



HAL
open science

Amélioration des systèmes de reconnaissance faciale par l'utilisation de caméra MSFA et des techniques d'intelligence artificielle

Médégnonmi Houssou

► **To cite this version:**

Médégnonmi Houssou. Amélioration des systèmes de reconnaissance faciale par l'utilisation de caméra MSFA et des techniques d'intelligence artificielle. Traitement du signal et de l'image [eess.SP]. Université Bourgogne Franche-Comté; Université d'Abomey-Calavi (Bénin), 2023. Français. NNT : 2023UBFCK061 . tel-04607861

HAL Id: tel-04607861

<https://theses.hal.science/tel-04607861v1>

Submitted on 11 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Université d'Abomey-Calavi
Ecole Doctorale de Sciences Exactes et Appliquées (ED-SEA)

Université de Bourgogne Franche Comté
Ecole Doctorale Science Physique pour
l'Ingénieur et Microtechniques (ED-SPIM)

N° d'ordre :

Thèse de Doctorat
Présentée pour l'obtention du
**Doctorat délivré conjointement par l'Université
d'Abomey-Calavi et l'Université de Bourgogne
Franche Comté**

Spécialité : Informatique

**Amélioration des systèmes de reconnaissance faciale par
l'utilisation de caméra MSFA et des techniques d'intelligence
artificielle**

Par

Médégnonmi Eléonore Elvire HOUSSOU

Soutenue publiquement le 18 /10/2023

Devant un jury composé de :

| | | | |
|-----------------------|-----------------------------|------------|---|
| Président : | Carlos OGOUYANDJOU | Professeur | Université d'Abomey-Calavi, Bénin |
| Rapporteurs : | Keivan ANSARI | Docteur | Institute of Color Science and Technology, Iran |
| | Mohamed-Chaker LARABI | Professeur | Université de Poitiers, France |
| Examineurs : | Sabir JACQUIR | Professeur | Université Paris Saclay, France |
| | Richard CHBEIR | Professeur | Université de Pau, France |
| Directeurs de thèse : | Pierre GOUTON | Professeur | Université de Bourgogne, France |
| | Guy DEGLA | Professeur | Université d'Abomey-Calavi, Bénin |
| Encadrant : | Amadou Tidjani SANDA MAHAMA | Docteur | Université d'Abomey-Calavi, Bénin |

Dédicaces

A mon père et à mon repère partis très top.

*A ma mère, mon guide, qui de façon extraordinaire a couvert mes arrières au cours de cette
thèse.*

A mes enfants et à mon compagnon, qui m'ont soutenu et encouragé durant ces années.

A mes frères et à tous mes amis.

Résumé

Ces dernières années, nous assistons à une émergence des systèmes biométriques utilisés dans le contrôle d'accès et la sécurité. La plus connue de ces applications biométriques reste la reconnaissance faciale qui ne cesse de connaître des progrès. Plusieurs approches existantes de la reconnaissance faciale utilisent des images acquises dans le spectre visible. Ces systèmes souffrent surtout de problèmes de variation de la lumière. Ces limitations sont améliorées avec les systèmes de reconnaissance dans l'infrarouge. D'autres approches de systèmes de reconnaissance faciale combinent le visible et l'infrarouge toujours dans le but d'améliorer leur performance de reconnaissance, elles sont dites hétérogènes. Les caméras multispectrales sont alors utilisées dans les systèmes de reconnaissance faciale hétérogènes. Aussi les méthodes d'intelligence artificielle plus précisément les méthodes d'apprentissage profond appelées Deep Learning (DL) en anglais ont favorisé d'énormes progrès en reconnaissance faciale. Dans le cadre de cette thèse, il est question d'exploiter les potentialités des algorithmes d'apprentissage profond et le réseau de filtre multispectraux (MSFA) pour développer et améliorer les systèmes de reconnaissance faciale. Quatre contributions sont apportées pour ces travaux de thèse. Premièrement, nous avons construit une base d'images multispectrales de visage pour la reconnaissance faciale avec les caméras MSFA one-shot qui opère de 650 à 950 nm. Deuxièmement nous avons développé un système robuste de reconnaissance faciale avec les caméras MSFA one-shot, la transformée discrète de curvelet et les réseaux de neurones convolutifs (CNN). Les résultats expérimentaux indiquent une exactitude de 100% avec le CNN VGG19. Ensuite une optimisation du système est faite par la réduction de dimensionalité. Le nombre de bandes des images est réduite de huit à trois bandes avec la transformée de Karhuen Loève. Quatrièmement nous avons effectué une étude comparative pour évaluer l'impact du demosaicing sur la reconnaissance faciale. Les résultats des expériences démontrent l'efficacité des méthodes proposées pour construire un système de reconnaissance faciale plus performant.

Mots clés : reconnaissance faciale, intelligence artificielle, images multispectrales, visible, infrarouge, MSFA, CNN.

Abstract

Recent years have seen the emergence of biometric systems for access control and security. The best-known of these biometric applications is facial recognition, which continues to make progress. Several existing approaches to facial recognition use images acquired in the visible spectrum. These systems suffer mainly from problems of light variation. Infrared facial recognition systems overcome these limitations. Other approaches to facial recognition systems combine the visible and infrared spectrums, again to improve recognition performance. Multispectral cameras are then used in heterogeneous face recognition systems. Artificial intelligence (AI) methods, more specifically Deep Learning (DL), have made enormous progress in face recognition. The purpose of this thesis is to exploit the potential of deep learning algorithms, and the multispectral filter array (MSFA) to develop and improve face recognition systems. Four contributions are made to this thesis work. Firstly, we have built a multispectral face image base for facial recognition with MSFA one-shot cameras operating from 650 to 950 nm. Secondly, we developed a robust facial recognition system using MSFA one-shot cameras, the discrete curvelet transform, and convolutional neural networks (CNN). Experimental results indicate 100% accuracy with the VGG19 CNN. Next, we optimized the system using dimensionality reduction. The number of image bands was reduced from eight to three using the Karhuen Loève Transform. Fourthly, we conducted a comparative study to assess the impact of demosaicing on facial recognition.

The experimental results demonstrate the effectiveness of the proposed methods in building a more efficient facial recognition system.

Keywords : facial recognition, artificial intelligence, multispectral images, visible, infrared, MSFA, CNN.

Remerciements

L'achèvement de ce travail mené sur plusieurs années procure une grande satisfaction. Il est l'occasion de se remémorer les différentes embûches qu'il a fallu surmonter mais surtout les personnes qui m'ont permis d'en arriver là.

Je voudrais tout d'abord remercier grandement mes directeurs de thèse, les Professeurs Pierre Gouton et Guy Degla pour m'avoir m'encadrer au cours de ces années. Je ne saurais assez les remercier pour leur disponibilité, leur soutien indéfectible et leur enthousiasme scientifique qui m'ont permis de mener à bien mes travaux.

J'exprime ma gratitude à mon encadrant le Dr Sanda Mahama Amadou Tidjani pour sa participation scientifique, sa disponibilité ainsi que ses encouragements, surtout dans les moments les plus difficiles.

J'adresse mes remerciements aux Professeurs Mohamed-Cheker Larabi, Sabir Jacquir, Chbeire Richard et aux Dr Keivan Ansari, Djara Tahirou, Pélagie HOUNGUE pour avoir accepté d'être rapporteurs et examinateur de mes travaux.

Je remercie également le Professeur Carlos Ogouyandjou qui a accepté d'évaluer ce travail en tant que Président du jury.

Je tiens à remercier particulièrement le Professeur Joel Tossa, pour toutes les démarches administratives relatives à cette cotutelle.

J'adresse mes remerciements au CEA-SMIA qui a financé cette thèse.

Je souhaite également remercier les membres des laboratoires ImViA et LMFA, l'administration des écoles doctorales ED-SPIM et ED-SEA, et l'IMSP pour leur pleine disponibilité et leur accompagnement.

Enfin, mes remerciements vont à l'endroit de tous mes collègues des deux laboratoires et particulièrement à Frantz Tossa, Esther Chabi Adjobo, Rose Waytehad Moskola, Vahid Mohamadi, Fernand Djoumessi, Jean Kacoutchy Ayikpa, Norbert Hounsou, avec qui, j'ai partagé une excellente ambiance de vie, de recherche et d'enseignement. Je nous souhaite plein succès.

Table des matières

| | |
|--|-------------|
| Résumé | ii |
| Abstract | iii |
| Remerciements | iv |
| Table des matières | v |
| Liste des tableaux | viii |
| Listes des figures | ix |
| Acronymes | xi |
| Introduction générale | 1 |
| 1- Contexte et motivation | 1 |
| 2- Problématique | 2 |
| 3- Objectif..... | 3 |
| 4- Contributions | 3 |
| 5- Liste des publications..... | 4 |
| 6- Organisation de la thèse | 5 |
| Chapitre 1 : Reconnaissance faciale et imagerie multispectrale | 6 |
| Introduction | 7 |
| 1.1 Fondamentaux de la biométrie | 7 |
| 1.1.1 <i>Architecture d'un système biométrique</i> | 8 |
| 1.1.2 <i>Mode de fonctionnement d'un système biométrique</i> | 9 |
| 1.1.3 <i>Avantages des systèmes biométriques</i> | 10 |
| 1.1.4 <i>Modalités biométriques</i> | 10 |
| 1.2 Reconnaissance de visage | 13 |
| 1.2.1 <i>Motivation</i> | 13 |
| 1.2.2 <i>Méthodes de reconnaissance de visage</i> | 13 |
| 1.2.2.1 <i>Méthodes classiques</i> | 13 |
| 1.1.2.2 <i>Intelligence artificielle</i> | 14 |
| 1.2.2.2.1 <i>Machine learning</i> | 15 |
| 1.2.2.2.2 <i>Deep learning</i> | 17 |
| 1.3 Imagerie multispectrale | 24 |
| 1.3.1 <i>Différents types de systèmes d'acquisition en imagerie multispectrale</i> | 24 |
| 1.3.2 <i>Système d'acquisition MSFA one-shot</i> | 25 |
| 1.4 Reconnaissance faciale multispectrale | 28 |
| 1.4.1 <i>Base d'images multispectrales</i> | 28 |

| | | |
|--|--|-----------|
| 1.4.2 | <i>Système de reconnaissance faciale multispectrale</i> | 29 |
| 1.4.3 | <i>Evaluation de la performance des systèmes de reconnaissance</i> | 29 |
| | Conclusion | 30 |
| Chapitre 2 Construction d'une base de données multispectrale avec une caméra MSFA one-shot | | 31 |
| | | |
| | Introduction | 32 |
| 2.1 | Imagerie multispectrale pour la reconnaissance faciale | 32 |
| 2.1.1 | <i>Motivation</i> | 32 |
| 2.1.2 | <i>Bases d'images utilisées en reconnaissance faciale multispectrale</i> | 33 |
| 2.2 | Système d'acquisition MSFA one shot | 37 |
| 2.2.1 | Imagerie multispectrale à filtre MSFA | 37 |
| 2.2.2 | <i>Description de la caméra EXIST</i> | 39 |
| 2.2.2.1 | <i>Capteurs CMOS</i> | 40 |
| 2.2.2.2 | <i>Conception et fabrication des filtres</i> | 41 |
| 2.3 | Protocole de collecte et description de la base d'images EXIST | 45 |
| | Conclusion | 47 |
| Chapitre 3 : Développement d'un système robuste de reconnaissance de faciale via les caméras one-shot | | 49 |
| | | |
| | Introduction | 50 |
| 3.1 | Systèmes de reconnaissance faciale multispectrale existants | 50 |
| 3.2 | Architecture d'un système de reconnaissance avec une caméra MSFA oneshot | 53 |
| 3.3 | Demosaicing des images de la base de données EXIST | 54 |
| 3.4 | Méthode proposée | 58 |
| 3.4.1 | <i>Fusion des images</i> | 58 |
| 3.4.1.1 | <i>Transformé discrète rapide de curvelets</i> | 58 |
| 3.4.1.2 | <i>Fusion des images avec la FDCT</i> | 60 |
| 3.4.2 | <i>Réseaux de neurones convolutifs</i> | 61 |
| 3.4.2.1 | <i>VGG19</i> | 61 |
| 3.4.2.2 | <i>ResNet 101</i> | 62 |
| 3.5 | Résultats et discussion | 64 |
| 3.5.1 | <i>Résultats expérimentaux</i> | 64 |
| 3.5.2 | <i>Discussion</i> | 68 |
| | Conclusion | 68 |
| Chapitre 4 : Amélioration des systèmes de reconnaissance faciale avec les caméras one shot par la réduction de dimensionalité | | 70 |
| | | |
| | Introduction | 71 |

| | |
|--|------------|
| 4.1 Réduction de dimensionalité pour la reconnaissance faciale | 71 |
| 4.1.1 Méthodes de projection linéaire | 72 |
| 4.1.2 Méthodes de projection non linéaire | 73 |
| 4.1.3 Réduction de dimensionnalité pour les systèmes de reconnaissance faciale | 75 |
| 4.2 Etude comparative des méthodes de réduction de dimensionalités | 76 |
| 4.2.1 Motivation | 76 |
| 4.2.2 Transformé de Karhuen-Loève | 76 |
| 4.2.3 Analyse en Composante Curviligne | 77 |
| 4.2.4 Analyse Factorielle | 78 |
| 4.3 Evaluation de la réduction de dimensionalité sur la reconnaissance faciale via les caméras MSFA one-nshot | 78 |
| 4.4 Résultats et discussion | 80 |
| Conclusion | 85 |
| Chapitre 5 : Optimisation de la qualité de l'image et du temps de calcul : analyse approfondie de quatre méthodes de demosaicing pour les algorithmes de reconnaissance faciale | 86 |
| Introduction | 87 |
| 5.1 Motivation | 88 |
| 5.2 Description des méthodes sélectionnées | 89 |
| 5.2.1 Interpolation Bilinéaire | 89 |
| 5.2.2 Transformée discrète en ondelette | 89 |
| 5.2.3 Arbre binaire | 91 |
| 5.2.4 Vecteur médian | 92 |
| 5.3 Métriques d'évaluation | 93 |
| 5.4 Résultats expérimentaux et discussion | 95 |
| 5.4.1 Base de données et configuration expérimentales | 95 |
| 5.4.2 Résultats et discussion | 96 |
| Conclusion | 102 |
| Conclusion et perspectives | 103 |
| Bibliographie | 105 |

Liste des tableaux

| | |
|---|-----|
| Tableau 1.1- Avantages et inconvénients des systèmes de reconnaissance faciale | 13 |
| Tableau 1.2- Quelques CNN les plus utilisés en reconnaissance faciale..... | 19 |
| Tableau 1.3- Quelques algorithmes d'AE pour la RF | 22 |
| Tableau 1.4- Quelques bases publiques d'images multispectrales..... | 28 |
| Tableau 2.1- Quelques bases d'images multispectrales..... | 36 |
| Tableau 3.1- Paramètres de VGG19 et ResNet101 au cours de la phase d'entraînement | 64 |
| Tableau 3.2- Résultats avec VGG19 | 66 |
| Tableau 3.3 – Résultats avec ResNet101 | 66 |
| Tableau 3.4- Taux de reconnaissance | 67 |
| Tableau 4.1- La moyenne des coefficients de corrélation de 80 images de la base de données EXIST. | 80 |
| Tableau 4.2- Cummul (%) des valeurs propres | 81 |
| Tableau 4.3- Moyenne des coefficients de corrélation avec TKL..... | 82 |
| Tableau 4.4- Moyenne des coefficients de corrélation avec AF | 82 |
| Tableau 4.5- Moyenne des coefficients de corrélation avec ACC | 83 |
| Tableau 4.6- Paramètres utilisés dans la procédure d'entraînement..... | 83 |
| Tableau 4.7 Résultats | 83 |
| Tableau 4.8- Résultats..... | 84 |
| Tableau 4.9- Comparaison avec autres méthodes | 84 |
| Tableau 5.1- Moyenne des coefficients de corrélation avec DWT | 98 |
| Tableau 5.2- Moyenne des coefficients de corrélation avec BI | 98 |
| Tableau 5.3- Moyenne des coefficients de corrélation avec BT | 99 |
| Tableau 5.4 Moyenne des coefficients de corrélation avec VM | 99 |
| Tableau 5.5- Paramètres utilisés dans la procédure de d'entraînement | 101 |
| Tableau 5.6- Accuracy | 101 |
| Tableau 5.7- Temps d'exécution | 101 |

Listes des figures

| | |
|---|----|
| Figure 1.1- Architecture d'un système biométrique | 9 |
| Figure 1.2- Mode de fonctionnement d'un système biométrique | 10 |
| Figure 1.3 - Modalités biométriques | 11 |
| Figure 1.4 - Modalités biométriques par type | 12 |
| Figure 1.5 - Architecture de base d'un CNN | 19 |
| Figure 1.6 -Tessellation hexagonale et rectangulaire | 26 |
| Figure 2.1-Spectre électromagnétique | 33 |
| Figure 2.2- Filtre de Bayer (CFA) | 37 |
| Figure 2.3- Schéma global du système d'imagerie MS avec MSFA. | 39 |
| Figure 2.4 Bandes d'images spectrales..... | 39 |
| Figure 2.5- a) capteur CMOS Viimagic 9220H, b) carte de pilotage face A, c) carte de pilotage face B..... | 41 |
| Figure 2.6- Réponse spectrale des huit filtres avant hybridation sur le capteur | 43 |
| Figure 2.7- Réponse spectrale du capteur hybride | 43 |
| Figure 2.8- Moxel du MSFA..... | 44 |
| Figure 2.9- Caméra MS one shot | 45 |
| Figure 2.10- Acquisition des images dans la salle noire..... | 46 |
| Figure 2.11- Quelques échantillons d'images de la base EXIST | 47 |
| Figure 3.1- Architecture d'un système de reconnaissance avec une caméra MSFA one-shot .. | 53 |
| Figure 3.2- Carte d'images obtenue après l'application de différents masques | 55 |
| Figure 3.3- Processus d'estimation de la valeur d'un pixel manquant dans une bande donnée. | 57 |
| Figure 3.4-Demosaicing des images brutes et stockage dans la base de données. | 57 |
| Figure 3.5- Méthode proposée pour la reconnaissance faciale avec une caméra one-shot | 58 |
| Figure 3.6- Architecture de VGG19 | 62 |
| Figure 3.7- Apprentissage résiduel : un élément constitutif | 62 |
| Figure 3.8- Architecture de ResNet101 | 63 |
| Figure 4.1- Quelques méthodes de réduction de dimensionalité | 75 |
| Figure 4.2- Système de reconnaissance faciale modifié | 79 |
| Figure 4.3-Echantillon d'images initiales | 80 |
| Figure 4.4- a) Images réduites avec le KLT, b) Images réduites avec le FA, c) Images réduites avec le CCA. | 82 |

| | |
|--|-----|
| Figure 5.1- Exemples d'images de la base de données EXIST | 95 |
| Figure 5.2- Images demosaiquées par méthode | 96 |
| Figure 5.3- Moyenne des NIQE par image et par méthode | 97 |
| Figure 5.4- Temps d'exécution des différentes méthodes | 97 |
| Figure 5.5- Moyenne des coefficients de corrélation pour les méthodes de demosaicing | 100 |

Acronymes

ACP : Analyse en Composantes Principales

ADL : Analyse Discriminante Linéaire

ADN : Acide DésoxyriboNucléique

AF: Analyse Factorielle

AOTF: Acousto-Optical Tunable Filter

ASIFT: Affine- Scale-Invariant Feature Transform

BT: Binary Tree

BTES: Binary Tree Edge Sensing

BI: Bilinear Interpolation

CASIA: Chinese Academy of Sciences

CFA: Color Filter Array

CMOS: Complementary Metal-Oxide-Semiconductor

CNN: Convolutional Neural Network

CT: Curvelet Transform

DL: Deep learning

DCNN: Deep Convolutional Neural Network

DWT: Discrete Wavelet Transform

DS: Down sampling

EBGM: Elastic Bunch Graph Matching

ECG: ElectroCardioGramme

ELANet : Efficient Lightweight Attention Networks

EO: Eigen Object

FDCT: Fast Discret Curvelet Transform

FWHM: Full-Width Half-Maximums

GAN: Generative Adversarial Network

GTNN: Gated Two stream Neural Network

ISOMAP: Isometric Mapping

HMM: Hidden Markov Models

HOG: Histogram of Oriented Gradients

IA: Intelligence Artificielle

KNN: K-Nearest Neighbors

LBP: Local Binary Pattern

LCTF: Liquid Crystal Tunable Filter
MCC: Matthews Correlation Coefficient
ND-NIVL: Notre Dame Near Infrared Visible Light Database
MDS: MultiDimensional Scale
ML: Maching learning
MSE: Means Square Error
MSER: Maximally Stable Extremal Regions
MSFA MultiSpectral Filter Array:
NIQE: Natural Image Quality Evaluator
NIR: Near InfraRed
PNSR: Peak Signal to Noise Ratio
POA: Probability of Appearance
PPL : Projections Préservant la Localité
ReLU: Rectified Linear Unit
RF : Reconnaissance Faciale
RNA : Réseaux de Neurones Artificiels (RNA)
RVB : Rouge Vert Bleu
SFA: Spectral Filter Array
SIFT: Scale-Invariant Feature Transform
SS: Séparation des Sources
SSIM: Structural Similarity Index for Measuring
SURF: Speeded Up Robust Features
SVM: Support Vector Machine
TKL: Transformé de Karhuen Loeve
USFFT: Unequally-Spaced fast Fourier transforms
VIS : Visible

Introduction générale

1- Contexte et motivation

La sécurité constitue un défi majeur pour toute nation. Les accès aux domaines, bâtiments et objets sont contrôlés par les systèmes d'identification et d'authentification. Les moyens classiques de sécurisation tels que le mot de passe, le badge, ou encore la carte à puce sont utilisés pour contrôler l'accès aux bâtiments, aux objets, aux zones et informations privées ou sensibles. Ces moyens, bien qu'efficaces présentent des limites. En effet, un mot de passe peut être oublié ou volé, une carte à puces piratée, un badge falsifié, volé ou utilisé par une autre personne. Les systèmes biométriques viennent alors améliorer précisément les limites de ces méthodes classiques. En effet, un système biométrique est un système automatique de mesure basé sur la reconnaissance de caractéristiques physiologiques et/ou comportementales propres à un individu. Il est caractérisé par l'unicité, son caractère public et la mesure de sa performance. Il existe plusieurs modalités biométriques que sont l'iris, le visage, l'empreinte digitale, l'ADN, l'électrocardiogramme (ECG)... La modalité du visage est beaucoup utilisée dans la biométrie car elle est universelle, non intrusive et facile à capturer dans les environnements contrôlés ou non. Le visage diffère des autres modalités par le fait qu'il constitue la modalité utilisée par les humains pour se reconnaître physiquement. La reconnaissance faciale (RF) a été l'objet de recherche intensive depuis environ une soixantaine d'année. Elle est utilisée dans plusieurs domaines dont le contrôle d'accès, les enquêtes criminelles, le diagnostic des maladies [1] . Les applications potentielles de la reconnaissance faciale sont nombreuses et diversifiées. Le premier module d'un système de reconnaissance est le module d'acquisition. Il est composé d'une ou de plusieurs caméras qui permettent de capturer l'image du visage à reconnaître.

Dans le même temps l'imagerie couleur devient de plus en plus populaire avec le développement de caméras présentant de bonne performance avec un coût de plus en plus faible. Les caméras couleurs sont utilisées dans plusieurs domaines tels que l'agriculture, la médecine, le contrôle d'accès aux bâtiments et aux objets tels que les aéroports et les smartphones. Malgré l'existence des systèmes biométriques et du coût moins élevé des caméras couleurs (RVB), les failles sont constatées au niveau de ces systèmes notamment la reconnaissance de vrais jumeaux, l'usurpation d'identité. En effet les visages présentent des caractéristiques similaires et sont astreints au vieillissement, ils varient dans le temps. L'imagerie multispectrale permet d'obtenir les caractéristiques distinctes du visage. Aussi les

outils traditionnels de reconnaissance utilisés dans le spectre du visible ont montré leurs limites [2], [3]. C'est dans ce contexte que nous avons réalisé notre thèse qui porte sur *l'Amélioration des systèmes de reconnaissance faciale par l'utilisation de caméra MSFA et des techniques d'intelligence artificielle.*

2- Problématique

La RF dans le spectre du visible a fait l'objet de nombreuses recherches, mais ses performances restent limitées surtout par la variation des conditions d'éclairage [4]–[6]. Bien que les images couleurs acquises pour la RF présentent des caractéristiques de texture, elles ne sont pas invariantes dans le temps. En effet, l'acquisition des images couleurs de visage est sensible à des paramètres tels que le changement d'éclairage, la variation de poses et d'expressions faciales, les occlusions, la fumée, le brouillard, le maquillage, l'aspect de vieillissement et le déguisement. Malheureusement, ces différents paramètres détériorent fortement la performance et la robustesse des systèmes de reconnaissance faciale dans le visible. L'imagerie infrarouge constitue une alternative à l'imagerie traditionnelle ou couleur pour surmonter ces limitations. En effet, les images acquises dans l'infrarouge sont peu sensibles à la variation de la lumière, de la couleur de la peau, le bronzage et les lentilles. Aussi, la performance des systèmes biométriques peut être améliorée en combinant plusieurs modalités biométriques ou en utilisant les images de bonne qualité acquises sur plusieurs spectres grâce à l'imagerie multispectrale.

L'utilisation de caméras multispectrales pour la RF est un défi majeur en termes de performance, de temps et de coût. Les systèmes d'acquisition multispectrale peuvent être regroupés en deux catégories : les systèmes d'acquisition composés de plusieurs caméras one-shot ou d'une caméra unique à plusieurs capteurs et celle à un seul coup avec le réseau de filtres spectraux ou MultiSpectral Filters Array (MSFA) en anglais. La première catégorie est un système complexe car il faut un alignement parfait des différentes caméras pour plusieurs prises de vue. Les systèmes d'acquisition multi cameras ou multi capteurs sont très lourds et gourmands en énergie. La deuxième catégorie, celle des systèmes d'acquisition MSFA one shot est composée d'une seule caméra à prise de vue unique. Elle est constituée fondamentalement d'un seul capteur avec un réseau de filtres spectraux (SFA) ou un réseau de filtres multispectraux (MSFA). Elle permet d'acquérir une seule image sur plusieurs bandes spectrales simultanément. Ce sont des systèmes d'acquisition très rapide qui opèrent en temps réel, et qui capturent même les scènes en mouvement. La plupart des systèmes de RF

multispectrale utilise le premier type de caméras. Ces systèmes de reconnaissance sont confrontés aux problèmes d'alignement des caractéristiques extraites et d'appariement étant donné que les images sont prises différemment. En outre, les systèmes de RF basés sur les méthodes traditionnelles d'extraction de caractéristiques et de correspondance présentent des performances de plus en plus faibles. Les méthodes d'Intelligence Artificielle pourraient améliorer la performance et la précision des systèmes biométriques.

Dans le but de résoudre les différents problèmes énumérés et d'exploiter les avantages des caméras MSFA one-shot pour la reconnaissance notre sujet de thèse porte sur l'***Amélioration des systèmes de reconnaissance faciale par l'utilisation de caméra MSFA et des techniques d'intelligence artificielle.*** La solution mise en place dans le cadre de notre thèse est un système de reconnaissance faciale composé d'une caméra MSFA one-shot et des réseaux de neurones.

3- Objectif

L'objectif général de cette thèse consiste à développer les outils robustes en vue d'améliorer les performances des systèmes de reconnaissance faciale avec les caméras multispectrales MSFA one-shot.

Les objectifs spécifiques qui se dégagent de cette problématique sont les suivants :

- Construction d'une base d'images de faces avec les caméras MSFA one-shot ;
- Identification de méthodes de demosaicing adaptées à la base ;
- Utilisation de l'imagerie multispectrale dans la reconnaissance de visage par fusion d'images visibles et du proche infrarouge ;
- Etude approfondie des informations contenues dans ces images multispectrales ;
- Exploitation de l'apprentissage profond pour le développement de systèmes robustes de reconnaissance faciale adaptés à ces types de caméras ;
- Optimisation du système proposé ;
- Evaluation de l'impact de l'utilisation des caméras MSFA one-shot sur la reconnaissance faciale.

4- Contributions

L'imagerie multispectrale pourrait améliorer les performances des systèmes de reconnaissance faciale. Aussi la plupart des systèmes de reconnaissance faciale multispectrale utilise les systèmes d'acquisition multi-capteurs ou multi-caméras. Le MSFA est une

extension du filtre coloré, en anglais Color Filters Array (CFA) de Bayer, utilisé dans les caméras numériques avec succès. Nous avons d'abord étudié et comparé quelques systèmes de RF qui existaient.

La première contribution de cette thèse est la mise en place d'une base d'images de faces multispectrales avec une caméra MSFA one-shot. Les images de cette base de données ont été collectées avec une caméra qui intègre un MSFA de huit filtres et qui opère dans les bandes spectrales de 650 à 950nm. Cette base d'images est constituée d'images de faces de 110 personnes.

Ensuite nous avons développé un système robuste de reconnaissance faciale qui constitue la deuxième contribution de cette thèse. Pour implémenter ce système, nous avons utilisé l'interpolation bilinéaire pour estimer les pixels manquants dans chaque bande. La nouvelle approche proposée est alors basée sur la fusion des images avec le FDCT et les réseaux de neurones VGG19 et ResNet 101.

La troisième contribution consiste à utiliser la réduction de dimensionalité pour améliorer le système de reconnaissance faciale proposé. Dans la méthodologie utilisée à cette fin, les méthodes de réduction de dimensionalité ont été introduites dans le système proposé avant la phase d'extraction de caractéristiques.

Enfin la quatrième contribution de cette thèse porte sur l'évaluation des méthodes de demosaicing sur la RF multispectrale avec les caméras MSFA one-shot. La méthodologie proposée utilise quatre algorithmes de demosaicing, la métrique d'évaluation de performance NIQE et les méthodes proposées dans les parties précédentes pour la reconnaissance faciale via les caméras MSFA.

5- Liste des publications

[1] M. E. E. Houssou, A. T. S. Mahama, P. Gouton, and G. Degla, "Robust Facial Recognition System using One Shot Multispectral Filter Array Acquisition System," *Int. J. Adv. Comput. Sci. Appl. IJACSA*, vol. 13, no. 1, Art. no. 1, 55/31 2022, doi: 10.14569/IJACSA.2022.0130104.

[2] M. E. E. Houssou, A. T. S. Mahama, P. Gouton, and G. Degla, "Improvement of a facial recognition system based on one shot camera," *Electronic imaging*, 2023. Doi: 10.2352/EI.2023.35.6.ISS-340

6- Organisation de la thèse

Cette thèse est organisée en cinq chapitres décrits comme suit :

Le chapitre 1 est consacré à la littérature générale sur la biométrie en particulier sur la reconnaissance faciale et l'imagerie multispectrale. Les concepts essentiels relatifs aux différents types de systèmes d'acquisition multispectrale et aux méthodes de reconnaissance faciale ont été présentés.

Le chapitre 2 présente une première base d'images de faces acquises avec une caméra MSFA one-shot. Nous avons décrit les différentes étapes de conception et de fabrication de notre caméra MSFA one shot qui intègre huit filtres spectraux. Les conditions d'acquisition et la procédure de mise en place de la base d'images ont été également présentées.

Le chapitre 3 fait l'objet d'une nouvelle approche de système de reconnaissance faciale proposée. Etant donné que les images acquises avec les caméras MSFA nécessitent le demosaicing avant leur utilisation, nous avons présenté une méthode de demosaicing basé sur l'interpolation bilinéaire. L'approche globale proposée combine la méthode de fusion FDCT des images multispectrales obtenues et les réseaux de neurones VGG19 et resNet101 pour la RF. Plusieurs métriques d'évaluation de performance telles que le Mean Squared Error (MSE), l'accuracy et le F-score ont été utilisées pour comparer cette approche à celles qui existaient.

Le chapitre 4 décrit une méthode d'optimisation du système de reconnaissance faciale proposé dans le chapitre 3. Cette optimisation du système est obtenue après une réduction de dimensionalité des images avant le module d'extraction de caractéristiques. Une comparaison de cette étude avec la première approche montre une diminution de temps de reconnaissance de 15% avec la Transformée de Karhuen-Loève.

Le chapitre 5 présente une évaluation des méthodes de demosaicing sur la RF avec les caméras one shot. Nous avons implémenté quatre méthodes de demosaicing à savoir l'arbre binaire, l'interpolation bilinéaire, la transformée en ondelettes discrètes et le vecteur médian pour reconstruire les images acquises avec notre caméra MSFA one shot. La métrique d'évaluation de la qualité des images sans référence NIQE a été utilisée pour comparer ces méthodes et le CNN VGG19 pour évaluer l'impact de ces méthodes sur la RF.

Chapitre 1

Reconnaissance faciale et imagerie multispectrale

Sommaire

| | |
|---|----|
| Introduction | 7 |
| 1.1 Fondamentaux de la biométrie | 7 |
| 1.1.1 Architecture d'un système biométrique | 8 |
| 1.1.2 Mode de fonctionnement d'un système biométrique | 9 |
| 1.1.3 Avantages des systèmes biométriques | 10 |
| 1.1.4 Modalités biométriques | 10 |
| 1.2 Reconnaissance de visage | 13 |
| 1.2.1 Motivation | 13 |
| 1.2.2 Méthodes de reconnaissance de visage | 13 |
| 1.2.2.1 Méthodes classiques | 13 |
| 1.2.2.2 Intelligence artificielle | 14 |
| 1.2.2.2.1 Machine learning | 15 |
| 1.2.2.2.2 Deep learning | 17 |
| 1.3 Imagerie multispectrale | 24 |
| 1.3.1 Différents types de systèmes d'acquisition en imagerie multispectrale | 24 |
| 1.3.2 Système d'acquisition MSFA one-shot | 25 |
| 1.4 Reconnaissance faciale multispectrale | 28 |
| 1.4.1 Base d'images multispectrales | 28 |
| 1.4.2 Système de reconnaissance faciale multispectrale | 29 |
| 1.4.3 Evaluation de la performance des systèmes de reconnaissance | 29 |
| Conclusion | 30 |

Introduction

La biométrie est une technologie innovante et émergente utilisée dans les contrôles d'accès. Elle étudie la reconnaissance d'un individu ou d'une personne à travers ses caractéristiques. Les systèmes biométriques sont robustes et basés sur des caractéristiques très représentatives appelées modalités ou traits biométriques. Il en existe plusieurs à savoir l'empreinte digitale, le signe de la paume, le visage, l'iris, la rétine, l'ADN, la voix, la géométrie de la main, la signature et la démarche, etc.

La reconnaissance faciale est une technique biométrique qui permet d'authentifier et d'identifier une personne à partir des traits de son visage. La RF est un sujet de recherche très actif en vision par ordinateur en général et en biométrie en particulier depuis environ 60 ans. Cet intérêt toujours renouvelé par les communautés de recherche a été marqué par le développement de plusieurs applications potentielles de la RF dans [5]–[13]. Ces différentes applications développées sont basées sur diverses méthodes de reconnaissance et systèmes d'acquisition. En fonction du système d'acquisition les systèmes de RF peuvent être classés en deux catégories : la RF hétérogène ou multispectrale et la RF mono spectrale. La RF hétérogène regroupe tous les systèmes qui acquièrent les images sur plusieurs spectres (visible, proche infrarouge, thermique) grâce à l'imagerie MS tandis que la RF mono spectrale utilise uniquement les images acquises dans un seul spectre donnée. Différents types de système d'acquisition multispectrale sont utilisés dans la RF multispectrale.

Ce chapitre présente l'état de l'art sur la reconnaissance faciale et l'imagerie multispectrale. Nous décrivons dans cette partie les bases de la biométrie et les méthodes [14] de la RF. Aussi les différents types de systèmes d'acquisition MS en occurrence ceux basés sur le MSFA sont détaillés.

1.1 Fondamentaux de la biométrie

La biométrie est la science qui étudie les variations biologiques à l'intérieur d'une population déterminée à l'aide d'outils mathématiques. Précisément il s'agit de la reconnaissance et de l'identification d'individus ou de personnes à l'aide d'une ou de plusieurs caractéristiques physiologiques et/ou comportementales. Le principal avantage des systèmes biométriques est que les caractéristiques biométriques sont uniques et par conséquent, difficiles à falsifier. Cependant, toutes les caractéristiques d'un être humain ne peuvent être utilisées comme modalité biométrique. En effet, toute caractéristique est dite biométrique lorsqu'elle présente les propriétés suivantes :

- Universalité : cette caractéristique doit être disponible chez toutes les personnes à identifier ;
- Unicité : la caractéristique ne doit pas être similaire pour toutes les personnes de la base de données ;
- Permanente : la caractéristique collectée doit être disponible durant toute la vie de la personne ;
- Collectable : la caractéristique doit être mesurable afin de permettre sa comparaison ;
- Acceptabilité : la caractéristique doit être facile à acquérir et disponible.

La vision par ordinateur élabore un ensemble de méthodes qui permettent de développer les applications biométriques. Un système biométrique est d'abord défini par son architecture et son mode de fonctionnement.

1.1.1 Architecture d'un système biométrique

Un système biométrique peut être représenté par quatre modules principaux[15], [16] :

1. La capture ou l'acquisition : ce module est responsable de l'acquisition des données biométriques d'un individu, cela peut être un appareil photo, un lecteur d'empreintes digitales, une caméra de sécurité ou d'un ensemble de caméras ;

2. L'extraction de caractéristiques : elle prend en entrée les données biométriques acquises par le module de capture. Le but est d'extraire seulement l'information pertinente afin de former une nouvelle représentation des données. Idéalement, cette nouvelle représentation est censée être unique pour chaque personne ;

3. La correspondance : elle permet de comparer l'ensemble des caractéristiques extraites avec le modèle enregistré dans la base de données du système lors de la phase d'enrôlement afin de déterminer le degré de similitude (ou de divergence) entre les deux ;

4. La décision : ce module vérifie l'identité affirmée par un utilisateur ou détermine l'identité d'une personne basée sur le degré de similitude entre les caractéristiques extraites et le(s) modèle(s) stocké(s).

L'architecture d'un système biométrique est illustrée dans la [Figure 1.1](#).

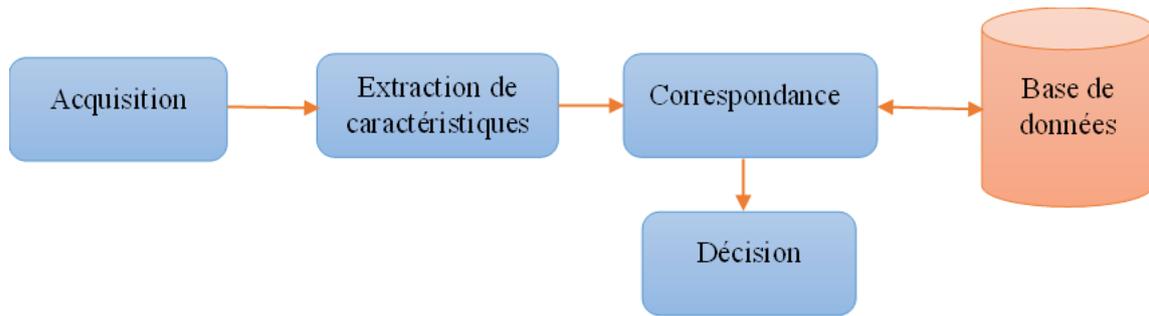


Figure 1.1- Architecture d'un système biométrique

1.1.2 Mode de fonctionnement d'un système biométrique

Un système biométrique fonctionne en trois modes à savoir l'enrôlement, l'identification et la vérification.

- L'enrôlement constitue la première étape de fonctionnement de tout système de reconnaissance. Au cours de cette phase, l'utilisateur est enregistré dans le système biométrique pour la première fois, la modalité biométrique est acquise et prétraitée. Une extraction des caractéristiques pertinentes de la modalité acquise est faite puis stockée dans une base de données en vue d'une utilisation ultérieure au cours de l'identification et de la vérification. Cet enregistrement peut s'accompagner de l'ajout d'informations biographiques dans la base de données ;

- En mode identification, les systèmes biométriques permettent à un individu de faire référence à ses données. L'utilisateur ne dévoile pas explicitement son identité. Cependant, l'affirmation implicite faite par l'utilisateur est qu'il est une personne déjà enrôlée par le système. Ainsi, cette donnée biométrique est comparée avec les modèles de tous les utilisateurs enrôlés dans la base de données.

- La vérification ou l'authentification consiste à contrôler si l'utilisateur est bien celui qu'il prétend être. Il affirme son identité et le système vérifie si cette affirmation est valide ou non. Lors de cette phase, l'identité est réclamée, une correspondance du modèle de vérification avec les modèles maîtres est effectuée pour vérifier la demande d'identité de la personne. Cette vérification fait référence à une correspondance 1 à 1.

La [Figure 1.2](#) résume les modes de fonctionnement d'un système biométrique

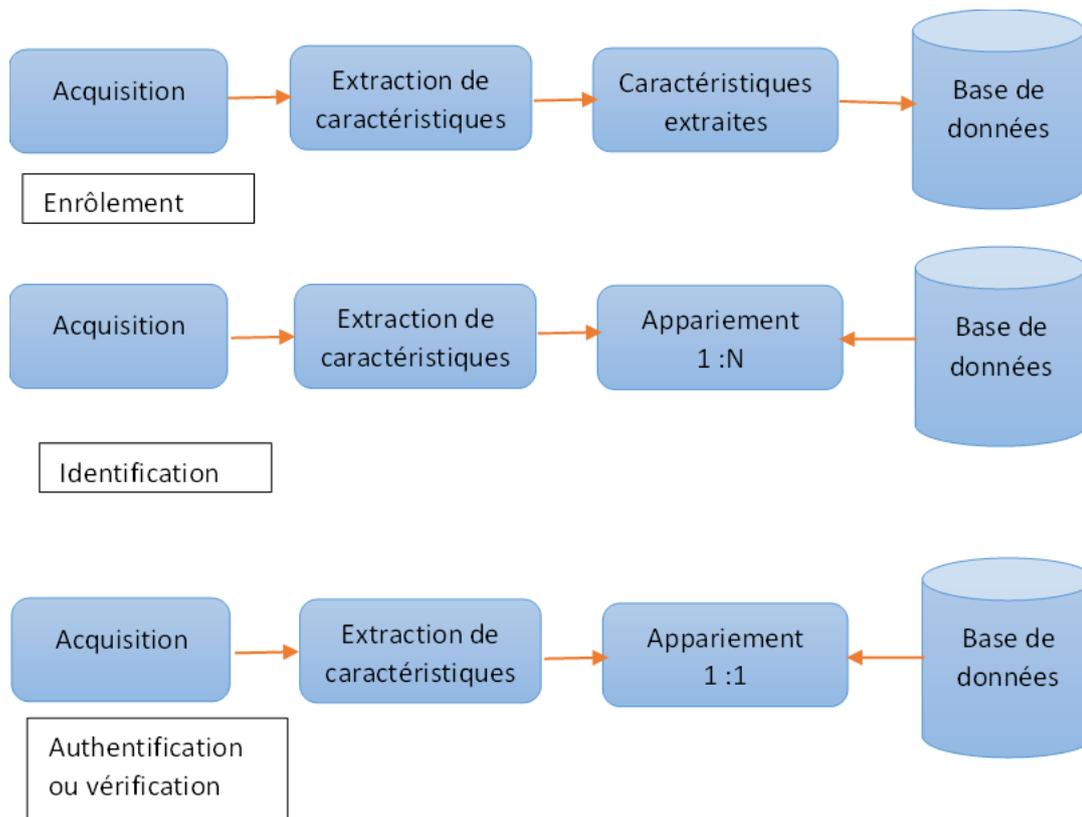


Figure 1.2- Mode de fonctionnement d'un système biométrique

1.1.3 Avantages des systèmes biométriques

La biométrie est la technologie la plus pertinente pour identifier et authentifier les personnes de manière fiable et rapide, en fonction de caractéristiques uniques. Elle présente plusieurs avantages tels que :

- Une haute sécurité assurée par des techniques peu coûteuses ;
- Une réduction des coûts de gestion de mot de passe ;
- La réduction de la fraude ;
- Une identification très rapide des personnes ;
- L'élimination des problèmes de pertes de codes d'accès et leur remplacement par des caractéristiques physiologiques ;
- La prévention des pertes de temps.

1.1.4 Modalités biométriques

La littérature présente plusieurs modalités biométriques. Elles peuvent être classées en trois grandes catégories que sont les modalités biométriques morphologiques, comportementales et biologiques. Anthony Ngozichukwuka et al. présentent une nouvelle

modalité qui est l'électrocardiogramme (ECG) dans [17]. Les Figure 1.4 et 1.4 présentent les différentes modalités biométriques et leur type.



Figure 1.3 - Modalités biométriques

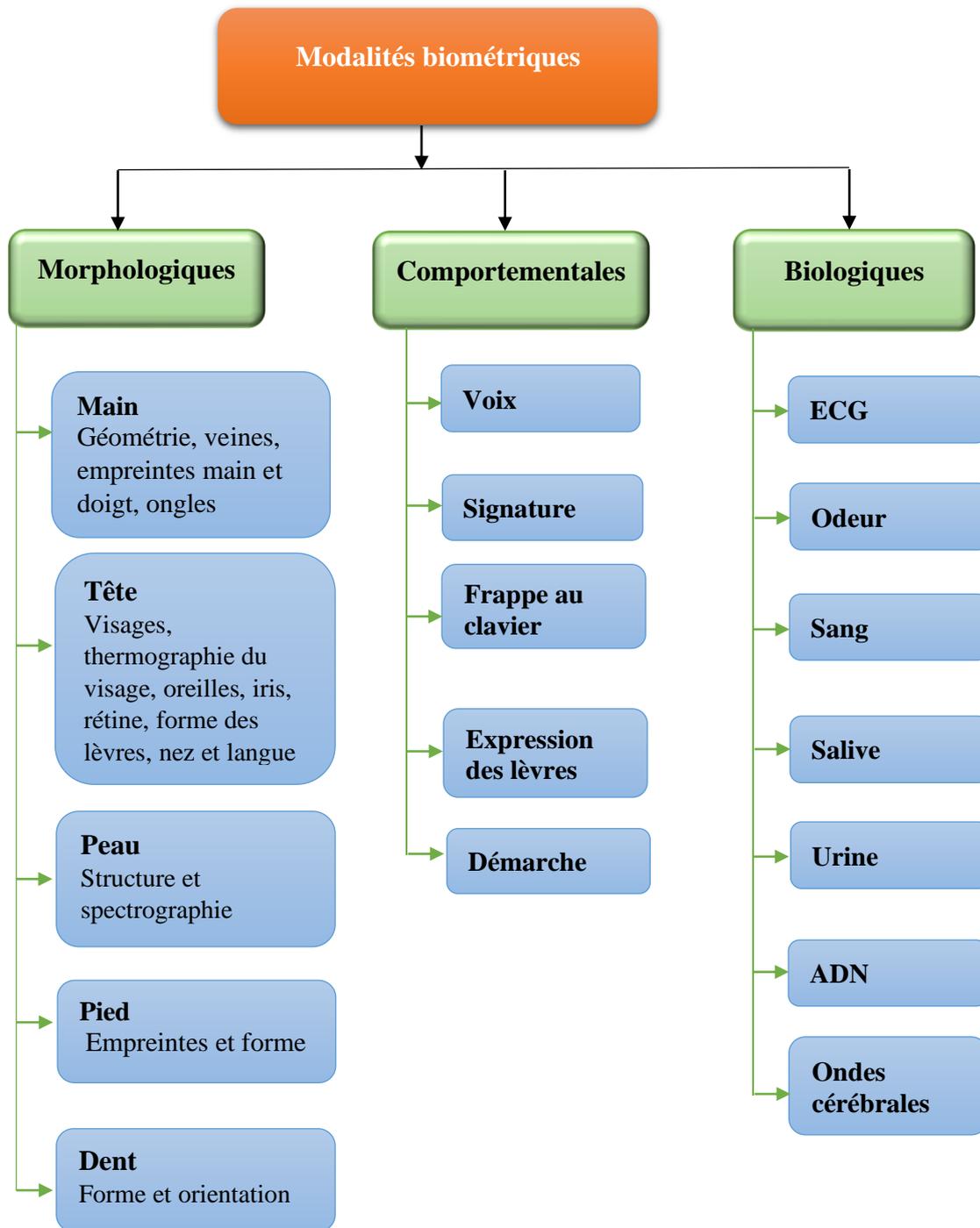


Figure 1.4 - Modalités biométriques par type

1.2 Reconnaissance de visage

1.2.1 Motivation

La RF est une technique biométrique qui utilise la face ou le visage d'une personne comme trait biométrique. La RF est très répandue et populaire. Ceci est dû au fait que le visage est une modalité biométrique non intrusive, et très facile à acquérir en comparaison aux autres modalités biométriques. Bien que la RF soit populaire, elle a également des limites. Le [Tableau 1.1](#) présente quelques avantages et inconvénients de la RF.

Tableau 1.1- Avantages et inconvénients des systèmes de reconnaissance faciale

| Avantages | Inconvénients |
|--|---|
| <ul style="list-style-type: none">- Bien accepté par le public- Peu intrusive- Pas de contact physique- Technique peu coûteuse- Plusieurs traits caractéristiques | <ul style="list-style-type: none">-Technologie sensible à l'expression du visage,-Technologie sensible à l'environnement : éclairage, position, fumée, neige, etc.- Difficultés à différencier de vrais jumeaux.- Sensible aux changements : barbe, moustache, lunettes, âge, tatouage, chirurgie... |

De nombreux algorithmes ou méthodes sophistiqués sont utilisés dans la RF. La littérature distingue les méthodes classiques, les détecteurs de points d'intérêt et les méthodes d'intelligence artificielle que sont celles d'apprentissage automatique et profond.

1.2.2 Méthodes de reconnaissance de visage

1.2.2.1 Méthodes classiques

Les méthodes classiques sont utilisées au cours des modules d'extraction de caractéristiques, de correspondance et de décision. Nous pouvons classer les différentes méthodes classiques en méthodes globales, méthodes locales, méthodes hybrides et les détecteurs descripteurs de points d'intérêts.

- Les *méthodes globales* utilisent toute la surface du visage comme source d'information sans tenir compte des caractéristiques locales comme les yeux, la bouche, le

nez. Elles sont relativement rapides à mettre en œuvre, mais sensibles aux variations d'illumination, de pose et d'expression faciale. Parmi les approches les plus importantes réunies au sein de cette classe on trouve : le Mélange Gaussien, le Modèle Surfaccique du Visage (3D), les ondelettes et autres méthodes telles que l'ACP, le SVM etc...

- Les *méthodes locales* aussi appelées méthodes à traits géométriques, à caractéristiques locales ou analytiques. Ces méthodes consistent à appliquer des transformations en des endroits spécifiques de l'image du visage. Précisément, ces méthodes extraient les caractéristiques locales de visage comme les yeux, le nez et la bouche, puis utilisent leur géométrie et/ou l'apparence comme donnée d'entrée au module de correspondance. L'énergie sera accordée aux détails locaux évitant ainsi le bruit engendré par les cheveux, les lunettes, les chapeaux, la barbe, etc. Mais leur difficulté se présente lorsqu'il s'agit de prendre en considération plusieurs vues du visage. Parmi les méthodes locales on peut citer : les Modèles de Markov Cachés (Hidden Markov Models), l'Algorithme Elastic Bunch Graph Matching (EBGM), l'Eigen Object (EO), l'appariement de gabarits.

- Les *méthodes hybrides* : elles combinent les deux types de méthodes, offrant ainsi potentiellement le meilleur des deux types de méthodes.

- Les *détecteurs et descripteurs de points d'intérêt* : un détecteur permet de déterminer et d'isoler les zones d'intérêt dans une image. Il s'agit des points ou des régions d'intérêt. Un point d'intérêt est défini par un changement brusque d'une caractéristique dans le voisinage d'un pixel. Il correspond alors à une discontinuité des niveaux de gris des pixels. Une région est définie comme une zone de pixels adjacents qui satisfait une certaine propriété souvent exprimée en termes d'homogénéité, de couleur et de saturation. Un descripteur a pour but de décrire les points d'intérêt obtenus par un détecteur. Il analyse le voisinage proche de chaque point d'intérêt et produit un vecteur caractéristique de cette zone. Ce vecteur est appelé vecteur descripteur et doit être unique pour chaque zone d'intérêt. Les détecteurs sont caractérisés par l'invariance et la discrimination. Plusieurs détecteurs et descripteurs ont fait leurs preuves en reconnaissance faciale, on peut citer entre autres : SIFT, ASIFT, MSER, HOG, KAZE, SURF.

1.1.1.2 Intelligence artificielle

L'Intelligence Artificielle (IA) est un processus d'imitation de l'intelligence humaine. Son but est de permettre à des ordinateurs de penser et d'agir exactement comme des êtres humains. L'IA est un ensemble de théories et de techniques mises en œuvre et exécutées dans un environnement informatique dynamique capable de traduire les raisonnements humains en

programmes informatiques. John McCarthy[18] est l'un des principaux pionniers de l'IA créée depuis 1949. L'IA est utilisée dans plusieurs domaines à savoir l'agriculture[19], [20], la médecine[21], [22], la reconnaissance de forme[23] [24], les banques et finances [25], [26], l'industrie, le commerce, IOT[27], [28] etc. Les applications d'IA sont utilisées pour simuler l'intelligence humaine afin de résoudre un problème ou de prendre une décision. L'IA offre les avantages de la permanence, de la fiabilité et de la rentabilité, tout en tenant compte de l'incertitude et de la rapidité dans la résolution d'un problème ou la prise d'une décision. L'IA englobe l'apprentissage automatique ou Machine Learning (ML) en anglais et l'apprentissage profond ou Deep Learning (DL).

1.2.2.2.1 Machine learning

L'apprentissage est un comportement humain naturel qui est devenu un aspect essentiel des machines. L'objectif principal du ML est de permettre aux ordinateurs d'apprendre sans assistance humaine. Ses méthodes peuvent être classées en quatre principales catégories, à savoir les approches d'apprentissage supervisé, les approches d'apprentissage non supervisé, les approches d'apprentissage semi-supervisé et les approches d'apprentissage par renforcement.

L'apprentissage supervisé est la tâche de ML consistant à apprendre une fonction de prédiction à partir d'exemples étiquetés. Les algorithmes de ML supervisé sont ceux qui nécessitent une assistance externe. L'ensemble des données d'entrée est subdivisé en ensemble de données d'apprentissage et de test. Une formulation standard de la tâche d'apprentissage supervisé est le problème de classification : l'apprenant doit apprendre (pour se rapprocher du comportement) d'une fonction qui mappe un vecteur dans l'une des classes en examinant plusieurs exemples d'entrée-sortie de la fonction. Lors de l'apprentissage supervisé, l'algorithme reçoit un ensemble de données qui est étiqueté avec des valeurs de sorties correspondantes sur lequel il va pouvoir s'entraîner et définir un modèle de prédiction. Cet algorithme pourra par la suite être utilisé sur de nouvelles données afin de prédire leurs valeurs de sorties correspondantes.

En RF, les algorithmes d'apprentissage supervisé sont généralement utilisés pour la classification et la régression. Les algorithmes d'apprentissage supervisé les plus utilisés sont : Support Vector Machine (SVM), l'arbre de décision, l'Analyse Discriminante, K-Nearest Neighbors (KNN), naïves bayes etc.

- La Machine à Vecteurs de Support (SVM) est une méthode de classification binaire par apprentissage supervisé, elle fut introduite par Vapnik et al. [29] en 1995. Cette méthode

permet de trouver le meilleur hyperplan qui sépare les ensembles de points dans un espace de grande dimension tout en minimisant le taux d'erreur total de classification ;

- L'Analyse Discriminante Linéaire (ADL) a été introduite par Fisher en 1936, et utilisée en RF par Belhumeur et al. en 1997. Cette méthode permet d'expliquer et de prédire l'appartenance d'un individu à une classe prédéfinie à partir de ses caractéristiques mesurées à l'aide de variables prédictives [30].

- Le KNN utilise la structure des plus proches voisins pour faire la prédiction. Cette approche peut être utilisée pour résoudre les problèmes de classification et de régression. Comme le SVM, il est beaucoup utilisé en RF[31].

Les approches d'apprentissage non supervisé s'opposent aux approches d'apprentissage supervisé, car elles ne nécessitent aucun processus de formation. Mais les approches d'apprentissage supervisé sont plus simples que les approches d'apprentissage non supervisé. Les algorithmes d'apprentissage non supervisé sont aussi beaucoup exploités en RF, parmi ces méthodes on peut citer l'Analyse en Composantes Principales (ACP). Beaucoup utilisé en RF, l'ACP est une approche statistique utilisée pour extraire les caractéristiques et pour réduire la dimensionalité des grandes données. L'application de l'algorithme de l'ACP (PCA) est née des travaux de M. Turk et al. au MIT Media Lab en 1991 [32]. Il est aussi connu sous le nom de Eigenfaces car il utilise des vecteurs propres de la matrice de covariance de l'ensemble des images de visages.

L'apprentissage semi-supervisé est la branche de ML qui consiste à combiner des données étiquetées et non étiquetées pour effectuer certaines tâches d'apprentissage. En effet, l'apprentissage semi-supervisé est conceptuellement situé entre l'apprentissage supervisé et non supervisé. Il permet d'exploiter les grandes quantités de données non étiquetées disponibles dans de nombreux cas d'utilisation.

L'apprentissage par renforcement est un algorithme qui apprend une politique sur la façon d'agir compte tenu d'une observation du monde. Chaque action a un certain impact sur l'environnement, et ce dernier fournit une rétroaction qui guide l'algorithme d'apprentissage. Par exemple pour un agent autonome comme un robot l'apprentissage par renforcement consiste à apprendre les actions à effectuer à partir d'expériences, de façon à optimiser une récompense quantitative au cours du temps. L'agent est plongé au sein d'un environnement et prend ses décisions en fonction de son état courant.

1.2.2.2.2 Deep learning

Au cours de cette dernière décennie, une série de progrès remarquables a impacté le ML avec des algorithmes d'apprentissage efficaces et très sophistiqués. L'une de ces progrès a été l'évolution des Réseaux de Neurones Artificiels (RNA) vers des architectures de réseaux neuronaux de plus en plus profondes avec des capacités d'apprentissage très améliorées. En effet, les RNA s'inspirent du fonctionnement des réseaux de neurones biologiques. L'utilisation principale des RNA pour la reconnaissance faciale est de combiner les caractéristiques extraites avec les différents outils d'extraction de caractéristiques et les RNA. L'un des premiers RNA utilisé en RF avait été développé par Aleksander et al. [33]. Ce système de RF nommé WISARD est un réseau adaptatif à une seule couche qui contient un réseau séparé pour chaque individu stocké. Plusieurs autres systèmes de RF basés sur les RNA ont été proposés dans la littérature. En 2011, MageshKumar et al. [34] ont développé un système de RF basé sur les ondelettes de Gabor, l'ACP, le LDA et un RNA à rétropropagation. Les caractéristiques des visages ont été extraites avec les ondelettes de Gabor puis réduites avec l'ACP. Ces caractéristiques ont été envoyées au classifieur qui est le réseau neuronal à rétropropagation pour la reconnaissance. Ce système avait fourni de bons résultats comparés à ceux qui existaient.

L'apprentissage profond ou deep learning est donc un domaine ou une extension de ML. C'est un concept d'apprentissage automatique basé sur les RNA et la notion de profondeur. En effet, DL utilise une grande quantité de données tandis que ML en utilise une petite. C'est l'approche d'apprentissage automatique la plus efficace, et rapide. Sa particularité est de découvrir des structures complexes dans de grands ensembles de données par l'utilisation d'algorithmes de rétropropagation. Ceci indique comment une machine doit modifier les paramètres internes utilisés pour calculer la représentation de chaque couche à partir de la représentation de la couche précédente.

Au cours de cette décennie les méthodes de DL ont fait une percée significative avec des performances exponentielles et très appréciables dans de nombreux domaines d'applications[35]–[37] tels que : la reconnaissance vocale, la détection, la reconnaissance d'objets, la détection du cancer, des flux de données, la reconnaissance de documents, la génomique, l'agriculture, la médecine et bien d'autres domaines. L'introduction du DL ou des réseaux de neurones profonds (Deep Neural Network DNN en anglais) dans les systèmes de reconnaissance faciale a largement amélioré leur performance. Les DNN sont très puissants et capables d'extraire efficacement des caractéristiques que les ANN et les méthodes

statistiques classiques n'arrivent pas à extraire [38]. Le DNN est une extension de l'ANN avec plusieurs couches de réseau de neurones cachées.

L'architecture des DNN est composée de plusieurs couches cachées et de plusieurs neurones par couche. Ces couches sont structurées et communiquent entre elles de façon que la couche suivante prend les résultats de la précédente en entrée. La littérature présente une grande variété d'algorithmes mise en œuvre pour l'apprentissage profond en reconnaissance faciale. Selon leur architecture et mode de fonctionnement on distingue :

- Les Réseaux de neurones convolutifs (CNN)
- Les Auto-Encodeurs (AE)
- Les réseaux génératifs contradictoires (GAN)
- Les réseaux hybrides
- L'apprentissage par renforcement profond

1.2.2.2.1 Les Réseaux de neurones convolutifs (CNN)

Les CNN sont constitués de plusieurs couches qui communiquent entre elles. Ce type de réseau de neurones diffère des ANN par l'extraction de caractéristiques. Le premier CNN a été développé par Lecun et al. [39]. La fonctionnalité des algorithmes de CNN en RF est basée sur deux parties : l'extraction de caractéristiques et la classification (ANN). En effet, le nom CNN vient de la convolution qui est une opération mathématique entre deux matrices que sont l'image de la face à reconnaître et le filtre de convolution. Cette opération répétée plusieurs fois permet d'extraire les caractéristiques les plus pertinentes de l'image dans un vecteur. A la fin de cette série d'opérations, le vecteur caractéristique obtenu constitue l'entrée de l'ANN. L'architecture de base du CNN comporte quatre couches : la couche convolutive, la couche de regroupement(pooling), la couche correction ReLU et la couche entièrement connectée. Les deux premières couches sont paramétrables. Cette architecture varie en fonction du problème à résoudre et devient un empilement de ces différentes couches. La [Figure 1.5](#) présente l'architecture de base d'un CNN.

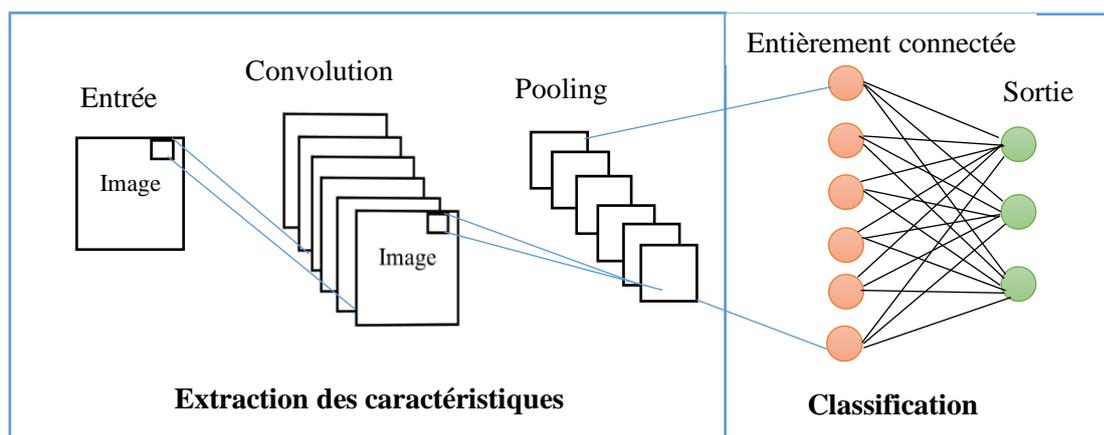


Figure 1.5 - Architecture de base d'un CNN

Tout CNN doit commencer par une couche de convolution et terminer par une couche entièrement connectée. Les couches intermédiaires peuvent s'empiler de différentes manières, à condition que la sortie d'une couche ait la même structure que l'entrée de la suivante. Par exemple, une couche fully-connected, qui renvoie toujours un vecteur, ne peut pas être placée avant une couche de pooling, puisque cette dernière doit recevoir une matrice 3D.

L'utilisation des CNN en reconnaissance faciale a démontré une excellente performance (jusqu'à 100% de précision) pour les systèmes de RF. De nombreux systèmes de RF développés sont basés sur les algorithmes de DL tels que VGGNet [40], GoogLeNet [41](22 couches), ResNet50 [42], LightCNN-v9 [6], SqueezeNet [43] et autres. Le Tableau 1.2 présente un aperçu de quelques méthodes de RF basés sur les CNN.

Tableau 1.2- Quelques CNN les plus utilisés en reconnaissance faciale

| Algorithmes | Base de données | Précision |
|----------------------------|---------------------------------|-------------|
| ResNet27 | LFW (Labeled Faces in the Wild) | 99,48 |
| ResNet50 | IJB (IARPA Janus Benchmark) | 0,995±0.001 |
| SqueezeNet ResNet50 | IJB | 0,996±0.001 |
| LigthCNN-V29 | LFW | 98.98 |
| Deep CNN | LFW | 99.77 |
| ReST | LFW | 93,4 |
| MTCNN | YouTubeFace | 95 |
| NAN | YouTubeFace | 95,72 |
| VGG16 | 300WLP | 98 |
| DDRL | YouTubeFace | 94,2 |
| PDA | VGGFace2 | 95, 32 |

Resnet50 est un CNN utilisé pour l'apprentissage supervisé, il contient 50 couches. Ce modèle a été testé sur un ensemble d'images appelé VGGFace2 [44] qui a été aussi utilisé pour tester le CNN SqueezeNet et ResNet-50. Wen et al. [45] ont proposé une nouvelle approche pour la RF basée sur l'invariance de l'âge. Les résultats expérimentaux ont été testés sur les bases de données d'âges à savoir MORPH Album2, FGNET, and CACD-VS. La précision de ce système varie en fonction de l'âge et de la base de données et peut atteindre jusqu'à 98.72%. Hu et al. [46] ont développé un CNN nommé GTNN (Gated Two stream Neural Network) pour améliorer les performances de la RF. Les auteurs ont fusionné les caractéristiques du visage avec les tenseurs non linéaires robustes pour la RF. Une précision de 99,65 % et de 99,94% était atteinte respectivement sur l'ensemble des données LFW et de CASIA NIR-VIS2. Ghazi et al. [47] ont utilisé les CNN pré entraînés VGG-Face net et CNN allégé pour développer une nouvelle approche de RF. Ce système prend en compte les changements d'éclairage de différentes intensités, l'occlusion supérieure et inférieure du visage et les angles de pose variables de la tête. Les auteurs ont testé leur méthode sur cinq ensembles de données à savoir : Yale, CMU PIE, Color FERET, FRGC et AR. Certains facteurs tels que l'illumination, l'occlusion, le désalignement et la pose de la tête ont réduit la précision de RF du modèle d'apprentissage profond. Le modèle VGG-Face avait obtenu une meilleure performance que Lightened CNN sur l'ensemble des données.

Chen et al. [48] ont proposé une nouvelle approche de reconnaissance spécifique aux visages de basse résolution. Cette méthode est basée sur les LightCNN-v9 et LightCNN-v29. La particularité de cette approche est qu'elle peut fonctionner de manière satisfaisante avec une résolution de visage inférieure à 32*32 pixels. L'étude expérimentale a permis d'atteindre une précision de 98,46 % sur LightCNN-v9 et 98,98 % sur LightCNN-v29. Wu et al. [49] ont proposé une méthode d'identification des visages de bout en bout en s'inspirant du transformateur spatial. Les auteurs ont introduit un module appelé Recursive Spatial Transformer (ReST) dans leur système de RF. Leur modèle comporte trois parties : les couches de convolution, le réseau DCNN (AlexNet modifié) avec une couche softmax et la couche de transformation spatiale. Wang et al. [50] ont proposé un système de RF basé sur les CNN qui intègre de multiples branches locales basées sur l'attention à différentes échelles. Cette architecture présente un regroupement bilinéaire hiérarchique pour combiner les caractéristiques à différentes échelles afin de mettre l'accent sur différentes régions faciales discriminées de manière automatique et adaptative. Zhang et al. dans [51] ont mis au point en 2022 une approche efficace appelée Efficient Lightweight Attention Networks (ELANet) pour la RF. Ce réseau d'attention efficace et léger permet d'améliorer la performance de RF en

tenant compte des problèmes de pose et de variation de l'âge des sujets à reconnaître. Par rapport aux réseaux légers récents, ELANet a amélioré ses performances de 1,83 %, 2,17 % et 0,92 % respectivement sur l'ensemble des données CPLFW, VGG2_FP et CALFW. Cette approche permet de relever le défi de l'impact des poses et des âges sur les performances de la RF avec peu de paramètres et d'efforts de calcul. En 2019, Lai et al. [52] ont conçu un système de RF appelé PoreNet. Ce système permet de détecter les pores à partir des images de visages pour la RF. En effet, Les images de visages à haute résolution permettent d'observer les caractéristiques faciales à l'échelle des pores. PoreNet est un détecteur/descripteur de pores basé sur les réseaux de neurones. C'est une version modifiée de HardNet développé en 2017 par Mishchuk et al. [53]. J. Yang et al. [54] ont présenté un réseau d'agrégation neuronale pour la reconnaissance vidéo des visages. Cette méthode basée sur les CNN prend en entrée une vidéo ou un ensemble d'images de visage d'une personne avec un nombre variable d'images de visage. L'extraction des caractéristiques est réalisée à l'aide du CNN GoogLeNet. Ces caractéristiques sont ensuite passées à deux blocs d'attention et des poids linéaires leur ont été attribués. Pour la vérification des visages, les auteurs avaient utilisé le réseau d'agrégation neuronale siamois. En 2019, Guo et al [55] avaient réalisé une enquête sur la RF basée sur l'apprentissage profond. Plusieurs modèles profonds basés sur les CNN populaires avaient été comparés. En 2021, Fuad et al. [38] ont présenté les avancées récentes en matière d'apprentissage profond pour la RF.

1.2.2.2.2 Les auto-encodeurs (AE)

La qualité des images à reconnaître impacte la performance des systèmes de RF. Plus l'image contient du bruit moins est performant le système de reconnaissance. L'AE est une technique de reconstruction d'une image à partir d'une image bruitée. Un AE est un réseau neuronal profond non supervisé basé sur l'apprentissage des caractéristiques, il encode et décode les données de grande taille de manière efficace. Cette technique comporte deux phases : l'encodage et le décodage avec une couche d'entrée, une couche de sortie et une ou plusieurs couches cachées. Lors de l'encodage, l'entrée est comprimée en une caractéristique de dimension inférieure avec une représentation significative. Ce processus est répété jusqu'à ce que la dimension requise soit atteinte. Dans la phase de décodage, le processus inverse est réalisé pour générer les caractéristiques essentielles à partir de l'étape d'encodage. Les AE sont utilisés en RF comme outils de réduction de dimensionalités des données (caractéristiques extraites et images) de grande taille. La littérature présente diverses variantes d'AE pour la RF. En 2017, Xu et al. [56] ont proposé un modèle de réseau neuronal appelé

réseaux d'auto-encodeurs couplés (CAN) pour traiter le problème de la RF non influencée par l'âge. Le CAN est composé de deux auto-encodeurs qui sont reliés. Cette approche s'adapte au processus complexe et non linéaire de vieillissement et de dé-vieillessement. Ce système intègre une méthode d'analyse factorielle non linéaire permettant de décomposer une image de visage donnée en trois composantes à savoir les caractéristiques d'identité, d'âge et de bruit. En 2018, Usman et al. [57] ont développé des AE profonds pour la sélection des caractéristiques et la réduction de la dimension pour la reconnaissance des expressions faciales. Les auteurs ont démontré que les caractéristiques extraites de l'AE empilé sont plus performantes que d'autres techniques de pointe de sélection de caractéristiques et de réduction des dimensions. Dans le but de régler les problèmes d'alignement en RF hétérogènes, Kantarci et al. [58] ont développé en 2020 un AE profond. Cette méthode permet d'apprendre la correspondance et d'évaluer l'impact de l'alignement entre les images du visible et du thermique. Le [Tableau 1.3](#) présente quelques algorithmes d'AE développés en RF.

Tableau 1.3- Quelques algorithmes d'AE pour la RF

| Algorithmes | Base d'images | Précision (%) |
|-------------------|---------------|---------------|
| CAN[56] | CACD-VS | 92,5 |
| U-net[58] | EURECOM | 88,33 |
| DC-SSDA[59] | CMU-PIE | 85 |
| D ² AE | CalebA | 87,82% |

1.2.2.2.3 Les réseaux antagonistes génératifs (GAN)

Comme son nom l'indique, les réseaux adversaires génératifs sont une approche de la modélisation générative utilisant des méthodes d'apprentissage profond non supervisé. Ils sont composés de deux modèles antagonistes : Le générateur et le discriminateur. Le générateur produit de nouvelles caractéristiques tandis que le discriminateur apprend à déterminer si les caractéristiques générées par le générateur sont réelles ou fausses. Les GANs ont été introduits en RF dans le but de synthétiser les visages d'un spectre à partir des images d'un spectre donné pour surmonter les problèmes liés à la variation des visages en fonction de l'âge, de la pose et autre. Il existe plusieurs de systèmes de RF basés sur les GANS parmi lesquels : TR-GAN [60], ACSFC(Adversarial Cross-spectral Face Completion) [61], Age-cGAN [62], DR-GAN [63].

1.2.2.2.4 Les réseaux hybrides

Un réseau hybride est la combinaison de deux ou plusieurs modèles génériques d'apprentissage automatique dans le but d'améliorer le modèle ou de résoudre un problème bien défini avec précision. Dans cette technique, un algorithme complète un autre algorithme pour résoudre les problèmes avec précision. Le choix du réseau hybride en RF dépend de l'optimisation du système qui doit être faite. Les algorithmes hybrides sont basés sur plusieurs types de combinaison telle que : CNN+GAN, CNN+AE, GAN+RL (REINFORCEMENT LEARNING). La littérature présente plusieurs systèmes de reconnaissance basés sur les réseaux hybrides parmi lesquels:

DeepID3 proposé Sun et al. [64] , ce réseau est une combinaison de VGG-net et GoogLeNet. Cette approche modifie l'architecture extérieure des réseaux et améliore la précision de la vérification et de l'identification des visages. Le réseau DeepID3 montre d'excellentes performances sur la base d'images LFW avec une précision de 99,52 %. Un réseau hybride nommé MDLFace a été proposé par Goswami et al. [65]. Cette approche est une combinaison d'une pile d'auto-encodeurs de débruitage et de machines de Boltzmann profondes pour la RF à l'aide des images les plus mémorables. Zhang et al. [66] ont aussi développé un système de RF hétérogène basé sur un réseau hybride. Ce système combine un DCNN et GAN pour la RF.

1.2.2.2.5 L'apprentissage profond par renforcement

L'apprentissage par renforcement est un domaine de l'apprentissage automatique qui s'intéresse à la manière dont les agents intelligents doivent agir dans un environnement afin de maximiser la notion de récompense cumulative. L'apprentissage profond par renforcement est une combinaison de l'apprentissage profond et de l'apprentissage par renforcement. Cette combinaison améliore la performance du système de RF qui l'implémente. Duong et al. [67] ont développé un système de reconnaissance pour le vieillissement du visage basé sur l'apprentissage profond par renforcement. Ce système se compose du CNN VGG19 pour l'extraction des caractéristiques et du modèle d'apprentissage par renforcement qui sélectionne les voisins afin d'exploiter la relation temporelle entre deux images consécutives. Cette approche génère un futur visage âgé à partir d'un jeune visage dans une image vidéo.

1.3 Imagerie multispectrale

L'imagerie MS regroupe l'ensemble des concepts et techniques utilisés pour capturer les images sur plusieurs spectres (visible et infrarouge) ou dans des gammes de longueurs d'onde spécifiques du spectre. En effet, l'imagerie couleur est un cas particulier d'imagerie multispectrale, où l'échantillonnage spectral vise à acquérir des primaires qui permettent de reconstruire les informations de couleur (RVB) en utilisant la théorie de la trichromie.

Les caractéristiques des images acquises dans chaque spectre diffèrent les unes des autres. Par exemple les images acquises dans le spectre du visible ne contiennent que les informations visibles à l'œil nu tandis que celles acquises dans l'infrarouge contiennent d'autres informations plus pertinentes. Les images MS contiennent des informations plus riches que celles en couleurs. En effet, cette combinaison de spectres rend disponible les informations relatives à chaque spectre. Au cours de cette décennie, le développement des systèmes d'imagerie MS a connu un grand essor en raison de leur fort potentiel dans de nombreux domaines d'application tels que la télédétection, l'astronomie, la physique, les musées, l'agriculture, la médecine, l'impression couleur haute précision, l'infographie, etc. Les images MS fournissent des informations sur 4 à 10 bandes par pixel.

Cette section présente brièvement les types de systèmes d'acquisition MS, plus précisément les systèmes d'acquisition one shot.

1.3.1 Différents types de systèmes d'acquisition en imagerie multispectrale

Les systèmes d'imagerie MS peuvent être classés en fonction des critères tels que leur capacité à produire des images instantanées, à utiliser un seul capteur ou non et par la technologie utilisée pour diviser la lumière. Ces systèmes d'acquisition peuvent être subdivisés en deux grandes catégories : les systèmes d'acquisition en mode scanning et les systèmes d'acquisition snapshot.

- Les systèmes d'acquisition **en mode scanning** : ce sont des caméras qui capturent l'image MS en plusieurs prises. Elles peuvent être subdivisées en trois catégories : les caméras à filtre accordable, les caméras à éclairage accordable et le système d'acquisition composé de plusieurs caméras à un coup. Les filtres accordables sont une technique qui permet de capturer chaque bande d'image MS avec un filtre, c'est un système de roue qui tourne en fonction de chaque filtre. Parmi les différents types de filtres accordables électroniquement, les deux plus utilisés sont les filtres accordables à Cristaux liquides ou LCTF Liquid Crystal Tunable Filter (LCTF) et les filtres accordables acousto-optiques, en anglais Acousto-Optical

Tunable Filter (AOTF) inventés par Harris et al. dans [68]. Le LCTF [69] incorpore des ralentisseurs de plaque d'onde à cristaux liquides accordés par l'application d'une tension électronique dans un filtre biréfringent de Lyot, tandis que les AOTF sont des cristaux biréfringents à l'état solide dont la réponse varie en fonction d'un champ acoustique appliqué. L'éclairage accordable est similaire aux filtres accordables.

- Les systèmes d'acquisition **instantanées(snapshot)** : ce sont des caméras qui capturent l'image MS en seule prise. Cette catégorie regroupe les systèmes d'acquisition multispectrale qui intègrent un seul capteur et plusieurs capteurs. Cette catégorie est subdivisée en plusieurs classes à savoir : les MSFA, les interféromètres, les capteurs accordables et les réseaux de lentilles filtrées. Les systèmes d'acquisition multispectrale et hyperspectral équipés de plusieurs capteurs utilisent les techniques de séparation des faisceaux.

Etant donné que nos travaux de recherches portent sur le développement d'une caméra MS one shot, nous allons plus nous focaliser sur les systèmes d'acquisition instantanée en particulier sur ceux qui intègrent un réseau de filtres spectraux.

1.3.2 Système d'acquisition MSFA one-shot

Les Caméras MSFA one-shot sont des systèmes d'acquisition MS instantanés qui sont équipés d'un SFA ou d'un MSFA. Elles intègrent un seul capteur et sont caractérisées par l'acquisition de l'image MS simultanément sur plusieurs spectres. Généralement, les caméras MSFA sont des systèmes compacts constitués fondamentalement d'un capteur, d'un SFA et d'une micro-lentille. La conception des caméras MSFA one-shot n'est pas aisée, elle commence par la conception des filtres spectraux avant leur fabrication, c'est un véritable défi. La conception et la fabrication des MSFA est le fruit d'un travail collaboratif entre les laboratoires de recherche et l'industrie. Les MSFA sont une extension du CFA de Bayer qui a explosé la performance des caméras digitales. Les réseaux de filtres colorés sont constitués de trois couleurs au plus tandis que les MSFA contiennent n Bandes. En effet, la conception des MSFA commence par la sélection et l'optimisation des bandes qui le constituent. Ensuite un arrangement des bandes est fait pour voir quelle distribution des bandes permet d'avoir les meilleures images et quelle tessellation optimale favorise l'utilisation de la technique du réseau de filtres à concevoir. La tessellation est la structure géométrique des bandes, il en existe deux types la rectangulaire et l'hexagonale. Nous présentons dans la [Figure 1.6](#) quelques MSFA associés aux différents types de tessellation.

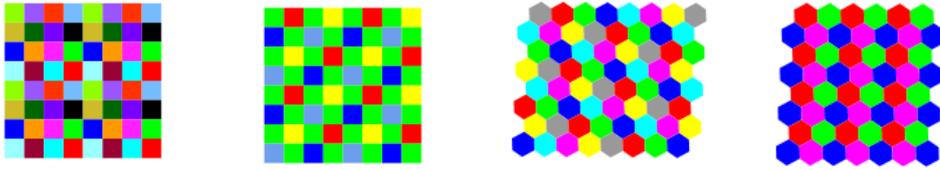


Figure 1.6 -Tessellation hexagonale et rectangulaire

L'implémentation matérielle des filtres multispectraux offrent certes des solutions mais souffre également de certains problèmes, tels que le manque de sensibilité, les fuites croisées entre les canaux, les mesures imprécises qui ne prennent pas en compte la caractérisation globale du système (capteur + filtres).

Beaucoup de recherches ont été menées pour la réalisation théorique du MSFA au cours de ces dernières années.

En 2004, Miao et al. [70], [71] ont proposé une méthode générique pour la génération des MSFA à partir d'un motif de damier avec des tesselles rectangulaires et hexagonales. En 2008, Brauers et al. [72] ont proposé une caméra multispectrale qui intègre les réseaux de filtres colorés de taille 3×2 pixels. Il s'agit d'un MSFA à 6 bandes qui permet d'accélérer l'interpolation linéaire. Hershey et Zhang [73] ont aussi développé une caméra intégrant à la fois des photodétecteurs de lumière visible et de lumière non visible dans un seul MSFA. En effet, la mosaïque est en grande partie la même que le modèle de Bayer, mais un vert dans les moxels 2×2 est remplacé par un élément de filtre non visible. Un MSFA superposé à un capteur peut être utilisé comme un spectromètre lorsque les filtres intégrés sont composés d'un nombre suffisant d'éléments filtrants à bande étroite. Ainsi, Wang et al. [74] ont conçu en 2007 un MSFA 8×16 qui comprend 128 filtres passe-bande étroits distincts pour capturer des spectres NIR (proche infrarouge). En 2009, Lu et al. [75] ont développé un MSFA en se basant sur le réseau de filtres colorés pour capturer à la fois les images du visible et du proche infrarouge. En 2022, Sawyer et al. [76] ont développé une boîte à outils pour la conception et l'optimisation généralisées de MSFA. Wu et al. [77] ont proposé des MSFA optimisés pour la reconstruction spectrale. Cette méthode de conception de MSFA est basée sur la représentation clairsemée, elle utilise les informations spectrales de l'image et optimise la sélection et la disposition spatiale des filtres. Cette approche est basée sur la recherche heuristique globale pour optimiser les MSFA tout en minimisant l'erreur estimée des MSFA. L'état de l'art montre que la conception d'un MSFA est spécifique à l'application en termes de nombre de bandes, de transmittances spectrales, ainsi que de la disposition géométrique. Son

coût de fabrication résulte de tous ces spécificités. Certains travaux de recherche montrent la conception et la fabrication du système multispectral.

Un MSFA à 23 bandes a été conçu et fabriqué en 2012 par Walls et al.[78]. Cet MSFA comporte 23 bandes de filtres Fabry-Pérot à bande étroite avec des FWHM (full-width half-maximums) de 22-46 nm, couvrant la gamme spectrale de 400 à 750 nm. Le MSFA fabriqué peut être intégré sur les capteurs d'image CMOS. En 2011, Gupta et al. [79] ont développé un autre capteur multispectral instantané qui utilise l'interféromètre de Fabry-Pérot. Le MSFA intégré dans le système est composé de 16 bandes avec un moxel 4x4. C'est un système qui couvre l'infrarouge à courte longueur d'onde de 1487 à 1769 nm avec une bande passante spectrale d'environ 10 nm.

Geelen et al. ont développé deux imageurs multispectraux [80], [81] respectivement en 2013 et en 2014. Le premier imageur intègre un capteur CMOS, les filtres en mosaïque et une duplication optique. Ce prototype d'imageur permet d'acquérir des images multispectrales de 256x256 pixels sur 32 bandes avec une vitesse d'environ 30 images par seconde dans des conditions de lumière du jour et jusqu'à 340 images par seconde dans des conditions typiques de l'environnement. Il couvre la gamme spectrale de 600-1000 nm. Le deuxième imageur multispectral instantané conçu en 2014 est basé sur des filtres optiques intégrés de manière monolithique au-dessus d'un capteur d'image CMOS standard. Le dépôt monolithique permet une grande souplesse de conception en tenant compte la résolution spatiale et spectrale.

En 2022, Mohammadi et al. [82] ont développé une caméra multispectrale instantanée équipée d'un MSFA pour l'agriculture utilisé pour la détection des mauvaises herbes ou l'estimation de la maturité des fruits. L'architecture matérielle de cet imageur est composée d'un MSFA de huit filtres et d'un capteur CMOS. La théorie de Fabry-Perot a été implémentée pour concevoir et fabriquer leur MSFA. Ce prototype de caméra couvre les bandes spectrales de 400 à 700 nm.

Les caméras instantanée MSFA produisent des images bruitées ou des mosaïques d'images. Chaque filtre est responsable de l'acquisition de l'image dans une bande donnée. Une estimation des pixels manquants dans une bande donnée est alors nécessaire pour reconstituer l'image dans cette bande. Ce processus est alors appelé demosaicing.

L'utilisation des différents types de systèmes d'acquisition en RF conduit à l'implémentation de la RF multispectrale.

1.4 Reconnaissance faciale multispectrale

De nombreuses études ont montré que les systèmes de RF dans le spectre visible présentent des limites. Au nombre de ces problèmes on peut citer : la reconnaissance des vrais jumeaux, la variation de l'âge, de pose, d'illumination et de la morphologie du visage, etc. La RF multispectrale constitue une alternative de résolution de ces problèmes. Au cours du développement des systèmes de RF, maintes bases d'images sont construites, certaines d'entre elles sont mises à disposition des chercheurs pour d'autres recherches.

1.4.1 Base d'images multispectrales

De nombreuses bases de données d'images multispectrales de face ont été mises en place pour les recherches en RF multispectrale. Ces bases de données multispectrales sont composées d'images acquises dans le visible et l'infrarouge. Plusieurs bases publiques d'images multispectrales sont présentées dans la littérature. Le [Tableau 1.4](#) présente quelques-unes de ces bases.

Tableau 1.4- Quelques bases publiques d'images multispectrales

| Base de données | Bandes spectrales |
|-------------------------------|-----------------------|
| CASIA-HFB | VIS, NIR |
| CASIA NIR-VIS 2.0 | VIS, NIR |
| Oulu-CASIA | VIS, NIR |
| NIVL | VIS, NIR |
| HK PolyU-HSFD | VIS, NIR |
| NIST Equinox | VIS, SWIR, MWIR, LWIR |
| Carl dataset | VIS, NIR, LWIR |
| BUAA-VisNir | VIS, NIR |
| Polarimétrie thermique | VIS, LWIR |
| UWA-HSFD | VIS, NIR |
| WVU NIR | VIS, NIR |
| UGCU | VIS, LWIR |
| ASSUMS | VIS, LWIR |
| RGB-D-T | VIS, LWIR |

1.4.2 *Système de reconnaissance faciale multispectrale*

Les systèmes de RF multispectrale sont basés sur des méthodes et combinaison de méthodes citées pour la RF. Ils peuvent être regroupés en quatre catégories en fonction des méthodes qu'ils implémentent : la fusion, l'apprentissage profond (Réseau neuronal profond et synthèse d'images), les méthodes de représentation des caractéristiques et les méthodes d'apprentissage par sous-espace couplé.

- Les *méthodes de fusion* sont un ensemble de techniques appliquées à la RF multispectrale. Il en existe trois types qui sont la fusion de pixels, la fusion de caractéristiques et la fusion de scores. La fusion de caractéristiques combine les informations pertinentes issues des différentes bandes d'images par le biais d'un extracteur de caractéristiques. La fusion de scores regroupe les scores issus de la classification (ou de la phase de correspondance) des différentes bandes d'images. Divers travaux de recherche ont utilisé les méthodes de fusion pour la reconnaissance faciale hétérogène.

- Les *méthodes d'apprentissage profond* regroupent les méthodes basées sur les CNN et les GAN. Ces méthodes de synthèse d'image transforment l'image d'une bande spectrale en une image d'une autre bande spectrale dans le but de les comparer. Par exemple une image du visible peut être synthétisée en une image infrarouge. Ceci permet de régler les problèmes d'appariement.

- Les *méthodes de représentation de caractéristiques* combinent les extracteurs de caractéristiques et les classifieurs. Les caractéristiques sont extraites dans chacune des bandes d'images, puis combinées et réduites avant d'être envoyées au classifieur.

- Les *méthodes d'apprentissage par sous-espace couplé* projettent les caractéristiques des différentes bandes spectrales dans un sous-espace commun. Ce dernier identifie les informations les plus pertinentes, en utilisant les redondances communes aux différentes bandes spectrales.

Les différentes méthodes de RF multispectrales seront présentées dans le [Chapitre 3](#) de cette thèse. Dans le but de comparer l'efficacité des systèmes de RF, des métriques d'évaluation de performance sont définies.

1.4.3 *Evaluation de la performance des systèmes de reconnaissance*

Idéalement un système biométrique en particulier un système de RF doit être robuste, efficace et capable de détecter les fraudes quel que soit les circonstances. Il est nécessaire d'utiliser des métriques pour comparer cette robustesse et l'efficacité ou la performance de ces

différents systèmes de RF. Parmi ces métriques d'évaluation de performance, les plus utilisées sont : le rang-N, le taux de fausses acceptations (FAR), le temps de calcul utilisé par l'algorithme et les métriques spécifiques aux méthodes d'IA telles que la précision, le recall, le F-score. La performance de l'appariement ou la correspondance est mesurée comme le pourcentage de tentatives d'identification pour lesquelles la prédiction de l'image de visage est renvoyée dans les N premiers résultats classés. Le rang 1 correspond alors au pourcentage d'identités prédites dont la correspondance est correcte (résultat le plus élevé). Le rang-10 correspond au pourcentage de prédictions d'images de visages qui correspondent correctement à leurs identités équivalentes dans les 10 premiers résultats les mieux notés. Le rang 1 est calculé en divisant le nombre total d'images correctement identifiées (Correctement identifiées) par le nombre total d'identifications effectuées (Tentatives d'identification). La Fausse Acceptation est un événement qui a lieu lorsqu'un système biométrique accepte une personne alors qu'elle n'est pas dans sa base d'utilisateurs. Le TFA (Taux de Fausse Acceptation) est la probabilité qu'un utilisateur inconnu soit identifié comme étant un utilisateur connu. Ce taux définit la sécurité du système biométrique. Le TFR (Taux de Faux Rejet) est la probabilité qu'un utilisateur connu soit rejeté par le système biométrique. Le temps de calcul de l'algorithme permet de mesurer aussi la performance du système dans le cas où les systèmes à comparer ont des rangs similaires, avec des valeurs de FAR fixes.

Vue que la solution proposée dans cette thèse utilise l'IA les métriques d'évaluation relatives à l'IA seront bien décrites dans les autres chapitres.

Conclusion

Dans ce chapitre, nous avons présenté les généralités sur les systèmes biométriques en particulier sur la reconnaissance faciale. L'intérêt de l'imagerie multispectrale et les différents types de systèmes d'acquisition multispectrale ont été décrits avec une précision sur la conception et le fonctionnement général des caméras MSFA one-shot. Les systèmes de RF existants ont été présentés dans ce chapitre. Aussi, un aperçu sur les méthodes de reconnaissance est résumé avec les différentes métriques d'évaluation de performance. Cet état de l'art a permis de constater que cette étude est la première sur le développement d'une caméra MSFA one-shot pour la RF. La reconnaissance faciale multispectrale est un défi surtout en ce qui concerne le développement d'une caméra MSFA one shot pour la reconnaissance faciale. Les chapitres 2, 3, 4 et 5 présenteront les outils nécessaires au développement d'une caméra MSFA one-shot pour la reconnaissance du visage.

Chapitre 2

Construction d'une base de données multispectrale avec une caméra MSFA one-shot

Sommaire

| | |
|---|----|
| Introduction | 32 |
| 2.1 Imagerie multispectrale pour la reconnaissance faciale | 32 |
| 2.1.1 Motivation | 32 |
| 2.1.2 Bases d'images utilisées en reconnaissance faciale multispectrale | 33 |
| 2.2 Système d'acquisition MSFA one shot | 37 |
| 2.2.1 Imagerie multispectrale à filtre MSFA | 37 |
| 2.2.2 Description de la caméra EXIST | 39 |
| 2.2.2.1 Capteurs CMOS | 40 |
| 2.2.2.2 Conception et fabrication des filtres | 41 |
| 2.3 Protocole de collecte et description de la base d'images EXIST | 45 |
| Conclusion | 47 |

Introduction

Les systèmes biométriques sont de plus en plus utilisés dans le contrôle d'accès aux objets (smartphones...) et aux lieux publics et privés. Le visage constitue l'une des modalités biométriques couramment utilisées à cause de son acquisition non intrusive. L'acquisition et l'utilisation de toute données biométriques nécessite l'accord préalable de l'utilisateur en particulier les images du visage. En fonction du type de caméras les conditions d'acquisition doivent être les mêmes lors de la phase d'inscription et celle d'identification/authentification. C'est une étape très importante dans la conception des systèmes de reconnaissance. La qualité des images impacte la performance du système. Il faut alors trouver un bon compromis entre le système d'acquisition et les méthodes de reconnaissance pour implémenter un système très performant. Les images acquises lors de l'inscription sont stockées dans une base pour les traitements ultérieurs. Les bases de données contiennent alors des informations qui permettent d'évaluer les méthodes développées ainsi donc la performance des systèmes de reconnaissance. Ces informations peuvent être les images ou des vecteurs caractéristiques extraites de ces images. La plupart des bases de données en reconnaissance faciale multispectrale contiennent des images de faces acquises sur une ou plusieurs spectres avec des systèmes d'acquisition multi-capteurs ou en mode scanning. Il existe plusieurs bases d'images de faces MS utilisées en reconnaissance faciale mais chacune d'elles présentent des spécificités.

Dans ce chapitre nous présentons la construction de EXIST, une nouvelle base d'images multispectrales de faces pour la RF avec les caméras MSFA one-shot. Cette base est mise en place avec une caméra MSFA qui opère dans le visible et le proche infrarouge.

2.1 Imagerie multispectrale pour la reconnaissance faciale

2.1.1 Motivation

La variation de la luminosité est l'un des problèmes auxquels les systèmes de RF sont confrontés. Les variations de condition d'éclairage détériorent les performances des systèmes de reconnaissance. L'imagerie MS se présente comme une alternative à l'imagerie traditionnelle utilisant des images couleurs. Les capteurs en imagerie multispectrale acquièrent les données d'image à des longueurs d'ondes spécifiques du spectre électromagnétique. La [Figure 2.1](#) présente le spectre électromagnétique.

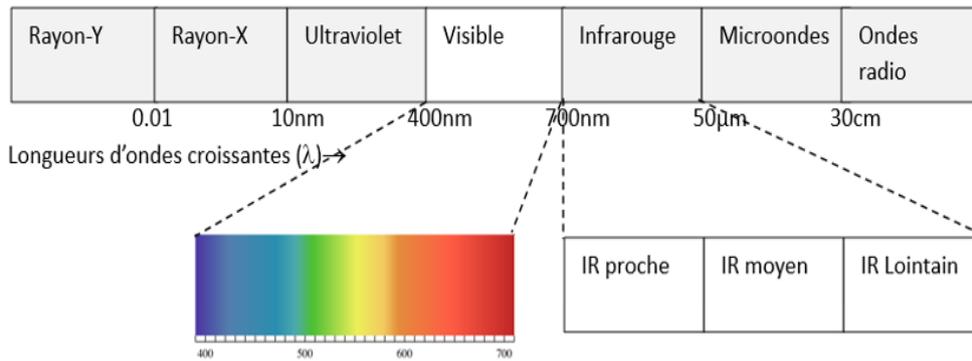


Figure 2.1-Spectre électromagnétique

Comme l'indique la figure 2.1, le spectre électromagnétique est subdivisé en cinq domaines : les rayons-X, les rayons-Y, l'ultraviolet, le visible, l'infrarouge, les microondes et les ondes radio. Les spectres du visible et de l'infrarouge sont utilisés en RF. La lumière visible utilise une petite partie du spectre électromagnétique et occupe la bande de longueur d'onde comprise entre 400 et 700nm. Le domaine infrarouge se subdivise en proche infrarouge (770-1400 nm) ou Near Infrared(NIR) , infrarouge de courte longueur d'onde (1,4-3 µm) ou Short Wavelength Infrared(SWIR), infrarouge de longueur d'onde moyenne (3-8 µm) ou Mid Wave infrared (MWIR) et infrarouge thermique (8-15 µm) ou Long Wavelength Infrared(LWIR).

Une image MS est une image acquise sur N bandes spectrales avec $4 \leq N \leq 10$. La littérature présente de nombreuses bases d'images MS utilisées pour tester les méthodes de RF multispectrale dans le Visible et NIR.

2.1.2 Bases d'images utilisées en reconnaissance faciale multispectrale

Le premier module d'un système de reconnaissance est le module d'acquisition. La performance d'un système de reconnaissance en dépend fortement. Les images acquises avec les caméras sont stockées dans les bases de données lors de la phase d'enrôlement. La plupart des bases d'images multispectrales de faces est collectée avec des caméras multispectrales en mode scanning ou d'un système d'acquisition constitué de plusieurs capteurs. La revue de littérature présente plusieurs bases d'images de visages, elles sont dites publiques mais vu le caractère personnel des images, ces bases ne sont accessibles qu'après la signature d'un protocole d'autorisation et d'utilisation entre le chercheur et l'auteur de la base d'images. Parmi les bases d'images de faces multispectrales nous pouvons citer :

La base de données multispectrales ASUMS : elle est composée d'images de faces du visible et du thermique. Elle a été collectée par l'Université d'État d'Alcorn [83] avec le système d'acquisition FLIR SC620 deux en un, dans lequel la caméra IR a une résolution originale de 640×480 pixels avec une plage spectrale de 7,5 à 13µm et la caméra pour le visible a une résolution originale de 2048×1536 pixels. Cette base de données est composée d'images de 96 personnes.

La base de données de visages BUAA-VisNir a été collectée par D. Huang et al. [84]. Elle contient 162 images/personne de 150 personnes. Les images ont été acquises dans le visible et le NIR avec 9 expressions faciales différentes.

V. Espinoza et al. [85], [86] ont utilisé un système d'acquisition multispectrale composé de deux caméras pour collecter les images multispectrales du visible, du proche infrarouge et du thermique de la base de données CARL. La caméra thermographique TESTO 880-3 dotée d'un détecteur non refroidi avec une plage de sensibilité spectrale de 8 à 14 µm a été utilisée pour capturer les images du visible de taille 640×480 et celles du thermique de taille 160×120 pixels. Les images infrarouges de taille 640×480 ont été acquises avec une caméra Logitech Quickcam messenger personnalisée E2500 doté un capteur d'image CMOS au silicium. Cette base de données est constituée de 7 380 images multispectrales de 41 personnes.

La base de données de visages CASIA(Chinese Academy of Sciences) HFB [87] est composée d'images du visible, du proches infrarouges et tridimensionnelles (3D). Elle a été collectée par deux caméras : la photo Canon A640 pour le visible et celle du proche infrarouge par une caméra conçue par CASIA pour la cause. CASIA HFB est composée de 992 images de taille 640 × 480 provenant de 100 sujets.

La base de données CASIA(Chinese Academy of Sciences) NIR-VIS 2.0 a été collectée de 2007 à 2010 par Stan Z. Li et al. [88], elle contient des images frontales du visible et du proche infrarouge de 735 sujets. Deux caméras ont été utilisées pour acquérir ces images. Les images couleur du visage sont capturées à l'aide de la caméra Canon A640 et une caméra fait maison avait été utilisé pour l'acquisition des images du proche infrarouge. Un filtre optique passe longue est intégré dans cette caméra et permet de capturer des images dans longueurs d'onde 720, 800, 850 et 880 nm. Cette base de données est constituée des images de taille 640×480 et des images recadrées de taille 128×128. La base de données CASIA NIR-VIS 2.0 est la plus accessible et utilisée dans les travaux de recherches relatifs à la RF multispectrale.

Aijing Yu et al. [89] ont proposé la base de données LAMP-HQ qui est une base d'images multispectrales du visible et du proche infrarouge. Elle contient 56 788 images du proche infrarouge et 16 828 images du visible de 573 sujets. La base de données LAMP-HQ a été collectée avec deux caméras : Canon-7D et AuthenMetric-CE31SE pour acquérir les images VIS et NIR respectivement. Les images sont de taille 960×720 .

La base de données Oulu-CASIA NIR&VIS a été mise en place par l'Académie chinoise des sciences. Elle contient des vidéos avec six expressions typiques : à savoir le bonheur, la tristesse, la surprise, la colère, la peur et le dégoût de 80 sujets. Les images de la base de données Oulu-CASIA NIR&VIS [90] ont été capturés avec deux systèmes d'imagerie (SN9C201 & 202) qui combinent une caméra NIR et une caméra visible. Ce système d'imagerie fonctionne à une fréquence de 25 images par seconde et capture des images NIR et visibles dans trois conditions d'éclairage différentes : éclairage intérieur normal, éclairage faible et éclairage sombre. La résolution de l'image acquises est de 320×240 pixels.

La base de données HK PolyU [91] de l'université Polytechnique de Hong Kong est constitué de cubes d'images hyper-spectrales de 48 sujets, qui ont été acquises à l'aide d'un VariSpec à filtre accordable à cristaux liquides (LCTF) sous une lumière halogène. La gamme spectrale s'étend de 400 à 720 nm avec un pas de 10 nm.

La base de données CMU (Carnegie Mellon University) a été collectée avec un prototype de caméra spectropolarimétrique développé par CMU. Elle contient les images de 54 sujets. La gamme hyper-spectrale est comprise entre 450 et 1090 nm avec une longueur de pas de 10 nm.

La base de données UWA [92] se compose de 120 cubes d'images hyperspectrales de 70 sujets acquis avec le VariSpec LCTF CRIs intégré à une caméra à focalisation photonique. Chaque cube d'image hyperspectrale contient 33 bandes couvrant la gamme spectrale de 400 à 720 nm avec un pas de 10 nm.

J.Bernhard et al. [93] ont publié la base de données Near Infrared Visible Light Database (ND-NIVL). C'est un ensemble de données composée d'images du visible et du NIR collectées avec deux caméras. Les images du visibles de taille 4288×2848 ont été capturées avec la caméra a Nikon D90, et celles de l'infrarouge de taille 4770×3177 avec la caméra infrarouge CFAIRS. ND-NIVL contient 2,341 VIS images and 22,264 NIR images de 574 sujets.

La base de données NVESD [94] est composée d'images du visible et du thermique. Ces images ont été acquises avec trois caméras : la caméra Basler Scout sc640 pour l'acquisition des images (659×494) du visible, et les caméras thermiques DRS MWIR et

DRS LWIR pour capturer respectivement les images (640×480) de l'infrarouge moyen et lointain. La base d'images NVESD est composée d'images de 50 personnes.

Polarimetric dataset [95] est une base d'images visibles et thermiques. L'imageur polarimétrique LWIR et une caméra visible ont été utilisés pour collecter les images respectivement du visible et du NIR de 111 personnes.

O. Nikisins et al. [96] ont mis en place la base d'images RGB-D-T qui est composée d'images couleurs, d'images de profondeur et d'images thermiques. RGB-D-T a été collectée avec la caméra Kinect RGB pour les images du visible, de profondeur et la caméra thermique AXIS Q1922 pour les images thermiques. Au total, elle est constituée de 45900 images de face provenant de 51 personnes. Les résolutions des images RVB, de profondeur et thermiques sont respectivement de 640×480 , 640×480 et 384×288 pixels.

NIST Equinox database [97] est composée d'images du visible, d'images infrarouge à ondes courtes, d'images infrarouges à onde moyennes et d'images infrarouge thermique appartenant à 95 personnes.

S.HU et al. [98] ont publié la base de données Polarimetric Thermal qui contient les images du visible et du thermique de 60 personnes. Le système d'acquisition est composé d'un imageur polarimétrique pour l'infrarouge à ondes longues et deux caméras pour l'acquisition des images du visible.

Le tableau 2.1 présente le résumé des bases d'images les plus utilisées en reconnaissance faciales.

Tableau 2.1- Quelques bases d'images multispectrales

| Base de données | Références | Bandes spectrales | Nombre de personnes |
|--------------------------|------------|--------------------------|---------------------|
| CASIA-HFB | [87] | VIS, NIR | 100 |
| CASIA NIR-VIS 2.0 | [87] | VIS, NIR | 725 |
| Oulu-CASIA | [90] | VIS, NIR | 80 |
| NIVL | [93] | VIS, NIR | 574 |
| HK PolyU-HSFD | [90] | VIS, NIR | 25 |
| NIST Equinox | [97] | VIS, SWIR, MWIR, LWIR | 95 |
| Carl dataset | [85], [86] | VIS, NIR, LWIR | 41 |

| | | | |
|-------------------------------|------|-----------|-----|
| BUAA-VisNir | [84] | VIS, NIR | 150 |
| Polarimétrie thermique | [95] | VIS, LWIR | 60 |
| UWA-HSFD | [92] | VIS, NIR | 70 |
| WVU NIR | | VIS, NIR | 103 |
| UGCU | [99] | VIS, LWIR | 84 |
| ASSUMS | [83] | VIS, LWIR | 96 |
| RGB-D-T | [96] | VIS, LWIR | 51 |
| Polarimetric dataset | [95] | VIS, LWIR | 111 |
| CASIA-HFB | [87] | VIS, NIR | 100 |

2.2 Système d'acquisition MSFA one shot

Ces dernières années, des efforts de recherche soutenus ont été réalisés dans le domaine des systèmes d'imagerie multispectrale intégrant les MSFA. Cette section présente l'imagerie multispectrale à filtre MSFA et la description du système d'acquisition one shot utilisé.

2.2.1 Imagerie multispectrale à filtre MSFA

La technologie des matrices de filtres colorés ou CFA a été largement utilisée dans l'industrie pour la fabrication des appareils photo numériques et des caméras car elle offre plusieurs avantages, tels qu'un faible coût de production, l'enregistrement exact et une forte capacité d'adaptation. L'imagerie couleur a donc été marquée par l'intégration de ces réseaux de filtres colorés [100] ou filtres de Bayer dans les caméras pour l'acquisition des images couleurs. Les CFA sont constitués de trois filtres de couleur primaire telle que rouge, vert et bleu. La [Figure 2.2](#) présente le CFA.

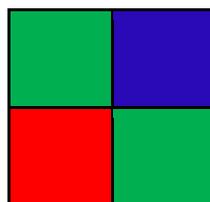


Figure 2.2- Filtre de Bayer (CFA)

L'impact de l'utilisation du CFA dans l'imagerie couleur a suscité l'intérêt des universitaires et des industriels pour étendre le CFA au MSFA. Un CFA a pour but de récupérer les informations relatives à la couleur, alors qu'une MSFA a pour objectif d'évaluer les propriétés de l'objet et/ou la mesure objective de la couleur. Le nombre de bandes dans une MSFA peut varier, et le choix des filtres peut être assez spécifique à l'application. Ces dernières années, des efforts de recherche soutenus ont été réalisés dans le domaine des systèmes d'imagerie multispectrale intégrant les MSFA [101], [102]. L'imagerie multispectrale utilisant un seul capteur avec MSFA est un moyen efficace d'acquérir des données spectrales et semble être la solution la plus adaptée pour les applications de vision temps réel sur les caméras embarquées. Elle a le potentiel de promouvoir un système d'imagerie multispectrale rapide et temps réel. L'imagerie multispectrale basée sur les MSFA est une technique compacte, robuste, pratique pour la capture et la reconstruction d'images spectrales instantanées. La plupart des MSFA est généralement conçue de manière ad hoc. La qualité des images et leur reconstruction dépendent fortement des sensibilités spectrales et de la disposition spatiales des canaux. L'influence des fonctions de sensibilité spectrale des canaux, la distribution spatiale des canaux et la méthode de reconstruction sont fortement corrélées lors de la conception d'un MSFA optimal. Il faut alors trouver un compromis entre ces trois éléments. Généralement, la conception d'un MSFA commence par la conception des filtres donc la définition d'un ensemble optimal de filtres. Pour une conception de MSFA optimal, certains travaux de recherche [77] optimisent les fonctions de sensibilité spectrale des canaux en se basant sur la fonction de transmittance théorique, d'autres optimisent soit les fonctions de sensibilité spectrale des canaux ou la disposition des canaux sur un motif périodique en mosaïque.

Ramanath et al. [103] ont présenté pour la première fois une extension du CFA pour l'acquisition des images MS. Leur MSFA était constitué de sept types d'éléments filtrants qui sont disposés en hexagone, de façon que chaque pixel d'une bande spectrale donnée soit entouré des pixels des six autres bandes. En général, un MSFA est constitué de motifs récurrents d'éléments filtrants. Chacun d'eux est sensible à une bande spectrale spécifique. Contrairement au CFA qui sont constitués de trois filtres, les MSFA contiennent quatre ou N filtres. Lors de la conception technique des caméras, les filtres sont soigneusement sélectionnés. L'architecture de la caméra MSFA one-shot montre que le MSFA est posé sur le capteur de la caméra de manière à le recouvrir. La lumière qui entre dans la caméra est donc filtrée par des filtres spectraux passe-bande sur chaque pixel.

Un MSFA peut être défini par son moxel qui est l'élément de mosaïque correspondant à l'occurrence d'un motif prédéfini et constitué d'un ensemble de filtres disposés géométriquement de manière relative [102]. Les moxels ou pixels multispectraux sont les plus petits motifs de la MSFA. Le système d'imagerie multispectrale avec MSFA est composé essentiellement des filtres MSFA, le capteur et d'une micro-lentille. Une vue d'ensemble de l'approche globale est présentée dans les figures 2.3 et 2.4.

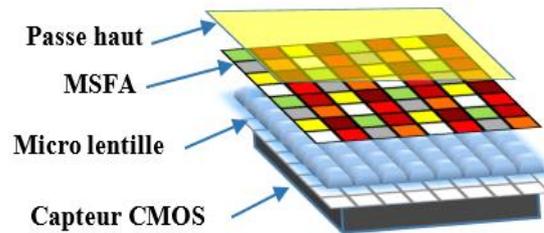


Figure 2.3- Schéma global du système d'imagerie MS avec MSFA.

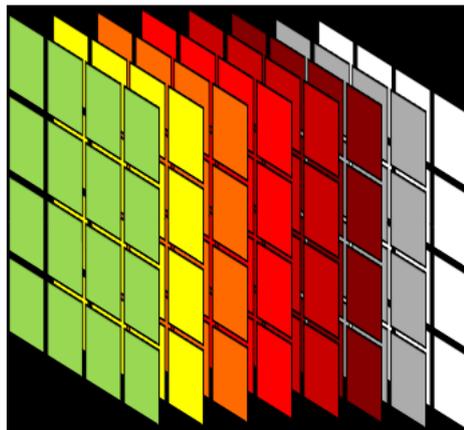


Figure 2.4 Bandes d'images spectrales

La réalisation du MSFA n'est aisée comme celle du CFA, elle est plein développement par la communauté industrielle. Parmi d'autres difficultés, la production et la fabrication de MSFA représentent un défi technique majeur. Des efforts soutenus ont donc été consacrés à la réalisation du MSFA. Une caméra MSFA one shot a été utilisée durant cette étude. La section suivante présente la description de cette caméra.

2.2.2 Description de la caméra EXIST

Nous avons utilisé une caméra one shot qui intègre un MSFA pour construire la base d'images MS pour la RF. Ce système d'acquisition one-shot est développé en collaboration

avec Silios [104] et le laboratoire d'Electronique, d'Informatique et d'Image (LE2I), actuellement Imagerie et Vision Artificielle(ImViA) dans le cadre du projet UE H2020 appelé EXIST (EXtended Image Sensing Technologies)[105]. L'objectif principal de ce prototype est de montrer l'efficacité de la nouvelle technologie de caméra au niveau de la biométrie en particulier la reconnaissance faciale. Il s'agit d'un compromis entre la résolution spatiale et la résolution spectrale. Notre caméra MSFA oneshot présente les caractéristiques suivantes :

- Application en temps réel, acquisition d'images MS en une seule fois ;
- Système compact, une simple combinaison d'un filtre et d'un capteur monochromatique (Équivalent à un appareil photo standard de l'utilisateur final) ;
- Robuste aux changements de lumière, le système multispectral tend à mesurer les propriétés physiques de la surface de l'objet. ;
- Large bande, le système couvre à la fois la zone visible et la zone NIR ;
- Système d'application à faible coût, le système est basé sur un capteur standard 2k couplé à des outils logiciels, ce qui réduira les coûts ;
- Système léger, très rapide et capable de capturer les scènes en mouvement.

La structure de notre caméra MSFA se compose des éléments suivants :

- Capteur CMOS ;
- Micro-lentille ;
- MSFA ;
- Carte électronique ;
- Carte caméra.

La carte électronique permet de piloter le capteur et la carte caméra sert à l'acquisition d'images. Le système d'acquisition MSFA est relié à l'ordinateur et piloté par un logiciel d'acquisition d'images.

2.2.2.1 Capteur CMOS

Le système d'imagerie multispectral one shot utilisé est équipé du capteur CMOS (Complementary Metal Oxide Semiconductor) Viimagic 9220H développé par Grass Valley. Un capteur CMOS est l'élément physique dont les performances ont un impact sur la qualité du système d'acquisition. L'avantage d'utiliser les capteurs CMOS est que le coût de leur fabrication est beaucoup moins élevé que celui des capteurs CCD (Charged coupled Device). En outre, les capteurs CMOS consomment moins d'énergie et permettent une acquisition plus

rapide de l'image [106], [107]. Ils offrent une résolution plus élevée, des cadences d'images supérieures que celui des CCD [108]. La facilité d'accès aux pixels disponibles dans les capteurs CMOS permet une grande flexibilité pour le traitement des données en temps réel. Tous ces avantages cités permettent d'obtenir des systèmes d'acquisition qui consomment moins d'énergie. Le choix de ce capteur a été fait en fonction des critères et contraintes imposés par la technologie de conception et de fabrication des filtres qui est COLOR SHADES. Parmi ces critères et contraintes on peut citer :

- La taille minimale des pixels est de 5 μm . Il s'agit d'une contrainte imposée par la technologie du filtre COLOR SHADES ;
- La résolution du capteur CMOS doit être suffisamment élevée pour compenser la perte liée au système MSFA ;
- La sensibilité spectrale du capteur doit être étendue au proche infrarouge.

La Figure 2.5 illustre le capteur CMOS et sa carte de pilotage.

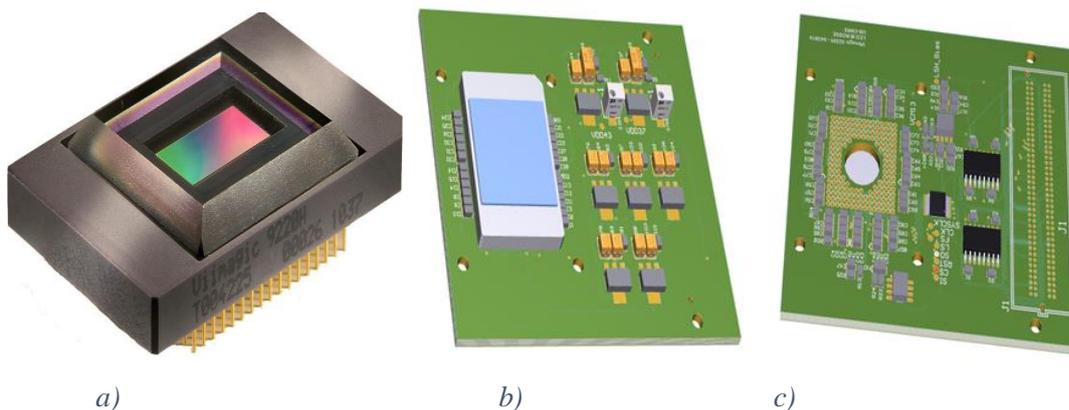


Figure 2.5- a) capteur CMOS Viimagic 9220H, b) carte de pilotage face A, c) carte de pilotage face B

Afin de corriger la linéarité du capteur et de mesurer la sensibilité spectrale, une caractérisation ou calibration du capteur CMOS a été effectuée avant de monter le réseau de filtres multi-spectraux. La calibration permet aussi d'identifier et d'éliminer les divers bruits qui existent. Le MSFA a été sélectionné avec soin en considérant une distribution régulière des pixels dans le moxel.

2.2.2.2 Conception et fabrication des filtres

L'extension du CFA au MSFA n'est pas simple. La plupart des MSFA existants sont généralement conçus de manière ad hoc. La conception des MSFA doit commencer par un algorithme de conception de filtre, il faut d'abord une sélection et une optimisation des filtres avant leur fabrication. Cependant, ces filtres optimisés sont souvent basés sur une fonction de

transmittance théorique simple, c'est-à-dire des courbes gaussiennes ou basées sur des gaussiennes. L'objectif est de concevoir et de fabriquer les filtres optimaux avec une balance d'énergie équilibrée. Pour ce faire, une sélection de filtres optimaux est d'abord réalisée en choisissant les filtres correspondants aux longueurs d'onde. Cette optimisation des filtres est basée sur un algorithme génétique qui fournit les meilleures bandes pour une application de l'utilisateur final qui est la reconnaissance faciale. Le pipeline de reconnaissance de visage développé pour la sélection des filtres optimaux est constitué des étapes suivantes : la préparation de la base de données, le calcul des images et l'ajout de bruit, l'extraction des caractéristiques, la détermination du score de similarité et l'analyse des données pour la sélection du filtre.

Notre réseau de filtres personnalisé a été construit et fabriqué en collaboration avec SILIOS technologies. La conception et la fabrication de ces filtres multispectraux sont basées sur la technologie exclusive COLOR SHADES de SILIOS. En effet, SILIOS Technologies a développé la technologie COLOR SHADES, qui utilise l'interféromètre Fabry-Perot pour fabriquer des filtres de transmittance multispectraux. Ceci permet d'obtenir une forme gaussienne pour chaque filtre. Cette technologie est basée sur la combinaison de procédés de dépôt de couches minces et de micro/nano gravure sur un substrat de silice fondue. Des étapes standards de micro-photolithographie sont utilisées pour définir la géométrie des cellules du filtre multispectral. A l'origine COLOR SHADES fournit des filtres passe-bande dans la gamme visible de 400 nm à 700 nm. SILIOS a développé des filtres dans la gamme NIR en collaboration avec le laboratoire LE2I (ImViA). Les résultats de recherche relative à la sélection des filtres optimaux ont été combinés avec la technologie classique d'interférence de film mince pour réaliser notre réseau de filtres. Cette technologie permet de choisir la gamme de longueurs d'onde, le nombre de longueurs d'onde (entre 2 et 16) et la taille de chaque cellule de filtre (jusqu'à la taille de pixel de l'imageur). Chaque filtre élémentaire est un filtre passe-bande dont la longueur d'onde centrale, la largeur spectrale et le niveau de transmission peuvent être fixés et adaptés sur demande.

L'intérêt principal de cette technologie est que les MSFA peuvent être conçus pour n'importe quelle géométrie d'imageur CMOS ou de détecteurs pixelisés au silicium. Le système MSFA a été intégré dans notre caméra avec des calculs matériels et logiciels dédiés, il permet de travailler en application temps réel avec 30 fps.

La [Figure 2.6](#) illustre la réponse spectrale des filtres avant leur hybridation sur le capteur.

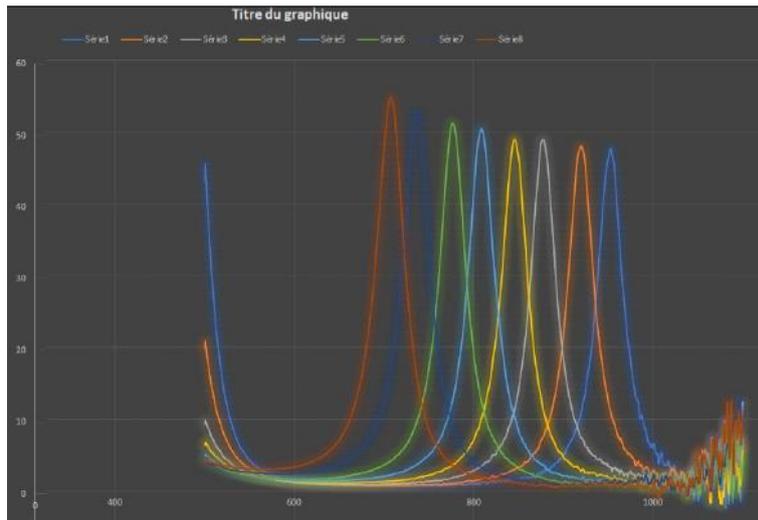


Figure 2.6- Réponse spectrale des huit filtres avant hybridation sur le capteur

Après la conception et la fabrication des filtres, l'hybridation du capteur a été faite par la superposition du réseau de filtres sur ce dernier. Afin d'évaluer l'impact de ce montage, la réponse spectrale du capteur hybride a été calculée. La Figure 2.7 présente la réponse spectrale du capteur hybride.

La réponse spectrale du réseau de filtres est estimée en considérant tout le capteur d'imagerie MS. La procédure de calibration se présente comme suit :

- La mesure de l'énergie générée par le monochromateur de 350 à 1100 nm
- Le système génère une lumière monochromatique tous les 5 nm de 350 à 1100 nm. Le capteur CMOS original (sans filtre monté) est donc éclairé. Le temps d'intégration de la caméra est fixé à une valeur constante. La réponse spectrale du capteur est ainsi mesurée.
- La deuxième mesure est effectuée avec le capteur d'imagerie MS. Un traitement d'image est effectué pour extraire de chaque moxel le pixel lié à un filtre spécifique.
- La réponse spectrale du filtre est obtenue à partir de la réponse du capteur CMOS.

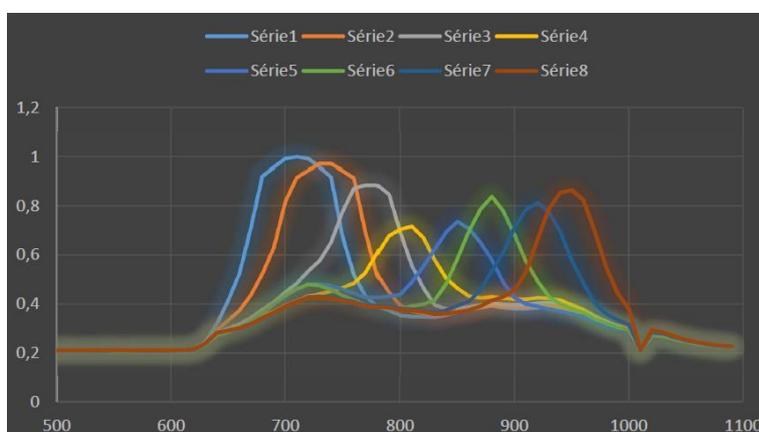


Figure 2.7- Réponse spectrale du capteur hybride

Les filtres utilisés permettent de surmonter les problèmes causés par les variations d'éclairage, de flou, de mouvement et le bruit SNR qui affectent sévèrement la performance des systèmes de reconnaissance faciale utilisant le CMOS.

Le MSFA conçu est composé de 8 filtres optimaux sélectionnés dans les longueurs d'onde respectivement de {685, 720, 770, 810, 835, 870, 895, 930} (en nm). Le système MSFA intégré dans notre système d'acquisition a un petit moxel 4x4 avec 2 pixels par bande et la taille du pas de filtre est de $5 \times 5 \mu\text{m}^2$.

La [Figure 2.8](#) présente la distribution spatiale du moxel final.

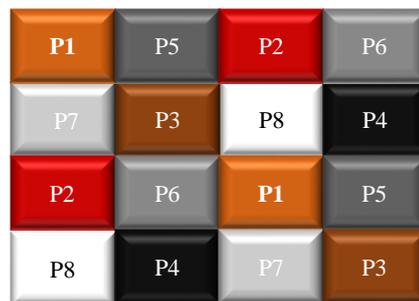


Figure 2.8- Moxel du MSFA

Ce système d'acquisition est un système multispectral à large bande, robuste et léger qui tend à mesurer les propriétés physiques de l'objet. Cette caméra est un système temps réel qui couvre les spectres du visible et du proche infrarouge, elle opère de 650 à 950 nm. À chaque acquisition, la caméra MSFA one shot fournit une meilleure résolution de l'image brute ou mosaïque de taille 2072 x 1104 pixels, dans laquelle les valeurs de chaque canal sont accessibles à chaque pixel selon le MSFA. Les valeurs manquantes des pixels sont ensuite estimées par le processus de demosaïcing. Afin de fournir une solution optimale à la perte de résolution spatiale inhérente au MSFA, des algorithmes spécifiques ont été développés pour le demosaïcing multispectral.

La [Figure 2.9](#) présente la caméra MSFA décrite ci-dessus.



Figure 2.9- Caméra MS one shot

2.3 Protocole de collecte et description de la base d'images EXIST

Le protocole de collecte de données utilisé pour acquérir la base de données EXIST a été validé par l'équipe de direction de la thèse. Les images multispectrales de visages ont été collectées pendant trois ans dans le laboratoire Imaging and Artificial Vision (ImViA) de la Faculté des Sciences et Technologies de l'Université de Bourgogne en France. La salle d'acquisition est une salle noire, dédiée à l'imagerie MS. Un lampadaire équipé d'une ampoule fluorescente compacte d'une puissance de 250 watts était placé devant le sujet à photographier pour éclairer la scène. Les images ont été prises avec différentes orientations de la lumière d'éclairage. La lumière est orientée de gauche à droite et vice versa pendant l'acquisition. La distance entre l'appareil photo et le sujet à acquérir est de 1 mètre. Le sujet est positionné en face de la caméra. La position relative de l'appareil photo est représentée sur la [Figure 2.10](#).



Figure 2.10- Acquisition des images dans la salle noire.

La base de données d'images MS a été acquise en hiver 2020, hiver 2021, printemps 2021 et 2022. Les participants étaient composés de résidents et d'étudiants internationaux. 75% des sujets ont accepté que leur photo soit utilisée dans le cadre de nos travaux de recherche. Il s'agit d'hommes et de femmes de tous les âges, noirs et blancs. Les participants sont originaires de différents continents.

La base de données d'images multispectrales nommée EXIST contient des images brutes de visages de 110 personnes. Ces images sont au format tiff et ont une taille 2072 x 1104 pixels.

La [Figure 2.11](#) montre quelques échantillons d'images de EXIST.

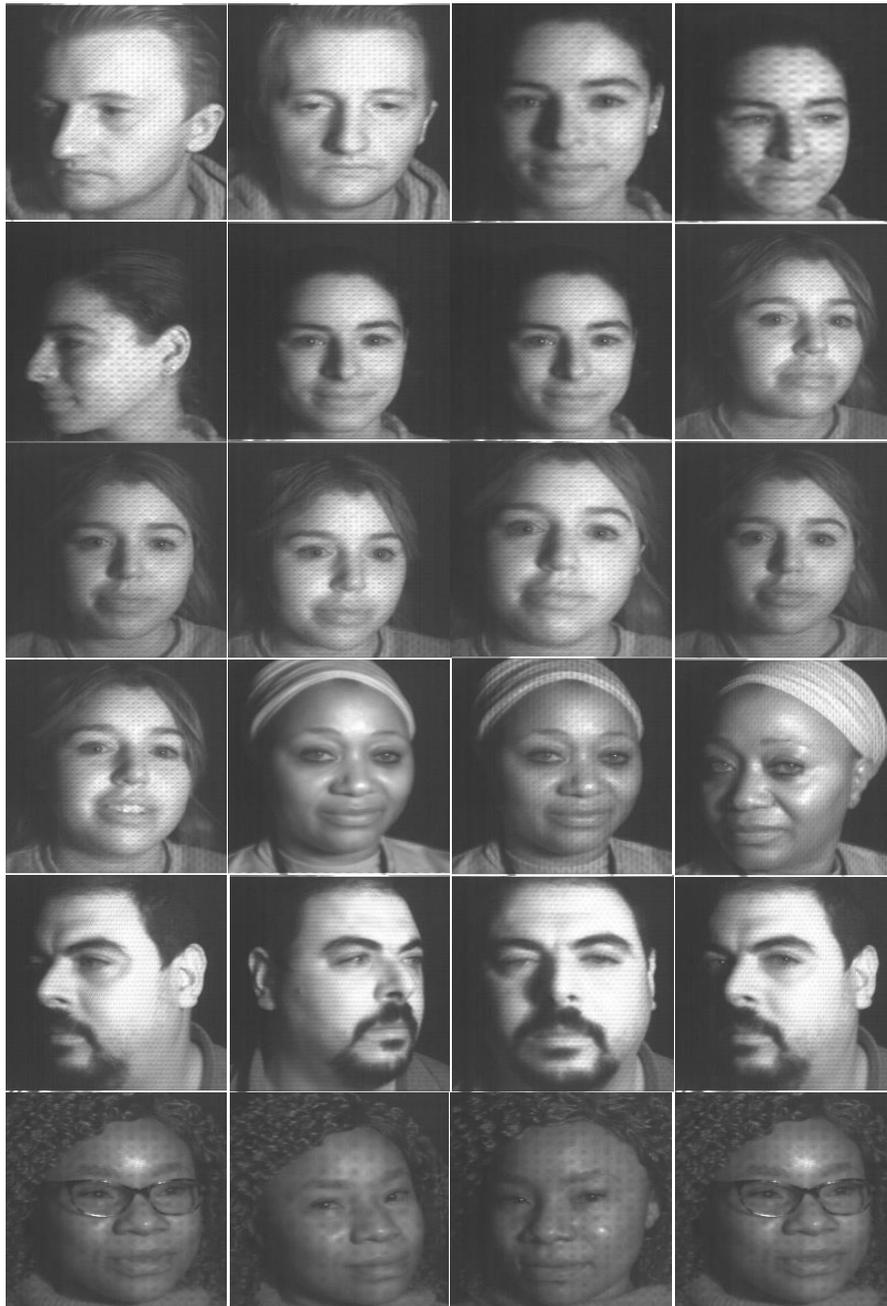


Figure 2.11- Quelques échantillons d'images de la base EXIST

Conclusion

Dans ce chapitre, nous avons présenté une nouvelle base de données d'images de faces de grande qualité pour la reconnaissance faciale. Cette base d'image est collectée avec une caméra one shot équipée d'un réseau de huit filtres spectraux. Les étapes de conception et de

fabrication de cette caméra ont été détaillées. Le protocole de collecte et la description de cette base d'images ont également été présentés. À notre connaissance, la base d'images EXIST est la première base d'images de faces acquises avec une caméra multispectrale one shot équipée d'un réseau de filtres multispectraux. Cette base d'images permettra de concevoir des systèmes de reconnaissance robustes et performants, et aussi d'améliorer ce type de caméras pour la reconnaissance faciale.

Chapitre 3

Développement de système robuste de reconnaissance faciale via les caméras one-shot

Sommaire

| | |
|--|----|
| Introduction | 50 |
| 3.1 Systèmes de reconnaissance faciale multispectrale existants | 50 |
| 3.2 Architecture d'un système de reconnaissance avec une caméra MSFA oneshot | 53 |
| 3.3 Demosaicing des images de la base de données EXIST | 54 |
| 3.4 Méthode proposée | 58 |
| 3.4.1 <i>Fusion des images</i> | 58 |
| 3.4.1.1 <i>Transformé discrète rapide de curvelets</i> | 58 |
| 3.4.1.2 <i>Fusion des images avec la FDCT</i> | 60 |
| 3.4.2 <i>Réseaux de neurones convolutifs</i> | 61 |
| 3.4.2.1 <i>VGG19</i> | 61 |
| 3.4.2.2 <i>ResNet 101</i> | 62 |
| 3.5 Résultats et discussion | 64 |
| 3.5.1 <i>Résultats expérimentaux</i> | 64 |
| 3.5.2 <i>Discussion</i> | 68 |
| Conclusion | 68 |

Introduction

La reconnaissance de visages dans le visible et le proche infrarouge a fait l'objet d'une grande attention ces dernières années. En effet, la performance des systèmes de RF dépend des spectres électromagnétiques dans lesquels les images du visage ont été acquises. La RF basée sur le spectre visible utilise généralement des caractéristiques de texture. Sa performance est affectée par la variation de la lumière et de pose, les occlusions, le vieillissement du visage. L'infrarouge présente plusieurs avantages par rapport au spectre visible : il n'est pas perceptible par l'œil humain et, en même temps, il est moins sensible aux variations de la lumière. La combinaison des deux spectres pourrait améliorer la performance des systèmes de RF. Les systèmes actuels d'imagerie MS utilisés pour la RF sont basés sur des caméras multiples dotées de plusieurs capteurs ou filtres. Ces systèmes d'acquisition sont normalement lents car ils prennent une image MS en plusieurs prises, ce qui les rend incapables d'acquérir les images en temps réel et de capturer des scènes en mouvement. Bien que les systèmes de RF implémentés avec ces caméras produisent de bonne performance, ils sont confrontés aux problèmes d'appariement de gabarits. Par ailleurs, les systèmes d'imagerie multispectrale instantanés qui intègrent un seul capteur avec un MSFA permettent ainsi d'avoir à chaque acquisition une image sur plusieurs spectres. Ces systèmes réduisent considérablement le temps d'acquisition des images MS et sont capables de capturer des scènes en mouvement en temps réel. Fort de cela, nous avons utilisé ce type de caméra qui contient un MSFA avec les bandes du visible et du proche infrarouge pour la RF.

Ce chapitre présente le développement d'une application robuste de reconnaissance faciale avec un système d'acquisition MSFA qui couvre le visible et le proche infrarouge. Un état de l'art sur les systèmes de RF multispectrale est d'abord fait. Ensuite la solution proposée est basée sur la transformé rapide discrète de curvelet et les CNNs.

3.1 Systèmes de reconnaissance faciale multispectrale existants

L'utilisation de plusieurs bandes spectrales améliore la performance du système de RF, ce qui explique l'intérêt que les chercheurs portent à la RF multispectrale ces dernières années[3], [109], [110]. Il existe plusieurs systèmes de RF dans le visible et le proche infrarouge, mais pour la plupart, les images ne sont pas acquises avec les caméras MSFA one-shot. Pour cette étude nous avons sélectionné les systèmes de RF les plus pertinents parmi lesquels :

Abood et al. [111] ont proposé un système de RF utilisant la fusion d'images multispectrales pour surmonter les limites de la reconnaissance faciale visible. Ce système de reconnaissance fusionne les caractéristiques des images du visible, du proche infrarouge et du thermique. Les auteurs avaient utilisé la transformée en ondelettes de Gabor pour l'extraction de caractéristiques et le SVM pour la classification. Les résultats expérimentaux atteignaient un taux de reconnaissance de 96,4% sur la base de données Carl.

Y.Jin et al. [112] ont développé une méthode d'apprentissage discriminative couplée pour la reconnaissance hétérogène des visages. Cette méthode représente les caractéristiques discriminantes du visage en construisant un vecteur propre de filtre optimal avec les pixels bruts de l'image. L'approche développée utilise le local ternaire pour coder les motifs locaux et la métrique du cosinus pour estimer les similitudes entre les images. La performance de cette méthode a été testée sur la base de données CASIA 2.0 NIR-VIS. Les résultats expérimentaux indiquent que le taux de reconnaissance exact est inférieur à 90%.

En 2021, R. He et al. [113] ont proposé un système d'Apprentissage Adversarial Couplé (CAL) pour les applications semi-supervisées de la RF hétérogène dans le visible et le proche infrarouge. Les images VIS ont été générées à partir des images VIS-NIR non appariées avec une série de réseaux neuronaux de bout en bout composés de 29 couches de convolution et des blocs résiduels (LightCNN-29). Les résultats expérimentaux ont indiqué une précision de rang 1 de 98,6 % à 99,6 % sur la base de données NIR-VIS 2.0 de CASIA.

A. Yu et al.[89] avaient mise en œuvre un système de reconnaissance de visage qui a utilisé les GAN pour générer les images du VIS à partir des images NIR. Afin de réduire l'écart de domaine entre les données NIR et VIS, un module d'attention a été développé par les auteurs. Un LightCNN-29 a été utilisé comme classificateur. Une précision de rang 1 de 94,94 % à 97,91 % a été obtenue.

Song et al. [114] ont développé un système d'apprentissage discriminatif de caractéristiques adversaires pour la reconnaissance faciale des images du VIS et du NIR. Afin de compenser l'écart de détection, les auteurs avaient appliqué les méthodes basées sur l'apprentissage contradictoire à la fois sur l'espace des pixels bruts et sur l'espace des caractéristiques compactes. Cette approche combine ResNet et Light CNN. L'expérience a été réalisée sur les bases de données : CASIA NIR-VIS 2.0, BUAA-VisNir et Oulu-CASIA NIR-VIS. Les résultats expérimentaux obtenus donnent respectivement une précision de rang 1 de 98,15%, 95,2% et 95,5%.

Afin de corriger les problèmes de désalignement entre les images visibles et les images NIR appariées, P. Zao et al [115] ont développé un système d'alignement automatique des images

NIR-VIS pour la reconnaissance hétérogène de visage. L'architecture proposée par les auteurs est basée sur les GAN. Cette méthode permet de générer des images NIR-VIS doubles sémantiquement alignées avec la même identité. Les caractéristiques ont été extraites avec un lighthCNN et deux réseaux codeurs pour les tâches de génération. Une précision de rang 1 proche de 99,9 % a été obtenue sur les ensembles de données CASIA NIR-VIS 2.0, Oulu-CASIA NIR-VIS et BUAA VIS-NIR.

M. Diarra et al. [116] ont proposé le modèle MS-FRHF (Reconnaissance multispectrale des visages à l'aide de caractéristiques hybrides) pour le visible et l'infrarouge thermique. Ils ont utilisé la base de données Robotics Intelligent System (IRIS) pour l'extraction des caractéristiques. Les points d'intérêt et la texture ont été extraites respectivement avec l'extracteur de points clés Maximally Stable Extremal Region (MSER) et la matrice de cooccurrence des niveaux de gris. Une fusion a été faite avec ACP. Les auteurs ont conclu que cette approche donne les meilleurs taux de reconnaissance que ceux obtenus dans le visible et dans l'infrarouge thermique.

En 2020, Zhihua Xie et al. [117] ont développé des méthodes de fusion basées sur le modèle binaire local et la transformée en cosinus discret pour la RF dans l'infrarouge et le visible. Les informations de basse fréquence sont d'abord extraites du proche infrarouge avec la transformée discrète et le Local Binary Patter(LBP) est appliqué pour représenter les caractéristiques discriminantes. Ensuite, les caractéristiques des images du visible ont été extraites avec le LBP et fusionnées. Les résultats expérimentaux ont montré que le taux de reconnaissance a été amélioré avec cette approche.

Guo et al. [118] ont proposé un système de reconnaissance faciale dans le visible et le proche infrarouge avec les réseaux de neurones profonds. Les caractéristiques visibles et proche infrarouge ont d'abord été extraites avec le réseau de neurones. La distance de cosinus a été utilisée pour déterminer le score de classification des deux ensembles suivis de leur fusion. Une précision de 99,89 % et 99,56 % a été obtenue respectivement sur les bases de données HIT LAB2 et SunWin Face.

Hu et al. [119] ont présenté un modèle d'apprentissage profond discriminant basé sur la perte de supervision conjointe et la fusion de caractéristiques multicouches pour la reconnaissance faciale. Cette approche intègre dans le réseau de neurones une fonction de perte appelée perte de dispersion afin d'améliorer le pouvoir discriminant des caractéristiques apprises en profondeur. Les caractéristiques extraites par les CNNs dans les différentes bandes visibles et proches infrarouges ont été ensuite fusionnées. La performance du système a été testée sur les bases d'images CASIA NIR-VIS 2.0 et Oulu-CASIA NIR-VI. Les résultats expérimentaux ont

donné une précision de rang 1 de 98,5% à 98,8% sur la base de données CASIA NIR-VIS 2.0, et de 98,5 % à 99,3 % sur la base de données Oulu CASIA NIR-VIS.

F. Wei et al. [91] ont développé une analyse de discrimination intra-spectrale et une analyse de corrélation inter-spectrale (IDICN) pour la reconnaissance faciale. Ce système a amélioré les performances de la reconnaissance multispectrale de visages en incluant des informations inter- et intra-spectrales. Cette approche consiste en un ensemble de réseaux neuronaux convolutifs profonds spécifiques à chaque spectre avec une couche de mise en commun des spectres. Ces réseaux extraient des caractéristiques liées à un ensemble de spectres, et la couche de mise en commun des spectres sélectionne un groupe de spectres ayant des capacités discriminantes. Les auteurs ont utilisé trois bases de données : celle de l'Université polytechnique de Hong Kong (HK PolyU), celle de l'Université Carnegie Mellon (CMU) et celle de l'Université d'Australie (UWA). Les résultats expérimentaux ont permis d'obtenir des taux de reconnaissance moyens de 99,76%, 100% et 99,85% respectivement sur les trois bases.

3.2 Architecture d'un système de reconnaissance avec une caméra MSFA oneshot

Les images acquises avec les caméras one-shot sont des images brutes. Chaque filtre du MSFA est responsable de l'acquisition dans une bande donnée. Il va falloir estimer la valeur des pixels manquants dans une bande donnée pour la reconstruction des images spectrales. Ce traitement est appelé *demosaicing*. Nous avons alors intégré ce module à la reconnaissance faciale avec une caméra MSFA one-shot. La [Figure 3.1](#) présente l'architecture générale d'un système de reconnaissance faciale avec une caméra MSFA one-shot.

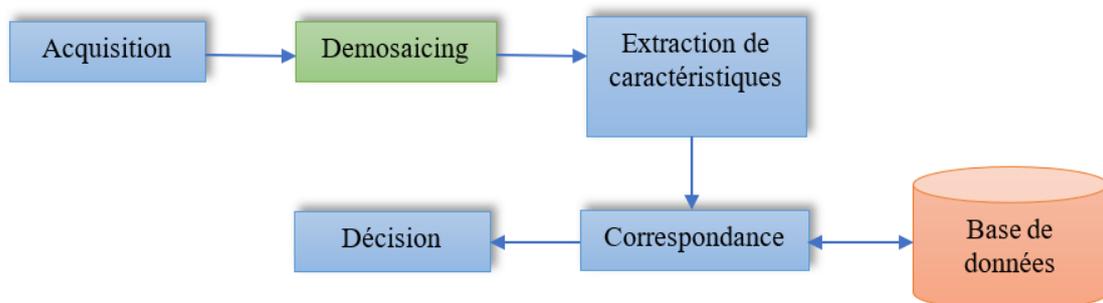


Figure 3.1- Architecture d'un système de reconnaissance avec une caméra MSFA one-shot

3.3 Demosaicing des images de la base de données EXIST

Les systèmes d'acquisition MSFA sont des caméras à prise de vue unique disposant d'un seul capteur qui échantillonne spatialement l'image entrante à l'aide d'un réseau de filtres spectraux. Par conséquent, chaque pixel ne contient que la valeur d'une couleur correspondant à une bande donnée. Les images acquises avec les caméras one-shot sont des mosaïques d'images. En effet, une mosaïque d'image ou une image brute I , de taille $X \times Y$ pixels, dans laquelle une seule bande $k \in \{1 \dots K\}$ fait référence à la valeur de chaque pixel p en fonction de la structure MSFA. Il est alors nécessaire de reconstruire l'image originale avant son utilisation. Ce processus de reconstruction est appelé demosaicing.

Le demosaicing est une technique qui consiste à estimer la valeur des pixels manquants dans une bande donnée. A l'issue de cette transformation le nombre d'images multispectrales obtenues est fonction du nombre de filtres qui composent le MSFA. Le MSFA intégré dans la caméra décrite ci-dessus est composé de huit filtres donc une image multipectrale de 8 bandes sera obtenue à l'issue du demosaicing.

Avant le demosaicing, une extraction de bandes a été effectuée. Elle consiste à multiplier l'image mosaïque par différents masques binaires $M^k(x,y)$, $k \in \{1, 2, \dots, K\}$. Ces masques ont la valeur 1 aux positions où la composante est disponible, et 0 aux autres positions. Chaque composante plane est obtenue après avoir multiplié l'image brute ou mosaïque par le masque correspondant M^k . Le masque est défini par l'équation 3.1

$$M^k_{(x,y)} = \begin{cases} 1 & \text{si } (x \bmod \sqrt{K}) + (y \bmod \sqrt{K}) \times \sqrt{K} = k \\ 0 & \text{sinon} \end{cases} \quad (3.1)$$

Dans notre cas, la multiplication de l'image mosaïque par chaque masque permet d'obtenir 8 plans d'images décalés dans lesquels une seule composante est disponible à chaque pixel. Chaque masque correspond à un plan d'image I'^k défini par l'équation 3.2

$$I'^k = I \odot M^k \quad (3.2)$$

Où \odot désigne le produit élément par élément et M^k est un masque binaire défini à chaque pixel P .

En considérant le moxel du MSFA intégré dans notre caméra one-shot on obtient les 8 plans d'images illustrées par la [Figure 3.2](#).

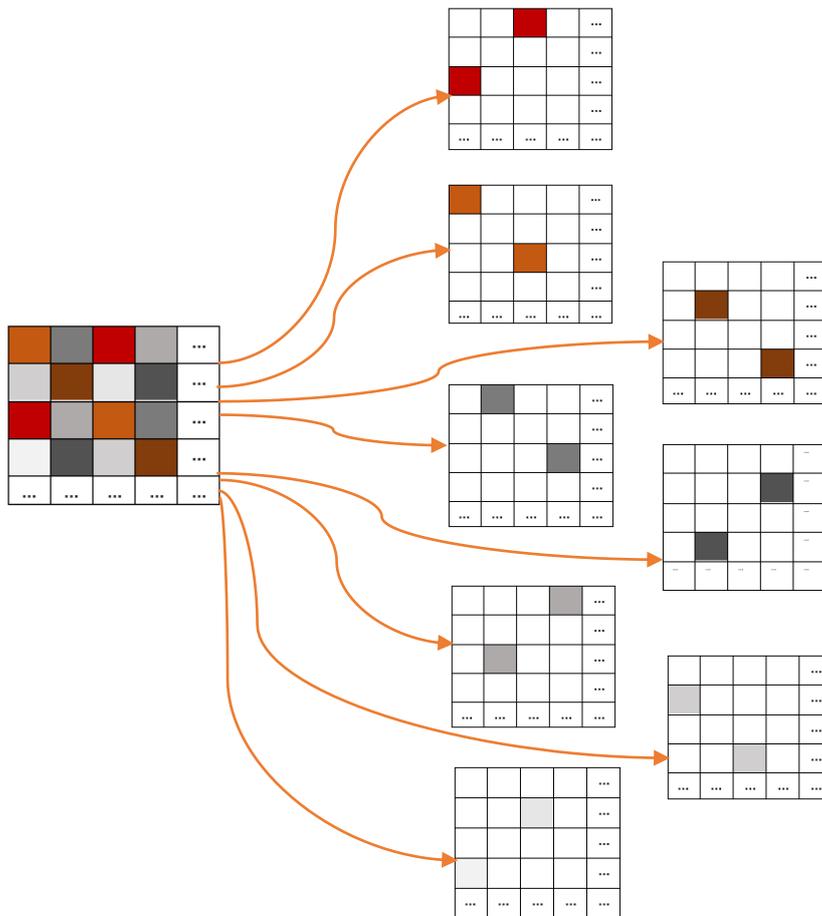


Figure 3.2- Carte d'images obtenue après l'application de différents masques

La littérature propose beaucoup de méthodes de demosaicing parmi lesquelles *l'Interpolation Bilinéaire*. Cette méthode de demosaicing a été utilisée pour estimer les pixels manquants dans chacune des bandes. Une étude comparative des différentes méthodes de demosaicing sur la base EXIST sera présentée dans le chapitre 5. L'interpolation bilinéaire constitue la base de plusieurs méthodes de demosaicing à savoir : l'arbre binaire, la transformée discrète en ondelette et le vecteur médian. L'interpolation bilinéaire est une technique de rééchantillonnage basée sur la moyenne pondérée par la distance des quatre valeurs des pixels les plus proches du pixel manquant (dont on veut estimer la valeur). Comme son nom l'indique, l'interpolation bilinéaire est une succession de deux interpolations linéaires chacune dans une direction. Les interpolations linéaires peuvent être effectuées dans plusieurs directions. Pour un pixel manquant P à la position (i, j) , l'interpolation linéaire est définie comme suit :

- Diagonal

$$P(i,j) = \frac{1}{4} \sum_{(m,n)=(-1,-1),(-1,1),(1,-1),(1,1)} p(i+m,j+n) \quad (3.4)$$

- Vertical

$$P(i,j) = \frac{1}{2} \sum_{(m,n)=(-1,0),(1,0)} p(i+m,j+n) \quad (3.5)$$

- Horizontal

$$P(i,j) = \frac{1}{2} \sum_{(m,n)=(0,-1),(0,1)} p(i+m,j+n) \quad (3.6)$$

Le demosaicing d'une image multispectrale par interpolation bilinéaire consiste également à convoluer chaque plan de composantes obtenues par un filtre H. Ce filtre est déterminé en fonction de la distance spatiale entre les pixels voisins et le pixel central dont on veut estimer la valeur. Deux filtres ont été utilisés pour effectuer la convolution, un filtre H₁ pour les bandes P1, P2, P3, P4, P5 et P6 puis un second noté H₂ pour les bandes P7 et P8. La bande de l'image interpolée est définie par l'équation 3.6

$$I^k = I'^k \odot H \quad (3.6)$$

Avec H=H₁ ou H=H₂

$$H_1 = \frac{1}{9} \begin{pmatrix} 1 & 2 & 3 & 2 & 1 \\ 2 & 4 & 6 & 4 & 2 \\ 3 & 6 & 9 & 6 & 3 \\ 2 & 4 & 6 & 4 & 2 \\ 1 & 2 & 3 & 2 & 1 \end{pmatrix} \quad H_2 = \begin{pmatrix} \frac{1}{5} & \frac{1}{4} & \frac{1}{3} & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & \frac{1}{5} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{3} & 1 & \frac{1}{3} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{4} & \frac{1}{3} & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & \frac{1}{5} & \frac{1}{4} & \frac{1}{5} \end{pmatrix}$$

Pour estimer la valeur du pixel manquant $P(i,j)$, l'image I^k est convoluée avec chacun des filtres H₁ et H₂. Si V₁ et V₂ sont les résultats de la convolution, alors la valeur du pixel manquant est définie par la moyenne de V₁ et V₂. Ce processus est utilisé pour estimer la valeur de chaque pixel manquant dans I'^k . La Figure 3.3 illustre l'estimation de la valeur des pixels manquants.

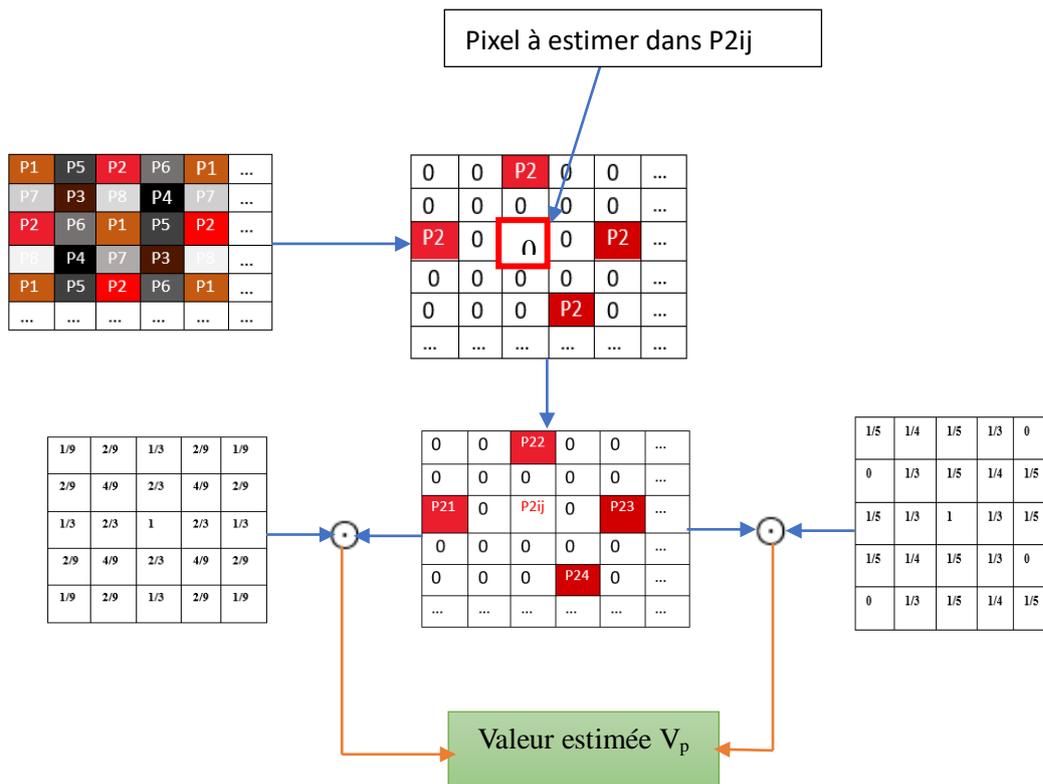


Figure 3.3- Processus d'estimation de la valeur d'un pixel manquant dans une bande donnée.

Le demosaicing de chaque image brute génère 8 images multispectrales correspondant respectivement aux longueurs d'onde {685, 720, 770, 810, 835, 870, 895, 930} nm. Ces images sont ensuite stockées dans la base d'images. La Figure 3.4 résume ce processus.

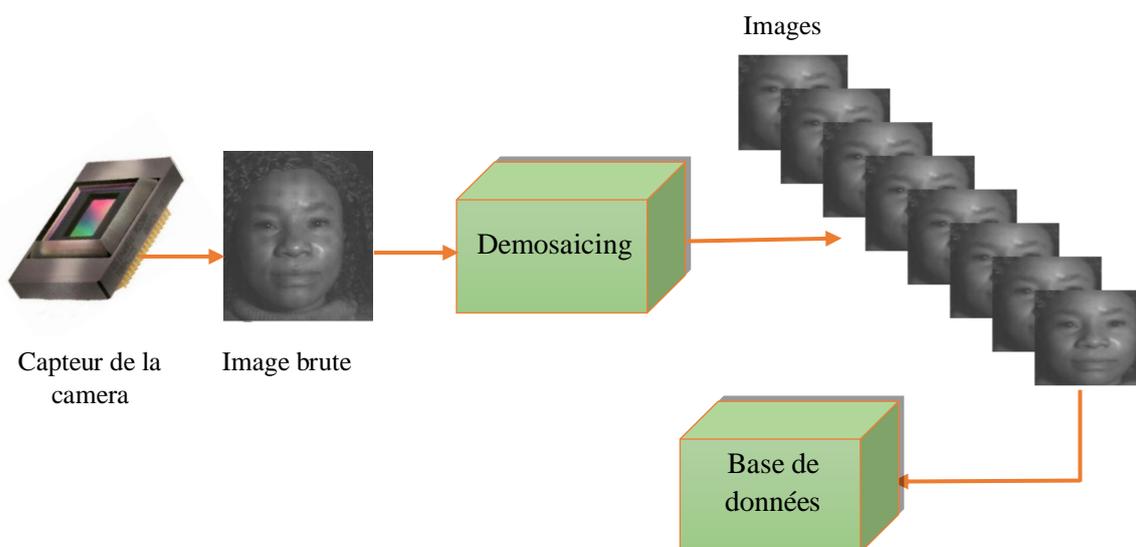


Figure 3.4-Demosaicing des images brutes et stockage dans la base de données.

3.4 Méthode proposée

Cette section décrit la méthode proposée pour la reconnaissance faciale via la caméra MSFA one shot décrite ci-dessus. Elle est composée de deux parties :

- La fusion des images avec la Transformée Discrète Rapide de Curvelet (TDRC)
- Les CNNs VGG19 et ResNet 101 pour l'extraction des caractéristiques et la classification

La Figure 3.5 illustre la méthode proposée :

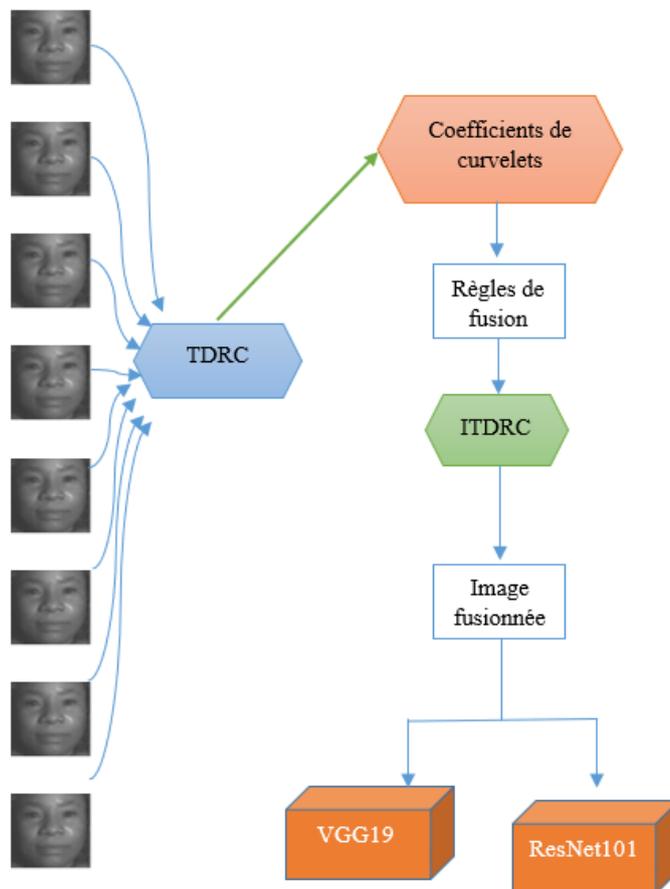


Figure 3.5- Méthode proposée pour la reconnaissance faciale avec une caméra one-shot

3.4.1 Fusion des images

3.4.1.1 Transformé discrète rapide de curvelets

Les ondelettes permettent d'extraire les informations importantes contenues dans les images et de faire leur fusion. Généralement elles sont des transformations qui permettent de séparer les informations contenues dans un signal en haute fréquence et en basse fréquence. Les ondelettes sont aussi utilisées sur les images et permettent de les décomposer en haute

fréquence et en basse fréquences pour une analyse objective de ces dernières. Les ondelettes permettent certes d'extraire les caractéristiques des images mais ne permettent pas de représenter de manière efficace les bords des images.

En 1999 J. Candles et al. [120] ont développé une transformée géométrique multi-échelles appelée Curvelet Transform(CT) en anglais . Cette transformée directionnelle multi-échelles définit une représentation clairsemée non-adaptative presque optimale des objets avec des bords. La CT permet de pallier les limitations relatives aux ondelettes, elle est mieux adaptée à l'analyse des bords (courbes et lignes) des images que les ondelettes. La CT est décrite par une association de trois méthodes :

- Une méthode d'analyse mieux spécifiques aux objets contenant des lignes droites avec discontinuités (ridgelets) ;
- Une pyramide de ridgelets fenêtrés ;
- Le filtrage passe-bande, une méthode permettant de séparer un objet en une série d'échelles disjointes.

Cependant, cet algorithme est complexe avec une grande redondance. Afin d'améliorer cette première génération de curvelets en terme de rapidité, de simplicité et de faciliter l'utilisation, une deuxième génération de la transformée en curvelets a été développée par les mêmes auteurs [121] en 2006. L'implémentation numérique de cette deuxième génération de transformée en curvelets est appelée Fast Discret Curvelet Transform (Transformée Discrète rapide de Curvelet). Il existe deux implémentations numériques de la FDCT : Curvelets via Unequally-Spaced fast Fourier transforms(USFFT), et Curvelets via Wrapping. La première est basée sur les transformations de Fourier rapides inégalement espacées (USFFT) et la seconde est basée sur l'enveloppement d'échantillons de Fourier spécialement sélectionnés (Wrapping). La différence entre les deux implémentations de la seconde génération de curvelet repose essentiellement sur le choix de la grille spatiale utilisée pour traduire les courbures à chaque échelle et angle. La version basée sur USFFT est caractérisée par une inclinaison et un alignement de la grille de traduction sur l'orientation du curvelet tandis que dans celle dite Wrapping la grille est la même pour chaque angle dans chaque quadrant avec une orientation appropriée du curvelet.

Cette deuxième génération de la transformé en curvelets s'exécute très rapidement et sont moins redondantes que la première. De l'exécution chacune des deux transformations numériques de la deuxième génération de la transformé en curvelets résulte une table contenant les coefficients de curvelets numériques indexés par un paramètre d'échelle, un paramètre d'orientation et un paramètre d'emplacement spatial.

Les deux implémentations digitales de la transformée en Curvelet sont des transformations numériques et linéaires. Elles prennent en entrée des tableaux cartésiens de la forme $f[t_1, t_2], 0 \leq t_1, t_2 < n$, et la sortie est une collection de coefficients $C^D(j, l, k)$ définie par :

$$C^D(j, l, k) = \sum_{0 \leq t_1, t_2 < n} f[t_1, t_2] \overline{\varphi_{j,l,k}^D[t_1, t_2]} \quad (3.7)$$

L'algorithme de la FDCT via USFFT se déroule en quatre principales étapes présentées comme suit :

- Extraction des coefficients de Fourier en appliquant la Transformée de Fourier Rapide bidimensionnelle (2D-FFT) à l'image dont on veut déterminer les coefficients de curvelets. Les coefficients de Fourier $\hat{f}[n_1, n_2]$ sont définis par :

$$\hat{f}[n_1, n_2] = , \text{pour } N/2 \leq n_1, n_2 \leq N/2$$

- Rééchantillonnage (ou interpoler) $\hat{f}[n_1, n_2]$ pour chaque paire échelle/angle (j, l), pour avoir des valeurs échantillonnées

$$\hat{f}[n_1, n_2 \text{ xtan}\theta] \text{ pour } (n_1, n_2) \in P_j$$

- Multiplication de l'objet interpolé \hat{f} par la fenêtre parabolique \tilde{U}_j , en localisant effectivement \hat{f} près du parallélogramme d'orientation θ_l , pour obtenir

$$\tilde{f}_{j,l}[n_1, n_2] = \hat{f}[n_1, n_2 - n_1 \tan\theta_l] \tilde{U}_j[n_1, n_2] \quad (3.8)$$

- Collection des coefficients discrets $C^D(j, l, k)$ en appliquant la transformé inverse rapide de Fourier à chaque $\tilde{f}_{j,l}$

L'implémentation logiciel de la deuxième génération de Curvelets est disponible sur Curveletlab [122].

3.4.1.2 Fusion des images avec la FDCT

Chaque image multispectrale étant composée de huit images, nous avons opté pour la fusion d'images avec la transformé discrète rapide de Curvelets. La fusion est composée de deux étapes : la méthode de fusion et la stratégie de fusion. Chacune des huit images a été décomposée avec la FDCT. Cela nous permet d'obtenir pour chaque image un tableau de

coefficients de curvelets. Plusieurs stratégies de fusion ont été testées à savoir la moyenne, le minimum et le maximum. Chacune d'elles a été utilisée pour fusionner les coefficients de curvelet cellule par cellule correspondant respectivement aux huit tableaux. La stratégie de fusion « *moyenne* » a permis d'avoir les meilleurs résultats. Une transformation inverse rapide des curvelets discrets a été appliquée pour reconstituer l'image fusionnée contenant ainsi les informations de toutes les bandes. Les images obtenues après fusion sont utilisées par les CNN VGG19 et ResNet 101 pour l'entraînement, les tests et la validation.

3.4.2 Réseaux de neurones convolutifs

3.4.2.1 VGG19

Nous avons utilisé VGG19 par transfert learning pour développer le système de reconnaissance proposé. En effet, VGG19 est une architecture de VGGNet qui est un CNN proposé par K. Simonyan et al. [40] en 2015 . L'architecture VGG a été largement utilisée en vision par ordinateur au cours des dernières années. Elle se compose de couches convolutionnelles empilées et de couches de mise en commun maximale. VGG19 contient 19 couches de poids consistant en 16 couches convolutionnelles avec 3 couches entièrement connectées et 5 couches de mise en commun. Comparé aux réseaux neuronaux convolutifs traditionnels, il a été amélioré en termes de profondeur de réseau. Il utilise une structure alternée de plusieurs couches convolutionnelles et de couches d'activation non linéaires, ce qui est mieux qu'une seule convolution. La structure en couches permet de mieux extraire les caractéristiques de l'image, d'utiliser le Maxpooling pour le sous-échantillonnage et de modifier l'unité linéaire (ReLU) comme fonction d'activation. L'architecture du VGG19 est détaillée dans la [Figure 3.6](#).

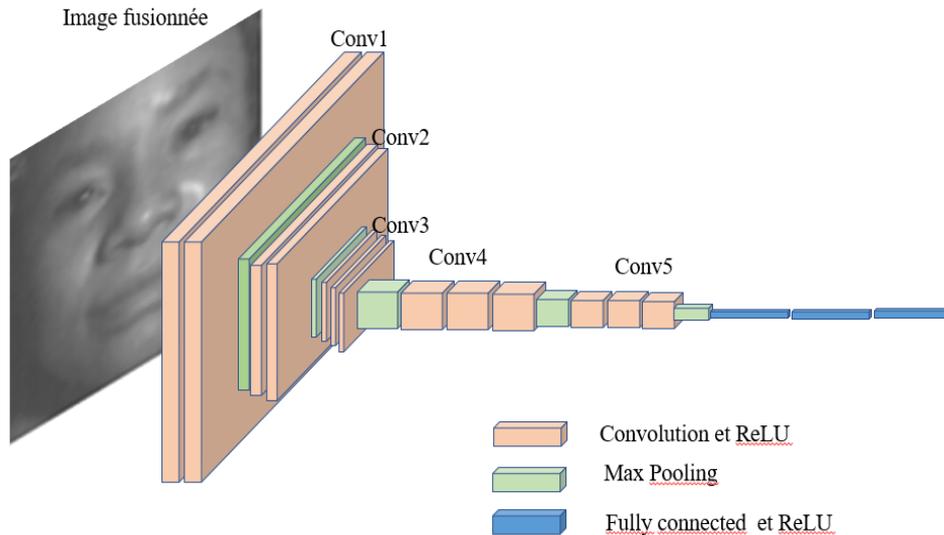


Figure 3.6- Architecture de VGG19

3.4.2.2 ResNet 101

Les réseaux résiduels appelés ResNets représentent un type puissant d'architecture de réseau neuronal convolutif. Ils sont largement adoptés, et utilisés dans diverses tâches comme la reconnaissance de visages. Pour résoudre la complexité des tâches de reconnaissance et améliorer la précision de la classification, les chercheurs augmentent la profondeur des réseaux neuronaux. Cependant, plus on va en profondeur, plus la formation du réseau neuronal devient difficile, et la précision commence à saturer, puis se détériore également. Pour surmonter ces problèmes, K. He et al. [42] ont introduit l'apprentissage résiduel. Les auteurs ont obtenu de bons résultats en évaluant les réseaux résiduels sur le jeu de données ImageNet comprenant 1000 classes d'objets avec une profondeur allant jusqu'à 152 couches. La Figure 3.7 présente le bloc d'apprentissage résiduel.

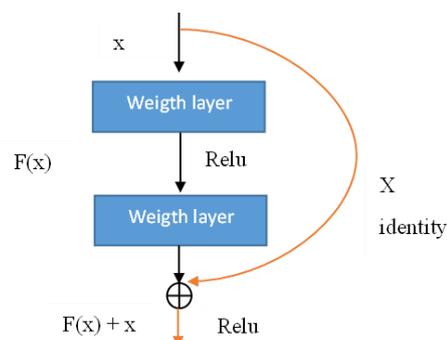


Figure 3.7- Apprentissage résiduel : un élément constitutif

ResNet101 est un réseau résiduel qui contient 101 couches. Il est constitué de couche de convolution, de couche de mise en commun, de bloc résiduel et de couche entièrement connectée. L'architecture de ResNet101 est illustrée par la figure [Figure 3.8](#).

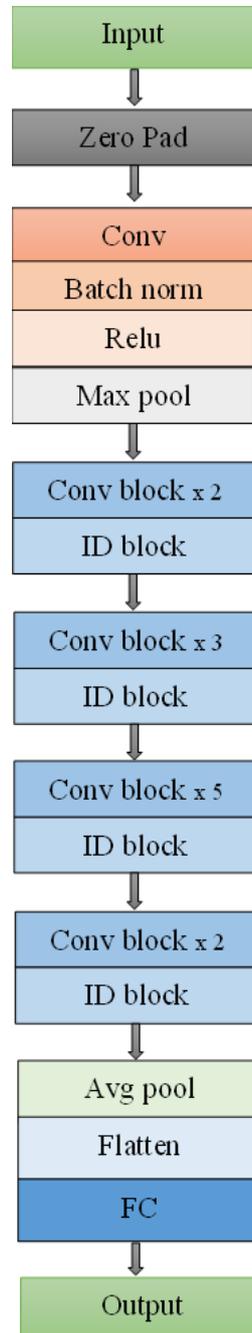


Figure 3.8- Architecture de ResNet101

Dans ce travail, nous avons modifié ce modèle et supprimé la couche FC qui comprend 1000 classes d'objets. Nous avons ajouté une nouvelle couche FC en fonction du nombre de classes d'images dont nous disposons.

3.5 Résultats et discussion

3.5.1 Résultats expérimentaux

La base d'images EXIST décrite dans le chapitre 2 est utilisée pour tester la méthode proposée. Les expérimentations sont réalisées sur Microsoft System windows, version 2010, avec deux ordinateurs. Le premier est équipé d'un processeur Intel (R) Core (TM) i7-8565U et d'une mémoire RAM de 8 Go. Le second possède une unité de traitement graphique (GPU) NVIDIA Quadro P400 avec 32 Go de mémoire vive (RAM). Tous les codes sont développés dans le langage de programmation Matlab 2020 et Python 3.7.

Le tableau 3.1 décrit les paramètres d'entraînement utilisés pour les deux CNNs.

Tableau 3.1- Paramètres de VGG19 et ResNet101 au cours de la phase d'entraînement

| Paramètres | VGG19 | ResNet 101 |
|---------------------------|------------------|------------|
| Taille du lot | 32 | 16 |
| Algorithme d'optimisation | SGD ^a | ADAM |
| Taux d'apprentissage | 0.0001 | 0.0001 |
| Numéro d'époque | 20 | 20 |

^a. Stochastic Gradient Descent

Pour analyser les résultats, les mesures d'évaluation de performances suivantes ont été calculées : accuracy, precision, recall, F1-score, Matthews Correlation Coefficient (MCC) et Means Square Error (MSE).

« accuracy » indique le pourcentage de prédictions correctes. Elle est définie par l'équation 3.9

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3.9)$$

Où TP (Vrais Positifs), TN (Vrais Négatifs), FP (Faux Positifs), FN (Faux Négatifs)

La *precision* est la proportion de vrais positifs sur l'ensemble des positifs détectés.

$$precision = \frac{TP}{TP + FP} \quad (3.10)$$

Le *recall* rappel est le nombre de vrais positifs qui sont correctement classés.

$$recall = \frac{TP}{TP + FN} \quad (3.11)$$

F1score est basé sur recall et precision. Il est défini comme suit

$$F1_{score} = \frac{2 * precision * recall}{precision + recall} \quad (3.12)$$

Le coefficient de corrélation de Matthews (MCC) est la méthode de calcul du coefficient de corrélation entre les valeurs réelles et celles prédites. Le MCC est un score plus informatif, il donne de meilleurs résultats dans l'évaluation de la classification binaire [27]. La gamme de valeurs du MCC est comprise entre -1 et 1. Un modèle avec un score de 1 est un modèle parfait et -1 est un mauvais modèle. Le MCC est définie par l'Equation (3.13).

$$MCC = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}} \quad (3.13)$$

L'erreur quadratique moyenne (MSE) permet de calculer l'erreur entre les valeurs prédites \hat{y} et les valeurs réelles y . L'équation de MSE est exprimée par l'Equation (3.14).

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (3.14)$$

Avec N le nombre d'échantillons que nous testons.

Les différentes métriques calculées permettent une meilleure interprétation des résultats. La précision d'un modèle prédit le pourcentage de toutes les personnes qui peuvent être reconnues par le modèle. Le Recall indique le nombre de prédictions correctes qui ont été effectivement reconnues.

Les stratégies de fusion telles que min, max et moyenne sont utilisées pour obtenir les images. Trois types d'expériences ont été réalisées : une première avec les images obtenues par FDCT et la stratégie de fusion min, une deuxième avec les images obtenues par FDCT et la stratégie de fusion max et une troisième avec la stratégie de fusion moyenne. Les tableaux 3.2 et 3.3 présentent les performances du modèle VGG19 et ResNet101 pour la fusion min, max et moyenne respectivement.

Tableau 3.2- Résultats avec VGG19

| Stratégies de Fusion | Métriques | | | | |
|-------------------------|-----------|----------|---------|--------|------|
| | Précision | MSE | F-score | Recall | MCC |
| Moyenne | 1.0 | 2.80e-07 | 1.0 | 1.0 | 1.0 |
| Min | 0.9849 | 7.96e-07 | 0.98 | 0.98 | 0.98 |
| Max | 0.9870 | 6.76e-07 | 0.98 | 0.98 | 0.98 |

Tableau 3.3 – Résultats avec ResNet101

| Stratégies de Fusion | Métriques | | | | |
|-------------------------|-----------|----------|---------|--------|------|
| | Précision | MSE | F-score | Recall | MCC |
| Moyenne | 0.99 | 4.91e-10 | 0.99 | 0.99 | 0.99 |
| Min | 0.97 | 7.98e-10 | 0.97 | 0.97 | 0.97 |
| Max | 0,96 | 91e-10 | 0,96 | 0,96 | 0,96 |

Notons que les deux modèles VGG19 et ResNet101 ont été entraînés avec les images des bases de données EXIST, en tenant compte des différents lots de taille 8, 16 et 32.

En comparant les différentes métriques calculées sur les deux réseaux de neurones, les résultats montrent que les performances de la méthode avec la règle de fusion moyenne sont supérieures à celles de la règle de fusion min et max.

Nous avons également comparé les résultats obtenus avec ceux de l'état de l'art. Le [Tableau 3.4](#) décrit la précision des systèmes de reconnaissance de l'état de l'art et celui de la méthode proposée.

Tableau 3.4- Taux de reconnaissance

| Système d'acquisition | Méthodes de reconnaissance | Précision |
|--|--|-------------------------------------|
| Visible camera + NIR camera + thermal camera | Gabor wavelet transform and Support Vector Machine (SVM) [111] | 96.4% |
| Visible camera and NIR camera | Local Ternary Patterns and cosine metrics [112] | 90% |
| Visible camera and NIR camera | LightCNN-29 [123] | 98.6% to 99.6% ^a |
| Visible camera and NIR camera | LightCNN-29 [89] | 94.94% to 97.91%. |
| Visible camera and NIR camera | ResNet and LightCNN-29[114] | 98.15%, 95.2% et 95.5% ^b |
| Visible camera and NIR camera | lighCNN and two encoder networks[115] | 99.9% |
| Visible camera and NIR camera | Neural Network and cosine distance[118] | 99.89% and 99.56% |
| Visible camera and NIR camera | CNN[119] | 98.5% to 98.8% et 98.5% to 99.3% |
| Hyper-spectral camera | deep convolutional neural networks[91] | 99.76%, 100% et 99.85 |
| EXIST camera | FDCT and VGG19 | 100% |
| EXIST camera | FDCT and ResNet 101 | 99% |

^a Résultats obtenus pour 2 valeurs ^b Résultats obtenus sur différentes bases de données

Le [Tableau 3.4](#) indique que la plupart des systèmes de reconnaissance faciale dans le domaine visible et du proche infrarouge utilise plusieurs caméras. Selon la base de données d'images utilisées, le taux de reconnaissance varie de 95,3% à 100%. Le taux de 100% est atteint avec l'algorithme de réseau neuronal convolutif profond. En général, les systèmes utilisant l'algorithme des réseaux neuronaux ont un taux proche de 100%. Dans ce cas, en fonction de l'algorithme de reconnaissance, le taux est compris entre 90% et 100%. Les systèmes de reconnaissance de visages qui utilisent un système d'acquisition intégrant des

réseaux de filtres multispectraux (MSFA) obtiennent d'aussi bons résultats que ceux qui utilisent plusieurs caméras.

3.5.2 Discussion

Dans ce chapitre, la RF basée sur la caméra MSFA one-shot a été démontrée. La plupart des études précédentes dans la littérature a utilisé des caméras multiples et des neurones à convolution pour extraire des caractéristiques pour l'identification des visages. Y.Jin et al. dans [112] ont utilisé des caméras multiples avec des motifs ternaires locaux et des métriques de cosinus, ils ont obtenu un taux de reconnaissance de 90%. Les systèmes de reconnaissance présentés dans la littérature qui sont basés sur la transformée en ondelettes de Gabor et le SVM atteignent un taux de reconnaissance de 96,4 % [12]. Les systèmes de reconnaissance basés sur les CNN [123],[89],[114],[118],[119] atteignent respectivement une précision de [98.6% - 99.6%], [94.94% - 97.91%] et [98.15%, 95.2%,95.5%], [99.89%, 99.56%], [98.5%-98.8%, 98.5%-99.3%] selon la base de données. De même, [91] a utilisé une caméra hyperspectrale avec des réseaux neuronaux et a obtenu des précisions de [99,76%, 100%, 99,6%] sur trois bases de données différentes. Les résultats et expériences de la reconnaissance faciale utilisant la caméra MSFA one-shot atteignent des précisions de 99% et 100% avec Resnet101 et VGG19 respectivement.

La comparaison des résultats montre que les différents algorithmes de reconnaissance faciale mis en œuvre donnent de bonnes performances en fonction des images et des méthodes utilisées. Le système proposé atteint l'une des meilleures performances. Mais aussi en raison de sa caméra et de ses algorithmes, c'est le seul système qui propose une acquisition d'images en temps réel.

Conclusion

Ce chapitre présente un nouveau système de reconnaissance faciale multispectrale utilisant des systèmes d'imagerie multispectrale à un coup avec un MSFA pour l'acquisition. Il s'agit d'un système d'acquisition à un coup qui fonctionne en temps réel sur les spectres du visible et du proche infrarouge. Le système de reconnaissance développé est basé sur la transformée rapide discrète de Curvelet, les CNNs ResNet 101 et VGG19. Les résultats expérimentaux montrent que les systèmes de reconnaissance de visages qui intègrent les caméras MSFA sont aussi performants que ceux utilisant des caméras multiples. Ce système

est cependant plus intéressant car il est plus économique, techniquement fiable et surtout équipé d'un système d'acquisition en temps réel.

Nous avons certes obtenu une précision de 100% avec le CNN VGG19, mais une optimisation du système développé sur la reconnaissance sera présentée dans les chapitres 4 et 5.

Chapitre 4

Amélioration des systèmes de reconnaissance faciale avec les caméras one shot par la réduction de dimensionalité

Sommaire

| | |
|--|----|
| Introduction | 71 |
| 4.1 Réduction de dimensionalité pour la reconnaissance faciale | 71 |
| 4.1.1 Méthodes de projection linéaire | 72 |
| 4.1.2 Méthodes de projection non linéaire | 73 |
| 4.1.3 Réduction de dimensionnalité pour les systèmes de reconnaissance faciale | 75 |
| 4.2 Etude comparative des méthodes de réduction de dimensionalités | 76 |
| 4.2.1 Motivation | 76 |
| 4.2.2 Transformé de Karhuen-Loève | 76 |
| 4.2.3 Analyse en Composante Curviligne | 77 |
| 4.2.4 Analyse Factorielle | 78 |
| 4.3 Evaluation de la réduction de dimensionalité sur la reconnaissance faciale via les caméras MSFA one-nshot..... | 78 |
| 4.4 Résultats et discussion..... | 80 |
| Conclusion..... | 85 |

Introduction

Ces dernières années, les caméras MSFA à prise unique ou one shot sont de plus en plus utilisées pour acquérir les images multispectrales. Dans les deux chapitres précédents, nous avons proposé une base d'images MS et un système de reconnaissance faciale basé sur ce type de caméra. Ce système présente certes une bonne performance, mais nous avons constaté que les images multispectrales acquises avec cette caméra sont fortement corrélées après le demosaicing par l'interpolation bilinéaire. Ces images présentent également des redondances d'information comme la plupart des images hyperspectrales. Une réduction de dimensionnalité est alors nécessaire pour diminuer ou supprimer cette redondance d'informations ce qui pourrait conduire à une optimisation du système développée. La réduction de dimensionnalité est un ensemble de techniques qui permettent de projeter une image initiale de dimension n pour obtenir une image finale de dimension p , avec $p < n$ tout en préservant ses informations pertinentes.

Ce chapitre propose une amélioration du système de reconnaissance faciale avec une caméra MSFA one shot par la Réduction de Dimensionnalité (RD). Un module de RD a été donc ajouté au système de reconnaissance faciale développé en vue de son optimisation. Quatre méthodes de RD ont été sélectionnées et appliquées au système développé. Une comparaison des performances des différentes méthodes de RD basées sur les valeurs propres et les résultats de la classification VGG19 est effectuée.

4.1 Réduction de dimensionnalité pour la reconnaissance faciale

La réduction de la dimensionnalité est un ensemble de méthodes visant à sélectionner ou à extraire un sous-ensemble optimal de caractéristiques pertinentes dans les données multidimensionnelles selon des critères bien définis. Elle permet non seulement de réduire ou de supprimer la redondance d'informations, le bruit, les informations non pertinentes mais aussi d'améliorer le traitement postérieur des données réduites. La RD permet aussi de réduire la complexité des algorithmes d'apprentissage ainsi que leurs temps d'exécution. De nombreuses méthodes de RD [124], [125] ont été développées et présentées dans la littérature. Cette pléthore d'algorithmes [126] peuvent être catégorisés en méthodes basées sur la projection et sur les réseaux de neurones, en méthodes de RD supervisées ou non supervisées. Cette classification est faite en fonction de la technique implémentée par ces méthodes. Diverses méthodes de RD basées sur la projection ont été proposées. En général, les méthodes de projection recherchent le meilleur sous-espace optimal en minimisant l'erreur de

projection. Les images d'un espace initial de dimension R^n sont projetées dans un espace final de dimension R^p tel que $p < n$. Les méthodes de projection rendent l'ensemble de données plus représentatif. La littérature distingue deux catégories de méthodes de projection :

- les méthodes de projection linéaires ;
- et les méthodes de projection non linéaires.

4.1.1 Méthodes de projection linéaire

Les méthodes de projection linéaire sont celles basées sur la projection des pixels de l'image initiale sur de nouveaux axes linéaires dans le sous-espace final. La littérature présente diverses méthodes de projection linéaire parmi lesquelles : la Transformé de Karhuen Loeve (TKL), l'Analyse en Composantes Principales (ACP) et ses variantes, l'Analyse Discriminante Linéaire (ADL), la séparation des sources, les Projections Préservant la Localité (PPL), l'échelle multidimensionnelle (MDS) et l'Analyse Factorielle (FA).

Karhuen et Loève ont proposé une méthode de projection linéaire appelée Karhuen Loeve Transform [127]. Cette méthode est basée sur le calcul des vecteurs propres. La TKL convertit un ensemble de variables corrélées en un ensemble de variables non corrélées. L'Analyse en Composantes Principales (ACP) [128] est une méthode statistique classique bien connue pour explorer la structure des données multidimensionnelles. C'est une technique linéaire de RD basée sur le calcul des valeurs propres comme TKL. L'objectif de l'ACP est de maximiser la variance des données projetées. L'ACP construit une représentation à faible dimension des données initiales en trouvant une base linéaire de dimensionnalité réduite pour les données de sorte que la variance des données soit la plus grande possible. L'Analyse Discriminante Linéaire(LDA) [129] extrait l'information discriminante en recherchant les directions de projection qui maximisent le rapport de la dispersion entre les classes. Le but principal de la LDA est d'utiliser la projection vers le meilleur espace vectoriel discriminant pour convertir les données initiales. La séparation des sources a été d'abord développée pour décomposer les mélanges de signaux. Elle utilise des moments d'ordre supérieur à 2 pour obtenir l'indépendance statistique des sources. Plusieurs algorithmes de RD basés sur la séparation des sources ont été développés, notamment l'algorithme FastICA ou l'algorithme d'identification aveugle de second ordre [130] ou Second Order Blind Identification(SOBI) en anglais. La Projection Préservant la Localité [131] est une approche de RD dont le principe est de construire un graphe de voisinage de l'ensemble des données. Cette approche de RD

utilise la notion de graphe du Laplacien, et une matrice de transformation qui fait correspondre les points de données à un sous-espace. L'Analyse Factorielle[132] décrit les relations de covariance entre de nombreuses variables en définissant les intercorrélations entre n variables et en considérant un ensemble de facteurs communs. L'échelle multidimensionnelle MDS [133] est une méthode statistique qui permet de cartographier les relations par paires de coordonnées. Contrairement à l'ACP, elle utilise une matrice de similarité/dissimilarité comme entrée. Le MDS génère des coordonnées qui résultent d'un ajustement linéaire optimal aux dissimilarités données entre les points dans la direction des moindres carrés. Le MDS est l'algorithme de base de la plupart des méthodes de projection non linéaire telles que Sammon' Mapping, ISOMAP et l'Analyse en Composantes Curvilignes (ACC).

Les méthodes de projection linéaire ne permettent pas de mettre en évidence la non-linéarité qui existe dans certaines données multidimensionnelles. Les méthodes linéaires de DR ne peuvent être appliquées avec succès aux données que lorsqu'elles sont linéairement séparables. Or, l'observation de certaines images multispectrales ou hyperspectrales montre qu'elles pourraient présenter une non-linéarité.

4.1.2 Méthodes de projection non linéaire

Les méthodes de projection non linéaire sont des techniques de RD basées sur les relations topologiques de voisinage entre les pixels les uns par rapport aux autres dans l'espace de projection. La littérature présente plusieurs méthodes de RD non linéaires parmi lesquelles : la cartographie de Sammon, l'intégration linéaire locale (LLE), la cartographie des caractéristiques isométriques ou Isometric Mapping en anglais (ISOMAP), les cartes propres laplaciennes, l'Analyse en Composantes Curvilignes (CCA) et ses variantes. La cartographie de Sammon [134] est une variante non linéaire de MDS. Cette méthode de RD est basée sur un mappage ponctuel de vecteurs spatiaux à n dimensions vers un espace p de dimension inférieure, de sorte que les informations inhérentes aux données soient maintenues. La cartographie de Sammon préserve la topologie du voisinage grâce à la conservation des distances euclidiennes en minimisant la fonction d'erreur par la descente du gradient. Roweis et al. [135] ont introduit l'intégration linéaire locale (LLE) qui est une méthode non linéaire de RD. Elle définit les représentations compactes des données à partir d'une variété sous-jacente. LLE détermine la représentation basée sur le poids de chaque point et de ses voisins en utilisant la linéarité locale du collecteur de données. Tenenbaum et al. [136] ont développé

une autre extension non linéaire du MDS appelée Isomap. Cette approche non linéaire de RD utilise la distance géodésique entre les données tandis que le MDS utilise la distance euclidienne. M. Belkin et al. [137] ont développé les Eigeimaps laplaciennes (LE). Comme LLE, Eigeimaps laplaciennes est une méthode non linéaire de RD qui utilise une approche locale. Il utilise également la linéarité locale pour trouver la représentation pondérée de chaque point et ses voisins. Pierre Demartines et al. [138] ont proposé l'Analyse en Composantes Curvilignes (CCA) pour traiter les données multidimensionnelles. Initialement appelée quantification et projection vectorielles (VQP), l'Analyse en Composantes Curvilignes est une méthode non linéaire de RD basée sur l'algorithme MDS. Cette approche reproduit et préserve la topologie locale de la distribution des individus dans un sous-espace réduit. Il est fréquemment interprété comme une extension non linéaire de l'ACP. L. Journaux et al. [139] ont proposé une variante de l'Analyse en Composantes Curvilignes nommée CCAint(PCA) pour réduire les images Landsat de 8 à 3 bandes. Cette approche combine l'ACP et l'Analyse en Composantes Curvilignes. L'ACP des images est d'abord extraite et ensuite l'Analyse en Composantes Curvilignes est appliquée aux images résultantes de cette transformation. C'est une méthode de RD qui combine les informations linéaires et non linéaires des données. En 2015, G. Cirrincione et al. [140] ont implémenté la méthode OnCCA qui est une variante du CCA. Cette variante est une implémentation temps réel du CCA. Elle hérite des mêmes caractéristiques que la CCA et suit des distributions non stationnaires de haute dimension. Les mêmes auteurs dans [141] ont proposé une amélioration de la méthode onCCA . Cette variante appelée Growing CCA (GCCA) présente une architecture incrémentale qui s'adapte aux données non stationnaires.

La [Figure 4.1](#) résume quelques méthodes de RD en RF.

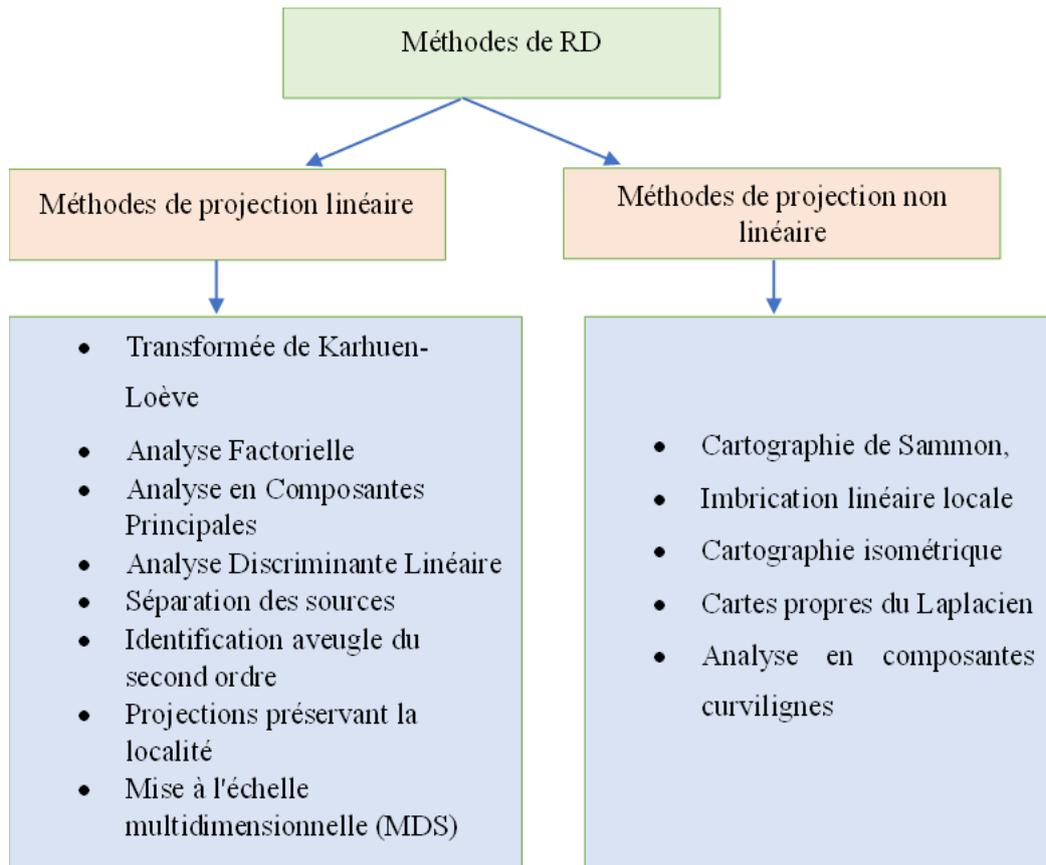


Figure 4.1- Quelques méthodes de réduction de dimensionnalité

4.1.3 Réduction de dimensionnalité pour les systèmes de reconnaissance faciale

La RD est utilisée à deux niveaux en RF : avant l'extraction de caractéristiques et avant le module de correspondance. Utilisée avant le module d'extraction de caractéristiques, la RD est appliquée aux images acquises directement. La RD permet aussi de réduire la taille des vecteurs caractéristiques extraites. Afin de mener à bien cette étude, nous avons identifié quelques systèmes de RF avec RD dans la littérature.

J. Ma et al. [142] ont utilisé l'ACP pour réduire la dimensionnalité des caractéristiques extraites pour la reconnaissance faciale. Les vecteurs caractéristiques ont été réduits de 3727 à 887 avec une amélioration de la précision avec le classifieur SVM de 87,97 à 90,35%.

Les auteurs dans [143] ont utilisé le 2DPCA pour réduire la taille des images (112x92) de la base de données ORL. Les classificateurs SVM et KNN avaient été utilisés pour évaluer la performance de cette méthode. Les précisions obtenues varient selon la taille des images réduites et le classifieur utilisé. K.Nivedita et al. [22] ont utilisé l'ACP et la fraction de bruit minimale pour réduire la dimensionnalité des images hyperspectrales de 425 à 9 bandes. Les

résultats des deux méthodes ont été comparés en utilisant les valeurs propres des images réduites.

4.2 Etude comparative des méthodes de réduction de dimensionalités

4.2.1 Motivation

Les méthodes linéaires de RD ne peuvent être appliquées avec succès aux images que lorsque celles-ci sont linéairement séparables. De plus, l'observation de certaines images multispectrales ou hyperspectrales montre qu'elles peuvent présenter des non-linéarités. Ceci nous a motivé à utiliser dans cette étude les méthodes de projection linéaire telles que la Transformée de Karhuen Loève (TKL), l'Analyse Factorielle (AF), et une méthode non linéaire qui est l'Analyse en Composantes Curvilignes (ACC). La TKL est une technique très efficace pour la réduction de la dimensionnalité des images hyperspectrales sans perte d'informations pertinentes [141]. C'est une transformation appropriée pour l'élimination des informations redondantes. La TKL est plus adaptée à la décorrélation spectrale, comparée à d'autres transformées [144]. L'analyse en composantes curvilignes est l'une des méthodes non linéaires de RD. C'est une technique qui utilise la topographie des données et les distances pondérées [141], [145]. L'Analyse Factorielle est sur les facteurs latents. En effet, la TKL cherche à maximiser la variance des nouvelles composantes tandis que l'AF trouve les variables latentes dans un ensemble de données. Nous avons donc choisi ces trois méthodes de DR en raison de cette motivation, du temps d'exécution et des résultats obtenus.

4.2.2 Transformé de Karhuen-Loève

La transformée de Karhuen Loève est une méthode linéaire optimale de RD des données. Sa spécificité est de transformer des composantes corrélées en composantes non corrélées tout en préservant l'information pertinente. Cette transformation s'effectue en deux principales étapes : le calcul de la matrice de covariance et la détermination des valeurs propres de cette matrice. La matrice de covariance est une matrice carrée de dimensions égales au nombre de bandes spectrales.

Soit X la matrice représentant les n bandes d'images de telle sorte que chaque bande d'image X_i $i \in \{1, \dots, n\}$ soit la colonne i de X et que chacune de ses lignes représente une observation. Pour déterminer la matrice de covariance, il faut d'abord calculer la moyenne des variables, elle est décrite dans l'équation (4.1) par

$$\bar{X} = \frac{1}{N} \sum_{j=1}^N X \quad (4.1)$$

Où N désigne le nombre de pixels de l'image X . La matrice de covariance est définie dans l'équation (4.2)

$$C_x = E[(X - \bar{X})(X - \bar{X})^T] \quad (4.2)$$

Où E représente l'espérance mathématique et T désigne la transposition. L'objectif de la TKL est de trouver une transformation linéaire Y qui permet de déterminer une matrice carrée $A(n \times n)$, telle que $Y = AX$,

où les vecteurs lignes $a_i (i \in \{1, 2, \dots, n\})$ de la matrice A sont les vecteurs propres respectivement associés aux valeurs propres $\lambda_i (i \in \{1, 2, \dots, n\})$ de la matrice de covariance. Les vecteurs lignes a_i sont les composantes finales. Ceci définit alors l'Equation (4.3)

$$C_x X = \lambda_i X \quad (4.3)$$

Une matrice diagonale des valeurs propres est obtenue après transformation de la matrice de covariance.

$$C_y = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & 0 & \dots & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_n \end{bmatrix}$$

4.2.3 Analyse en Composante Curviligne

L'analyse en composantes curvilignes est une méthode de RD non linéaire basée sur le MDS. Il s'agit d'un réseau neuronal auto-organisé qui effectue la quantification vectorielle du sous-modèle dans l'ensemble de données (espace d'entrée) et la projection non linéaire (P) de ces vecteurs de quantification vers un espace de sortie, fournissant ainsi un déploiement révélateur du sous-modèle. L'ACC est un réseau de neurones composé de deux couches : une entrée et une sortie. Ce réseau prend en entrée le vecteur $x_i, i \in \{1, 2, \dots, N\}$ avec n la dimension l'image multispectrale ($n = 8$ dans le cas de cette étude) et la sortie est un vecteur de dimension p tel que $p \leq n$. La distance euclidienne entre $x_i: X_{ij} = d(x_i, x_j)$ et $y_i: Y_{ij} = d(y_i, y_j)$ est calculée. Selon l'auteur [15], il existe mathématiquement une fonction d'erreur qui caractérise la différence de topologies entre l'espace initial et le sous-espace de projection final. L'objectif principal de l'ACC est de minimiser cette fonction d'erreur définie par :

$$E = \frac{1}{2} \sum_i \sum_{j(i \neq j)} (X_{ij} - Y_{ij})^2 F(Y_{ij}, \lambda_y) \quad (4.5)$$

Où λ_y désigne le déploiement d'échelle et $F(Y_{ij}, \lambda_y)$ est une fonction décroissante définie par :

$$F(Y_{ij}, \lambda_y) = \begin{cases} 1 & \text{if } Y_{ij} \leq \lambda_y \\ 0 & \text{if } Y_{ij} > \lambda_y \end{cases} \quad (4.6)$$

4.2.4 Analyse Factorielle

L'analyse factorielle est une méthode linéaire de RD. Cette approche détermine les variables ou facteurs sous-jacents qui définissent l'intercorrélation d'un ensemble de variables observées. Ces variables sont ensuite définies comme des combinaisons linéaires des facteurs. Mathématiquement, la méthode d'analyse factorielle peut être décrite comme suit :

Soit le nombre de variables (X_1, X_2, \dots, X_p) , m le nombre de facteurs sous-jacents (F_1, F_2, \dots, F_m) et X_j la variable qui représente les facteurs latents.

$$X_j = a_{j1}F_1 + a_{j2}F_2 + \dots + a_{jm}F_m + e_j \quad (4.5)$$

Où $j = 1, 2, \dots, P (P = 8)$, e_j le facteur unique et a_j la charge des facteurs définie par les corrélations entre les variables, et le facteur.

Pour toutes les paires de variables (X_i, X_{i+1}) , l'objectif de l'AF est de trouver les facteurs tels que la corrélation entre les paires soit nulle. Une fois qu'une matrice de corrélation est calculée, les charges factorielles sont ensuite analysées pour voir quelles variables sont chargées sur quels facteurs. Essentiellement, l'AF est décrite par l'équation suivante :

$$R_{m \times m} - U_{m \times m}^2 = F_{m \times p} F'_{p \times m} \quad (4.6)$$

Où $R_{m \times m}$ fait référence à la matrice de corrélation, $U_{m \times m}^2$ représente la matrice diagonale des variances uniques de chaque variable, et $F_{m \times p}$ représente les charges factorielles communes.

4.3 Evaluation de la réduction de dimensionalité sur la reconnaissance faciale via les caméras MSFA one-shot

Les données volumineuses nécessitent la réduction de dimensionalité pour améliorer la performance du système de reconnaissance en terme score et de temps. La réduction de la dimensionalité peut se faire directement sur les données justes après leur acquisition ou après l'extraction des caractéristiques sur les vecteurs caractéristiques.

Le système de reconnaissance faciale amélioré est présenté dans la [Figure 4.2](#) .

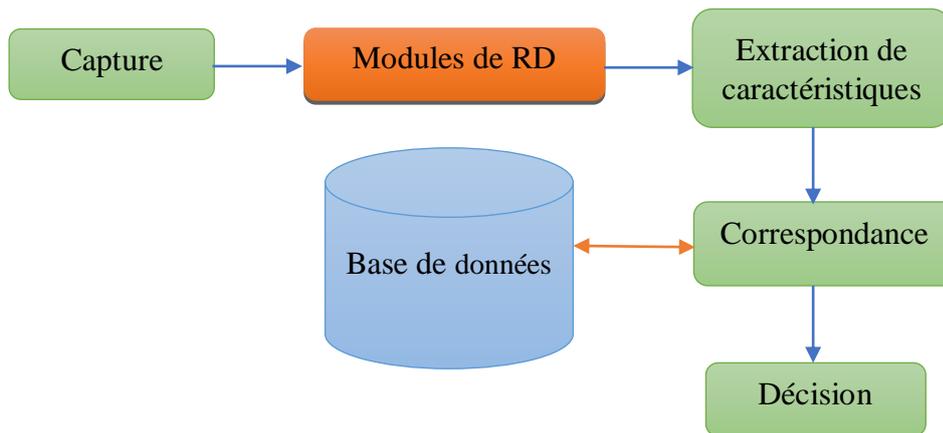


Figure 4.2- Système de reconnaissance faciale modifié.

L'évaluation de la méthode est décrite en deux étapes :

- La réduction des bandes spectrales ;
- L'utilisation du réseau neuronal VGG19 pour l'extraction de caractéristiques et la classification.

Dans la première étape, les trois méthodes de RD à savoir la TKL, l'ACC et l'AF ont été appliquées à chacune des images de la base de données. Chaque méthode prend en entrée une image multispectrale X composée de 8 bandes $X_i, i \in \{1, \dots, n\}, n = 8$. Elle renvoie après transformation une image de sortie X composée de p bandes telles que $X_i, i \in \{1, \dots, p\}, p < n$. Une comparaison des trois méthodes de RD utilisées est effectuée. Pour ce faire, les valeurs propres des composantes principales sont d'abord calculées pour déterminer la dimension intrinsèque des p composantes. La corrélation entre les p composantes réduites est calculée pour évaluer la méthode de réduction de bande qui fournit les composantes les moins corrélées.

La deuxième étape est basée sur l'extraction de caractéristiques et la classification des images réduites. Le CNN VGG19 est utilisé pour l'extraction de caractéristiques et la classification des images réduites. L'architecture de VGG19 a été décrite dans le chapitre 3. Les précisions obtenues avec VGG19 pour KLT, CCA et FA sont comparées afin de déterminer quelle méthode DR fournit une meilleure précision. Aussi une comparaison des temps de classification des images par méthode est faite.

4.4 Résultats et discussion

Avant de passer aux résultats, la Figure 4.3 montre une des images MS de la base de données EXIST après le processus de demosaicing.

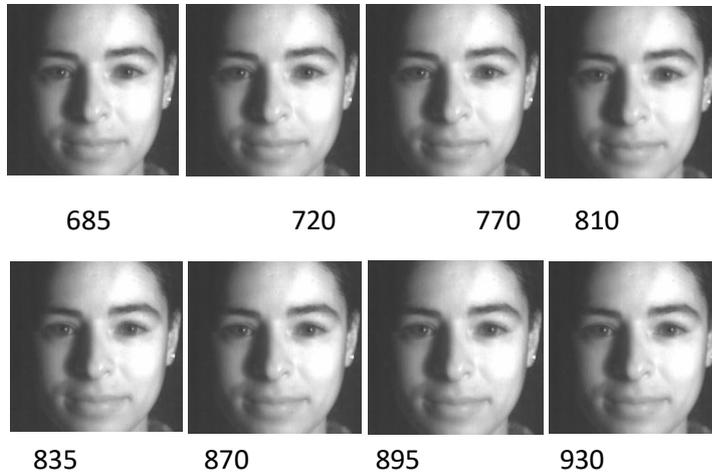


Figure 4.3-Echantillon d'images initiales

Comme décrit ci-dessus, les images de la base de données EXIST sont un ensemble d'images multispectrales (8 bandes) de visages. Ces images contiennent des informations redondantes, elles sont donc fortement corrélées. Le tableau 4.1 montre la moyenne des coefficients de corrélation des images (80) de la base de données EXIST.

Tableau 4.1- La moyenne des coefficients de corrélation de 80 images de la base de données EXIST.

| Bandes | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|--------|------|------|------|------|------|------|------|------------|
| 1 | 1 | 0.98 | 0.97 | 0.97 | 0.97 | 0.97 | 0.96 | 0.95 |
| 2 | 0.98 | 1 | 0.98 | 0.97 | 0.98 | 0.97 | 0.96 | 0.95 |
| 3 | 0.97 | 0.98 | 1 | 0.98 | 0.98 | 0.98 | 0.97 | 0.96 |
| 4 | 0.97 | 0.97 | 0.98 | 1 | 0.98 | 0.98 | 0.97 | 0.96 |
| 5 | 0.97 | 0.98 | 0.98 | 0.98 | 1 | 0.98 | 0.97 | 0.96 |
| 6 | 0.97 | 0.97 | 0.98 | 0.98 | 0.98 | 1 | 0.97 | 0.967 6 |
| 7 | 0.96 | 0.96 | 0.97 | 0.97 | 0.97 | 0.97 | 1 | 0.97 |
| 8 | 0.95 | 0.95 | 0.95 | 0.96 | 0.96 | 0.96 | 0.97 | 1 |

Les résultats montrent que le coefficient de corrélation entre les bandes est très proche de 1, ce qui confirme la forte corrélation entre les 8 bandes d'images de la base EXIST. Chacune

des trois méthodes de RD a été appliquée à chacune des images de la base de données. Les résultats obtenus avec ces trois méthodes de RD en termes de valeurs propres ont conduit au choix de la dimension intrinsèque (n=3). Le tableau 4.2 présente les valeurs propres cumulées des composantes obtenues.

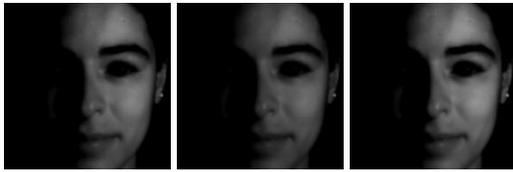
Tableau 4.2- Cummul (%) des valeurs propres

| Images | $\% \lambda_8$ | $\%(\lambda_1 + \lambda_2)$ | $\%(\lambda_1 + \lambda_2 + \lambda_3)$ | $\%(\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4)$ | $\%(\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5)$ | $\%(\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 + \lambda_6)$ | $\%(\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 + \lambda_6 + \lambda_7)$ | $\sum_{i=1}^8 \lambda_i$ |
|--------|----------------|-----------------------------|---|---|---|---|---|--------------------------|
| 1 | 98.7871 | 99.2067 | 99.4666 | 99.6246 | 99.7375 | 99.8284 | 99.9175 | 100 |
| 2 | 98.9550 | 99.5210 | 99.6432 | 99.7568 | 99.8394 | 99.9009 | 99.9528 | 100 |
| 3 | 98.7893 | 99.1991 | 99.4962 | 99.6478 | 99.7509 | 99.8452 | 99.9242 | 100 |
| 4 | 98.3726 | 99.1835 | 99.4152 | 99.5851 | 99.7479 | 99.8476 | 99.9250 | 100 |
| 5 | 96.7474 | 97.8150 | 98.3909 | 98.7893 | 99.1697 | 99.4563 | 99.7359 | 100 |
| 6 | 97.4816 | 98.0294 | 98.5301 | 98.8920 | 99.2075 | 99.4907 | 99.7587 | 100 |
| 7 | 96.8745 | 97.8027 | 98.3479 | 98.7962 | 99.1518 | 99.4776 | 99.7404 | 100 |
| 8 | 98.7797 | 99.1073 | 99.4012 | 99.5984 | 99.7124 | 99.8253 | 99.9161 | 100 |
| 9 | 98.7846 | 99.3688 | 99.5236 | 99.6436 | 99.7567 | 99.8558 | 99.9321 | 100 |
| 10 | 98.2215 | 98.9021 | 99.2353 | 99.4131 | 99.5878 | 99.7432 | 99.8767 | 100 |

La comparaison des différentes valeurs propres permet de sélectionner les trois dernières qui sont celles dont les valeurs sont les plus élevées. La première valeur propre concentre globalement plus de 96% de l'information de l'image. La sélection des trois premières concentre globalement plus de 99% de l'information de l'image. Les huit bandes sont alors réduites à 3 bandes. La [Figure 4.4](#) illustre les images réduites à 3 bandes.



a)



b)



c)

Figure 4.4- a) Images réduites avec le KLT, b) Images réduites avec le FA, c) Images réduites avec le CCA.

Les tableaux 4.3, 4.4 et 4.5 montrent les nouveaux coefficients de corrélation obtenus après la réduction de dimension.

Tableau 4.3- Moyenne des coefficients de corrélation avec TKL

| Bande | 1 | 2 | 3 |
|-------|--------|--------|--------|
| 1 | 1 | 0.0097 | 0.0341 |
| 2 | 0.0097 | 1 | 0.0861 |
| 3 | 0.0341 | 0.0861 | 1 |

Tableau 4.4- Moyenne des coefficients de corrélation avec AF

| Bande | 1 | 2 | 3 |
|-------|------|--------|--------|
| 1 | 1 | 1 | 0.97 |
| 2 | 1 | 1 | 0.9876 |
| 3 | 0.97 | 0.9876 | 1 |

Tableau 4.5- Moyenne des coefficients de corrélation avec ACC

| Bande | 1 | 2 | 3 |
|----------|--------|--------|--------|
| 1 | 1 | 0.9708 | 0.9584 |
| 2 | 0.9708 | 1 | 0.8630 |
| 3 | 0.9584 | 0.8630 | 1 |

La comparaison des coefficients de corrélation entre les bandes d'images réduites montre que la transformée de Karhuen Loeve donne les meilleurs résultats de décorrélation.

Ensuite, le système de reconnaissance faciale basé sur le réseau neuronal convolutif pré-entraîné VGG19 est utilisé pour extraire les caractéristiques et classifier les images réduites obtenues avec KLT, CCA et FA. Les ensembles de données étaient identiques pour toutes les méthodes. Les ensembles de données d'entraînement et de test ont été aléatoirement sélectionnées. Les tableaux 4.6 et 4.7 décrivent les paramètres d'apprentissage de VGG19 et les précisions obtenues.

Tableau 4.6- Paramètres utilisés dans la procédure d'entraînement

| Paramètres | Valeurs |
|----------------------------------|-----------------------------|
| Taille du lot | 32 |
| Algorithme d'optimisation | Stochastic Gradient Descent |
| Taux d'apprentissage | 0.0001 |
| Numéro d'époque | 20 |

Tableau 4.7 Résultats

| Méthodes | KLT +VGG19 | CCA +VGG19 | FA +VGG19 |
|-----------------|------------|------------|-----------|
| Accuracy | 100% | 90% | 89,4% |

En comparant les précisions obtenues avec les différentes méthodes de DR, la méthode KLT+VGG19 est la plus élevée et égale à 100%.

Pour vérifier l'impact de la réduction de dimensionnalité sur la RF avec les systèmes d'acquisition MSFA, nous comparons les précisions des images réduites et celle des images initiales. Pour ce faire le système développé dans le chapitre précédant a été utilisé. Le tableau 4.8 présente la précision et le temps d'exécution de l'algorithme.

Tableau 4.8- Résultats

| Methodes | KLT +VGG19 | CCA +VGG19 | FA +VGG19 | FDCT+VGG19 |
|--------------------|------------|------------|-----------|------------|
| Accuracy | 100% | 90% | 89,4% | 100% |
| Temps (min) | 3 | 6 | 7 | 6 |

Une précision de validation de 100% est obtenue avec les images réduites par KLT. Le temps d'exécution des images obtenues avec KLT + VGG19 est réduit de 15% par rapport à celui des images obtenues avec le FDCT + VGG19.

La réduction de bande des images multispectrales a permis d'optimiser le système de reconnaissance faciale avec les cameras MSFA one-shot.

Nous avons également comparé les résultats obtenus avec ceux de l'état de l'art. Le tableau 4.9 présente cette comparaison.

Tableau 4.9- Comparaison avec autres méthodes

| Base de données | Bandes initiales | Méthodes de RD | Bandes réduites | Classifieurs | Précision |
|---------------------------|-----------------------|----------------|-----------------|--------------|-------------|
| Orl(face database) | 112x92 | 2DPCA | 4,3 | KNN | 92%, 98,3% |
| orl | 112x92 | 2DPCA | 4,3 | SVM | 90%,90% |
| Feature vector | 3727 | PCA | 887 | SVM | 90.35 |
| Exist | 2072x 1104 x 8 | KLT | 3 | VGG19 | 100% |

La réduction de dimensionnalité peut être appliquée pour réduire la taille de l'image ou du vecteur caractéristique. Les méthodes de réduction de dimensionnalité améliorent la précision des classifieurs et du CNN.

Une comparaison des résultats obtenus montre que la précision varie en fonction des classifieurs et de la taille des images. Les résultats obtenus avec la base de données EXIST permettent de réduire l'image MS de 8 à 3 bandes.

Conclusion

Après avoir tiré la conclusion de la nécessité de réduire les bandes des images acquises avec nos caméras MSFA one-shot, une étude comparative de deux méthodes de projection linéaire KLT, FA et d'une méthode de projection non linéaire a été réalisée. Le calcul des coefficients de corrélation entre les 3 bandes des images réduites obtenues avec la TKL montre qu'elles sont moins corrélées. La TKL a permis une réduction optimale du nombre de bandes spectrales de 8 à 3. Cette transformée a donné les meilleurs résultats de reconnaissance faciale en termes de temps et de précision. La réduction de dimensionnalité avec la transformé de Karhuen Loève a conduit à une optimisation du système de reconnaissance faciale proposée.

Les travaux futurs se concentreront sur l'impact des algorithmes de demosaicing sur les systèmes de reconnaissance faciale avec les caméras MSFA one-shot.

Chapitre 5

Optimisation de la qualité de l'image et du temps de calcul : analyse approfondie de quatre méthodes de demosaicing pour les algorithmes de reconnaissance faciale

Sommaire

| | |
|---|-----|
| Introduction | 87 |
| 5.1 Motivation | 88 |
| 5.2 Description des méthodes sélectionnées | 89 |
| 5.2.1 Interpolation Bilinéaire | 89 |
| 5.2.2 Transformée discrète en ondelette | 89 |
| 5.2.3 Arbre binaire | 91 |
| 5.2.4 Vecteur médian | 92 |
| 5.3 Métriques d'évaluation | 93 |
| 5.4 Résultats expérimentaux et discussion | 95 |
| 5.4.1 Base de données et configuration expérimentales | 95 |
| 5.4.2 Résultats et discussion | 96 |
| Conclusion | 102 |

Introduction

L'utilisation de l'imagerie multispectrale est devenue de plus en plus populaire dans de nombreux domaines au cours de cette décennie, dans le but de surmonter les limitations de l'imagerie couleur. Il existe plusieurs types de systèmes d'acquisition multispectraux, parmi lesquels celui des caméras à prise de vue unique qui intègrent un MSFA. Ces systèmes d'acquisition sont très rapides, légers et capables d'acquérir des scènes en mouvement. Mais ces caméras sont fabriquées avec, au mieux, un logiciel de correction du positionnement des filtres, mais pas de logiciel de demosaicing. D'où la nécessité d'identifier un algorithme de demosaicing approprié en termes de qualité d'image, de temps de calcul, de facteur de décorrélation et de score de reconnaissance. En effet, Certaines méthodes de demosaicing [146]–[149] sont développées pendant la conception théorique du MSFA, mais les contraintes industrielles de la fabrication du MSFA font que les méthodes de demosaicing efficaces sont développées après la fabrication du MSFA. Les méthodes de demosaicing ont un impact sur la qualité de l'image et donc sur la performance du système de RF avec une caméra MSFA à prise de vue unique.

Pour cette étude, nous disposons déjà d'une caméra MSFA et nous voulons déterminer les méthodes de demosaicing appropriées en utilisant ses images pour la reconnaissance faciale. Ce système d'acquisition est principalement équipé d'un capteur CMOS viiamagic, d'un MSFA avec huit filtres, d'un micro-lentille, d'une carte électronique pour piloter le capteur et d'une carte caméra pour l'acquisition des images. Les bandes MSFA ont été sélectionnées théoriquement à l'aide d'un algorithme génétique combiné à une application de reconnaissance faciale [105]. Ce système d'acquisition couvre la gamme spectrale de 650 à 950 nm.

Ce chapitre présente une étude comparative de quatre méthodes pertinentes de demosaicing pour le processus de reconnaissance faciale, en utilisant des images acquises avec une caméra MSFA à prise de vue unique conçue dans notre laboratoire. Pour atteindre cet objectif, nous avons sélectionné quatre méthodes de demosaicing à savoir :

- L'interpolation bilinéaire ;
- La transformée en ondelettes discrète ;
- L'arbre binaire ;
- Le vecteur médian.

Chacune de ces méthodes a été adaptée aux images multispectrales acquises avec notre caméra MSFA. Les évaluations ont d'abord été réalisées à l'aide de la mesure de performance

NIQE et du coefficient de corrélation. Ensuite, les images démosaïquées ont été utilisées pour entraîner le réseau neuronal VGG19 afin de déterminer la méthode de demosaicing qui contient le mieux les caractéristiques pertinentes pour la RF et le meilleur temps de calcul.

5.1 Motivation

La littérature distingue trois catégories de méthodes de demosaicing : l'interpolation de pixels, la transformation de fréquence et la probabilité d'occurrence ou Probability of Appearance (POA) en anglais. Les méthodes d'interpolation de pixels consistent à utiliser la valeur ou le poids des pixels voisins pour estimer la valeur d'un pixel manquant. L'interpolation bilinéaire comprend les méthodes telles que l'interpolation bilinéaire pondérée et l'arbre binaire. La transformation de fréquence consiste à utiliser des ondelettes pour extraire des informations essentielles des pixels voisins, afin d'estimer la valeur du pixel manquant. La transformée en ondelettes discrètes est l'une des méthodes de transformation de fréquence. La probabilité d'occurrence fait référence aux méthodes qui déterminent la probabilité d'occurrence d'une bande et d'un pixel dans une bande sélectionnée afin d'estimer la valeur du pixel manquant. Ces méthodes incluent les approches basées sur l'arbre binaire et la détection des contours. En général, les méthodes de demosaicing dépendent du moxel du MSFA, mais il existe maintenant des méthodes génériques de demosaicing qui peuvent être adaptées à n'importe quel type de MSFA. La transformée en ondelettes discrète, l'interpolation bilinéaire, l'arbre binaire et le vecteur médian ont été sélectionnés pour cette étude comparative. Ces quatre méthodes de demosaicing ont été choisies parce que chacune d'elle présente des caractéristiques intéressantes pour l'étude. Les raisons de cette sélection sont les suivantes :

- L'interpolation bilinéaire utilise la valeur ou les poids des pixels voisins pour estimer un pixel manquant dans une bande donnée. La plupart des algorithmes de demosaicing combine leur technique de demosaicing avec l'interpolation bilinéaire [150]–[152].
- La méthode de détection des contours basée sur un arbre binaire est une approche générique qui utilise la notion de Probabilité d'apparition pour sélectionner la bande et le pixel dans la bande choisie avec précision. Cette méthode combine la POA et l'interpolation bilinéaire basée sur les informations de détection des contours [153] pour déterminer la valeur du pixel manquant ;
- La transformée en ondelettes discrète [16] est une technique qui utilise les informations de fréquence et l'interpolation bilinéaire pondérée pour approximer la valeur du pixel manquant ;

- Le filtrage médian vectoriel [154] est une méthode de demosaicing qui utilise les opérations vectorielles et le concept de pseudo-pixel. Cette méthode regroupe les pixels voisins en fonction de la taille du moxel pour évaluer la valeur du pixel manquant.

5.2 Description des méthodes sélectionnées

5.2.1 Interpolation Bilinéaire

Le demosaicing basé sur l'interpolation bilinéaire est la méthode la plus simple pour calculer la valeur des pixels manquants. D'autres méthodes de demosaicing combinent leurs spécificités avec l'interpolation bilinéaire[155]. L'interpolation bilinéaire approxime la valeur de chaque pixel manquant au moyen d'une moyenne pondérée par la distance de ses pixels voisins. Comme son nom l'indique, l'interpolation bilinéaire est une succession de deux interpolations linéaires. Elle peut être effectuée dans plusieurs directions. Pour un pixel manquant $P(i, j)$ à la position (i, j) , l'interpolation linéaire est définie par les équations 3.4, 3.5, et 3.6.

Pour cette étude, un filtre de convolution H est utilisé pour le demosaicing. Ce filtre est défini en termes de distance spatiale entre les voisins du pixel central. La bande d'image interpolée I^k est définie par l'équation (3.6) avec

$$H = \begin{pmatrix} 5 & 5 & 1 & 1 & 6 & 6 & 8 & 8 & 1 & 1 & 6 & 6 \\ 5 & 5 & 1 & 1 & 6 & 6 & 8 & 8 & 1 & 1 & 6 & 6 \\ 4 & 4 & 8 & 8 & 3 & 3 & 4 & 4 & 5 & 5 & 3 & 3 \\ 4 & 4 & 8 & 8 & 3 & 3 & 4 & 4 & 5 & 5 & 3 & 3 \\ 2 & 2 & 7 & 7 & 5 & 5 & 2 & 2 & 7 & 7 & 8 & 8 \\ 2 & 2 & 7 & 7 & 5 & 5 & 2 & 2 & 7 & 7 & 8 & 8 \\ 5 & 5 & 1 & 1 & 6 & 6 & 8 & 8 & 1 & 1 & 6 & 6 \\ 5 & 5 & 1 & 1 & 6 & 6 & 8 & 8 & 1 & 1 & 6 & 6 \\ 4 & 4 & 8 & 8 & 3 & 3 & 4 & 4 & 5 & 5 & 3 & 3 \\ 4 & 4 & 8 & 8 & 3 & 3 & 4 & 4 & 5 & 5 & 3 & 3 \\ 2 & 2 & 7 & 7 & 5 & 5 & 2 & 2 & 7 & 7 & 8 & 8 \\ 2 & 2 & 7 & 7 & 5 & 5 & 2 & 2 & 7 & 7 & 8 & 8 \end{pmatrix}$$

5.2.2 Transformée discrète en ondelette

La transformée en ondelettes discrète ou Discrete Wavelet Transform (DWT) en anglais est une transformée qui décompose un signal donné en un certain nombre d'ensembles, où chaque ensemble est une série temporelle de coefficients. Appliquée à une image, la DWT transforme l'image en différentes bandes de fréquences. La DWT est utilisée

pour décomposer les images en une série de sous-bandes avec différentes composantes de fréquence.

Plusieurs méthodes basées sur la DWT ont été proposées dans [156]–[160] pour le demosaicing du CFA au fil des dernières années. Xingbo et al. [149] avaient étendu l'application du DWT au demosaicing MSFA. Cette approche englobe le concept des images sous échantillonnées, l'ondelette de Haar, la règle de "remplacement" pour l'estimation des sous-bandes à haute fréquence et l'interpolation bilinéaire pour l'estimation des sous-bandes à basse fréquence. Cette méthode est applicable à toutes les MSFA avec des motifs de mosaïque réguliers, quel que soit le nombre de canaux. L'algorithme se déroule en trois principales étapes décrites comme suit :

a. Estimation des coefficients à haute fréquence :

L'image est d'abord décomposée en K images sous-échantillonnées (Down Sampling). Ensuite chacune des images sous échantillonnée est décomposée en sous-bandes de fréquence spatiale par DWT à l'aide de l'ondelette de Haar. Enfin une estimation des coefficients des images DS manquantes dans les sous-bandes à haute fréquence est faite en utilisant la règle du "remplacement".

b. Estimation des coefficients de basse fréquence

L'image est d'abord décomposée en K images sous-échantillonnées (ou Down Sampling DS). Ensuite chacune des images sous échantillonnée est décomposée en sous-bandes de fréquence spatiale par DWT à l'aide de l'ondelette de Haar. Enfin une estimation des coefficients des images DS manquantes dans les sous-bandes à basse fréquence est faite en utilisant la règle du "remplacement".

c. Reconstitution de l'image décomposée

Recomposer les composantes de basse fréquence et haute fréquence puis calculer la transformée en ondelettes discrète inverse pour reconstruire l'image.

Cette méthode a été testée par les auteurs avec le CFA et deux MSFA de 4 et 8 bandes du visible. Étant donné la compatibilité de cette méthode avec toutes les MSFA ayant des motifs réguliers, quel que soit le nombre de canaux, nous l'avons modifiée pour notre MSFA à 8 bandes décrite ci-dessus avec 2 pixels par bande. Pour cet algorithme, les longueurs d'onde des bandes ont été modifiées en tenant compte des longueurs d'onde suivantes {685, 720, 770, 810, 835, 870, 895, 930} nm pour les huit bandes.

5.2.3 Arbre binaire

Lidan Miao et al [151]–[153] avaient proposé une méthode générique de demosaicing basée sur l'arbre binaire. Cette approche utilise spécifiquement la méthode de détection des bords de l'arbre binaire appelée Binary Tree Edge Sensing (BTES) en anglais. Les pixels manquants sont estimés progressivement en utilisant toutes les informations de corrélation des bords de toutes les bandes spectrales. Le demosaicing d'une MSFA basé sur un arbre binaire peut être adapté à n'importe quelle MSFA 4×4 . L'algorithme générique de demosaicing se compose de trois modules interconnectés : la sélection des bandes, la sélection des pixels et l'interpolation [20].

- **Sélection des bandes** : ce module définit l'ordre d'interpolation des différentes bandes spectrales, en utilisant la POA. En effet, les bandes spectrales ont des POA différents, et la bande spectrale ayant la POA la plus élevée contient les informations les plus détaillées. La sélection de bandes équivaut à choisir des nœuds feuilles (bandes spectrales) à différents niveaux de l'arbre. Les nœuds situés au même niveau ont les mêmes POA, et les nœuds les plus profonds ont les POA les plus bas. Le processus de sélection des bandes est décrit comme suit :

Tout d'abord, la bande ayant la POA la plus élevée est sélectionnée, c'est-à-dire le nœud feuille au premier niveau de l'arbre binaire. Ensuite, pour les niveaux comportant plus d'une feuille, les bandes du niveau suivant sont sélectionnées au hasard. Enfin, ce processus est répété jusqu'au dernier niveau de l'arbre. La bande contenant le plus d'informations sur les bords est interpolée en premier, et les informations sur les bords de la première bande interpolée sont utilisées pour estimer les pixels des autres bandes.

- **Sélection des pixels** : ce module détermine l'ordre dans lequel les pixels de chaque bande spectrale sont interpolés. Les valeurs des pixels sont estimées progressivement. Cet algorithme utilise l'arbre binaire pour la sélection des pixels. Il prend en entrée les motifs de feuilles sélectionnés lors de la sélection des bandes et interpole pour chacun d'eux d'abord les informations de bande manquantes aux emplacements des pixels où se trouve son motif frère, puis l'algorithme remonte d'un niveau dans l'arbre binaire pour trouver le motif frère de son motif parent. Si ce dernier est un nœud interne, les motifs de feuilles du sous-arbre situé en dessous de ce motif frère sont examinés. Cette étape est répétée jusqu'à ce que la racine de l'arbre soit atteinte.

- **L'interpolation** est utilisée pour estimer la valeur des pixels manquants sélectionnés dans une bande donnée lors des deux premières étapes. La valeur estimée d'un pixel P à la

position (i, j) est calculée par la somme pondérée de ces quatre pixels voisins. Les poids de ces quatre pixels est estimé sur la base de l'ampleur de leurs bords. Les poids de deux pixels voisins dans les directions verticale et horizontale sont calculés par les équations 5.1 et 5.2.

Verticalement :

$$W_{m,n} = \left(1 + |P_{m+2,n} - P_{m,n}| + |P_{m-2,n} - P_{m,n}| + \frac{1}{2} |P_{m-1,n-1} - P_{m+1,n-1}| + \frac{1}{2} |P_{m-1,n+1} - P_{m+1,n+1}| \right)^{-1} \quad (5.1)$$

avec $m \in \{i - 1, i + 1\}, n = j$

Horizontalement :

$$W_{m,n} = \left(1 + |P_{m,n+2} - P_{m,n}| + |P_{m,n-2} - P_{m,n}| + \frac{1}{2} |P_{m+1,n-1} - P_{m+1,n+1}| + \frac{1}{2} |P_{m-1,n-1} - P_{m-1,n+1}| \right)^{-1} \quad (5.2)$$

avec $m = i, n \in \{j - 1, j + 1\}$

La Valeur estimée $\hat{P}_{i,j}$ du pixel à la position (i, j) est définie dans l'Equation 5.3.

$$\hat{P}_{i,j} = \frac{\sum_{s,t(|s+t|=1)} W_{i+s,j+t} P_{i+s,j+t}}{\sum_{s,t(|s+t|=1)} W_{i+s,j+t}} \quad (5.3)$$

L'algorithme avait été développé pour 7 bandes. Étant donné que le MSFA utilisé comporte 8 bandes, cet algorithme a été modifié pour 8 bandes, en tenant compte de la distribution du moxel de notre MSFA.

5.2.4 Vecteur médian

Gupta et al. [161] ont proposé un algorithme de demosaicing basé sur le vecteur médian pour le CFA. Cette approche sélectionne le vecteur de couleur qui minimise la somme des distances entre les pixels voisins pour estimer les couleurs manquantes. Cette approche de demosaicing est basée sur la notion de pseudo-pixel, qui est définie par un regroupement de valeurs de pixels rouges, verts et bleus voisins (horizontalement et verticalement) pour le CFA.

Xingbo et al. [154] ont étendu cette technique au demosaicing MSFA. Selon les auteurs, cette méthode est basée sur deux caractéristiques spécifiques : premièrement, les pseudo-pixels sont formés en fonction de la dimension d'un moxel. Deuxièmement, les pseudo-pixels sont ceux qui sont connectés horizontalement, verticalement et diagonalement.

Le vecteur médian \underline{x}_{VM} de $\underline{x}_1, \dots, \underline{x}_N$ est défini comme suit :

$$\underline{x}_{VM} \in \{\underline{x}_i | i = 1, \dots, N\}$$

$$\text{Pour tout } j = 1, \dots, N \quad \sum_{i=1}^N \|\underline{x}_{VM} - \underline{x}_i\|_2 \leq \sum_{i=1}^N \|\underline{x}_j - \underline{x}_i\|_2 \quad (5.4)$$

L'algorithme de calcul du vecteur médian pour le demosaicing des images multispectrales se déroule comme suit :

- Pour chaque vecteur \underline{x}_i , calculer la somme des distances à tous les autres vecteurs à l'aide de la norme L^1 -norm or L^2 -norm, comme indiqué dans l'équation (5.5).

$$S_i = \sum_{j=1}^N \|\underline{x}_i - \underline{x}_j\|, i = 1, \dots, N \quad (5.5)$$

- Le vecteur médian est \underline{x}_{min} correspondant à S_{min} , qui est le minimum de S_i .

Cette méthode a été testée par les auteurs avec le CFA, un MSFA à 4 bandes et un MSFA à 8 bandes. Pour cet algorithme, les longueurs d'onde des bandes ont été modifiées en considérant les longueurs d'onde suivantes {685, 720, 770, 810, 835, 870, 895, 930} nm pour les huit bandes de notre MSFA.

5.3 Métriques d'évaluation

Le choix des métriques d'évaluation appropriées est une étape cruciale pour tout système de reconnaissance. En effet, ces mesures permettent d'évaluer la performance du système et de déterminer s'il peut produire de bons résultats sur de nouvelles données.

Afin d'identifier l'algorithme de demosaicing approprié, les évaluations ont été réalisées sur la base de trois critères : la qualité de l'image, le facteur de décorrélation et le taux de reconnaissance en utilisant respectivement le NIQE, le coefficient de corrélation et l'accuracy comme mesures de performance.

D'après la littérature [162]–[164], il existe deux types de métriques de performance pour la qualité de l'image : les mesures de la qualité avec référence complète, telles que le rapport signal/bruit maximal ou Peak Signal to Noise Ratio (PSNR), l'indice de similarité structurelle pour la mesure (SSIM), l'erreur quadratique moyenne (MSE) et les mesures de la qualité sans référence, telles que le NIQE, le BRISQE et le PIQE. Étant donné que la base de données d'images EXIST ne contient que des images brutes sans références, la mesure de performance NIQE est utilisée pour évaluer la qualité des images reconstruites à l'aide des quatre méthodes sélectionnées.

NIQE a été développé par des chercheurs de l'Université du Texas[165]. C.Kawan et al.[166] ont utilisé le *NIQE* pour comparer les méthodes de demosaicing des images Mastcam. Le *NIQE* est une mesure de qualité d'image sans référence qui utilise les caractéristiques statistiques de l'image d'entrée pour évaluer la qualité de l'image. Pour calculer la valeur *NIQE*, l'image est divisée en petites portions et les caractéristiques sont modélisées comme des distributions gaussiennes multivariantes (MVG). Le *NIQE* est défini dans l'équation 5.6.

$$D(v_1, v_2, \Sigma_1, \Sigma_2) = \sqrt{((v_1 - v_2)^T \left(\frac{\Sigma_1 + \Sigma_2}{2} \right)^{-1} (v_1 - v_2))} \quad (5.6)$$

Où v_1, v_2 et Σ_1, Σ_2 sont les vecteurs moyens et la matrice de covariance.

Une fois la qualité de l'image évaluée avec *NIQE*, l'inter-corrélation des images reconstruites est évaluée en calculant le coefficient de corrélation. Ce coefficient identifie le degré de similarité entre une paire d'images. Les images sont identiques lorsque le coefficient de corrélation est égal à 1. Pour calculer ce coefficient de corrélation, la méthode décrite dans [167] a été utilisée et définie par l'équation 5.7 :

$$CC = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \bar{A})^2)(\sum_m \sum_n (B_{mn} - \bar{B})^2)}} \quad (5.7)$$

Où A, B représentent deux images de taille (m,n) et \bar{A}, \bar{B} leur moyenne respective.

En outre, pour identifier la meilleure méthode de demosaicing avec la meilleure qualité d'image, le facteur de décorrélation, les images demosaiced ont été utilisées pour former quatre modèles basés sur le CNN VGG19. La métrique accuracy est utilisée pour évaluer les images après la classification. Pour rappel l'accuracy indique le pourcentage de prédictions correctes. Elle est définie dans l'équation 5.8.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5.8)$$

où TP (True Positive), TN (True Negative), FP (False Positive), FN (False Negative)

5.4 Résultats expérimentaux et discussion

5.4.1 Base de données et configuration expérimentales

Les expériences ont été menées sur la base d'images EXIST. Pour rappel, EXIST est une base de données d'images multispectrales qui a été collectée avec la caméra MSFA one-shot décrite ci-dessus dans le chapitre 2 de cette thèse. L'ensemble des données EXIST est composé de 2100 images brutes de visages de 110 sujets. Chaque image multispectrale a une taille de 2072 x 1104. La [Figure 5.1](#) montre quelques images de la base de données EXIST.

Les expériences sont réalisées sur Microsoft System Windows, version 2010, avec deux ordinateurs. Le premier est équipé d'un processeur Intel (R) Core (TM) i7-8565U et d'une mémoire RAM de 8 Go. Le second est équipé d'une unité de traitement graphique (GPU) NVIDIA Quadro P400 avec 32 Go de mémoire vive (RAM). Tout le code est développé dans le langage de programmation Matlab 2022 et Python 3.8.



Figure 5.1- Exemples d'images de la base de données EXIST

5.4.2 Résultats et discussion

Pour réaliser le test, les étapes suivantes ont été suivies pour chaque méthode :

- a) demosaicing des images brutes ;
- b) calcul des valeurs NIQE des images obtenues et du temps de demosaicing ;
- c) calcul des coefficients de corrélation des images obtenues après demosaicing ;
- d) reconnaissance des images démosaïquées ;
- e) Calcul du temps de reconnaissance.

Ces algorithmes conduisent aux résultats présentés dans la Figure 5.2 .

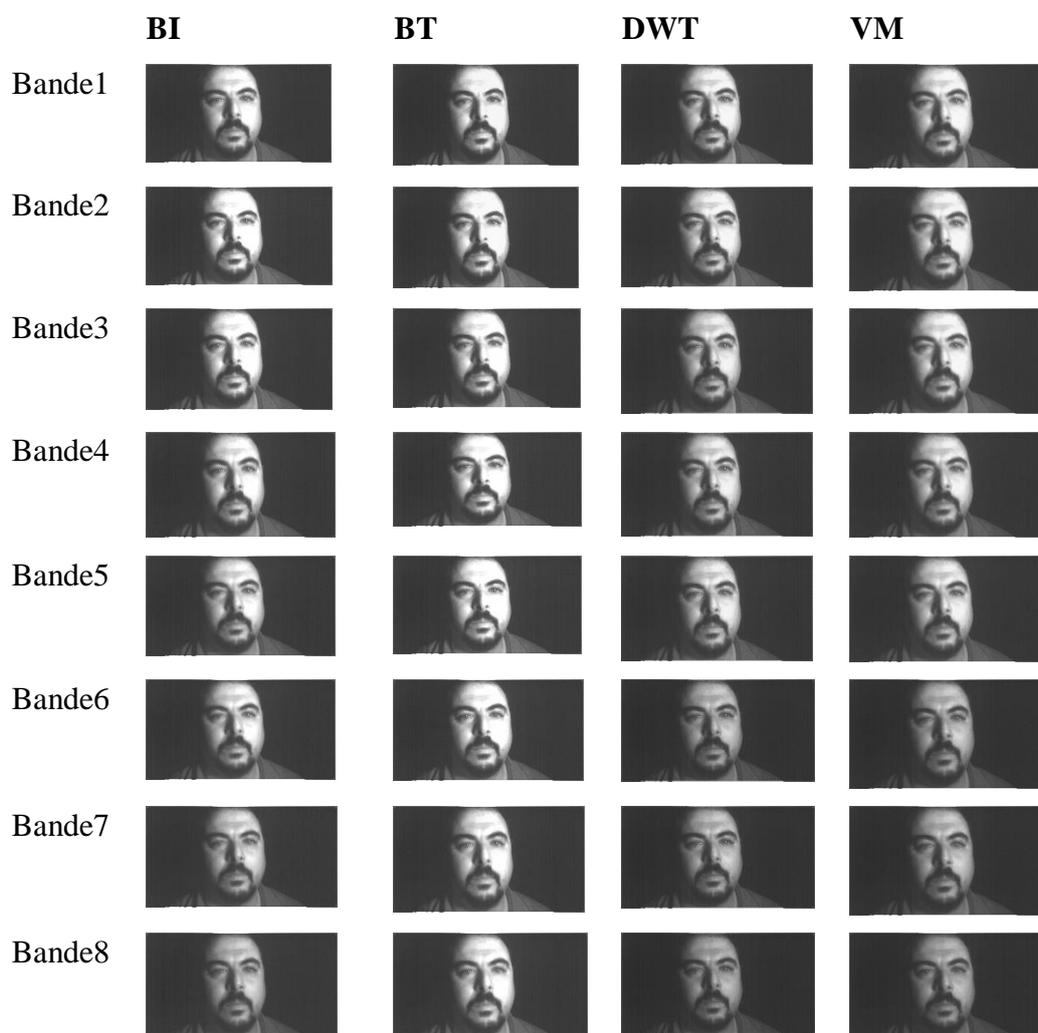


Figure 5.2- Images demosaiquées par méthode

Visuellement, nous pouvons constater que l'arbre binaire présente de meilleures images en termes de netteté des différentes parties du visage que les trois autres méthodes de demosaicing. Mais l'évaluation visuelle seule n'est pas suffisante. La métrique NIQE a donc

été utilisée pour confirmer ou infirmer cette évaluation visuelle. La qualité de l'image varie en fonction de la valeur du NIQE. Plus la valeur du NIQE est petite, meilleure est la qualité de l'image. La Figure 5.3 montre la valeur NIQE des images par méthode et par bande.

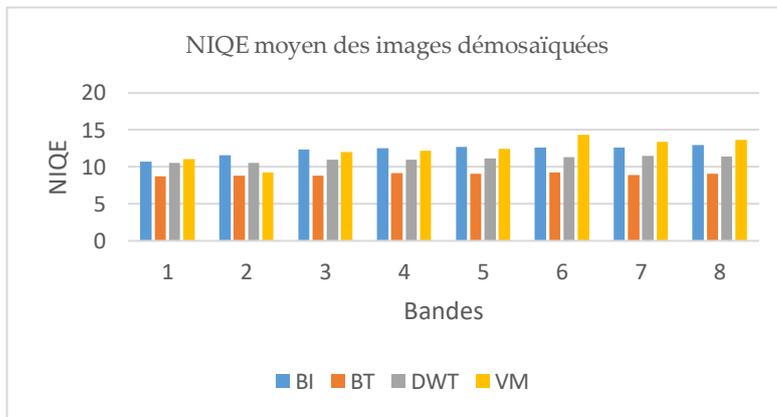


Figure 5.3- Moyenne des NIQE par image et par méthode

Pour toutes les méthodes de demosaicing, la valeur NIQE des images correspondantes varie entre 8,75 et 14,38 en fonction de la bande et de la méthode de demosaicing. Les valeurs NIQE des images obtenues par interpolation bilinéaire se situent entre 10,68 et 12,97, celles de BT entre 8,75 et 9,29, celles de DWT entre 10,51 et 11,42 et celles de VM entre 11,68 et 13,66.

L'analyse de la valeur NIQE par bande et par méthode de demosaicing montre que l'arbre binaire permet d'obtenir la valeur NIQE la plus faible (moyenne de 8,99). Sur la base de cette valeur NIQE, l'arbre binaire fournit des images de meilleure qualité que les autres. Cette analyse confirme l'observation visuelle faite après les résultats expérimentaux.

Des simulations temporelles ont ensuite été effectuées pour comparer les quatre méthodes. La figure 5.4 montre le temps d'exécution de chacune d'entre elles.

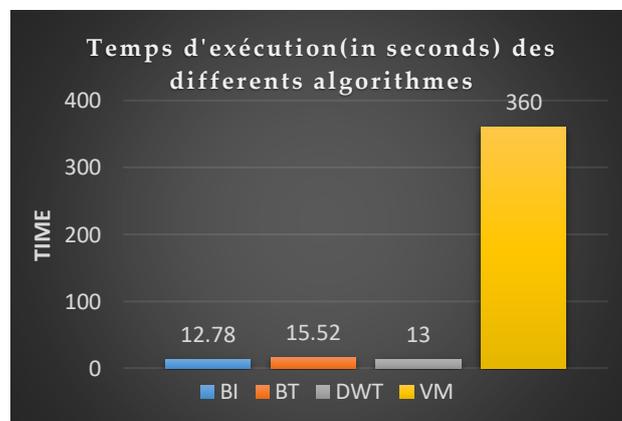


Figure 5.4- Temps d'exécution des différentes méthodes

L'interprétation de la figure 5.4 montre que le temps d'exécution minimal d'un algorithme de demosaicing est de 12,78 secondes et qu'il est obtenu avec l'interpolation bilinéaire. Ce temps est légèrement inférieur à celui de l'arbre binaire. Le vecteur médian est la méthode dont le temps d'exécution est le plus long. Le croisement du temps d'exécution et de la valeur NIQE des algorithmes de demosaicing montre que l'arbre binaire est la méthode de demosaicing qui permet d'avoir des images de meilleure qualité dans un temps raisonnable.

Pour étudier le facteur de décorrélation des images démosaïquées, nous avons calculé leurs coefficients de corrélation. Les tableaux 5.1, 5.2, 5.3 et 5.4 présentent la moyenne des coefficients de corrélation des images par méthode.

Tableau 5.1- Moyenne des coefficients de corrélation avec DWT

| Bande | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|--------------|----------|----------|----------|----------|----------|----------|----------|----------|
| 1 | 1 | 0,99 | 0,99 | 0,99 | 0,98 | 0,98 | 0,98 | 0,97 |
| 2 | 0,99 | 1 | 0,99 | 0,99 | 0,99 | 0,98 | 0,98 | 0,98 |
| 3 | 0,99 | 0,99 | 1 | 0,99 | 0,99 | 0,99 | 0,98 | 0,98 |
| 4 | 0,98 | 0,99 | 0,99 | 1 | 0,99 | 0,99 | 0,98 | 0,98 |
| 5 | 0,98 | 0,99 | 0,99 | 0,99 | 1 | 0,99 | 0,98 | 0,98 |
| 6 | 0,98 | 0,98 | 0,99 | 0,99 | 0,99 | 1 | 0,98 | 0,98 |
| 7 | 0,98 | 0,98 | 0,98 | 0,98 | 0,98 | 0,98 | 1 | 0,99 |
| 8 | 0,97 | 0,98 | 0,98 | 0,98 | 0,98 | 0,98 | 0,99 | 1 |

Tableau 5.2- Moyenne des coefficients de corrélation avec BI

| Bande | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|--------------|----------|----------|----------|----------|----------|----------|----------|----------|
| 1 | 1 | 0,98 | 0,97 | 0,97 | 0,97 | 0,97 | 0,96 | 0,95 |
| 2 | 0,98 | 1 | 0,98 | 0,97 | 0,98 | 0,97 | 0,96 | 0,95 |
| 3 | 0,97 | 0,98 | 1 | 0,98 | 0,98 | 0,98 | 0,97 | 0,96 |
| 4 | 0,97 | 0,97 | 0,98 | 1 | 0,98 | 0,98 | 0,97 | 0,96 |
| 5 | 0,97 | 0,98 | 0,98 | 0,98 | 1 | 0,98 | 0,97 | 0,96 |
| 6 | 0,97 | 0,97 | 0,98 | 0,98 | 0,98 | 1 | 0,97 | 0,9676 |
| 7 | 0,96 | 0,96 | 0,97 | 0,97 | 0,97 | 0,97 | 1 | 0,97 |
| 8 | 0,95 | 0,95 | 0,95 | 0,96 | 0,96 | 0,96 | 0,97 | 1 |

Tableau 5.3- Moyenne des coefficients de corrélation avec BT

| Bande | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|--------------|----------|----------|----------|----------|----------|----------|----------|----------|
| 1 | 1,00 | 0,99 | 0,98 | 0,98 | 0,98 | 0,98 | 0,98 | 0,97 |
| 2 | 0,99 | 1,00 | 0,99 | 0,98 | 0,98 | 0,98 | 0,98 | 0,98 |
| 3 | 0,98 | 0,99 | 1,00 | 0,99 | 0,99 | 0,99 | 0,98 | 0,98 |
| 4 | 0,98 | 0,98 | 0,99 | 1,00 | 0,99 | 0,99 | 0,98 | 0,98 |
| 5 | 0,98 | 0,98 | 0,99 | 0,99 | 1,00 | 0,99 | 0,98 | 0,98 |
| 6 | 0,98 | 0,98 | 0,99 | 0,99 | 0,99 | 1,00 | 0,99 | 0,98 |
| 7 | 0,98 | 0,98 | 0,98 | 0,98 | 0,98 | 0,99 | 1,00 | 0,98 |
| 8 | 0,97 | 0,98 | 0,98 | 0,98 | 0,98 | 0,98 | 0,98 | 1,00 |

Tableau 5.4 Moyenne des coefficients de corrélation avec VM

| Bande | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|--------------|----------|----------|----------|----------|----------|----------|----------|----------|
| 1 | 1,00 | 0,99 | 0,99 | 0,99 | 0,99 | 0,99 | 0,99 | 0,99 |
| 2 | 0,99 | 1,00 | 0,99 | 0,98 | 0,98 | 0,98 | 0,98 | 0,98 |
| 3 | 0,98 | 0,99 | 1,00 | 0,99 | 0,99 | 0,98 | 0,98 | 0,98 |
| 4 | 0,98 | 0,98 | 0,99 | 1,00 | 0,99 | 0,99 | 0,98 | 0,98 |
| 5 | 0,98 | 0,98 | 0,99 | 0,99 | 1,00 | 0,99 | 0,98 | 0,98 |
| 6 | 0,98 | 0,98 | 0,98 | 0,99 | 0,99 | 1,00 | 0,98 | 0,98 |
| 7 | 0,98 | 0,98 | 0,98 | 0,98 | 0,98 | 0,98 | 1,00 | 0,98 |
| 8 | 0,98 | 0,98 | 0,98 | 0,98 | 0,98 | 0,98 | 0,98 | 1,00 |

Chaque tableau indique le coefficient de corrélation moyen pour chaque méthode. Cependant, pris séparément, ils ne permettent pas d'apprécier l'inter-corrélation entre les bandes. Afin de comparer le degré de similarité entre les bandes, le coefficient de corrélation moyen est regroupé par bande. La [Figure 5.5](#) montre le coefficient de corrélation moyen entre chaque bande selon les méthodes de demosaicing.

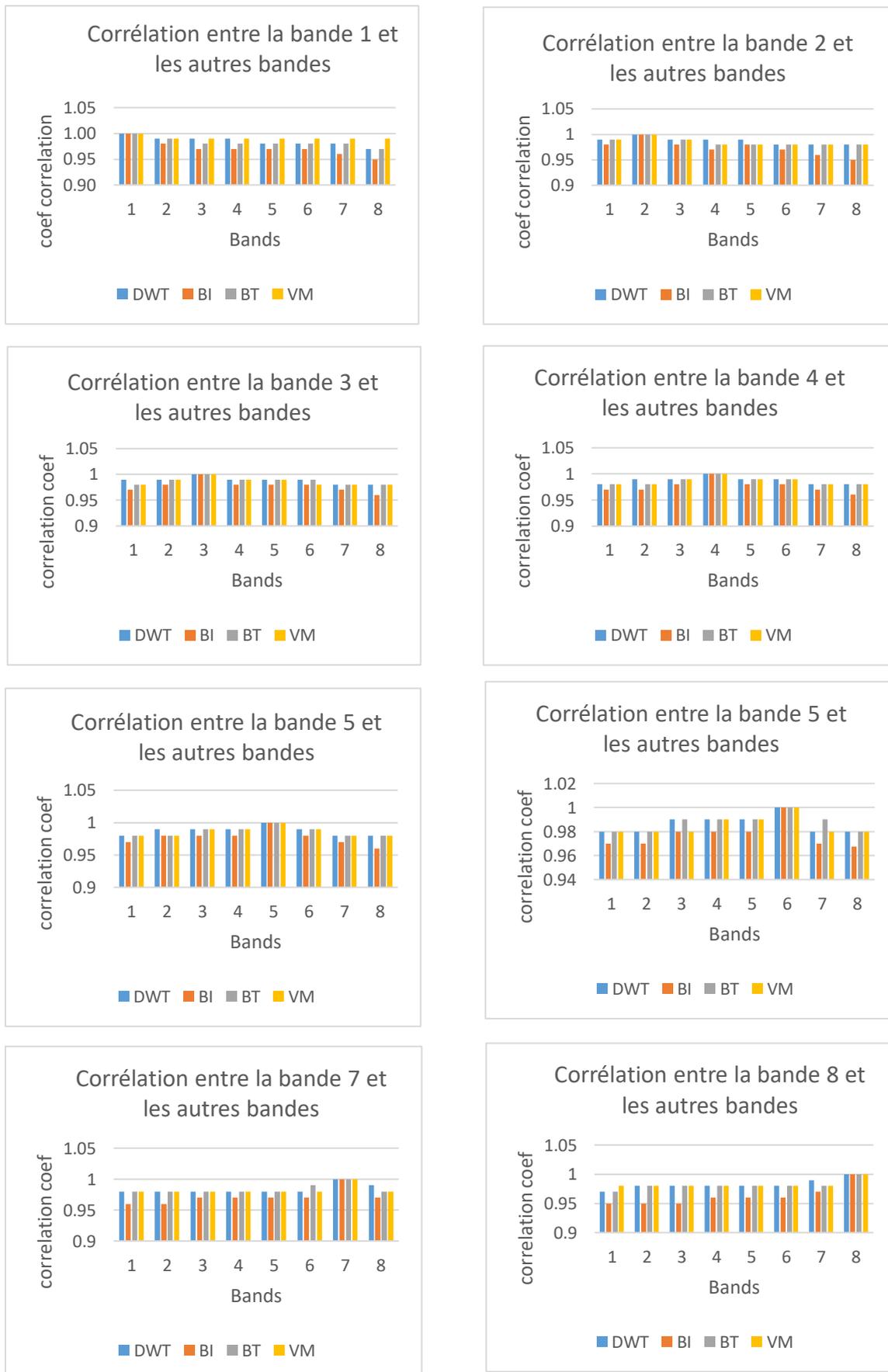


Figure 5.5- Moyenne des coefficients de corrélation pour les méthodes de demosaicing

Les images démosaïquées sont fortement corrélées avec un coefficient de corrélation moyen de 0,9 pour toutes les méthodes. L'analyse de ces différents chiffres nous permet de déduire que le facteur d'intercorrélation entre les différentes bandes est plus faible pour l'interpolation bilinéaire que pour les autres méthodes de demosaicing. L'interpolation bilinéaire permet d'avoir alors des images moins corrélées que les autres méthodes.

Après le processus de demosaicing et la comparaison, le CNN VGG19 a été utilisé pour l'extraction des caractéristiques et la classification des images obtenues par méthode de demosaicing. Les ensembles de données étaient identiques pour toutes les méthodes et sélectionnés aléatoirement. Les tableaux 5.5, 5.6 et 5.7 décrivent respectivement les paramètres d'apprentissage du VGG19, la précision et le temps de reconnaissance pour chaque image par méthode.

Tableau 5.5- Paramètres utilisés dans la procédure de d'entraînement

| Paramètres | Valeurs |
|----------------------------------|---------|
| Taille du lot | 32 |
| Algorithme d'optimisation | SGD |
| Taux d'apprentissage | 0.0001 |

Tableau 5.6- Accuracy

| Accuracy | BI | DWT | BT | VM |
|--------------|-----|-----|-----|----|
| VGG19 | 100 | 89 | 100 | 80 |

L'analyse des résultats du tableau 5.6 montre que les précisions obtenues varient entre 80% et 100%. Les images obtenues avec l'interpolation bilinéaire et l'arbre binaire ont une meilleure précision de 100%.

Tableau 5.7- Temps d'exécution

| Time | BI | DWT | BT | VM |
|--------------|----|-----|----|----|
| VGG19 | 3 | 4 | 2 | 6 |

La comparaison des résultats du tableau 5.7 montre que le temps d'exécution des images démosaïquées varie entre 2 et 6 s. Les images démosaïquées avec l'arbre binaire sont reconnues plus rapidement que celles des autres méthodes.

En conclusion, les différents résultats montrent que la méthode de demosaicing basée sur l'arbre binaire est la meilleure de ces quatre méthodes en ce qui concerne la qualité de l'image, le temps de calcul et la précision pour la reconnaissance faciale.

Conclusion

Nous avons proposé une évaluation des méthodes de demosaicing pour la reconnaissance faciale via des caméras MSFA one-shot afin d'identifier un algorithme de demosaicing approprié en termes de qualité d'image, de temps de calcul et de facteur de décorrélation. Cette étude utilise et compare quatre algorithmes de demosaicing sur une base d'images brutes acquises avec une caméra MSFA one-shot. La méthode de demosaicing par arbre binaire a été utilisée pour obtenir les images de meilleure qualité avec le meilleur temps de calcul et la meilleure précision pour la reconnaissance faciale. Le demosaicing a un impact sur le système de reconnaissance faciale en termes de qualité d'image et de temps. Plus les images sont de bonne qualité, plus le système de reconnaissance faciale est précis. Le temps de demosaicing minimal est de 12,78 secondes avec l'interpolation bilinéaire ce qui est énorme comparé au temps de demosaicing du CFA qui est de quelques millisecondes.

Une optimisation de ce système sera effectuée dans notre prochain travail afin de le rendre le système temps réel comme le CFA.

Conclusion et perspectives

L'imagerie multispectrale à filtre MSFA est utilisée pour l'acquisition d'images temps réel. Il est possible aujourd'hui de développer un tel type de système d'acquisition pour la reconnaissance faciale dans le visible et le proche infrarouge. Cependant, la conception et la fabrication d'une caméra MSFA one shot n'est pas une tâche triviale. Dans le cadre de cette thèse, l'objectif général était d'utiliser les MSFA et les approches d'apprentissage profond pour déployer une caméra MSFA one shot pour la reconnaissance faciale dans le visible et le proche infrarouge dans le but d'améliorer les systèmes de RF hétérogènes. Plus précisément, il était premièrement question de résoudre le problème d'indisponibilité de base d'images MS acquises avec les caméras MSFA. Deuxièmement, nous avons implémenté un système robuste de RF avec notre caméra MSFA. Pour ce faire, les méthodes de RF basées sur l'apprentissage profond et adaptées sont proposées. Pour finir, le système développé a été optimisé et une étude de l'impact du demosaicing sur la RF a été réalisée.

Pour atteindre les objectifs visés, quatre principales contributions correspondant aux objectifs spécifiques cités ont été proposées.

Premièrement, en ce qui concerne le manque d'images multispectrales acquises avec les caméras MSFA, une base d'images a été construite avec un système d'acquisition MSFA one shot qui opère de 650 à 950 nm. Cette base d'images MS est la première construite avec caméra MSFA pour la RF.

Deuxièmement, nous avons utilisé l'apprentissage profond pour proposer la méthode curvelet + CNN pour la RF. L'architecture des systèmes de RF a été présentée en tenant compte du module de demosaicing. La méthode proposée a permis d'atteindre de bonne performance de reconnaissance avec une précision satisfaisante de 100% obtenue avec le CNN VGG19.

Ensuite dans le but d'optimiser le système développé, plusieurs méthodes de RD ont été sélectionnées et testées sur la base d'images mise en place au cours de cette thèse. Parmi ces méthodes de RD, la TKL, l'AF et l'ACC ont été présentées. Au terme de cette étude, il a été montré que le TKL a permis d'optimiser le système avec une précision de 100% et une diminution du temps d'exécution de 15% par rapport au système de RF sans RD.

Pour finir une étude comparative des méthodes de demosaicing a été réalisée dans le but d'évaluer l'impact du demosaicing sur la RF. Pour ce faire les méthodes génériques de demosaicing à savoir l'interpolation bilinéaire, l'arbre binaire, la transformée discrète en ondelettes et le vecteur médian ont été choisies et adaptées à notre MSFA pour cette étude.

Les résultats expérimentaux ont montré que l'arbre binaire permet d'obtenir une meilleure qualité d'image moins corrélées et un meilleur score de reconnaissance.

Bien que les objectifs de cette thèse aient été atteints et que les solutions proposées aient produit de bons résultats, des extensions et améliorations de nos travaux sont envisagées. Il a été montré dans ces chapitres que le système proposé et optimisé atteint une précision de 100%. Cependant c'est seulement l'acquisition des images qui est temps réel, le processus de demosaicing n'est pas temps réel comme celui du CFA. L'une des perspectives de cette thèse est le développement d'algorithmes de demosaicing temps réel pour la reconnaissance faciale multispectrale. Dans nos travaux futurs nous allons nous intéresser aussi à la reconnaissance faciale multimodale avec les caméras MSFA one shot. Une autre perspective de cette thèse est l'étude de la RF basée sur la variation de l'âge avec les caméras MSFA one shot.

Bibliographie

- [1] Jiaqi Qiang, Danning Wu, Hanze Du, Huijuan Zhu, Shi Chen, and Hui Pan, “Review on Facial-Recognition-Based Applications in Disease Diagnosis,” *Bioengineering*, vol. 9, no. 7, Art. no. 7, Jul. 2022, doi: 10.3390/bioengineering9070273.
- [2] Diarra Mamadou, “Extraction et fusion de points d’intérêt et textures spectraux pour l’identification, le contrôle et la sécurité,” These de doctorat, Bourgogne Franche-Comté, 2018. Accessed: Apr. 14, 2023. [Online]. Available: <https://www.theses.fr/2018UBFCK031>
- [3] Lius Lopes Chambino, José Silvestre Silva, and Alexandre Bernardino, “Multispectral Facial Recognition: A Review,” *IEEE Access*, vol. 8, pp. 207871–207883, 2020, doi: 10.1109/ACCESS.2020.3037451.
- [4] Shwetank Arya, Neeraj Pratap, and Karamjit Bhatia, “Future of Face Recognition: A Review,” *Procedia Comput. Sci.*, vol. 58, pp. 578–585, Jan. 2015, doi: 10.1016/j.procs.2015.08.076.
- [5] Krishna Rao Kakkirala, Srinivasa Rao Chalamala, and Santosh Kumar Jami, “Thermal Infrared Face Recognition: A Review,” in *2017 UKSim-AMSS 19th International Conference on Computer Modelling & Simulation (UKSim)*, Apr. 2017, pp. 55–60. doi: 10.1109/UKSim.2017.38.
- [6] Shuowen Hu, Nathaniel Short, Benjamin S. Riggan, Matthew Chasse, and M. Saquib Sarfraz, “Heterogeneous Face Recognition: Recent Advances in Infrared-to-Visible Matching,” in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, May 2017, pp. 883–890. doi: 10.1109/FG.2017.126.
- [7] Shwetank Arya, Neeraj Pratap, and Karamjit Bhatia, “Future of Face Recognition: A Review,” *Procedia Comput. Sci.*, vol. 58, pp. 578–585, Jan. 2015, doi: 10.1016/j.procs.2015.08.076.
- [8] Mohamed Abdul-Al, George Kumi Kyeremeh, Raed A. Abd-Alhameed, Rami Qahwaji, Mohammad M. Abdul-Atty, Naser Ojaroudi Parchin, Jonathan Rodriguez and Ahmed S.I. Amar, “Types of Infrareads Focusing on Face Recognition: Promises, Advances and Challenges,” in *2022 International Telecommunications Conference (ITC-Egypt)*, Jul. 2022, pp. 1–6. doi: 10.1109/ITC-Egypt55520.2022.9855672.
- [9] Usman Cheema, Mobeen Ahmad, Dongil Han, and Seungbin Moon, “Heterogeneous Visible-Thermal and Visible-Infrared Face Recognition Using Cross-Modality

- Discriminator Network and Unit-Class Loss,” *Comput. Intell. Neurosci.*, vol. 2022, p. e4623368, Mar. 2022, doi: 10.1155/2022/4623368.
- [10] Fangzheng Huang, Xikai Tang, Chao Li, and Dayan Ban, “Near-infrared and visible light face recognition: a comprehensive survey,” *Soft Comput.*, Jun. 2023, doi: 10.1007/s00500-023-08366-8.
- [11] Pedro Martins, José Silvestre Silva, and Alexandre Bernardino, “Multispectral Facial Recognition in the Wild,” *Sensors*, vol. 22, no. 11, Art. no. 11, Jan. 2022, doi: 10.3390/s22114219.
- [12] Dorra Mahouachi, Moulay A. Akhloufi, “Recent Advances in Infrared Face Analysis and Recognition with Deep Learning,” *AI*, vol. 4, no. 1, Art. no. 1, Mar. 2023, doi: 10.3390/ai4010009.
- [13] Athreya V Shet, Chinmay BS, Anvithkumar A Shetty, T. Shankar, Hemavathy R, Ramakanth P, “Face Detection and Recognition in Near Infra-Red Image,” in *2022 6th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*, Dec. 2022, pp. 1–6. doi: 10.1109/CSITSS57437.2022.10026378.
- [14] Paramjit Kaur, Kewal Krishan, Suresh K. Sharma, Tanuj Kanchan, “Facial-recognition algorithms: A literature review,” *Med. Sci. Law*, vol. 60, no. 2, pp. 131–139, Apr. 2020, doi: 10.1177/0025802419893168.
- [15] Benoît Vibert, “Contributions à l’évaluation de systèmes biométriques embarqués,” PhD Thesis, Normandie, 2017.
- [16] Mohamad El-Abed, “Évaluation de système biométrique,” PHD thesis, Université de Caen, 2011. Accessed: Mar. 03, 2023. [Online]. Available: <https://theses.hal.science/tel-01007679>
- [17] Anthony Ngozichukwuka Uwaechia, D. A. Ramli, “A Comprehensive Survey on ECG Signals as New Biometric Modality for Human Authentication: Recent Advances and Future Challenges,” *IEEE Access*, vol. 9, pp. 97760–97802, 2021, doi: 10.1109/ACCESS.2021.3095248.
- [18] John McCarthy, “Generality in artificial intelligence,” *Commun. ACM*, vol. 30, no. 12, pp. 1030–1035, Dec. 1987, doi: 10.1145/33447.33448.
- [19] Lefteris Benos, Aristotelis C. Tagarakis, Georgios Dolias, Remigio Berruto, Dimitrios Kateris, Dionysis Bochtis, “Machine Learning in Agriculture: A Comprehensive Updated Review,” *Sensors*, vol. 21, no. 11, Art. no. 11, Jan. 2021, doi: 10.3390/s21113758.

- [20] Konstantinos G. Liakos, Patrizia Busato, Dimitrios Moshou, Simon Pearson, Dionysis Bochtis, “Machine Learning in Agriculture: A Review,” *Sensors*, vol. 18, no. 8, Art. no. 8, Aug. 2018, doi: 10.3390/s18082674.
- [21] Kee Yuan Ngiam, Ing Wei Khor, “Big data and machine learning algorithms for health-care delivery,” *Lancet Oncol.*, vol. 20, no. 5, pp. e262–e273, May 2019, doi: 10.1016/S1470-2045(19)30149-4.
- [22] Nivedita Priyadarshini, V. Sivashankari, Sulochana Shekhar, Karuppusamy Balasubramani, “Comparison and Evaluation of Dimensionality Reduction Techniques for Hyperspectral Data Analysis,” *Proceedings*, vol. 24, no. 1, Art. no. 1, 2019, doi: 10.3390/IECG2019-06209.
- [23] Pin Wang, En Fan, Peng Wang, “Comparative analysis of image classification algorithms based on traditional machine learning and deep learning,” *Pattern Recognit. Lett.*, vol. 141, pp. 61–67, Jan. 2021, doi: 10.1016/j.patrec.2020.07.042.
- [24] Ana Cristina Nunes Rodrigues, Ana Cristina Nunes Rodrigues, Rui Manuel Sousa Mendes, André Gonçalves Araújo, Micael Santos Couceiro, António José Figueiredo, “Using Artificial Intelligence for Pattern Recognition in a Sports Context,” *Sensors*, vol. 20, no. 11, Art. no. 11, Jan. 2020, doi: 10.3390/s20113040.
- [25] Israel González-Carrasco, Jose Luis Jiménez-Márquez, Jose Luis. López-Cuadrado, Belén Ruiz-Mezcua, “Automatic detection of relationships between banking operations using machine learning,” *Inf. Sci.*, vol. 485, pp. 319–346, Jun. 2019, doi: 10.1016/j.ins.2019.02.030.
- [26] Martin Leo, Suneel Sharma, Koilakuntla Maddulety, “Machine Learning in Banking Risk Management: A Literature Review,” *Risks*, vol. 7, no. 1, Art. no. 1, Mar. 2019, doi: 10.3390/risks7010029.
- [27] Iqbal H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions,” *SN Comput. Sci.*, vol. 2, no. 3, p. 160, Mar. 2021, doi: 10.1007/s42979-021-00592-x.
- [28] D. Praveen Kumar, Tarachand Amgoth, Chandra Sekhara Rao Annavarapu, “Machine learning algorithms for wireless sensor networks: A survey,” *Inf. Fusion*, vol. 49, pp. 1–25, Sep. 2019, doi: 10.1016/j.inffus.2018.09.013.
- [29] Corinna Cortes and Vladimir Vapnik, “Support-vector networks,” *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.

- [30] Peter N. Belhumeur, João P. Hespanha, and David J. Kriegman, “Eigenfaces vs. Fisherfaces: recognition using class specific linear projection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997, doi: 10.1109/34.598228.
- [31] Pallabi Parveen, Bhavani Thuraisingham, “Face Recognition Using Multiple Classifiers,” in *2006 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI’06)*, Nov. 2006, pp. 179–186. doi: 10.1109/ICTAI.2006.59.
- [32] Matthew Turk, Alex Pentland, “Eigenfaces for Recognition,” *J. Cogn. Neurosci.*, vol. 3, no. 1, pp. 71–86, Jan. 1991, doi: 10.1162/jocn.1991.3.1.71.
- [33] Igor Aleksander, W. V. Thomas, P. A. Bowden, “WISARD: a radical step forward in image recognition,” *Sens. Rev.*, vol. 4, no. 3, pp. 120–124, Mar. 1984, doi: 10.1108/eb007637.
- [34] C. MageshKumar, R. Thiyagarajan, S. P. Natarajan, S. Arulselvi, and G. Sainarayanan, “Gabor features and LDA based face recognition with ANN classifier,” in *2011 International Conference on Emerging Trends in Electrical and Computer Technology*, Mar. 2011, pp. 831–836. doi: 10.1109/ICETECT.2011.5760234.
- [35] Shaveta Dargan, Munish Kumar, Rohit Ayyagari, Gulshan Kumar, “A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning,” *Arch. Comput. Methods Eng.*, vol. 27, no. 4, pp. 1071–1092, Sep. 2020, doi: 10.1007/s11831-019-09344-w.
- [36] Shi Dong, Pin Wang, Khushnood Abbas, “A survey on deep learning and its applications,” *Comput. Sci. Rev.*, vol. 40, p. 100379, May 2021, doi: 10.1016/j.cosrev.2021.100379.
- [37] Christian Janiesch, Patrick Zschech, Kai Heinrich, “Machine learning and deep learning,” *Electron. Mark.*, vol. 31, no. 3, pp. 685–695, Sep. 2021, doi: 10.1007/s12525-021-00475-2.
- [38] Tahmid Hasan Fuad, Awal Ahmed Fime, Delowar Sikder, Akil Raihan Iftee, Jakaria Rabbi, Mabrook S. Al-Rakhami, Abdu Gumaiei, Ovishake Sen, Mohtasim Fuad, Nazrul Islam, “Recent Advances in Deep Learning Techniques for Face Recognition,” *IEEE Access*, vol. 9, pp. 99112–99142, 2021, doi: 10.1109/ACCESS.2021.3096136.
- [39] Yann LeCun, Yoshua Bengio, Geoffrey Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
- [40] Karen Simonyan and Andrew Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *ArXiv14091556 Cs*, Apr. 2015, Accessed: Oct. 13, 2021. [Online]. Available: <http://arxiv.org/abs/1409.1556>

- [41] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, “Going deeper with convolutions,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, pp. 1–9. doi: 10.1109/CVPR.2015.7298594.
- [42] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, “Deep Residual Learning for Image Recognition,” *ArXiv151203385 Cs*, Dec. 2015, Accessed: Oct. 13, 2021. [Online]. Available: <http://arxiv.org/abs/1512.03385>
- [43] Brett Koonce, “SqueezeNet,” in *Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization*, B. Koonce, Ed., Berkeley, CA: Apress, 2021, pp. 73–85. doi: 10.1007/978-1-4842-6168-2_7.
- [44] Qiong Cao, Li Shen, Weidi Xie, Omkar M. Parkhi, Andrew Zisserman, “VGGFace2: A Dataset for Recognising Faces across Pose and Age,” in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, May 2018, pp. 67–74. doi: 10.1109/FG.2018.00020.
- [45] Yandong Wen, Zhifeng Li, Yu Qiao, “Latent Factor Guided Convolutional Neural Networks for Age-Invariant Face Recognition,” *2016 IEEE Conf. Comput. Vis. Pattern Recognit. CVPR*, pp. 4893–4901, Jun. 2016, doi: 10.1109/CVPR.2016.529.
- [46] Guosheng Hu, Yang Hua, Yang Yuan, Zhihong Zhang, Zheng Lu, Sankha S. Mukherjee, Timothy M. Hospedales, Neil M. Robertson, Yongxin Yang, "Attribute-Enhanced Face Recognition with Neural Tensor Fusion Networks," *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 3764-3773, doi: 10.1109/ICCV.2017.404.
- [47] Mostafa Mehdipour Ghazi, Hazim Kemal Ekenel, “A Comprehensive Analysis of Deep Learning Based Representation for Face Recognition.” *arXiv*, Jun. 09, 2016. doi: 10.48550/arXiv.1606.02894.
- [48] Jin Chen, Jun Chen, Zheng Wang, Chao Liang, and Chia-Wen Lin, “Identity-Aware Face Super-Resolution for Low-Resolution Face Recognition,” *IEEE Signal Process. Lett.*, vol. 27, pp. 645–649, 2020, doi: 10.1109/LSP.2020.2986942.
- [49] Wanglong Wu, Meina Kan, Xin Liu, Yi Yang, Shiguang Shan, Xilin Chen, “Recursive Spatial Transformer (ReST) for Alignment-Free Face Recognition,” in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017, pp. 3792–3800. doi: 10.1109/ICCV.2017.407.
- [50] Qiangchang Wang, Tianyi Wu, He Zheng, Guodong Guo, "Hierarchical Pyramid Diverse Attention Networks for Face Recognition," *2020 IEEE/CVF Conference on*

- Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 8323–8332, doi: 10.1109/CVPR42600.2020.00835.
- [51] Peng Zhang, Feng Zhao, Peng Liu, Mengwei Li, “Efficient Lightweight Attention Network for Face Recognition,” *IEEE Access*, vol. 10, pp. 31740–31750, 2022, doi: 10.1109/ACCESS.2022.3150862.
- [52] Shun Cheung Lai, Minna Kong, Kin Man Lam, Dong Li, “High-Resolution Face Recognition Via Deep Pore-Feature Matching,” in *2019 IEEE International Conference on Image Processing (ICIP)*, Sep. 2019, pp. 3477–3481. doi: 10.1109/ICIP.2019.8803686.
- [53] Anastasiya Mishchuk, Dmytro Mishkin, Filip Radenovic, Jiri Matas, “Working hard to know your neighbor’s margins: Local descriptor learning loss” arXiv, Jan. 12, 2018. doi: 10.48550/arXiv.1705.10872.
- [54] Jiaolong Yang, Peiran Ren, Dongqing Zhang, Dong Chen, Fang Wen, Hongdong Li, Gang Hua, “Neural Aggregation Network for Video Face Recognition.” arXiv, Aug. 02, 2017. doi: 10.48550/arXiv.1603.05474.
- [55] Guodong Guo, Na Zhang, “A survey on deep learning based face recognition,” *Comput. Vis. Image Underst.*, vol. 189, p. 102805, Dec. 2019, doi: 10.1016/j.cviu.2019.102805.
- [56] Chenfei Xu, Qihe Liu, Mao Ye, “Age invariant face recognition and retrieval by coupled auto-encoder networks,” *Neurocomputing*, vol. 222, no. C, pp. 62–71, Jan. 2017, doi: 10.1016/j.neucom.2016.10.010.
- [57] Muhammad Usman, Siddique Latif, Junaid Qadir, “Using Deep Autoencoders for Facial Expression Recognition,” arXiv, Jan. 25, 2018. doi: 10.48550/arXiv.1801.08329.
- [58] Alperen Kantarcı, Hazım Kemal Ekenel, “Thermal to Visible Face Recognition Using Deep Autoencoders,” arXiv, Feb. 10, 2020. doi: 10.48550/arXiv.2002.04219.
- [59] Lele Cheng, Jinjun Wang, Yihong Gong, Qiqi Hou, “Robust Deep Auto-encoder for Occluded Face Recognition,” in *Proceedings of the 23rd ACM international conference on Multimedia*, in MM ’15. New York, NY, USA: Association for Computing Machinery, Oct. 2015, pp. 1099–1102. doi: 10.1145/2733373.2806291.
- [60] Landry Kezebou, Victor Oludare, Karen Panetta, Sos Agaian, “TR-GAN: thermal to RGB face synthesis with generative adversarial network for cross-modal face recognition,” in *Mobile Multimedia/Image Processing, Security, and Applications 2020*, SPIE, Apr. 2020, pp. 158–168. doi: 10.1117/12.2558166.

- [61] Ran He, Jie Cao, Lingxiao Song, Zhenan Sun, and Tieniu Tan, “Adversarial Cross-Spectral Face Completion for NIR-VIS Face Recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 5, pp. 1025–1037, May 2020, doi: 10.1109/TPAMI.2019.2961900.
- [62] Grigory Antipov, Moez Baccouche, Jean-Luc Dugelay, “Face aging with conditional generative adversarial networks,” *2017 IEEE Int. Conf. Image Process. ICIP*, pp. 2089–2093, Sep. 2017, doi: 10.1109/ICIP.2017.8296650.
- [63] Luan Tran, Xi Yin, and Xiaoming Liu, “Disentangled Representation Learning GAN for Pose-Invariant Face Recognition,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 1283–1292. doi: 10.1109/CVPR.2017.141.
- [64] Yi Sun, Ding Liang, Xiaogang Wang, Xiaoou Tang, “DeepID3: Face Recognition with Very Deep Neural Networks” arXiv, Feb. 03, 2015. doi: 10.48550/arXiv.1502.00873.
- [65] Gaurav Goswami, Romil Bhardwaj, Richa Singh, Mayank Vatsa, “MDLFace: Memorability augmented deep learning for video face recognition,” in *IEEE International Joint Conference on Biometrics*, Sep. 2014, pp. 1–7. doi: 10.1109/BTAS.2014.6996299.
- [66] Wuming Zhang, Zhixin Shu, Dimitris Samaras, Liming Chen, “Improving Heterogeneous Face Recognition with Conditional Adversarial Networks.” arXiv, Sep. 13, 2017. doi: 10.48550/arXiv.1709.02848.
- [67] Chi Nhan Duong, Khoa Luu, Kha Gia Quach, Nghia Nguyen, Eric Patterson, Tien D. Bui, Ngan Le, “Automatic Face Aging in Videos via Deep Reinforcement Learning,” *2019 IEEE CVF Conf. Comput. Vis. Pattern Recognit. CVPR*, pp. 10005–10014, Jun. 2019, doi: 10.1109/CVPR.2019.01025.
- [68] Stephen E. Harris, R. W. Wallace, “Acousto-Optic Tunable Filter*,” *JOSA*, vol. 59, no. 6, pp. 744–747, Jun. 1969, doi: 10.1364/JOSA.59.000744.
- [69] Steven C Gebhart, Reid C Thompson, Anita Mahadevan-Jansen, “Liquid-crystal tunable filter spectral imaging for brain tumor demarcation,” *Appl. Opt.*, vol. 46, no. 10, pp. 1896–1910, Apr. 2007, doi: 10.1364/AO.46.001896.
- [70] Lidan Miao, Hairong Qi, W.E. Snyder, “A generic method for generating multispectral filter arrays,” in *2004 International Conference on Image Processing, 2004. ICIP '04.*, Oct. 2004, pp. 3343–3346 Vol. 5. doi: 10.1109/ICIP.2004.1421830.

- [71] Lidan Miao, Hairong Qi, “The design and evaluation of a generic method for generating mosaicked multispectral filter arrays,” *IEEE Trans. Image Process.*, vol. 15, no. 9, pp. 2780–2791, Sep. 2006, doi: 10.1109/TIP.2006.877315.
- [72] Johannes Brauers, Til Aach, “A Color Filter Array Based Multispectral Camera,” Jan. 2006.
- [73] John Hershey, Zhengyou Zhang, “Multispectral digital camera employing both visible light and non-visible light sensing on a single image sensor,” US20060066738A1, Mar. 30, 2006 Accessed: Sep. 19, 2022. [Online]. Available: <https://patents.google.com/patent/US20060066738A1/en>
- [74] Shao-Wei Wang, Changsheng Xia, Xiaoshuang Chen, Wei Lu, Ming Li, Haiqian Wang, Weibo Zheng, Tao Zhang, “Concept of a high-resolution miniature spectrometer using an integrated filter array,” *Opt. Lett.*, vol. 32, no. 6, pp. 632–634, Mar. 2007, doi: 10.1364/OL.32.000632.
- [75] Yue M. Lu, Clément Fredembach, Martin Vetterli, Sabine Süsstrunk, “Designing color filter arrays for the joint capture of visible and near-infrared images,” 2009.
- [76] Travis W. Sawyer, Michaela Taylor-Williams, Ran Tao, Ruqiao Xia, Calum Williams, Sarah E. Bohndiek, “Opti-MSFA: a toolbox for generalized design and optimization of multispectral filter arrays,” *Opt. Express*, vol. 30, no. 5, pp. 7591–7611, Feb. 2022, doi: 10.1364/OE.446767.
- [77] Renjie Wu, Yuqi Li, Xijiong Xie, Zhijie Lin, “Optimized Multi-Spectral Filter Arrays for Spectral Reconstruction,” *Sensors*, vol. 19, no. 13, Art. no. 13, Jan. 2019, doi: 10.3390/s19132905.
- [78] K. Walls, Q. Chen, J. Grant, S. Collins, D. R. S. Cumming, and T. D. Drysdale, “Narrowband multispectral filter set for visible band,” *Opt. Express*, vol. 20, no. 20, pp. 21917–21923, Sep. 2012, doi: 10.1364/OE.20.021917.
- [79] Neelam Gupta, Philip R. Ashe, Songsheng Tan, “Miniature snapshot multispectral imager,” *Opt. Eng.*, vol. 50, no. 3, p. 033203, Mar. 2011, doi: 10.1117/1.3552665.
- [80] Bert Geelen, Nicolaas Tack, Andy Lambrechts, “A snapshot multispectral imager with integrated tiled filters and optical duplication,” in *Advanced Fabrication Technologies for Micro/Nano Optics and Photonics VI*, SPIE, Mar. 2013, pp. 173–185. doi: 10.1117/12.2004072.
- [81] Bert Geelen, Nicolaas Tack, Andy Lambrechts, “A compact snapshot multispectral imager with a monolithically integrated per-pixel filter mosaic,” in *Advanced*

- Fabrication Technologies for Micro/Nano Optics and Photonics VII*, SPIE, Mar. 2014, pp. 80–87. doi: 10.1117/12.2037607.
- [82] Vahid Mohammadi, S. Guillome Sodjinou, Kossi Khruma Katakpe, M. Rossé, pierre Gouton, *Development of a Multi-spectral Camera for Computer vision applications*. Václav Skala - UNION Agency, 2022. doi: 10.24132/CSRN.3201.4.
- [83] Yufeng Zheng, “Orientation-based face recognition using multispectral imagery and score fusion,” *Opt. Eng.*, vol. 50, p. 7202, Nov. 2011, doi: 10.1117/1.3643729.
- [84] D. Huang, Jia Sun, and Y. Wang, “The BUAA-VisNir face database instructions,” *Sch. Comput Sci Eng Beihang Univ Beijing China Tech Rep IRIP-TR-12-FR-001*, vol. 3, p. 3, 2012.
- [85] Virginia Espinosa-Duró, Marcos Faundez-Zanuy, Jiří Mekyska, “A new face database simultaneously acquired in visible, near infrared and thermal spectrum,” *Cogn. Comput.*, vol. 5, no. 1, pp. 119–135, Mar. 2013, doi: 10.1007/s12559-012-9163-2.
- [86] Virginia Espinosa-Duró, Marcos Faundez-Zanuy, Jiří Mekyska, “A New Face Database Simultaneously Acquired in Visible, Near-Infrared and Thermal Spectrums”, *Cognitive Computation*, vol. 5, no. 1, pp. 119-135, 2013
- [87] Stan Z. Li, Zhen Lei, Meng Ao, “The HFB Face Database for Heterogeneous Face Biometrics research,” in *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Miami, FL: IEEE, Jun. 2009, pp. 1–8. doi: 10.1109/CVPRW.2009.5204149.
- [88] Stan Z. Li, Dong Yi, Zhen Lei and Shengcai Liao, “The CASIA NIR-VIS 2.0 Face Database,” in *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, OR, USA: IEEE, Jun. 2013, pp. 348–353. doi: 10.1109/CVPRW.2013.59.
- [89] Aijing Yu, Haoxue Wu, Huaibo Huang, Zhen Lei, Ran He, “LAMP-HQ: A Large-Scale Multi-pose High-Quality Database and Benchmark for NIR-VIS Face Recognition,” *Int. J. Comput. Vis.*, vol. 129, no. 5, pp. 1467–1483, May 2021, doi: 10.1007/s11263-021-01432-4.
- [90] Guoying Zhao, Xiaohua Huang, Matti Taini, Stan Z. Li, Matti Pietikäinen, "Facial expression recognition from near-infrared videos", *Image and Vision Computing*, Volume 29, Issue 9, 2011, Pages 607-619, ISSN 0262-8856, <https://doi.org/10.1016/j.imavis.2011.07.002>.
- [91] Fei Wu, Xiao-Yuan Jing, Xiwei Dong, Ruimin Hu, Dong Yue, Lina Wang, Yi-Mu Ji, Ruchuan Wang, Guoliang Chen,” *IEEE Trans. Cybern.*, vol. 50, no. 3, pp. 1009–1022, Mar. 2020, doi: 10.1109/TCYB.2018.2876591.

- [92] Muhammad Uzair, Zohaib Khan, Arif Mahmood, Faisal Shafait, Ajmal Mian, March 28, 2023, "UWA Hyperspectral Face Database", IEEE Dataport, doi: <https://dx.doi.org/10.21227/8714-kx37>
- [93] John Bernhard, Jeremiah Barr, Kevin W. Bowyer, Patrick Flynn, "Near-IR to visible light face matching: Effectiveness of pre-processing options for commercial matchers," in *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, Arlington, VA, USA: IEEE, Sep. 2015, pp. 1–8. doi: 10.1109/BTAS.2015.7358780.
- [94] Shuowen Hu, Jonghyun Choi, Alex L. Chan, and William Robson Schwartz, "Thermal-to-visible face recognition using partial least squares," *JOSA A*, vol. 32, no. 3, pp. 431–442, Mar. 2015, doi: 10.1364/JOSAA.32.000431.
- [95] He Zhang, Benjamin S. Riggan, Shuowen Hu, Nathaniel J. Short, Vishal M. Patel, "Synthesis of High-Quality Visible Faces from Polarimetric Thermal Faces using Generative Adversarial Networks." arXiv, Dec. 12, 2018. doi: 10.48550/arXiv.1812.05155.
- [96] Olegs Nikisins, Kamal Nasrollahi, Modris Greitans, and Thomas B. Moeslund, "RGB-D-T Based Face Recognition," in *2014 22nd International Conference on Pattern Recognition*, Aug. 2014, pp. 1716–1721. doi: 10.1109/ICPR.2014.302.
- [97] Georges Bebis, Aglika Gyaourova, Saurabh Singh, Ioannis Pavlidis, "Face recognition by fusing thermal infrared and visible imagery," *Image Vis. Comput.*, vol. 24, no. 7, pp. 727–742, Jul. 2006, doi: 10.1016/j.imavis.2006.01.017.
- [98] Shuowen Hu; Nathaniel J. Short; Benjamin S. Riggan; Christopher Gordon; Kristan P. Gurton, Matthew Thielke, Prudhvi Gurram, Alex L. Chan, "A Polarimetric Thermal Database for Face Recognition Research," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2016, pp. 187–194. doi: 10.1109/CVPRW.2016.30.
- [99] A. Seal, D. Bhattacharjee, M. Nasipuri, and D. Basu, "UGC-JU face database and its benchmarking using linear regression classifier," *Multimed. Tools Appl.*, vol. 74, Jan. 2013, doi: 10.1007/s11042-013-1754-8.
- [100] Rastislav Lukac, K. N. Plataniotis, "Color filter arrays: design and performance analysis," *IEEE Trans. Consum. Electron.*, vol. 51, no. 4, pp. 1260–1267, Nov. 2005, doi: 10.1109/TCE.2005.1561853.
- [101] Kazuma Shinoda, Shu Ogawa, Yudai Yanagi, Madoka Hasegawa, Shigeo Kato, Masahiro Ishikawa, Hideki Komagata, Naoki Kobayashi , "Multispectral filter array

- and demosaicking for pathological images,” in *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, Hong Kong: IEEE, Dec. 2015, pp. 697–703. doi: 10.1109/APSIPA.2015.7415362.
- [102] Pierre-Jean Lapray, Xingbo Wang, Jean-Baptiste Thomas, Pierre Gouton, “Multispectral Filter Arrays: Recent Advances and Practical Implementation,” *Sensors*, vol. 14, no. 11, pp. 21626–21659, Nov. 2014, doi: 10.3390/s141121626.
- [103] R. Ramanath, W. E. Snyder, G. L. Bilbro, and W. A. Sander, “Robust Multispectral Imaging Sensors for Autonomous Robots,” North Carolina State University: Raleigh, NC, USA, Technical Report, 2001.
- [104] “Custom and Off-the-shelf Multispectral Imaging | SILIOS Technologies,” *silios*. <https://www.silios.com/multispectral-imaging> (accessed Mar. 07, 2023).
- [105] Porteur UBFC : Gouton, “EXIST (EXtended Image Sensing Technologies), H2020 (ECSEL),” au 12/2018 2015.
- [106] Sanket Mehta, Arpita Patel, Jagrat Mehta, “CCD or CMOS Image sensor for photography,” in *2015 International Conference on Communications and Signal Processing (ICCSP)*, Apr. 2015, pp. 0291–0294. doi: 10.1109/ICCSP.2015.7322890.
- [107] Munir El-Desouki, M. Jamal Deen, Qiyin Fang, Louis Liu, Frances Tse, David Armstrong, “CMOS Image Sensors for High Speed Applications,” *Sensors*, vol. 9, no. 1, Art. no. 1, Jan. 2009, doi: 10.3390/s90100430.
- [108] “CAM-4073M-FR_CMOS_Whitepaper.pdf.” Accessed: Mar. 09, 2023. [Online]. 2012, Available:
https://www.wapps.grassvalley.com/docs/WhitePapers/broadcast/cameras/ldk3000plus/CAM-4073M-FR_CMOS_Whitepaper.pdf
- [109] Younghyeon Park, Byeungwoo Jeon, “An Acquisition Method for Visible and Near Infrared Images from Single CMYK Color Filter Array-Based Sensor,” *Sensors*, vol. 20, no. 19, p. 5578, Sep. 2020, doi: 10.3390/s20195578.
- [110] Xianglong Chen, Haipeng Wang, Yaohui Liang, Yin Meng, Shifeng Wang, “A Novel Infrared and Visible Image Fusion Approach Based on Adversarial Neural Network,” *Sensors*, vol. 22, no. 1, Art. no. 1, Jan. 2022, doi: 10.3390/s22010304.
- [111] Ziad Abood, Ghada Karam, Eram Rafid, *Face Recognition Using Fusion of Multispectral Imaging*. 2017, p. 112. doi: 10.1109/AIC-MITCSA.2017.8722957.
- [112] Yi Jin, Jiwen Lu, Qiuqi Ruan, “Coupled Discriminative Feature Learning for Heterogeneous Face Recognition,” *IEEE Trans. Inf. Forensics Secur.*, vol. 10, no. 3, pp. 640–652, Mar. 2015, doi: 10.1109/TIFS.2015.2390414.

- [113] Ran He, Yi Li, Xiang Wu, Lingxiao Song, Zhenhua Chai, Xiaolin Wei, “Coupled adversarial learning for semi-supervised heterogeneous face recognition,” *Pattern Recognit.*, vol. 110, p. 107618, Feb. 2021, doi: 10.1016/j.patcog.2020.107618.
- [114] Lingxiao Song, Man Zhang, Xiang Wu, Ran He, “Adversarial Discriminative Heterogeneous Face Recognition,” *Proc. AAAI Conf. Artif. Intell.*, vol. 32, no. 1, Art. no. 1, Apr. 2018.
- [115] Pengcheng Zhao, Fuping Zhang, Jianming Wei, Yingbo Zhou, Xiao Wei, “SADG: Self-Aligned Dual NIR-VIS Generation for Heterogeneous Face Recognition,” *Appl. Sci.*, vol. 11, no. 3, Art. no. 3, Jan. 2021, doi: 10.3390/app11030987.
- [116] Mamadou Diarra, Pierre Gouton, ADOU Kablan Jérôme, “Multispectral face recognition using hybrid feature,” *Electron. Imaging*, vol. 2017, no. 18, pp. 200–203, Jan. 2017, doi: 10.2352/ISSN.2470-1173.2017.18.COLOR-061.
- [117] Zhihua Xie, Ling Shi, Yi Li, “Two-Stage Fusion of Local Binary Pattern and Discrete Cosine Transform for Infrared and Visible Face Recognition,” in *Emerging Trends in Intelligent and Interactive Systems and Applications*, M. Tavana, N. Nedjah, and R. Alhaji, Eds., in *Advances in Intelligent Systems and Computing*. Cham: Springer International Publishing, 2021, pp. 967–975. doi: 10.1007/978-3-030-63784-2_117.
- [118] Kai Guo, Shuai Wu, and Yong Xu, “Face recognition using both visible light image and near-infrared image and a deep network,” *CAAI Trans. Intell. Technol.*, vol. 2, no. 1, pp. 39–47, Mar. 2017, doi: 10.1016/j.trit.2017.03.001.
- [119] Weipeng Hu, Haifeng Hu, “Discriminant Deep Feature Learning based on joint supervision Loss and Multi-layer Feature Fusion for heterogeneous face recognition,” *Comput. Vis. Image Underst.*, vol. 184, pp. 9–21, Jul. 2019, doi: 10.1016/j.cviu.2019.04.003.
- [120] Emanuel Candès, David Donoho, “Curvelets: A Surprisingly Effective Nonadaptive Representation for Objects with Edges,” 2000. Accessed: Feb. 21, 2023. [Online]. Available: <https://www.semanticscholar.org/paper/Curvelets%3A-A-Surprisingly-Effective-Nonadaptive-for-Cand%C3%A8s-Donoho/0843c2f5cbd7dda30cbeed73d95b77d068408476>
- [121] Emanuel Candès, L. Demanet, David Donoho, and L. Ying, “Fast Discrete Curvelet Transforms,” *Multiscale Model. Simul.*, vol. 5, no. 3, pp. 861–899, Jan. 2006, doi: 10.1137/05064182X.
- [122] “Curvelet Home.” <http://www.curvelet.org/> (accessed Feb. 22, 2023).

- [123] Ran He, Yi Li, Xiang Wu, Lingxiao Song, Zhenhua Chai, Xiaolin Wei, “Coupled adversarial learning for semi-supervised heterogeneous face recognition,” *Pattern Recognit.*, vol. 110, p. 107618, Feb. 2021, doi: 10.1016/j.patcog.2020.107618.
- [124] Xuan Huang, Lei Wu, Yinsong Ye, “A Review on Dimensionality Reduction Techniques”, *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 33, No. 10, 1950017 (2019) <https://www.worldscientific.com/doi/10.1142/S0218001419500174>
- [125] Mateus Espadoto, Rafael M. Martins, Aandreas Kerren, Nina S. T. Hirata, Alexandru C. Telea, “Toward a Quantitative Survey of Dimension Reduction Techniques,” *IEEE Trans. Vis. Comput. Graph.*, vol. 27, no. 3, pp. 2153–2173, Mar. 2021, doi: 10.1109/TVCG.2019.2944182.
- [126] Guoqing Chao, Yuan Luo, and Weiping Ding, “Recent Advances in Supervised Dimension Reduction: A Survey,” *Mach. Learn. Knowl. Extr.*, vol. 1, no. 1, Art. no. 1, Mar. 2019, doi: 10.3390/make1010020.
- [127] Jae Woong Soh, Hyun-Seung Lee, and Nam Ik Cho, “An image compression algorithm based on the Karhunen Loève transform,” in *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Dec. 2017, pp. 1436–1439. doi: 10.1109/APSIPA.2017.8282257.
- [128] Ian Jolliffe, “A 50-year personal journey through time with principal component analysis,” *J. Multivar. Anal.*, vol. 188, p. 104820, Mar. 2022, doi: 10.1016/j.jmva.2021.104820.
- [129] Haifeng Zhao, Zheng Wang, Feiping Nie, “A New Formulation of Linear Discriminant Analysis for Robust Dimensionality Reduction,” *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 4, pp. 629–640, Apr. 2019, doi: 10.1109/TKDE.2018.2842023.
- [130] Pedram Ghamis, Naoto Yokoya, Jun Li, Wenzhi Liao, Sicong Liu, Javier Plaza, Behnood Rasti, Antonio Plaza, “Advances in Hyperspectral Image and Signal Processing: A Comprehensive Overview of the State of the Art,” *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 37–78, Dec. 2017, doi: 10.1109/MGRS.2017.2762087.
- [131] Rong Wang, Feiping Nie, Richang Hong, Xiaojun Chang, Xiaojun Yang, Weizhong Yu, “Fast and Orthogonal Locality Preserving Projections for Dimensionality Reduction,” *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 5019–5030, Oct. 2017, doi: 10.1109/TIP.2017.2726188.
- [132] An Gie Yong, S. Pearce, “A Beginner’s Guide to Factor Analysis: Focusing on Exploratory Factor Analysis,” 2013. doi: 10.20982/TQMP.09.2.P079.

- [133] Jianzhong Wang, "Classical Multidimensional Scaling. In: Geometric Structure of High-Dimensional Data and Dimensionality Reduction". Springer, Berlin, Heidelberg, 2012, https://doi.org/10.1007/978-3-642-27497-8_6
- [134] Benyamin Ghojogh, Ali Ghodsi, Fakhri Karray, Mark Crowley, "Multidimensional Scaling, Sammon Mapping, and Isomap: Tutorial and Survey," arXiv, arXiv:2009.08136, Sep. 2020. doi: 10.48550/arXiv.2009.08136.
- [135] SAM T. Roweis and Lawrence K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000, doi: 10.1126/science.290.5500.2323.
- [136] Joshua B. Tenenbaum, Vin de Silva, and John C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000, doi: 10.1126/science.290.5500.2319.
- [137] Mikhail Belkin and Partha Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, Jun. 2003, doi: 10.1162/089976603321780317.
- [138] Pierre Demartines and Jeanny Hérault, "Curvilinear component analysis: a self-organizing neural network for nonlinear mapping of data sets," in *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 148–154, Jan. 1997, doi: 10.1109/72.554199.
- [139] Ludovic Journaux, Irène Foucherot, and Pierre Gouton, "Reduction of the number of spectral bands in Landsat images: a comparison of linear and nonlinear methods," *Opt. Eng.*, vol. 45, no. 6, p. 067002, Jun. 2006, doi: 10.1117/1.2212108.
- [140] Giansalvo Cirrincione, Jeanny Hérault, Vincenzo Randazzo, "The on-line curvilinear component analysis (onCCA) for real-time data reduction," in *2015 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2015, pp. 1–8. doi: 10.1109/IJCNN.2015.7280318.
- [141] Giansalvo Cirrincione, Vincenzo Randazzo, Eros Pasero, "The Growing Curvilinear Component Analysis (GCCA) neural network," *Neural Netw.*, vol. 103, pp. 108–117, 2018, doi: 10.1016/j.neunet.2018.03.017.
- [142] Ji Ma, Yuyu Yuan, "Dimension reduction of image deep feature using PCA," *J. Vis. Commun. Image Represent.*, vol. 63, p. 102578, Aug. 2019, doi: 10.1016/j.jvcir.2019.102578.
- [143] Ala Eldin Omer, Adil Khurran, "Facial recognition using principal component analysis based dimensionality reduction," in *2015 International Conference on Computing*,

- Control, Networking, Electronics and Embedded Systems Engineering (ICCNEEE)*, Sep. 2015, pp. 434–439. doi: 10.1109/ICCNEEE.2015.7381408.
- [144] Yongjian Nian, Ke Xu, Jianwei Wan, Ling Wang, Mi He, “Block-based KLT compression for multispectral images,” *Int. J. Wavelets Multiresolution Inf. Process.*, vol. 14, no. 04, p. 1650029, Jul. 2016, doi: 10.1142/S0219691316500296.
- [145] M. Lennon, G. Mercier, M.-C. Mouchot, and L. Hubert-Moy, “Curvilinear component analysis for nonlinear dimensionality reduction of hyperspectral images,” in *Image and Signal Processing for Remote Sensing VII*, SPIE, Jan. 2002, pp. 157–168. doi: 10.1117/12.454150.
- [146] Sofiane Mihoubi, Olivier Losson, Benjamin Mathon, Ludovic Macaire, “Multispectral demosaicing using intensity-based spectral correlation,” in *2015 International Conference on Image Processing Theory, Tools and Applications (IPTA)*, Nov. 2015, pp. 461–466. doi: 10.1109/IPTA.2015.7367188.
- [147] Medha Gupta, Vishwas Rathi, and Puneet Goyal, “Adaptive and Progressive Multispectral Image Demosaicking,” *IEEE Trans. Comput. Imaging*, vol. 8, pp. 69–80, 2022, doi: 10.1109/TCI.2022.3140554.
- [148] Medha Gupta, Puneet Goyal, “Demosaicing Method for Multispectral Images Using Derivative Operations,” *Am. J. Math. Manag. Sci.*, vol. 40, no. 2, pp. 163–176, Apr. 2021, doi: 10.1080/01966324.2021.1939206.
- [149] Xingbo Wang, Jean-Baptiste. Thomas, Jon Yngve Hardeberg, Pierre Gouton, “Discrete wavelet transform based multispectral filter array demosaicking,” in *2013 Colour and Visual Computing Symposium (CVCS)*, Sep. 2013, pp. 1–6. doi: 10.1109/CVCS.2013.6626274.
- [150] Medha Gupta, “Generalizing Spectral Difference Method for Multispectral Image Demosaicking and Analyzing the Role of MSFA Patterns,” in *2020 9th International Conference System Modeling and Advancement in Research Trends (SMART)*, Dec. 2020, pp. 416–420. doi: 10.1109/SMART50582.2020.9337090.
- [151] Lidan Miao, Hairong Qi, Rajeev Ramanath, Wesley Snyder, “Binary Tree-based Generic Demosaicking Algorithm for Multispectral Filter Arrays,” *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3550–3558, Nov. 2006, doi: 10.1109/TIP.2006.877476.
- [152] Lidan Miao, Hairong Qi, Rajeev Ramanath, “A Generic Binary Tree-Based Progressive Demosaicking Method for Multispectral Filter Array,” in *2006 International Conference on Image Processing*, Oct. 2006, pp. 3221–3224. doi: 10.1109/ICIP.2006.312909.

- [153] Lidan Miao, Hairong Qi, Rajeev Ramanath, “Generic MSFA mosaicking and demosaicking for multispectral cameras,” in *Digital Photography II*, SPIE, Feb. 2006, pp. 88–97. doi: 10.1117/12.642366.
- [154] Xingbo Wang, Jean-Baptiste. Thomas, Jon Yngve Hardeberg, Pierre Gouton, “Median filtering in multispectral filter array demosaicking,” in *Digital Photography IX*, SPIE, Feb. 2013, pp. 103–112. doi: 10.1117/12.2005256.
- [155] Norbert Hounsou, Amadou Tidjani Sanda Mahama, Pierre Gouton, “Extension of luminance component based demosaicking algorithm to 4- and 5-band multispectral images,” *Array*, vol. 12, p. 100088, Dec. 2021, doi: 10.1016/j.array.2021.100088.
- [156] J. Driesen, Paul Scheunders, “Wavelet-based color filter array demosaicking,” in *2004 International Conference on Image Processing, 2004. ICIP '04.*, Oct. 2004, pp. 3311–3314 Vol. 5. doi: 10.1109/ICIP.2004.1421822.
- [157] Giuseppe. Spampinato, Arcangelo Bruna, Giuseppe Sanguedolce, Edoardo Ardizzone, Marco L. Cascia, “Improved color interpolation using discrete wavelet transform,” in *Image and Video Communications and Processing 2005*, SPIE, Mar. 2005, pp. 753–760. doi: 10.1117/12.585200.
- [158] Jan Aelterman, Bart Goossens, Aleksandra Pižurica, Wilfried Philips, “Locally adaptive complex wavelet-based demosaicing for color filter array images,” in *Wavelet Applications in Industrial Processing VI*, SPIE, Jan. 2009, pp. 115–126. doi: 10.1117/12.807347.
- [159] Min Cheol Kim, Joon Hyeon Park, Myung Hoon Sunwoo, “Multilevel Feature Extraction Using Wavelet Attention for Deep Joint Demosaicking and Denoising,” *IEEE Access*, vol. 10, pp. 77099–77109, 2022, doi: 10.1109/ACCESS.2022.3192451.
- [160] Bahadır Kürşat Gunturk, Yucel Altunbasak, Russell M. Mersereau, “Color plane interpolation using alternating projections,” in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 2002, p. IV-3333-IV–3336. doi: 10.1109/ICASSP.2002.5745367.
- [161] Maya R. Gupta, Ting Chen, “Vector color filter array demosaicing,” presented at the Photonics West 2001 - Electronic Imaging, M. M. Blouke, J. Canosa, and N. Sampat, Eds., San Jose, CA, May 2001, p. 374. doi: 10.1117/12.426974.
- [162] Alain Horé, Djemel Ziou, “Image Quality Metrics: PSNR vs. SSIM,” in *2010 20th International Conference on Pattern Recognition*, Aug. 2010, pp. 2366–2369. doi: 10.1109/ICPR.2010.579.

- [163] Zhou Wang and Alan C. Bovik, “Reduced- and No-Reference Image Quality Assessment,” *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 29–40, Nov. 2011, doi: 10.1109/MSP.2011.942471.
- [164] Maksim Siniukov, Dmitriy Kulikov, Dmitriy Vatolin, “Applicability limitations of differentiable full-reference image-quality metrics,” in *2023 Data Compression Conference (DCC)*, Mar. 2023, pp. 1–1. doi: 10.1109/DCC55655.2023.00082.
- [165] Anish Mittal, Rajiv Soundararajan and Alan C. Bovik, “Making a ‘Completely Blind’ Image Quality Analyzer,” *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013, doi: 10.1109/LSP.2012.2227726.
- [166] Chiman Kwan, Bryan Chou, James F. Bell III, “Comparison of Deep Learning and Conventional Demosaicing Algorithms for Mastcam Images,” *Electronics*, vol. 8, no. 3, Art. no. 3, Mar. 2019, doi: 10.3390/electronics8030308.
- [167] A. G. Asuero, A. Sayago, and A. G. González, “The Correlation Coefficient: An Overview,” *Crit. Rev. Anal. Chem.*, vol. 36, no. 1, pp. 41–59, Jan. 2006, doi: 10.1080/10408340500526766.