep Learning mit Python und Keras: Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek*. MITP-Verlags GmbH & Co. KG, 2018.

[62]Nielsen, Michael A. *Neural networks and deep learning*. Vol. 2018. San Francisco, CA, USA:: Determination press, 2015.

[63] Pascal Lamblin, Entrainement de reseaux de neurones profonds,19 avril 2007.

[64] Hugo Larochelle,audela de la machine de boltzmann restreinte,University of toronto, 2009.

[65]Chollet, Francois. "Building auto encoders in keras." *The Keras Blog* (2016).

[66] Hochreiter, S., &Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780

[67] Sun, Rémy. "Apprentissage profond et acquisition de représentations latentes de séquences peptidiques." (2017).

[68] Boukerroucha Ouafa et Harbi abdenacer : " Séléction et extraction des attributs à l'aide de l'apprentissage profond : application sur des données médicales réelles ", 2018.Master en Informatique.

[69]Radzi, Syafeeza Ahmad, MOHAMED KHALIL HANI, and Rabia Bakhteri. "Finger-vein biometric identification using convolutional neural network." *Turkish Journal of Electrical Engineering & Computer Sciences* 24.3 (2016): 1863-1878.

[70]Das, Rig, et al. "Convolutional neural network for finger-vein-based biometric identification." *IEEE Transactions on Information Forensics and Security* 14.2 (2018): 360-373.

[71] Fang, Yuxun, Qiuxia Wu, and Wenxiong Kang. "A novel finger vein verification system based on two-stream convolutional network learning." *Neurocomputing* 290 (2018): 100-107.

[72]Wu, Jianxin. "Introduction to convolutional neural networks." *National Key Lab for Novel Software Technology. Nanjing University. China* 5 (2017): 23.

[73] deep learning LISA lab, University of Montreal ; September01,2015

[74]Agarap, Abien Fred. "Deep learning using rectified linear units (relu)." *arXiv preprint arXiv:1803.08375* (2018).

[75] Mokri, Mohammed Zakaria. *Classification des images avec les réseaux de neurones convolutionnels*. Diss. 09-01-2018, 2017.

[76]Kim, Wan, Jong Min Song, and Kang Ryoung Park. "Multimodal biometric recognition based on convolutional neural network by the fusion of finger-vein and finger shape using near-infrared (NIR) camera sensor." *Sensors* 18.7 (2018): 2296.

[77]Banerjee, Anupam, et al. "ARTeM: a new system for human authentication using finger vein images." *Multimedia Tools and Applications* 77.5 (2018): 5857-5884.

[78] James, Gareth, et al. *An introduction to statistical learning*. Vol. 112. New York: springer, 2013.

[79] Syazana-Itqan, K., et al. "A review of finger-vein biometrics identification approaches." *Indian J. Sci. Technol* 9.32 (2016).

[80]Yang, Wenming, et al. "Comparative competitive coding for personal identification by using finger vein and finger dorsal texture fusion." *Information sciences* 268 (2014): 20-32.

[81]Williams, Jennifer, et al. "Dnn multimodal fusion techniques for predicting video sentiment." *Proceedings of Grand Challenge and Workshop on Human Multimodal Language (Challenge-HML)*. 2018.

[82]Yang, Jinfeng, and Yihua Shi. "Finger–vein ROI localization and vein ridge enhancement." *Pattern Recognition Letters* 33.12 (2012): 1569-1579.

[83] A. Karpathy and T.Leung, "Karpathy_Large-scale Video Classification 2014 CVPR paper," Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 10–20, 2018, doi: 10.1109/CVPR.2014.223.

[84] Ismail, Boucherit, and Ould Zmirli Mohamed. "Finger Vein Image Enhancement Technique based on Gabor filter and Discrete Cosine Transform." *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)* 11.2 (2019): 43-48.

[85]Gupta, Puneet, Saurabh Srivastava, and Phalguni Gupta. "An accurate infrared hand geometry and vein pattern based authentication system." *Knowledge-Based Systems* 103 (2016): 143-155.

[86] Kono, Miyuki, Hironori Ueki, and Shin-ichiro Umemura. "Near-infrared finger vein patterns for personal identification." *Applied Optics* 41.35 (2002): 7429-7436.

[87] Hong, Hyung Gil, Min Beom Lee, and Kang Ryoung Park. "Convolutional neural network-based finger-vein recognition using NIR image sensors." *Sensors* 17.6 (2017): 1297.

[88] Xie, Shan Juan, et al. "Feature component-based extreme learning machines for finger vein recognition." *Cognitive Computation* 6.3 (2014): 446-461.

[89]Shazeeda, Shazeeda, and Bakhtiar Affendi Rosdi. "Finger vein recognition using mutual sparse representation classification." *IET Biometrics* 8.1 (2018): 49-58.

[90]Qiu, Shirong, et al. "Finger-vein recognition based on dual-sliding window localization and pseudo-elliptical transformer." *Expert Systems with Applications* 64 (2016): 618-632.

[91] Van, Hoang Thien, Thanh Tuan Thai, and Thai Hoang Le. "Robust finger vein identification base on discriminant orientation feature." *2015 Seventh International Conference on Knowledge and Systems Engineering (KSE)*. IEEE, 2015.

Production Scientifique

- ISMAIL, Boucherit; MOHAMED, Ould Zmirli. Enhancement of Finger Vein Patterns Extracted by Maximum Curvature Method. In: *2018 International Conference on Applied Smart Systems (ICASS)*. IEEE, 2018. p. 1-4.

- ISMAIL, Boucherit; MOHAMED, Ould Zmirli. Finger Vein Image Enhancement Technique based on Gabor filter and Discrete Cosine Transform. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 2019, 11.2: 43-48.

- ISMAIL, Boucherit; et al. Finger Vein Identification Using Deeply-Fused Convolutional Neural Network. *Journal of King Saud University-Computer and Information Sciences*, 2020.