



Etude de nouveaux caractères biométriques de la main dans un contexte Télécom

Julien Doublet

► To cite this version:

Julien Doublet. Etude de nouveaux caractères biométriques de la main dans un contexte Télécom. Traitement du signal et de l'image [eess.SP]. Université de Caen, 2009. Français. NNT: . tel-01073225

HAL Id: tel-01073225

<https://theses.hal.science/tel-01073225>

Submitted on 10 Oct 2014

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 de CAEN/BASSE-NORMANDIE

U.F.R.: Sciences

ÉCOLE DOCTORALE: SIMEM

THÈSE

présentée par

Mr Julien Doublet

et soutenue

le 14/12/2009

En vue de l'obtention du

DOCTORAT de l'UNIVERSITÉ de CAEN

Spécialité: Informatique et applications

(Arrêté du 7 août 2006)

Étude de nouveaux caractères biométriques de la main dans un contexte Télécom

MEMBRES du JURY

<i>Rapporteurs :</i>	Bernadette DORIZZI	Professeur	Télécom & Management Sud Paris
	Atila BASKURT	Professeur	INSA de Lyon
<i>Examineurs :</i>	Hubert CARDOT	Professeur	Université de Tours
	Christophe ROSENBERGER	Professeur	ENSICAEN
	Olivier LEPETIT	Ingénieur	Orange LABS
<i>Directrice :</i>	Marinette REVENU	Professeur	ENSICAEN

Remerciements

J'aimerais tout d'abord remercier ma directrice de thèse, Marinette Revenu, pour m'avoir guidé et encouragé tout au long de ces années de recherche. Sa patience, sa disponibilité et ses encouragements durant la longue période d'écriture du manuscrit m'ont permis de finaliser ce document. Je la remercie très très vivement.

Je remercie les rapporteurs de cette thèse Bernadette Dorizzi et Atilla Baskurt pour l'intérêt qu'ils ont porté à mon travail. Particulièrement, je pense au travail d'un samedi après-midi qui a permis d'éclaircir les zones d'ombres restantes dans ce rapport.

Merci également aux autres membres du jury qui ont accepté de juger ce travail : Hubert Cardot, Christophe Rosenberger et particulièrement Olivier Lepetit pour m'avoir guidé avec beaucoup de liberté dans cette recherche. Sa gentillesse et son soutien m'ont beaucoup apporté et ont contribué grandement au plaisir durant ces trois années à France Télécom. Dans cette période difficile en ce moment à France Télécom, cette thèse est également la sienne.

Je remercie également Jérôme Laudouar, Patrick Picard et Sébastien Marie pour m'avoir accueilli au sein du laboratoire Monétique et Transaction électronique (qui doit avoir changé de nom aujourd'hui je suppose!).

Une dédicace spéciale pour Marie, Vincent, Jean-Jacques, Julien, Pierre et toute l'équipe du laboratoire que j'ai eu le plaisir de côtoyer et pour l'ambiance qu'ils apportaient dans les bureaux, à la pause café ou en dehors du travail. Mention à part à Jérôme et Jean-Damien pour les nuits et les lendemains de nuits Caennaises mémorables.

Un remerciement spécial également à Natacha, Guillaume, Romuald, Julie, Nicolas, Séverine, Sylvain et Pascal pour tout durant toutes les années d'études qui m'ont conduit jusqu'à la thèse et qui font que Caen me manque quand même un peu. Les soirées et les matchs de foot ne sont plus tout à fait pareil.

Je tiens à remercier également toute ma famille et en particulier ma mère qui m'a toujours encouragé et soutenu dans mes décisions même lorsqu'elles sont à l'opposé de ces désirs.

Enfin, mes remerciements vont à Amélie pour son soutien constant, sa joie contagieuse et l'événement merveilleux qui s'approche de jours en jours...

Table des matières

Chapitre 1	
Introduction	1
1.1 Contexte de la recherche	1
1.2 Notre contribution	2
1.3 Plan du mémoire	3
Chapitre 2	
La reconnaissance biométrique	5
2.1 Introduction	5
2.2 Apports de la biométrie	7
2.3 Les systèmes biométriques	7
2.3.1 Décomposition en modules	7
2.3.2 Enrôlement, authentification et identification	8
2.4 Évaluation et taux d'erreurs des systèmes biométriques	10
2.5 Présentation de différentes modalités biométriques	13
2.5.1 ADN (Acide désoxyribonucléique)	14
2.5.2 Démarche	14
2.5.3 Dynamique de la frappe au clavier	15
2.5.4 Empreinte digitale	15
2.5.5 Empreinte palmaire	16
2.5.6 Géométrie de la main	16
2.5.7 Iris	16
2.5.8 Rétine	16
2.5.9 Signature dynamique	17
2.5.10 Thermographie de la main	17
2.5.11 Visage	17
2.5.12 Voix	17
2.5.13 Autres Systèmes	18

2.5.14	Multi-biométrie	18
2.6	Avantages et inconvénients des systèmes biométriques	18
Partie I Segmentation de la main		21
Chapitre 3 État de l'art de la segmentation de main 23		
3.1	Introduction à la segmentation	23
3.2	Contexte de la détection de la main	24
3.3	Détection de la main dans les systèmes biométriques	26
3.4	Détection de la main par modélisation de la couleur de la peau	27
3.4.1	Représentation couleur	27
3.4.1.1	Espace <i>RVB</i> - Rouge, Vert, Bleu	28
3.4.1.2	Espace <i>YUV</i> et <i>YC_rC_b</i>	30
3.4.1.3	Espaces <i>HSV</i> et <i>HSL</i> - Teinte, Saturation, Valeur ou Lumiance	30
3.4.1.4	Espaces <i>CIE XYZ</i> et <i>RVB normalisé</i>	31
3.4.1.5	Espaces <i>CIE L*a*b*</i> et <i>CIE L*u*v*</i>	33
3.4.2	Méthode de modélisation de la couleur de la peau	34
3.4.2.1	Définition explicite de la couleur de la peau	34
3.4.2.2	Modélisation non-paramétrique de la couleur de la peau . . .	36
3.4.2.3	Modélisation paramétrique	38
3.5	Détection de la main intégrant des connaissances à priori	40
3.5.1	Modèle de forme	40
3.5.2	Approches par classification de zones d'image	41
3.6	Comparaison des différentes méthodes de détection	43
3.7	Extraction de la zone d'intérêt : la paume	45
3.7.1	Segmentation de la paume sous forme carrée	45
3.7.1.1	Extraction des points importants	45
3.7.1.2	Extraction de la paume	45
3.7.2	Segmentation de la paume de façon circulaire	46
3.8	Conclusion sur les approches de détection de la main	46

Chapitre 4

Fusion d'information pour la segmentation de la main dans une image complexe 49

4.1	La segmentation pour la reconnaissance biométrique sans contact	49
4.2	Segmentation couleur multi-résolution des zones de peau par apprentissage	51
4.2.1	Apprentissage de la couleur de la peau	51
4.2.1.1	Perceptron multi-couche	51
4.2.1.2	Machine à vecteurs de support	52
4.2.2	Détection multi-résolution	54
4.2.3	Post-traitement sur la carte de probabilité	55
4.3	Utilisation de connaissance de forme à priori	57
4.3.1	Apprentissage de la forme d'une main	58
4.3.2	Recherche d'une main à partir du modèle	61
4.3.2.1	Initialisation du modèle de forme	61
4.3.2.2	Évolution du contour du modèle	62
4.4	Évaluation de la méthode de segmentation	64
4.4.1	Base de données d'évaluation	64
4.4.2	Évaluation de la détection de peau	65
4.4.3	Évaluation de la segmentation par contour actif de forme	67
4.5	Conclusion sur la détection de la main dans une image complexe	68

Partie II Reconnaissance de la main

73

Chapitre 5

État de l'art des systèmes de reconnaissance biométrique par la main 75

5.1	Introduction à la reconnaissance de la main	75
5.2	Reconnaissance biométrique basée sur la forme de la main	76
5.3	Authentification s'appuyant sur la paume de la main	78
5.3.1	Approche basée sur les lignes de la main	78
5.3.2	Approche par décomposition en sous-espaces	79
5.3.3	Approche par codage élémentaire	80
5.3.3.1	Analyse par transformée de Fourier	81
5.3.3.2	Analyse par ondelettes	83
5.3.3.3	Analyse par filtrage de Gabor	86
5.3.4	Approche par apprentissage	88

5.3.5	Autres approches	89
5.4	Multi-biométrie par fusion de méthodologies	90
5.5	Comparaison des différentes approches	91
5.6	Conclusion sur la reconnaissance par la main	93

Chapitre 6

Reconnaissance de la main sans contact

95

6.1	Extraction et normalisation de la paume	95
6.2	Combinaison des informations de forme et de texture	96
6.2.1	Extraction des caractéristiques	97
6.2.2	Evaluation des performances	99
6.3	Authentification par filtre de Gabor circulaire et distance d'édition	102
6.3.1	Extraction des caractéristiques	102
6.3.2	Comparaison des données extraites	103
6.3.3	Evaluation des performances	105
6.4	Classification par estimation de distribution non linéaire	106
6.4.1	Extraction des caractéristiques	106
6.4.1.1	Extraction par banc de filtres de Gabor	106
6.4.1.2	Extraction par analyse robuste des niveaux de gris	106
6.4.2	Comparaison par estimation de distribution non linéaire	109
6.4.2.1	Estimation de la distribution	110
6.4.2.2	Fonction de reconnaissance individuelle	111
6.4.3	Évaluation des performances	112
6.5	Etude sur la robustesse aux paramètres extérieurs	113
6.5.1	Description des filtres log-normaux	114
6.5.2	Evaluation des performances suivant les perturbations extérieures	115
6.5.2.1	Les changements de luminosité et de contraste	116
6.5.2.2	Le flou Gaussien et le bruit blanc	116
6.5.2.3	Les translations, rotations et zooms	116
6.5.2.4	Résumé des performances	119
6.6	Conclusion sur la reconnaissance de la main sans contact	119

Chapitre 7

Conclusion et perspectives

121

Annexe A

Physiologie de la main

A.1 Anatomie de la main	125
A.2 Physiologie de la peau	126
Bibliographie	129

Table des figures

1.1	Acquisition sans contact (a) vue de profil et (b) vue de 3/4	2
2.1	Exemples de capteurs biométriques	8
2.2	Diagramme d'utilisation biométrique	9
2.3	Courbes de distribution des imposteurs et des authentiques	11
2.4	Courbe ROC	12
2.5	Les cinq classes principales des empreintes digitales : (a) arche, (b) boucle gauche, (c) boucle droite, (d) ogive, (e) spirale.	13
2.6	Marché de la biométrie	14
2.7	Les principales caractéristiques biométriques	15
3.1	Modèle cinétique de la main	26
3.2	Deux images de mains collectées par (a) un capteur spécifique et (c) un scanner digital puis segmentées par la méthode d'Otsu (b) et (d).	28
3.3	Courbe d'absorption des cônes	29
3.4	Représentation cubique du domaine <i>RVB</i>	29
3.5	Représentation cône du domaine <i>TSV</i>	31
3.6	Diagramme de chromacité <i>xy</i> schématisé	32
3.7	Les fonctions colorimétriques du domaine (a) <i>RVB</i> et (b) <i>XYZ</i>	33
3.8	Représentation sphérique du <i>CIE L*a*b*</i>	34
3.9	Répartition des pixels de peau dans le domaine (a) <i>HSV</i> et (b) <i>YC_BC_R</i>	35
3.10	Limitations de la surface représentant la couleur de peau dans le domaine (a) <i>HSV</i> et (b) <i>YC_BC_R</i>	36
3.11	Structuration hiérarchique de modèles de main	41
3.12	Localisation de visage par claissifieur	42
3.13	Points importants d'une main	46
3.14	Extraction de la paume	48
4.1	Exemples d'acquisitions avec (a) une Webcam (b) un appareil photo numérique (c) un téléphone mobile	50
4.2	Réseau de neurones (a) Réseau de type 1 avec N=3 dans le domaine RGB (b) Réseau de type 3 avec N=2, N2=3 dans le domaine RGB+TSL	52
4.3	Calcul de la carte de probabilité (pixels par pixels)	52
4.4	Classification par les SVMs	53
4.5	Carte de probabilité pyramidale	55
4.6	Post-Traitement (a) Carte de probabilités (b) Carte binarisée (c) Ajustement du patch (d) Carte binarisée avec patch	57
4.7	Schéma d'une main annotée avec 60 points	58

4.8	Mains avant (a) et après alignement (b)	60
4.9	Variations selon les trois principales valeurs propres	61
4.10	Exemple de zone mal segmentée avec une forte courbure	62
4.11	Extraction automatique des landmarks (a) forme de main (b) diagramme de distance (c) points caractéristiques	62
4.12	Détermination du gradient (a) Image originale (b) Di Zenzo dans l'espace <i>RVB</i> (c) Di Zenzo dans l'espace de peau (d) Di Zenzo ajusté	64
4.13	Evolution du contour aux itérations 0, 5, 10, 20, 30 et 40	65
4.14	Taux d'erreurs suivant les constantes W et G	69
4.15	Taux de divergences suivant les constantes W et G	70
4.16	Qualité de détection de main et vitesse d'exécution suivant le nombre de points caractéristiques	71
4.17	Exemples de détection de la main	72
5.1	Caractérisation de la main par axes dans le système de Jain et al. [53]	76
5.2	<i>Eigenpalms</i> : Six vecteurs propres avec les six valeurs propres les plus importantes [19]	80
5.3	Six vecteurs de bases de l' <i>ACI</i> [19]	80
5.4	Exemples de représentation dans le domaine fréquentiel	82
5.5	Représentation des caractéristiques de la paume	83
5.6	Schéma de Décomposition en ondelette	84
5.7	Décomposition en ondelettes	85
5.8	Fonction de Gabor 2D	86
5.9	Architecture dans le réseau de neurones de [46]	89
5.10	Opérateur <i>LBP</i>	89
5.11	Extraction des zones des doigts et de la paume	91
5.12	Caractéristiques des doigts et de la paume	94
6.1	Extraction de la paume	96
6.2	Dimension de la main	98
6.3	Hand size	99
6.4	Courbe ROC	100
6.5	Distributions après fusion	101
6.6	Fonctions de Gabor (a) orientée et (b) circulaire	103
6.7	Deux exemples de notre base de données et les paumes correspondantes	105
6.8	Résultat des tests d'authentification : distribution des distances de comparaison (a) et courbe ROC (b)	107
6.9	Un exemple de restauration d'une image bruitée : (a) image bruitée (b) paume restaurée	108
6.10	Décomposition et normalisation de la paume restaurée de la figure 6.9 en neuf sous-images	110
6.11	Taux de reconnaissance avec différents noyaux	113
6.12	Filtre bidimensionnel log-normal en coordonnées log-polaires	114
6.13	Exemple de filtrages : (a) image originale, (b) image filtrée par filtre log-normal et (c) image filtrée par filtre de Gabor	115
6.14	Exemple de bruits et de flous : (a) image originale, (b) image ayant subi un flou Gaussien d'écart type 9 et (c) image bruitée avec une dispersion de 20	117

6.15	Exemple de rotation 3D : (a) image originale et (b) image tournée de $45^\circ C$ en slant et en tilt	118
A.1	Anatomie de la main : (a) squelette et (b) structure interne	126
A.2	Schémas de coupe de (a) la peau et de (b) l'épiderme	127

Liste des tableaux

2.1	Comparaisons des systèmes biométriques	19
3.1	Comparaison des méthodes de détection de la couleur de la peau. VP désigne les Vrais Positif et FP les Faux Positifs	44
4.1	Taux d'erreur avec (a) l'estimation par Bayes et (b) la somme de gaussiennes pour la détection de peau	66
4.2	Taux d'erreur de classification avec un réseau de neurones	67
4.3	Taux d'erreur de classification dans l'espace RVB et l'espace couleur de peau avec un réseau de neurones et les SVMs.	67
4.4	Temps de calcul moyen pour les images de taille 454 * 341.	67
4.5	Ratio entre la moyenne de l'intensité du gradient de la main et du fond	68
5.1	Comparaison des méthodes de reconnaissance de la main	92
6.1	Exemple de matrice de score	104
6.2	Comparaison du système proposé avec les méthodes de Zhang et al.	106
6.3	Comparaison des performances du système suivant le nombre d'exemples d'entraînement et le seuil d'ajustement τ	113
6.4	Référence pour l'évaluation de la robustesse aux variations	115
6.5	Exemple de distance selon la luminosité et le contraste entre une image de référence et l'image ayant subi la perturbation	116
6.6	Exemple de distance selon le bruit et le flou entre une image de référence et l'image ayant subi la perturbation	117
6.7	Robustesse à la rotation	117
6.8	Robustesse au zoom	118
6.9	Robustesse au translation	118
6.10	Robustesse au translation	118

Chapitre 1

Introduction

Sommaire

1.1	Contexte de la recherche	1
1.2	Notre contribution	2
1.3	Plan du mémoire	3

1.1 Contexte de la recherche

Dans le contexte télécom, la principale modalité biométrique utilisée est la reconnaissance vocale. Elle présente l'avantage d'être associée logiquement à la téléphonie et ne nécessite aucun module matériel supplémentaire à l'équipement traditionnel pour fonctionner.

Cependant, deux autres modalités ont fait dernièrement leur apparition dans le secteur télécom : la reconnaissance par empreintes digitales et la reconnaissance faciale. La première méthode permet une reconnaissance plus fiable de l'utilisateur mais nécessite un équipement ad-hoc. La reconnaissance par le visage est utilisée dans les scénarios de vidéo conférence ou classiquement pour l'identification lors de l'ouverture d'une session. Grâce à la qualité accrue des capteurs sur la téléphonie mobile cette méthode ne nécessite, comme la reconnaissance par la voix, aucun accessoire supplémentaire.

En France, la CNIL (commission national de l'informatique et des libertés) émet de nombreuses réserves pour l'utilisation de la biométrie dans les applications civiles. En particulier, les biométries du visage et des empreintes digitales qui servent dans les applications gouvernementales et policières sont quasi systématiquement rejetées pour des applications commerciales. Par contre, la reconnaissance par la main est autorisée fréquemment pour ce genre d'application (utilisation dans des cantines scolaires par exemple). Partant de ce constat et des bonnes performances de reconnaissance, nous avons étudié cette modalité de reconnaissance en ajoutant les contraintes liés aux contextes Télécom.

Les services envisageables sont divers. La liste non exhaustive suivante identifie les principaux types d'utilisation possible :

- Contrôle d'accès aux locaux
- Sécurisation des systèmes d'informations
- Sécurisation des communications



FIGURE 1.1 – Acquisition sans contact (a) vue de profil et (b) vue de 3/4

- Utilisation dans un cadre administratif
- Sécurisation de transaction
- Vérification de présence

Dans le cadre de la téléphonie mobile, la sécurisation liée aux demandes de micro-transactions électroniques devient de plus en plus importante et pourra également être associée à la biométrie. Dans ces différents scénarios, l'acquisition de l'image de la main est effectuée dans l'espace ce qui implique une grande liberté d'utilisation. Lors de l'acquisition, l'utilisateur place sa main face au capteur qui effectue une capture de celle-ci. La reconnaissance de la main doit ainsi fonctionner dans un environnement complexe (arrière plan non-défini, luminosité variable et liberté de positions). Dans le cas d'une acquisition par une caméra fixe de type webcam, le visage de l'utilisateur peut également être présent dans l'image. Un exemple d'acquisition de paume avec un téléphone portable illustrant ce type d'acquisition est montré à la Figure 1.1.

La reconnaissance de la paume de la main est étudiée depuis plusieurs années et un grand nombre d'algorithmes d'authentification ont été proposés. Dans tous ces systèmes, l'acquisition de l'image de la main est effectuée avec un contact entre la main et le support d'acquisition. L'objectif de cette thèse est de définir des nouveaux processus pour permettre la reconnaissance de la main à partir d'un capteur à bas coût sans contact pour une utilisation dans l'univers des télécoms. Les caméras envisagées sont des capteurs de téléphones mobiles ou des webcams.

1.2 Notre contribution

L'objectif de notre travail consiste à définir et mettre en oeuvre de nouvelles méthodes pour la détection et la reconnaissance de main.

Pour cela, dans un premier temps, une méthode de détection de la main robuste et sans notion de suivi est proposée [29]. Cette méthode s'appuie sur une fusion de connaissances à priori de la couleur de peau et de la forme de la main. La couleur de peau est modélisée grâce à un réseau de neurones. Afin de répondre aux impératifs de travail en temps réel, cette méthode est appliquée dans un contexte multi-résolution. Même si notre méthode de segmentation par la couleur de la peau s'avère après expérimentations plus efficace que celle des systèmes antérieurs, elle ne

permet pas une détection suffisamment robuste de la main pour une reconnaissance biométrique efficace.

Dans le but d'accomplir cette segmentation, nous avons proposé d'affiner cette recherche en nous appuyant sur la méthode proposée par Cootes et al.[20] utilisant la notion de forme : les modèles actifs de formes (ASM). Afin de traiter les problèmes de divergence du modèle actif, notre modèle est déformé suivant un modèle de couleur de peau et des contraintes sur les contours en deux dimensions. Comparé aux autres approches, notre modèle effectue une détection robuste et précise des contours de la main permettant l'application d'algorithmes de reconnaissance biométrique.

La reconnaissance biométrique de la main a été largement étudiée ces dernières années. Les méthodes proposées les plus performantes se basent sur les travaux de reconnaissance par l'iris de Daugman [23] qui utilisent une analyse de la texture basée sur les filtres de Gabor orientés. Afin d'améliorer les performances de reconnaissance, des méthodes utilisant plus d'informations tout en gardant la notion de rapidité d'exécution sont définies.

Nous avons tout d'abord proposé une méthode s'appuyant sur une fusion des données de forme et de texture avec la même taille de référence que les systèmes antérieurs [28, 30]. Ensuite, la méthode de Zhang et al. [136] (similaire à celle proposée par Daugman pour l'iris) est généralisée par l'utilisation d'un filtre de Gabor circulaire et la définition d'un processus de comparaison s'appuyant sur une distance d'édition robuste [32]. Les taux de reconnaissance obtenus montrent le bien fondé de cette approche dans notre contexte. Afin d'optimiser la comparaison des données connues, des méthodes de reconnaissance par projection dans un espace non linéaires ont également été avancées [31, 33]. Elles s'appuient sur une estimation de la distribution non-linéaire des pixels dans la paume de la main sans notion de contre-exemples par machine à vecteurs de support. Les pixels de la paume de la main sont des données dérivées de la notion de texture [31] ou directement de l'énergie normalisée [33]. Chacun de ces systèmes montre l'optimisation de critères de comparaison avec un seuil de comparaison personnalisé.

1.3 Plan du mémoire

Ce mémoire est organisé en six chapitres distincts :

1. Le second chapitre décrit le contexte et les systèmes biométriques.
2. Le troisième chapitre effectue un état de l'art des méthodes de détection de la main dans un fond complexe sans utiliser la notion de suivi. Des méthodes d'apprentissage de formes ou de données utilisées par la suite sont également décrites.
3. Le quatrième chapitre présente notre méthode de détection de la main combinant des informations de forme et de couleurs. Des expérimentations qualitatives et quantitatives sont également formulées.
4. Le cinquième chapitre présente un état de l'art des méthodes de reconnaissance de la main basée sur la forme et sur la paume. Une étude comparative des taux de reconnaissance des

principales méthodes est également proposée.

5. Le sixième chapitre énumère nos différentes propositions de systèmes biométriques basées sur la paume ou sur la fusion de la paume et de la forme de la main. Il montre ensuite dans une phase d'évaluation les performances des méthodes proposées comparées à l'état de l'art.
6. Finalement, le septième chapitre conclut ce document et propose des perspectives de travail pour améliorer la reconnaissance du système.

Chapitre 2

La reconnaissance biométrique

Sommaire

2.1	Introduction	5
2.2	Apports de la biométrie	7
2.3	Les systèmes biométriques	7
2.3.1	Décomposition en modules	7
2.3.2	Enrôlement, authentification et identification	8
2.4	Évaluation et taux d'erreurs des systèmes biométriques	10
2.5	Présentation de différentes modalités biométriques	13
2.5.1	ADN (Acide désoxyribonucléique)	14
2.5.2	Démarche	14
2.5.3	Dynamique de la frappe au clavier	15
2.5.4	Empreinte digitale	15
2.5.5	Empreinte palmaire	16
2.5.6	Géométrie de la main	16
2.5.7	Iris	16
2.5.8	Rétine	16
2.5.9	Signature dynamique	17
2.5.10	Thermographie de la main	17
2.5.11	Visage	17
2.5.12	Voix	17
2.5.13	Autres Systèmes	18
2.5.14	Multi-biométrie	18
2.6	Avantages et inconvénients des systèmes biométriques	18

2.1 Introduction

De tout temps, les hommes ont utilisé les caractéristiques de visage, de voix ou de démarche pour se reconnaître. Au milieu du XIX siècle, Alphonse Bertillon a énoncé l'idée d'utiliser des mesures corporelles basées sur les empreintes digitales afin de résoudre des crimes. Très rapidement, dès 1893, the Home Ministry Office acceptait l'idée que deux individus ne pouvaient avoir les mêmes empreintes digitales. Après cette énonciation, de nombreux scientifiques ont recherché des méthodes afin de créer un livre d'empreintes digitales des criminels et ont formé les premiers

experts de la reconnaissance par les empreintes. Malgré des méthodes ingénieuses afin d'augmenter l'efficacité des recherches d'individus présents dans ce livre, les demandes, trop nombreuses, ont entraîné des identifications inefficaces car beaucoup trop tardives. Des recherches ont ainsi été effectuées afin de pouvoir acquérir et stocker une représentation numérique des empreintes digitales. Cette transcription a permis des études et des projets d'identification automatique/semi-automatique ces dernières décennies. Les applications biométriques ont cependant réellement émergées en raison d'une utilisation intensive dans les applications civiles.

L'utilisation de plus en plus importante de systèmes informatiques ou de services électroniques nécessite une plus grande sécurité d'accès pour éviter les usages abusifs ou frauduleux. Ce besoin est sensible dans différents domaines comme le paiement électronique, l'accès à des ressources informatiques, l'accès à des locaux protégés, ... Tous ces systèmes doivent authentifier ou identifier la personne désirant accéder à un service, c'est-à-dire vérifier ou déterminer l'exactitude de son identité. Pour ces accès, les méthodes d'authentification traditionnelles sont basées sur la connaissance d'une information (code confidentiel et mot de passe) ou la possession d'un objet (clé, badge, carte à puce,...). Ces méthodes présentent des inconvénients : les objets peuvent être volés ou perdus, les codes et mots de passe peuvent être oubliés, divulgués, espionnés (car il n'est pas toujours facile de les saisir à l'abri des regards indiscrets) et finalement rejoués. En pratique les personnes choisissent, en effet, souvent des mots de passe aisément devinables (comme des mots du dictionnaire ou des dates de naissance), un mot de passe unique pour sécuriser des accès différents, les écrivent parfois à proximité de leur ordinateur, voire sur les cartes d'accès, les rendant quasiment inutiles. En conséquence, il est souvent nécessaire par sécurité de les changer régulièrement.

En comparaison, en se référant à l'être, la biométrie pallie ces défauts en utilisant les caractéristiques physiques de la personne pour vérifier son identité. Le code confidentiel est alors remplacé par une référence contenant les paramètres extraits du caractère biométrique retenu pour l'authentification. Cette référence peut être mémorisée sous la forme d'un fichier. Authentifier la personne revient donc à comparer un caractère biométrique dont elle est obligatoirement porteuse à une référence enregistrée au préalable et calculée à partir de ce même caractère biométrique. Une seconde approche, l'identification, consiste à comparer le caractère biométrique de la personne à un ensemble de références afin de déterminer l'identité de celle-ci. Toutes les mesures biologiques ne peuvent pas être utilisées comme une caractéristique biométrique. Afin d'être utilisée dans des systèmes sécurisés, une caractéristique doit satisfaire les conditions suivantes [25] :

- **Universelle** : chaque personne doit posséder la caractéristique.
- **Distinctive** : deux personnes doivent avoir des caractéristiques différentes.
- **Permanente** : la caractéristique doit être invariante au cours du temps.
- **Collectionnable** : la caractéristique doit pouvoir être acquise.

En pratique, d'autres clauses doivent être incluses. Ainsi, un système biométrique doit également remplir les conditions suivantes :

- **Performance** : le système biométrique doit permettre une reconnaissance précise, rapide et robuste à des changements opérationnels et environnementaux.
- **Acceptabilité** : la caractéristique biométrique doit être bien acceptée par les utilisateurs.
- **Infalsifiabilité** : le système doit être difficilement utilisable par des méthodes frauduleuses et des attaques.

2.2 Apports de la biométrie

La biométrie constitue une réponse à la complexité croissante des interfaces des services. Outre son côté ergonomique, la biométrie apporte un certain nombre d'avantages par rapport aux méthodes plus classiques d'authentification dont le secteur des télécoms pourrait tirer profit :

- Pas de risque de perte ou de vol, contrairement à la clé ou au code PIN,
- Simplification d'usage (pas de mémorisation nécessaire, le client porte ses droits d'accès sur lui),
- Caractéristiques plus difficilement imitables qu'un code,
- Outil ergonomique de personnalisation (la biométrie est le dernier maillon de la chaîne d'accès à un système informatique, elle « touche » de ce fait le client),
- Non délégation des droits cédés à un utilisateur : l'opérateur maîtrise les droits cédés à un utilisateur qui ne peut déléguer à une tierce personne (droits d'accès sur un intranet ou abonnement mobile d'entreprise) ; ce qui apparaît comme une contrainte pour la personne est un avantage pour l'opérateur,
- Non-répudiation d'une transaction (puisque'elle est non délégable, la signature biométrique donne l'identité du responsable d'une transaction),
- Apport de nouvelles fonctionnalités (par exemple la sécurisation par l'empreinte digitale des appels sortants sur un mobile, qui serait trop contraignante avec le code PIN, est possible avec l'usage de la biométrie),
- Élément de différenciation vis-à-vis du marché (la biométrie peut être vue par certains comme un plus commercial).

En pratique, la biométrie peut remplacer entièrement les systèmes classiques (codes et objets), ou bien fonctionner en complément de ceux-ci pour s'intégrer à un schéma existant. Il y a par exemple des systèmes qui comparent l'empreinte digitale d'une personne avec celle contenue dans une carte à puce.

2.3 Les systèmes biométriques

2.3.1 Décomposition en modules

Un système biométrique est un système de reconnaissance de formes se divisant en quatre modules :

- **Module d'acquisition** : il capture la donnée biométrique d'un individu. Des exemples de capteurs biométriques sont présentés à la Fig. 2.1.
- **Module d'extraction des caractéristiques** : il détermine un ensemble de caractéristiques biométriques à partir de la donnée acquise.
- **Module de comparaison** : il compare les caractéristiques extraites durant la phase de reconnaissance avec des caractéristiques enregistrées afin de déterminer un score de similarité. Ce module est aussi un module de décision dans lequel un utilisateur indique son identité qui est confirmé ou non (authentification). Dans un autre mode, l'identifiant d'un utilisateur est établi (identification) à partir du score de similarité.
- **Module de base de données** : il stocke les références biométriques des utilisateurs enrôlés. Suivant l'application, la référence est stockée dans une base de données centrale ou, pour plus de sécurité, enregistrée dans une carte individuelle.

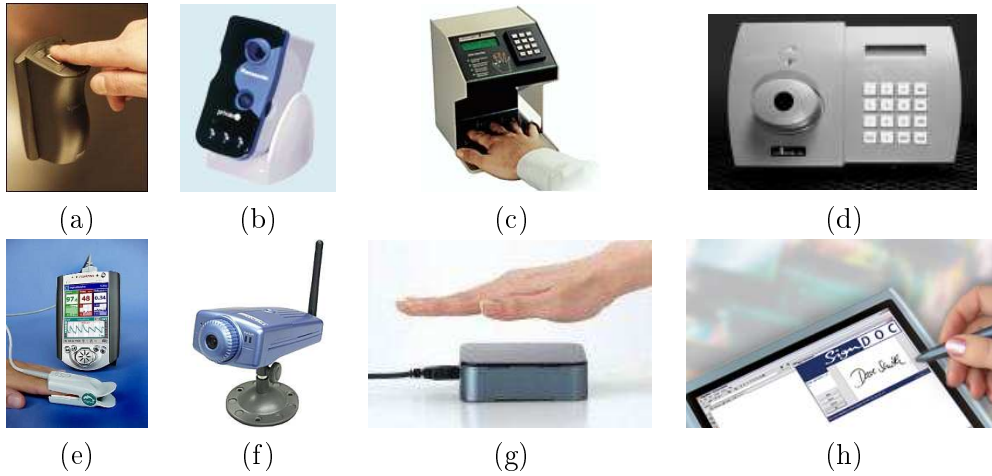


FIGURE 2.1 – Exemples de capteurs biométriques : (a) empreinte digitale, (b) iris (c) géométrie de la main, (d) rétine, (e) pulsation, (f) visage, (g) thermographie de la main et (h) signature.

2.3.2 Enrôlement, authentification et identification

Un système biométrique peut être utilisé suivant trois modes d'utilisation : l'enrôlement, l'authentification et l'identification (Fig. 2.2).

En phase d'enrôlement, la donnée biométrique est capturée par le module d'acquisition. Suivant la qualité de l'acquisition, la donnée biométrique peut être rejetée et une autre capture désirée. Si la donnée est de bonne qualité, les caractéristiques biométriques extraites sont nommées référence ou signature. Généralement, plusieurs références sont enregistrées pour chaque individu et celles-ci peuvent être mises à jour pendant l'utilisation. Pour des raisons de sécurité, les références de chaque individu sont cryptées. Cette sécurité permet de rendre impossible la reconstitution des caractéristiques physiques et limite l'utilisation d'une référence à l'application visée.

En mode authentification, le système vérifie si un utilisateur est bien celui qu'il prétend être en comparant la donnée biométrique capturée et sa référence stockée par le module de base de données. Classiquement, un individu désirant être reconnu précise son identité référant à une signature dans la base de données. Cette signature est comparée à la donnée biométrique capturée ce qui conduit à une comparaison un pour un. Le système détermine si la demande est fondée ou non.

Le problème d'authentification peut être formulé par :

Soit un vecteur d'entrée de caractéristiques X_E défini à partir du module d'extraction des caractéristiques et une identité déclarée I_d , déterminer si $A(X_E, I_d)$ est vrai ou faux où vrai indique que l'identité déclarée est exacte (il s'agit du bon utilisateur) et faux que l'identité déclarée est erronée (il s'agit d'un imposteur). Soit X_{I_d} la référence biométrique de l'utilisateur I_d , $A(X_E, I_d)$ peut être défini par :

$$A(X_E, I_d) = \begin{cases} \text{vrai} & \text{si } S(X_E, X_{I_d}) \geq s \\ \text{faux} & \text{sinon} \end{cases} \quad (2.1)$$

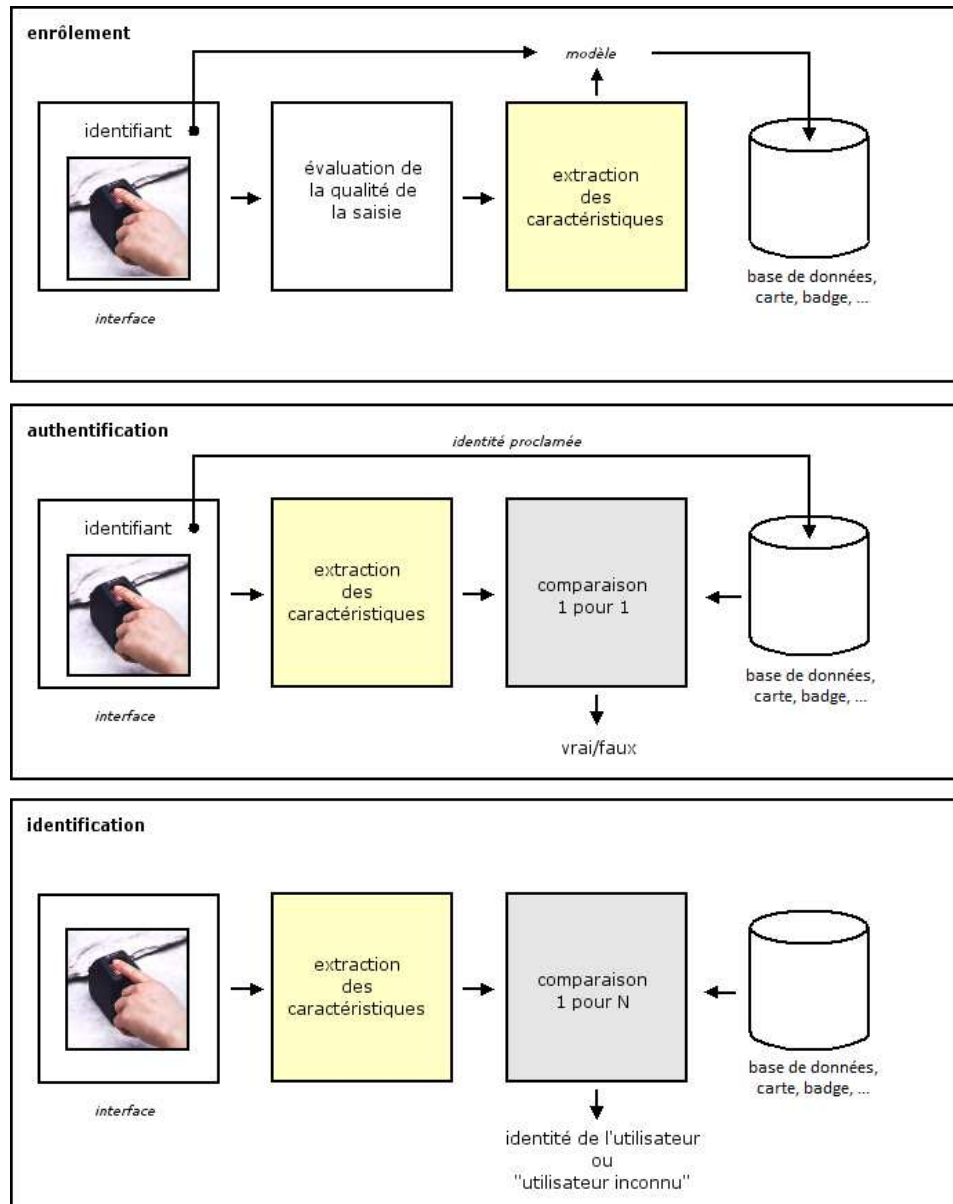


FIGURE 2.2 – Diagramme d'utilisation biométrique

où S est la mesure de similarité entre les vecteurs X_E et X_{I_d} , et s est un seuil fixé. La valeur $S(X_E, X_{I_d})$ est nommée comme le score de similarité entre deux caractéristiques biométriques. Il est intéressant de remarquer que les caractéristiques biométriques provenant d'un même individu ne sont jamais parfaitement identiques ce qui explique l'introduction du seuil s .

En mode identification, le système doit reconnaître un individu en cherchant parmi les références de tous les individus dans la base de données. Ainsi, le système recherche l'identité d'un individu sans qu'il la précise ce qui conduit à une comparaison un contre N où N est le nombre d'utilisateurs enregistrés dans la base. Le système doit également indiquer une erreur si l'utilisateur n'est pas enregistré dans la base. L'identification est utilisée dans deux contextes : elle peut être utilisée pour déterminer l'identité d'un individu à priori inconnu (comme lors de l'identification par empreinte digitale dans les affaires policières) ou pour faciliter l'identification d'un individu qui n'a plus à préciser un identifiant ou présenter une preuve d'identité (telle qu'une carte personnelle).

Comme le mode d'authentification, le problème d'identification peut être formalisé.

Soit un vecteur d'entrée de caractéristiques X_E , l'identification revient à déterminer l'identité I_k , $k \in \{0, 1, 2, \dots, N\}$ où $I_1, I_2, \dots, I_N$ sont les identités des utilisateurs enregistrés dans le système et I_0 indique le cas d'échec où aucune identité ne peut être indiquée pour l'utilisateur. La fonction d'identification $I(X_E)$ peut ainsi être exprimée par :

$$I(X_E) = \begin{cases} I_k & \text{si } \left[\max_{(1 \leq k \leq N)} (S(X_E, X_{I_k})) \right] \geq s \\ I_0 & \text{sinon} \end{cases} \quad (2.2)$$

où X_{I_k} est la référence biométrique correspondant à l'identifiant I_k et s est un seuil fixé.

2.4 Évaluation et taux d'erreurs des systèmes biométriques

Les taux d'erreur sont l'une des caractéristiques d'un système biométrique. A cause des conditions ambiantes (température, luminosité...), de la qualité d'acquisition (bruit du capteur, positionnement de l'utilisateur...) et des changements physiques ou comportementales de l'utilisateur (coupures, prise de poids...), deux caractéristiques biométriques provenant de la même personne ne sont pas exactement identiques. La comparaison entre deux caractéristiques biométriques retourne donc une valeur de comparaison. Plus la valeur est importante et plus il est probable que les deux caractéristiques proviennent de la même personne. Comme indiqué dans les équations (2.1) et (2.2), un seuil s permet d'effectuer la décision de classement : si le score de comparaison est supérieur à s alors les deux caractéristiques proviennent de la même personne, sinon elles sont déclarées comme provenant de personnes différentes. Afin d'évaluer la qualité d'un système biométrique, la distribution des scores de comparaison déterminée à partir de couples d'exemples provenant de la même personne nommée distribution des authentiques est spécifiée. De même, la distribution des imposteurs est spécifiée à partir de couples d'exemples provenant de personnes différentes. Ces deux courbes de distribution sont illustrées à la Fig. 2.3.

Un système biométrique se caractérise par deux types d'erreur, les faux rejetés et les faux acceptés :

- **Faux accepté** : deux caractéristiques biométriques provenant de deux personnes différentes sont indiquées comme provenant de la même personne.

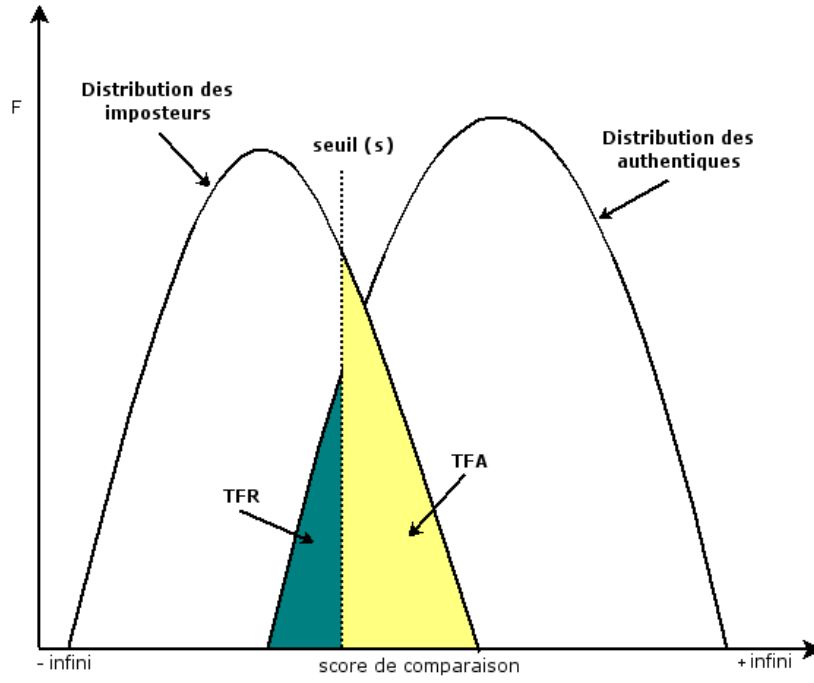


FIGURE 2.3 – Courbes de distribution des imposteurs et des authentiques

- **Faux rejeté** : deux caractéristiques biométriques provenant de la même personne sont classées comme provenant de deux personnes différentes.

Les deux taux d'erreur correspondants, le taux de faux acceptés (TFA) et le taux de faux rejetés (TFR) respectivement, sont ambivalents. Il est ainsi nécessaire d'établir un compromis entre ces deux taux d'erreur. Ce compromis est effectué en ajustant le seuil s ; si s augmente pour avoir un système plus sécurisé alors le TFR augmente et le TFA baisse. Inversement, si le seuil s est diminué afin d'avoir un système plus tolérant alors le TFR baisse et le TFA augmente. Les performances des systèmes biométrique sont reportés sur un graphe nommé Receiver Operating Curve (ROC) déterminé par l'ensemble des points du TFA contre les points du TFR en faisant varier le seuil s (voir Fig. 2.4). Une courbe ROC peut être exprimée par :

$$ROC(s) = (TFA(s), TFR(s)) \quad (2.3)$$

Dans le mode authentification, les courbes de distribution des imposteurs et des authentiques peuvent être déterminées. Une base de vecteurs de caractéristiques biométriques provenant d'un ensemble d'utilisateurs d'identifiant I_k , $k \in \{1, 2, \dots, N\}$ est collectée où N est le nombre d'utilisateurs dans la base. Pour chaque identifiant I_k , N_k acquisitions sont effectuées. Soit $X_{I_k}^i$ le vecteur des caractéristiques provenant de l'acquisition i de l'individu I_k , la courbe de distribution des imposteurs (CDI) et la courbe de distribution des authentiques (CDA) sont définies formellement par :

$$CDI(x) = p(x = S(X_{I_k}^i, X_{I_l}^j) | k \neq l) \quad (2.4)$$

$$CDA(x) = p(x = S(X_{I_k}^i, X_{I_k}^j) | i \neq j) \quad (2.5)$$

où $S(X, Y)$ mesure la similarité entre les vecteurs X et Y , $(k, l) \in \{1, 2, \dots, N\}^2$ et $(i, j) \in \{1, 2, \dots, N_k\}^2$. Les TFA et les TFR peuvent également être formulés en utilisant les probabilités

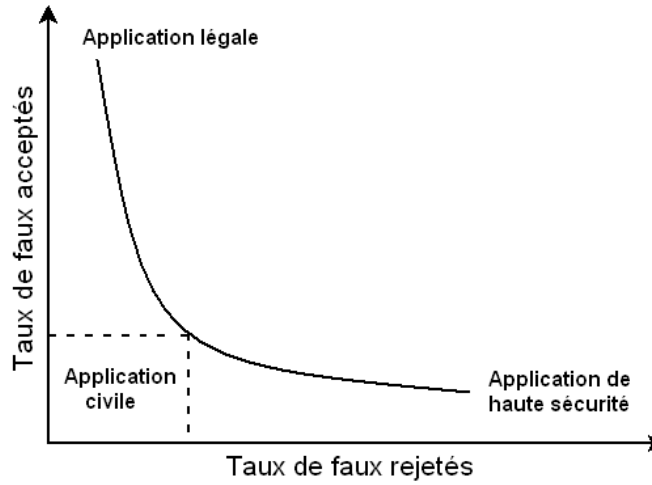


FIGURE 2.4 – Courbe ROC

conditionnelles. Ainsi, ces deux taux d'erreurs peuvent être exprimés par :

$$TFA(s) = p(s \leq S(X_{I_k}^i, X_{I_l}^j) | k \neq l) \quad (2.6)$$

$$TFR(s) = p(s > S(X_{I_k}^i, X_{I_l}^j) | i \neq j) \quad (2.7)$$

où $S(X, Y)$ mesure la similarité entre les vecteurs X et Y , $(k, l) \in \{1, 2, \dots, N\}^2$ et $(i, j) \in \{1, 2, \dots, N_k\}^2$.

Dans le mode identification, le TFA et le TFR peuvent être déterminés de façon similaire en utilisant l'équation (2.2). Dans ce mode, les taux d'erreur dépendent du nombre d'identités N dans la base de données. Ainsi, le taux de faux acceptés et le taux de faux rejetés sont désignés par TFA_N et TFR_N respectivement. Ils peuvent être formulés par :

$$TFA_N(s) = p(s \leq \left[\max_{1 \leq k \leq N} (S(X_{I_k}^i, X_{I_l}^j)) \right] \wedge S(X_{I_l}^j, X_{I_l}^m) \leq \left[\max_{1 \leq k \leq N} (S(X_{I_k}^i, X_{I_l}^j)) \right] | k \neq l \wedge j \neq m) \quad (2.8)$$

$$TFR_N(s) = p(s \leq S(X_{I_l}^j, X_{I_l}^m) < \left[\max_{1 \leq k \leq N} (S(X_{I_k}^i, X_{I_l}^j)) \right] | k \neq l \wedge j \neq m) + p(s > S(X_{I_k}^i, X_{I_l}^j) | i \neq j) \quad (2.9)$$

Des approximations peuvent être effectuées afin de déterminer le TFA_N et le TFR_N à partir du TFA et du TFR respectivement. De bons résultats d'approximation peuvent être obtenus lorsque $N \cdot TFA(s) < 0.1$, ainsi :

$$TFA_N(s) \simeq 1 - (1 - TFA(s))^N \simeq N \cdot TFA(s) \quad (2.10)$$

$$TFR_N(s) \simeq TFR(s) \quad (2.11)$$

Dans les bases de données importantes composées de plusieurs milliers de références, des algorithmes de classification sont utilisés. Ces méthodes permettent de classer et d'indexer les

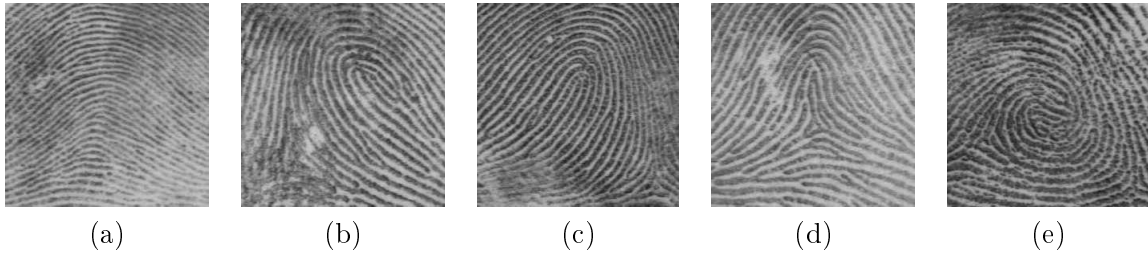


FIGURE 2.5 – Les cinq classes principales des empreintes digitales : (a) arche, (b) boucle gauche, (c) boucle droite, (d) ogive, (e) spirale.

caractéristiques biométriques dans des groupes distincts afin d'effectuer des recherches d'identification dans une seule partie de la base de données. Des exemples de classes pour les empreintes digitales sont illustrées à la Fig. 2.5. Les algorithmes d'indexation possèdent un taux d'erreur de classement (TEC) indiquant le pourcentage de mauvaise classification des caractéristiques biométriques. Soit P le pourcentage de l'espace de recherche dans la base de données utilisée pour l'identification, le TFA_N et le TFR_N peuvent être définis par :

$$TFA_N(s) \simeq 1 - (1 - TFA(s))^{N \cdot P} \simeq N \cdot P \cdot TFA(s) \quad (2.12)$$

$$TFR_N(s) \simeq TEC + (1 - TEC) \cdot TFR(s) \quad (2.13)$$

Il est à noter que dans les équations (2.11) et (2.13), une approximation est effectuée en supposant qu'il n'y a pas d'imposteur dont le score de comparaison est supérieur au seuil s et au score de comparaison avec la référence de l'individu. Dans les équations (2.10) et (2.12), il est supposé qu'il n'y a pas d'imposteur dont le score de comparaison est supérieur au seuil s et inférieur au score de comparaison avec la référence de l'individu.

La sécurité souhaitée d'un système biométrique est très dépendante de l'utilisation. Dans des systèmes de très haute sécurité tels que les centrales nucléaires par exemple, il est impératif qu'aucun imposteur ne puisse se faire autoriser l'accès par erreur ; le TFA doit être le plus faible possible. *A contrario*, dans les systèmes d'identification par les empreintes digitales dans les enquêtes policières, le TFR doit être très proche de zéro (même s'il est nécessaire ensuite que les fausses détections soient examinées manuellement) afin de limiter le risque de ne pas sélectionner un individu présent dans la base. Un TFR faible est également souhaité dans les applications grand public pour privilégier les côtés ergonomique et d'acceptation du service. Plus généralement, dans les applications civiles un compromis est effectué entre les deux erreurs. Le compromis absolu, lorsque le taux de faux acceptés et le taux de faux rejetés sont identiques, est nommé le niveau d'égale erreur du système (EER).

2.5 Présentation de différentes modalités biométriques

Les applications de biométrie se sont développées de plus en plus ces dernières décennies. Elles sont réparties en trois groupes majeurs d'utilisation :

- **Commercial** : utilisation pour la sécurisation d'accès physique ou virtuel tel que le contrôle d'accès physique, la sécurisation des transactions électroniques, l'accès à Internet ou à un réseau local, les cartes de crédit, etc.
- **Gouvernemental** : application dans les cartes d'identité, le permis de conduire, la carte de sécurité sociale, le passeport, etc.

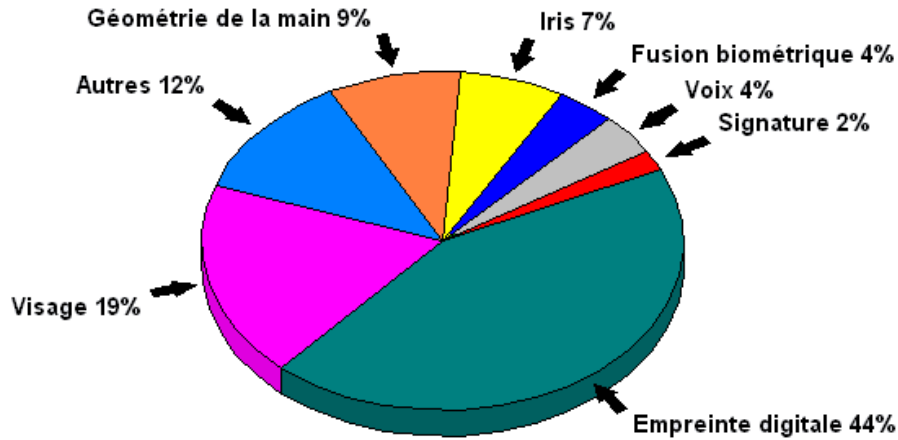


FIGURE 2.6 – Marché de la biométrie

- **Sécuritaire** : reconnaissance dans les investigations criminelles, l'identification terroriste, etc.

Pour chacune de ces utilisations, les contraintes biométriques sont différentes. Aucune biométrie unique ne peut répondre à toutes ces exigences. Ainsi, le type de biométrie utilisée dépend essentiellement de l'application envisagée. La Fig 2.6 indique la décomposition du marché de la biométrie suivant le type de technologie utilisée. Les méthodes biométriques sont divisées en trois catégories de technologies :

- **Analyses biologiques** : ADN, odeur, ...
- **Analyses comportementales** : dynamique de la frappe au clavier, démarche, signature,...
- **Analyses morphologiques** : empreintes digitales, géométrie de la main, iris, visage,...

Un bref descriptif de différents traits biométriques illustrés à la Fig. 2.7 donnés ci-dessous présente les avantages et les inconvénients de chaque modalité [54].

2.5.1 ADN (Acide désoxyribonucléique)

L'ADN est un code unique pour chaque individu, excepté pour les jumeaux homozygotes. Il peut être extrait à partir de sang, de peau ou de cheveux. L'identification par ADN est ainsi une méthode très répandue dans les utilisations sécuritaires. Cette reconnaissance possède trois limites majeures : il est facile de voler un "morceau" d'ADN pouvant être utilisé de façon frauduleuse, le processus de reconnaissance est lent et coûteux et l'ADN contient des informations telles que les maladies génétiques qui pourraient être utilisées à mauvais escient.

2.5.2 Démarche

La démarche est supposée unique pour chaque individu. Elle est clairement peu caractéristique et peu stable dans le temps mais est suffisamment caractéristique pour une application

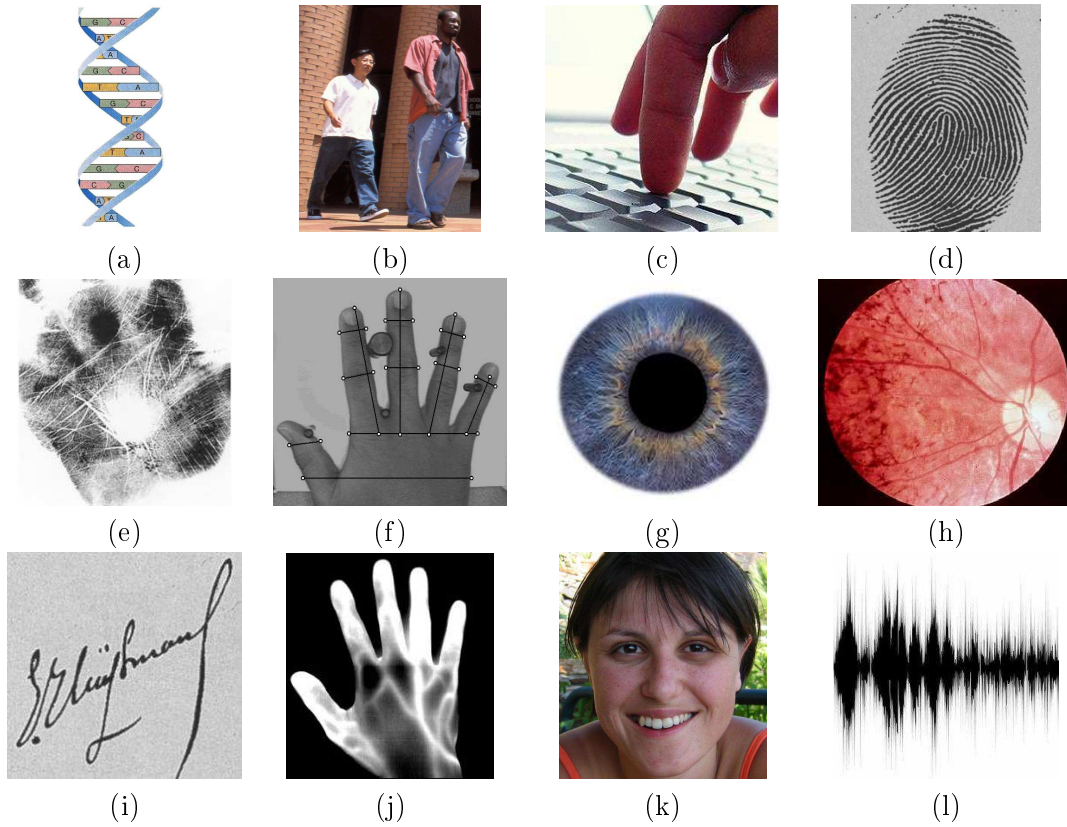


FIGURE 2.7 – Les principales caractéristiques biométriques : (a) ADN, (b) Démarche, (c) Dynamique de la frappe au clavier, (d) Empreinte digitale, (e) Empreinte palmaire, (f) Géométrie de la main, (g) Iris, (h) Rétine, (i) Signature, (j) Thermographie de la main, (k) Visage, et (l) Voix

de faible sécurité. Deux méthodes d'acquisition sont utilisées. L'une s'appuie sur une séquence d'images d'une personne marchant d'où sont extraites les mouvements de points importants (coude, genou, poignet, etc.). L'autre méthode d'acquisition, utilisée dans les téléphones mobiles expérimentalement, emploie un accéléromètre permettant d'obtenir les changements de vitesse ainsi que l'orientation 3D de ces changements lors de la marche de l'utilisateur.

2.5.3 Dynamique de la frappe au clavier

La dynamique de la frappe au clavier est le processus qui analyse la façon de frapper au clavier en surveillant le nombre et le rythme de frappe par seconde. Cette caractéristique n'est pas unique à chaque individu mais présente un fort confort d'utilisation tout en supprimant l'acquisition d'un équipement supplémentaire. Cette méthode est une biométrie comportementale, elle n'est donc pas stable au cours du temps ce qui peut limiter son utilisation.

2.5.4 Empreinte digitale

Les empreintes digitales sont la technologie biométrique la plus mature. Cette technologie est utilisée depuis plus d'un siècle et de très nombreuses recherches ces dernières décennies ont permis d'obtenir de très bons taux de reconnaissance. Une empreinte digitale est une structure composée de vallées et de crêtes, de points caractéristiques appelés minuties (bifurcation, fin

de lignes, lac, etc.) et de pores de la peau uniques à chaque individu (même pour les jumeaux homozigotes). La précision de la reconnaissance par empreinte digitale permet l'utilisation dans des bases de données de grandes tailles (de plusieurs millions d'individus). Le faible coût et la petite taille des capteurs permettent en plus une utilisation dans les téléphones et les ordinateurs portables. Cependant, une partie de la population ne peut pas utiliser la reconnaissance par les empreintes digitales à cause de l'âge, de facteurs génétiques, environnementaux ou de travail (les travailleurs manuels n'ont souvent plus d'empreintes digitales).

2.5.5 Empreinte palmaire

La paume de la main est une technologie récente. Elle contient comme les empreintes digitales des vallées, des rides, etc. sur une surface cependant plus large. Cette technologie est ainsi supposée plus caractéristique que les empreintes digitales. Deux types de résolution sont utilisées pour la reconnaissance par la paume : la haute et la faible résolution. En haute résolution, toutes les caractéristiques de la main peuvent être obtenues ce qui permet une reconnaissance de haute performance. Dans les applications de moyenne sécurité, un capteur de faible résolution et donc de faible prix peut être utilisé pour extraire les lignes principales et les rides de la paume. Le chapitre 5 précise les méthodes d'extraction et de comparaison de caractéristiques des systèmes de reconnaissance palmaire.

2.5.6 Géométrie de la main

La géométrie de la main ou des doigts est une méthode physique de reconnaissance très simple s'appuyant sur la taille de la paume, la largeur, la longueur et la courbure des doigts. Bien que les méthodes d'extraction et de comparaison de caractéristiques soient très simples, la géométrie de la main permet une reconnaissance relativement robuste et de nombreux systèmes sont installés pour des contrôles physiques. L'avantage de cette technique est qu'elle est complètement robuste aux facteurs environnementaux tels que l'humidité ainsi qu'à des facteurs individuels tels que des coupures ou la peau sèche. Les principales limitations consistent à extraire des informations géométriques pertinentes même lorsque les utilisateurs portent des bagues ou sont limités dans leurs mouvements à cause d'une arthrite par exemple.

2.5.7 Iris

L'iris est la région annulaire colorée de l'œil située derrière la cornée. La texture visuelle de l'iris est unique et très stable au cours du temps ; elle est stabilisée dès deux ans. La précision et la vitesse de la reconnaissance par l'iris permettent une identification efficace sur de très larges bases de données. Chaque iris est distinct entre deux individus même pour les vrais jumeaux et il est très difficile de modifier son iris ou de faire un faux (contrairement à l'empreinte digitale par l'exemple). La limite de cette approche vient du coût du capteur et d'une utilisation non intuitive du capteur d'acquisition (même si ces deux facteurs tendent à largement s'améliorer).

2.5.8 Rétine

Le réseau sanguin de la rétine est très riche en structures et est réputé caractéristique pour chaque individu. La reconnaissance par la rétine requiert une coopération de l'utilisateur plus forte que l'iris avec un contact entre le capteur et l'œil. Le réseau vasculaire permet également de connaître des données médicales telles que l'hypertension. Cependant, ce principe de reconnaissance est réputé être le système le plus sécurisé puisqu'il est impossible de changer le réseau

vasculaire de la rétine sans danger physique et qu'il est également très difficile de répliquer un réseau rétinien.

2.5.9 Signature dynamique

La signature d'un individu est utilisée dans le domaine des transactions commerciales ou des documents gouvernementaux. Il est cependant très difficile de caractériser parfaitement la signature d'un individu qui est très variable au cours du temps selon l'état d'esprit de l'utilisateur. Généralement, la vitesse et la géométrie de la signature sont utilisées comme référence biométrique. Le principal inconvénient de cette méthode vient des imitateurs professionnels pouvant duper le système.

2.5.10 Thermographie de la main

Les motifs de peau (de la main, d'un doigt ou du visage) capturées par infrarouge de longueur d'onde proche de celle de la chaleur corporelle sont caractéristiques de chaque individu. L'avantage principal de ce type de méthode est l'acquisition qui s'effectue sans contact mais qui peut s'avérer très difficile dans le cas d'environnement incontrôlé pour lequel la température est très variable. Cependant, les veines du dos de la main peuvent être obtenues avec une longueur d'onde adéquate afin de caractériser un individu. Le prix très élevé des capteurs infrarouges reste prohibitif pour une utilisation grand public. Le dernier système basé sur cette méthode et développé par la société Fujitsu est cependant utilisé à grande échelle dans certaines banques Japonaises pour l'authentification d'utilisateurs lors du retrait d'argent aux automates bancaires.

2.5.11 Visage

La reconnaissance par le visage est la méthode d'identification la plus naturelle ; c'est celle que tout le monde utilise tous les jours pour se reconnaître. De très nombreuses recherches se concentrent sur ce secteur biométrique très porteur car pouvant être effectué à distance avec ou sans la coopération de l'utilisateur. La reconnaissance du visage reste cependant l'une des méthodes biométriques dont l'extraction des caractéristiques est la plus difficile. En effet, le visage dans l'image d'acquisition est très variable à cause des occlusions partielles (lunettes, foulards, etc), des changements d'illumination, des expressions du visage (joie, tristesse, etc.) et des positions d'acquisition changeantes. Les caractéristiques sont définies suivant deux approches. La première est une méthode géométrique s'appuyant sur la localisation et la forme des yeux, des lèvres ou du nez. D'autres méthodes plus robustes s'appuient sur des études complètes du visage principalement par approches statistiques.

2.5.12 Voix

La voix est la seule technologie biométrique existante s'appuyant sur une combinaison des approches physiques et comportementales. La voix est basée sur la forme et la taille des tractus vocaux (cordes vocales, pharynx, larynx, luette et lèvres). Ces caractéristiques physiques sont invariantes pour chaque individu mais très variables au cours du temps à cause de l'âge, de raisons médicales (gros rhume par exemple) ou de l'état émotionnel. De plus, la voix n'est pas très distinctive entre chaque individu et ne permet pas d'identifier deux jumeaux. Cependant, la reconnaissance par la voix reste la méthode la plus adéquate pour l'authentification biométrique

par téléphone.

2.5.13 Autres Systèmes

D'autres systèmes biométriques font actuellement l'objet de recherche. De manière non-exhaustive voici une liste de caractéristiques utilisées dans des applications potentielles très diverses : l'odeur, les dents, le rythme cardiaque, la forme des lèvres, les signaux électriques du cerveau, etc.

2.5.14 Multi-biométrie

Une autre approche biométrique consiste à effectuer de la multi-biométrie en combinant plusieurs caractéristiques. Par exemple, des méthodes se basant sur le visage et la voix, la géométrie et la paume de la main sont fréquemment utilisées pour limiter les faiblesses de certains traits biométriques et renforcer la décision de reconnaissance. La fusion des informations s'effectue à trois niveaux distincts : fusion à la représentation (fusion des données extraites), fusion à des niveaux intermédiaires (fusion aux niveaux des comparaisons) et fusion à la décision (utilisation de la décision de chaque procédé). Il a été montré que la fusion biométrique améliore le taux de comparaison en moyenne et que le pire des cas de fusion correspond au plus faible taux d'erreur des systèmes fusionnés.

2.6 Avantages et inconvénients des systèmes biométriques

Une comparaison des différentes techniques biométriques suivant les sept critères énoncés initialement (universalité, unicité, permanence, facilité de collecte, performance, acceptabilité et infalsifiabilité) est présentée au tableau 2.1. La qualité accordée à chaque critère est répertoriée en trois catégories, haute, moyenne et faible désignées par les notations ,  et  respectivement. Ce tableau provenant de l'étude [54] indique que les biométries comportementales (signature ou frappe au clavier) présentent des performances de reconnaissance faibles contrairement aux données intrusives telles que les méthodes basées sur la rétine ou l'ADN. Un bon compromis entre les données provient des traits biométriques de l'empreinte palmaire ou digitale ainsi que de la thermographie ou la géométrie de la main. Ces quatre méthodes présentent l'avantage de ne posséder aucun critère négatif. Cependant d'autres critères non mentionnés doivent également être pris en compte pour l'évaluation d'un système biométrique. Le temps d'enrôlement, le temps de comparaison (dépendant de la complexité des algorithmes), la taille des références, la résistance à l'environnement (froid, humidité, éclairage, etc.), la facilité d'intégration, les besoins de maintenance, le coût et la taille des capteurs sont autant d'éléments indiquant l'intérêt d'un système biométrique. Aucune étude comparative exhaustive n'existe afin d'évaluer les différents systèmes suivant ces différentes contraintes.

Caractères	U	N	P	F	E	A	I
ADN	+	+	+	-	+	-	-
Démarche	=	-	-	+	-	+	=
Frappe sur le clavier	-	-	-	=	-	=	=
Empreinte digitale	=	+	+	=	+	=	+
Empreinte palmaire	=	+	+	=	+	=	+
Géométrie de la main	=	=	=	+	=	=	=
Iris	+	+	+	=	+	-	+
Rétine	+	+	=	-	+	-	+
Signature	-	-	-	+	-	+	+
Thermographie de la main	=	=	=	=	=	=	+
Visage	+	-	=	+	-	+	-
Voix	=	-	-	=	-	+	-

TABLE 2.1 – Comparaisons des systèmes biométriques suivant sept critères : Universalité (U), Unicité (N), Permanence (P), Facilité de collecte (F), Performance (E), Acceptabilité (A), Infalsifiabilité (I).

Première partie

Segmentation de la main

Chapitre 3

État de l'art de la segmentation de main

Sommaire

3.1	Introduction à la segmentation	23
3.2	Contexte de la détection de la main	24
3.3	Détection de la main dans les systèmes biométriques	26
3.4	Détection de la main par modélisation de la couleur de la peau	27
3.4.1	Représentation couleur	27
3.4.1.1	Espace <i>RVB</i> - Rouge, Vert, Bleu	28
3.4.1.2	Espace <i>YUV</i> et <i>YC_rC_b</i>	30
3.4.1.3	Espaces <i>HSV</i> et <i>HSL</i> - Teinte, Saturation, Valeur ou Luminance	30
3.4.1.4	Espaces <i>CIE XYZ</i> et <i>RVB normalisé</i>	31
3.4.1.5	Espaces <i>CIE L*a*b*</i> et <i>CIE L*u*v*</i>	33
3.4.2	Méthode de modélisation de la couleur de la peau	34
3.4.2.1	Définition explicite de la couleur de la peau	34
3.4.2.2	Modélisation non-paramétrique de la couleur de la peau	36
3.4.2.3	Modélisation paramétrique	38
3.5	Détection de la main intégrant des connaissances à priori	40
3.5.1	Modèle de forme	40
3.5.2	Approches par classification de zones d'image	41
3.6	Comparaison des différentes méthodes de détection	43
3.7	Extraction de la zone d'intérêt : la paume	45
3.7.1	Segmentation de la paume sous forme carrée	45
3.7.1.1	Extraction des points importants	45
3.7.1.2	Extraction de la paume	45
3.7.2	Segmentation de la paume de façon circulaire	46
3.8	Conclusion sur les approches de détection de la main	46

3.1 Introduction à la segmentation

La segmentation est un processus de partitionnement d'une image en régions disjointes [74]. Elle peut être définie de manière identique en recherchant les bords limitant ses régions. Haralick et Shapiro [48] ont défini les caractéristiques d'une bonne segmentation par la formulation

suivante : "Les régions doivent être uniformes et homogènes suivant la couleur et/ou la texture. L'intérieur des régions doit-être simple et sans de nombreux petits trous. Les régions adjacentes de la segmentation doivent avoir des caractéristiques suffisamment différentes. Le bord des régions doit être simple, lisse et précis". Cette définition a été formellement définie. Ainsi, soit $\mathcal{I}$ une image et $\mathcal{H}$ le critère d'homogénéité. La segmentation de $\mathcal{I}$ est une partition P de $\mathcal{I}$ en un ensemble de N régions $\mathcal{R}_n$, $n = 1, \dots, N$ suivant les quatre contraintes suivantes :

$$\left\{ \begin{array}{l} \bigcup_{n=1}^N \mathcal{R}_n = \mathcal{I} \\ \forall (n, m) \in (1, \dots, N)^2 \text{ tels que } n \neq m, \mathcal{R}_n \cap \mathcal{R}_m = \emptyset \\ \forall n \in (1, \dots, N), \mathcal{H}(\mathcal{R}_n) = \text{vrai} \\ \forall (n, m) \in (1, \dots, N)^2 \text{ tels que } \mathcal{R}_n \text{ et } \mathcal{R}_m \text{ adjacentes, } \mathcal{H}(\mathcal{R}_n \cup \mathcal{R}_m) = \text{faux} \end{array} \right. \quad (3.1)$$

Les deux premières conditions formulent que l'ensemble des régions couvre la globalité de l'image et que les régions sont disjointes respectivement. La troisième condition indique que chaque région satisfait le critère d'homogénéité $\mathcal{H}$. Finalement la quatrième condition exprime que deux régions adjacentes sont "suffisamment différentes" et ainsi ne peuvent pas être fusionnées en une seule région satisfaisant $\mathcal{H}$.

La segmentation est l'une des tâches les plus importantes dans les applications de traitement d'images. En effet dans de très nombreux systèmes, elle représente la première étape de traitement d'une image. Suivant les images et les applications, le critère d'homogénéité est défini soit par la couleur (ou le niveau de gris si l'information couleur n'est pas présente dans l'image), soit par la notion de texture. Initialement, les techniques de segmentation ont été proposées sur des images en niveau de gris afin de limiter la complexité de calcul. Avec l'augmentation de la puissance des machines, cette limitation n'est plus réellement d'actualité. Cependant d'autres contraintes précisées au paragraphe 3.2 sont appliquées pour limiter la complexité de la tâche de la segmentation, en particulier pour celle de segmentation de la main qui nous concerne.

3.2 Contexte de la détection de la main

Les nouvelles interfaces d'interactions Homme-Machine utilisent tous les moyens possibles de communication et de manipulation naturelles de l'Homme (gestes, paroles, etc.). Ces nouvelles méthodes permettent une meilleure interaction avec la machine mais induisent des problèmes de compréhension des entrées qui sont beaucoup plus complexes que les entrées couramment utilisées (clavier ou souris). De toutes les parties du corps humain, la main est l'élément le plus efficace et le plus naturel pour manipuler un objet et pour communiquer sans utiliser le sens auditif.

Les méthodes les plus robustes et efficaces pour détecter les mouvements de la main s'appuient sur des capteurs sensoriels (les *data gloves*) [106]. Ces gants permettent grâce à de nombreux capteurs de parfaitement identifier la position de la main et des doigts dans l'espace en 3D. Grâce à cet équipement, des interfaces intuitives sont mise en place. Dans notre contexte, ce type de capteurs n'est pas utilisable puisqu'ils sont très onéreux, requièrent une calibration complexe et surtout ils sont assez encombrants et peu naturels. De plus notre objectif est de permettre une reconnaissance biométrique sans capteur spécifique, en utilisant donc une webcam ou la caméra

d'un téléphone mobile. De nombreuses méthodes s'appuyant sur le traitement d'images pour la détection de main dans une scène complexe ont ainsi été développées. La majorité de ces approches ont été répertoriées dans plusieurs états de l'art sur le domaine pour l'estimation de la position, la modélisation de la main ou la reconnaissance des gestes [37, 85, 128, 127, 119].

L'estimation de la pose de la main peut être divisée en deux ensembles de méthodes : les approches basées sur des modèles de formes et les approches basées sur un ensemble de vues. Le premier ensemble consiste à utiliser un modèle 2D ou 3D de la main. Le modèle est ensuite projeté dans l'image et une fonction d'erreur détermine la qualité de comparaison entre les éléments de l'image et le modèle. Un algorithme itératif permet ensuite d'adapter le modèle afin que l'erreur soit minimale. Dans la seconde approche, le problème est un problème de classification à partir d'un ensemble de références. Les références sont définies à partir d'un ensemble de mains annotées et classées et permettent la mise en place d'un classifieur. Cette méthode est principalement utilisée pour la reconnaissance de gestes. Dans les deux cas, l'estimation de la pose de la main dépend initialement de la segmentation de celle-ci car le temps d'exécution de ces méthodes est trop important pour être appliqué sur une image entière dans des conditions de temps réel. La faible texture de la main implique une segmentation de celle-ci en se basant uniquement sur la notion de formes et de couleurs. Ces deux caractéristiques sont directement déduites des données anatomiques de la main et de la physiologie de la peau décrites en annexe.

Elles sont supposées caractéristiques par rapport à l'ensemble de tous les objets existants sur une image. Ainsi, l'objectif des méthodes de détection de main dans une image consiste à définir un processus permettant de caractériser tous les types de main (de différents tons de peau, de tailles variables, ...) sans effectuer de détections erronées.

Afin de détecter la main, les deux ensembles de méthodes de détection de la main (approches basées sur des modèles de formes et approches basées sur un ensemble de vues) utilisent un processus d'apprentissage pour modéliser la forme ou la couleur de la main. Ensuite à partir d'un ou des deux modèles, un autre processus permet la recherche de mains dans une image fixe ou le suivi dans une séquence d'images. Ce processus de détection de la main doit répondre à des contraintes générales de segmentation et des limites particulières liées à l'anatomie de la main. Toutes ces contraintes peuvent être exprimées par [37] :

1. **Problème de grande dimension** : La main est un objet complexe et articulé comprenant plus de 25 degrés de liberté déterminés à partir de son anatomie. Le modèle cinétique simplifié introduit en [69] et utilisé dans de nombreux systèmes, comprend 27 degrés de liberté (Fig. 3.1).
2. **Environnement complexe et non-contrôlé** : Pour une utilisation sans contrainte, les systèmes de détection de la main doivent fonctionner sans connaissance à priori de l'arrière plan et de la luminosité. Ces deux éléments extérieurs sont très influents sur la qualité de la segmentation et restent pour le moment un challenge important dans le traitement d'images. Les méthodes de segmentation basées sur la couleur de la peau détaillées dans le paragraphe 3.4 sont par exemple très sensibles à de trop grands changements de luminosité [77].
3. **Occlusions** : Une des difficultés majeures de la segmentation réside dans la gestion des

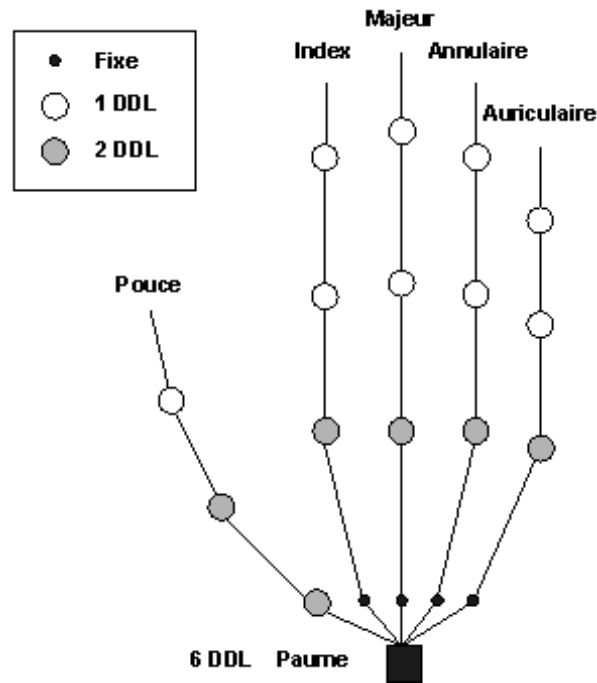


FIGURE 3.1 – Modèle cinétique de la main

occlusions partielles ou totales, les occlusions totales n'étant détectables que dans les méthodes utilisant une notion de suivi. Comme la main est un objet articulé comprenant de nombreux degrés de liberté, le résultat des projections dans une image 2D présente de nombreuses auto-occlusions provenant de doigts pliés (poing fermé par exemple) ou d'une main dont la paume est non-parallèle au plan de l'image. Ces circonstances rendent le problème des occlusions très complexes.

4. **Vitesse de traitement** : Dans une utilisation pratique, la notion de vitesse de traitement est bien sûr un critère majeur. La main pouvant être déplacée très rapidement en translation (5 m/s) et en rotation (300 degrés/s) [37]. Ainsi, le critère de vitesse s'impose automatiquement dans le suivi de la main principalement dans des applications d'interactions Homme-Machine ne supportant pas de latences.

L'ensemble des contraintes pouvant difficilement être satisfaites simultanément, des relâchements sont effectués suivant l'utilisation souhaitée en imposant des limites environnementales ou des restrictions à l'utilisateur. Par exemple afin de limiter la seconde contrainte, le fond peut être supposé fixe ou stable et la main être le seul objet ayant une couleur de peau. D'autres restrictions portent sur la minimisation des occlusions et des degrés de liberté en assurant que la paume est parallèle au plan du capteur.

3.3 Détection de la main dans les systèmes biométriques

Dans les systèmes biométriques se basant sur la main (forme et paume), la première étape consiste à détecter la main dans une image simple. Dans ces systèmes, la main est acquise à l'aide de trois types de capteurs différents : des scanners CCD spécialisés pour la reconnaissance de la

main (applications commerciales), des appareils photos numériques et des scanners d'images. Les scanners spécialisés sont des scanners d'images sur lesquels sont ajoutés des guides pour obtenir une position standard de la main. Dans toutes les méthodes proposées dans la littérature [135], le fond est uni ce qui permet une détection de la main facilitée. Deux exemples d'acquisitions collectées par un scanner CCD spécialisé et un scanner digital sont présentés à la figure Fig.3.2.

A partir des acquisitions, l'image d'entrée est transformée en une image en niveau de gris par une simple moyenne des différentes composantes couleurs dans le domaine RVB (Cf. 3.4.1.1). Ensuite, une simple binarisation à partir d'un seuil fixe S est généralement appliquée sur les images. Ce seuil doit être fixé suivant chaque type de capteurs. Pour une image $I(x, y)$, l'image binarisée $I_{bin(S)}(x, y)$ est ainsi donnée par :

$$I_{bin}(x, y) = \begin{cases} 0 & \text{si } I(x, y) < S \\ 1 & \text{sinon} \end{cases} \quad (3.2)$$

D'autres systèmes utilisent une méthode se basant sur la binarisation par seuillage en utilisant une méthode de définition automatique du seuil. La méthode d'Otsu [84] est une méthode classique de binarisation déterminant le seuil optimal en s'appuyant sur une recherche de la maximisation de la variance interclasse de l'histogramme. Deux exemples de segmentation utilisant cette méthode sont présentés à la Fig 3.2. D'autres méthodes de binarisation peuvent également être utilisées. Un état de l'art décomposant ces différentes méthodes en six catégories (méthodes basées sur l'histogramme, le regroupement de niveaux de gris, l'entropie, les attributs d'objets, la cohérence spatiale et l'adaptation locale) est proposé par Sezgin et Sankur [96].

3.4 Détection de la main par modélisation de la couleur de la peau

Dans les dernières décennies, la détection de la couleur de la peau est souvent utilisée dans le traitement d'images pour la détection de visages et de mains [50, 131, 105]. Elle est ainsi employée dans de nombreuses applications telles que la *motion capture*, les interactions Homme-Machine, la biométrie du visage, la surveillance ou l'indexation d'images.

Pour chacune de ces applications, de très nombreuses méthodes ont été proposées dans la littérature. Cependant dans tous les systèmes, la problématique de segmentation par la peau peut être décomposée en trois sous-problèmes :

- le choix de l'espace couleur,
- la méthode de modélisation de la distribution de la couleur de peau,
- la façon de regrouper des pixels de peau (avec ou sans utiliser les relation spatiales).

Ces trois sous-problèmes sont développés dans les prochains paragraphes de cette section.

3.4.1 Représentation couleur

De très nombreuses représentations de la couleur sont couramment utilisées. La plus connue, l'espace RVB décrit ci-dessous, est inspirée du système visuel humain. Notre système visuel utilise les cônes et les bâtonnets (récepteurs situés sur la rétine) afin de transformer le signal de la lumière en signal électrique. Dans l'œil humain, il y a trois types de cônes différents réagissant principalement aux couleurs bleu, vert et rouge selon la longueur d'onde du signal (Fig. 3.3).

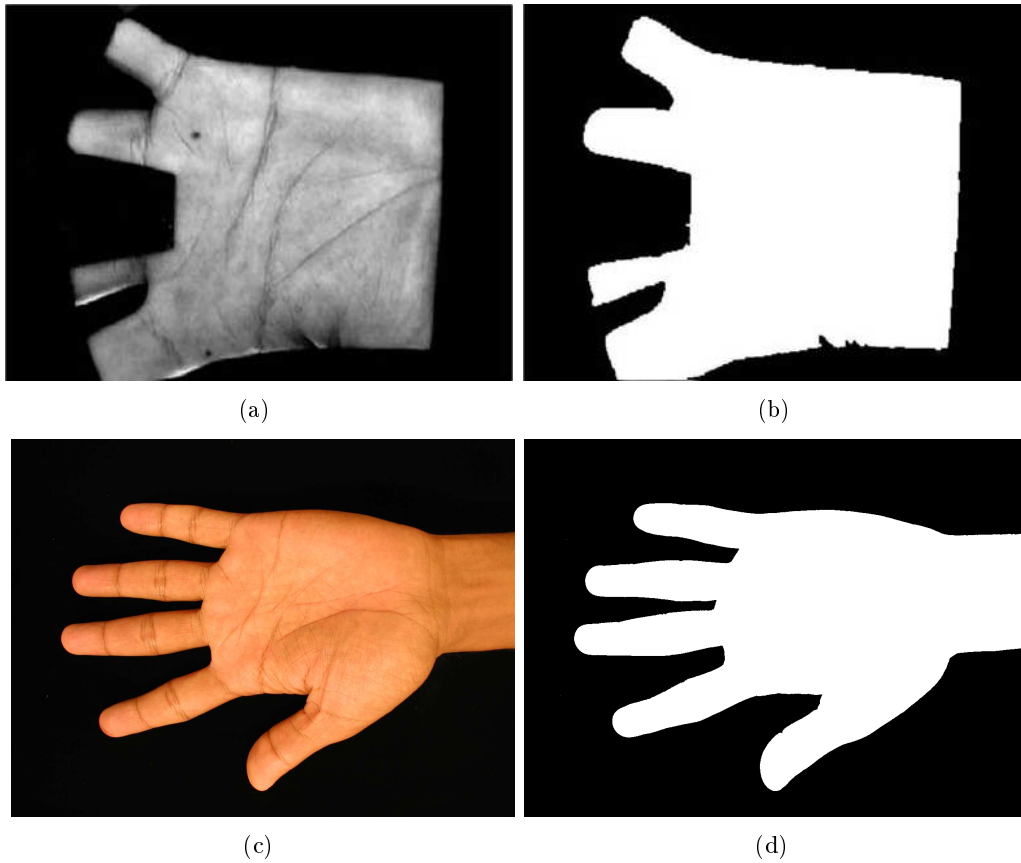


FIGURE 3.2 – Deux images de mains collectées par (a) un capteur spécifique et (c) un scanner digital puis segmentées par la méthode d'Otsu (b) et (d).

Comme il y a trois types de photorécepteurs, trois composants numériques sont suffisants pour décrire la couleur pour l'œil humain. Dans ce paragraphe, les domaines couleurs les plus utilisés pour représenter le ton de peau sont présentés.

3.4.1.1 Espace RVB - Rouge, Vert, Bleu

Le domaine RVB est l'espace de couleur original dans les applications d'affichage lorsque la couleur est décrite par trois ensembles de couleurs (Rouge, Vert, Bleu). Cet espace est le domaine le plus utilisé pour le stockage d'images numériques. Schématiquement, il est très souvent représenté sous la forme d'un espace orthonormé cubique (Fig. 3.4).

Le point d'origine O est le point noir, le sommet opposé est le point blanc W . Le vecteur OW qui relie le point noir au point blanc est l'axe achromatique, c'est-à-dire l'axe des niveaux de gris. Les sommets RVB sont l'emplacement des trois primaires et les trois sommets restants sont occupés par les couleurs secondaires Cyan, Jaune et Magenta. Par convention en colorimétrie chaque primaire peut varier de la valeur 0 à la valeur 1. Cependant les différentes composantes du domaine RVB sont très corrélées, ainsi ce domaine est perceptuellement non-uniforme : une variation d'une composante engendre une variation de couleur non-relative à la variation.

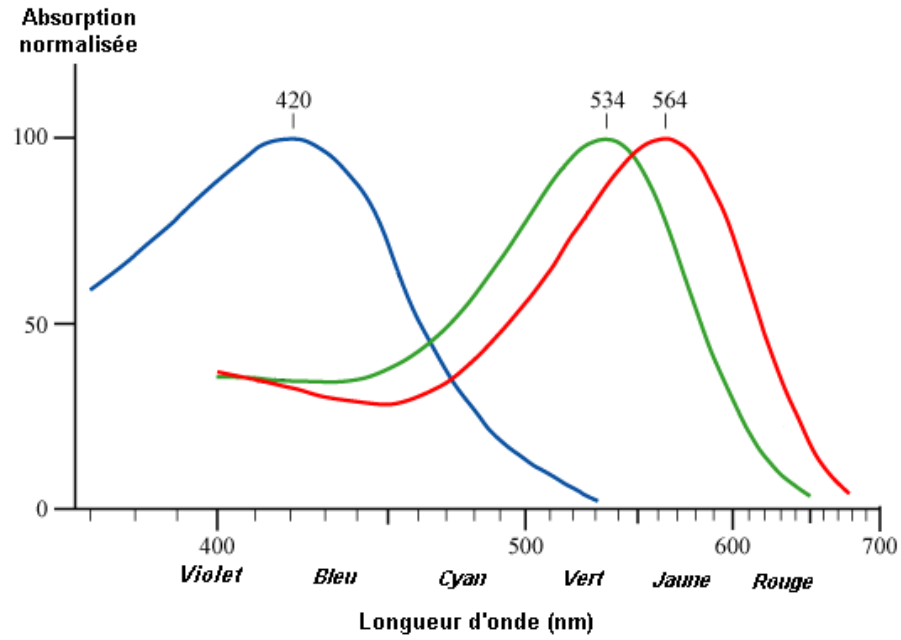


FIGURE 3.3 – Courbe d'absorption des cônes

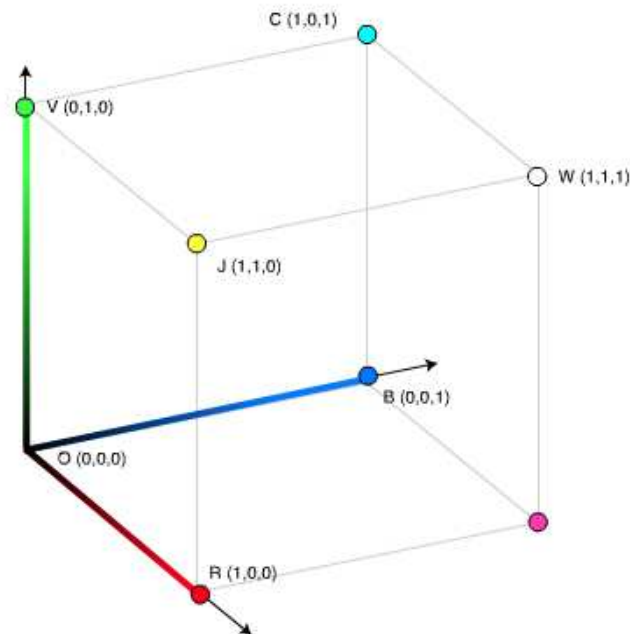


FIGURE 3.4 – Représentation cubique du domaine RVB

3.4.1.2 Espace YUV et YC_rC_b

Ces espaces ont été introduits par le standard allemand PAL et sont également utilisés dans le standard NTSC et SECAM. La couleur est représentée par la luminance Y et les composantes couleurs UV ou C_rC_b sont déterminées par combinaison linéaire des composantes RVB . Ainsi pour les anciennes télévisions en noir et blanc seule l'information de luminance était utilisée, les informations de couleurs ont été utilisées lors du passage en mode couleur. Un autre avantage des espaces YUV et YC_rC_b est qu'une partie de l'information couleur peut être ignorée afin de réduire la largeur de bande utilisée. En effet l'œil humain a peu de sensibilité à la couleur, l'information de luminosité provenant du canal de luminance a bien plus d'impact sur l'image perçue que celle des composantes couleurs. Ainsi, la quantité de données utilisées pour coder l'information de chrominance est moins importante. L'œil effectue ensuite une extrapolation sur une grande partie de la couleur. Par conséquent, les signaux résultant des composantes U et V peuvent être sensiblement comprimés comme dans le standard $NTSC$.

Il est également intéressant de noter que le domaine YC_rC_b fait partie du standard de compression JPEG2000 [109]. Finalement, les espaces YUV et YC_rC_b sont déterminés à partir du domaine RVB par :

$$\begin{cases} Y &= 0.299R + 0.587V + 0.114B \\ U &= -0.147R - 0.289V + 0.436B = 0.493 \times (B - Y) \\ V &= 0.615R - 0.515V - 0.100B = 0.877 \times (R - Y) \end{cases} \quad (3.3)$$

$$\begin{cases} Y &= 0.299R + 0.587V + 0.114B \\ C_r &= -0.169R - 0.331V + 0.500B \\ C_b &= 0.500R - 0.419V - 0.081B \end{cases} \quad (3.4)$$

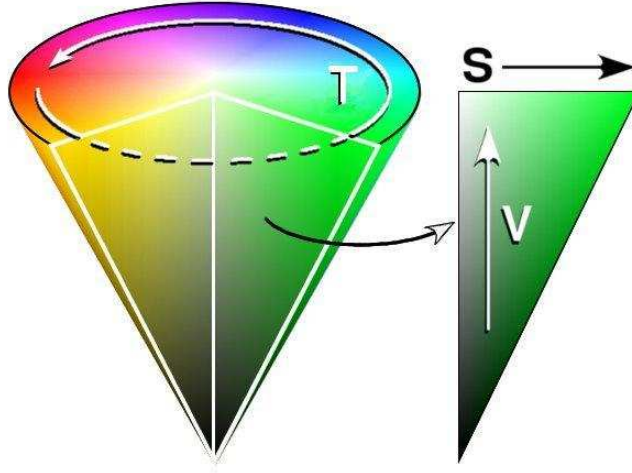
3.4.1.3 Espaces HSV et HSL - Teinte, Saturation, Valeur ou Luminance

Ces espaces couleur décrivent les couleurs avec des données intuitives basées sur la représentation de peintres concernant la teinte, la saturation et le ton. La teinte définit la couleur dominante et la saturation indique l'intensité de la couleur et le ton sa brillance. Les espaces HSV et HSL peuvent être représentés par un cône (Fig. 3.5). Ils sont définis par une transformation non linéaire du domaine RVB .

Soit $V^* \in [0, 1]$ la valeur, $L \in [0, 1]$ la luminosité, $S \in [0, 1]$ la saturation, $H \in [0, 360]$, $(R, V, B) \in [0, 1]^3$, $Max = MAX(R, V, B)$ et $Min = MIN(R, V, B)$. La teinte H est définie de façon identique dans les deux espaces par :

$$H = \begin{cases} \text{indéfini} & \text{si } Max = Min \\ 60 \times (V - B) / (Max - Min) & \text{si } Max = R \text{ et } V \geq B \\ 60 \times (V - B) / (Max - Min) + 360 & \text{si } Max = R \text{ et } V < B \\ 60 \times (B - R) / (Max - Min) + 120 & \text{si } Max = V \\ 60 \times (R - V) / (Max - Min) + 240 & \text{si } Max = B \end{cases} \quad (3.5)$$

La notion de saturation est définie de façon différente selon l'espace couleur. Ainsi, la saturation et la valeur du domaine HSV sont définies par (3.6) et la saturation et la luminance du domaine HSL sont définies par (3.7) :


 FIGURE 3.5 – Représentation cône du domaine TSV

$$\begin{cases} S &= \begin{cases} 0 & \text{si } Max = 0 \\ 1 - Max/Min & \text{sinon} \end{cases} \\ V^* &= Max \end{cases} \quad (3.6)$$

$$\begin{cases} S &= \begin{cases} 0 & \text{si } Max + Min = 0 \text{ ou } Max = Min \\ (Max - Min)/(Max + Min) & \text{si } 0 < Max + Min \leq 1 \\ (Max - Min)/(2 - Max + Min) & \text{sinon} \end{cases} \\ L &= (Max + Min)/2 \end{cases} \quad (3.7)$$

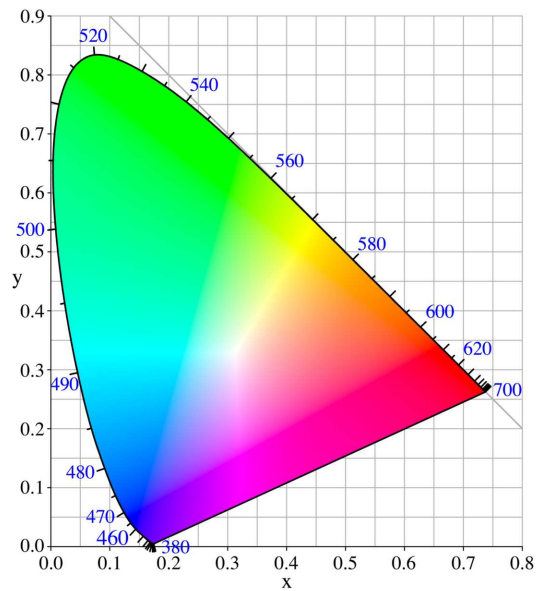
D'autres domaines sont de la même famille tels le domaine HSI ainsi que les espaces TSV et TSL introduits par Terrillon et al.[111, 110]. Ces deux derniers domaines se basent cependant sur le RVB normalisé afin d'être robuste aux changements d'intensité de la lumière comme indiqué ci-dessous.

3.4.1.4 Espaces $CIE XYZ$ et RVB normalisé

L'espace $CIE XYZ$ est défini par l'étude la perception de la couleur déterminée par la Commission Internationale de l'Éclairage (CIE) en 1931. Le paramètre Y est désigné comme la mesure de similarité tandis que la couleur est spécifiée par les paramètres x et y définis à partir des données X, Y, Z par :

$$x = \frac{X}{X + Y + Z}, y = \frac{Y}{X + Y + Z} \quad (3.8)$$

Où les valeurs X, Y, Z sont explicitées dans l'équation 3.9. Un paramètre induit z peut être défini par $x + y + z = 1$. Comme vu précédemment, l'œil a trois types de cônes sensibles à différentes longueurs d'ondes dont les couleurs résultantes visibles sont représentables par une figure en 3 dimensions (Cf. Fig 3.4 et Fig. 3.5). Dans l'espace $CIE XYZ$, l'ensemble des couleurs est uniquement représenté par les deux composantes x et y . Une représentation classique de


 FIGURE 3.6 – Diagramme de chromacité xy schématisé

ce domaine correspond à un diagramme de chromacité xy schématisé à la Fig. 3.6. La courbe extérieure est appelé le "spectral locus" avec la longueur d'onde notée en nanomètres. Cependant, la couleur réelle n'est pas complètement définie par ce diagramme puisque la notion de luminosité doit y être ajoutée.

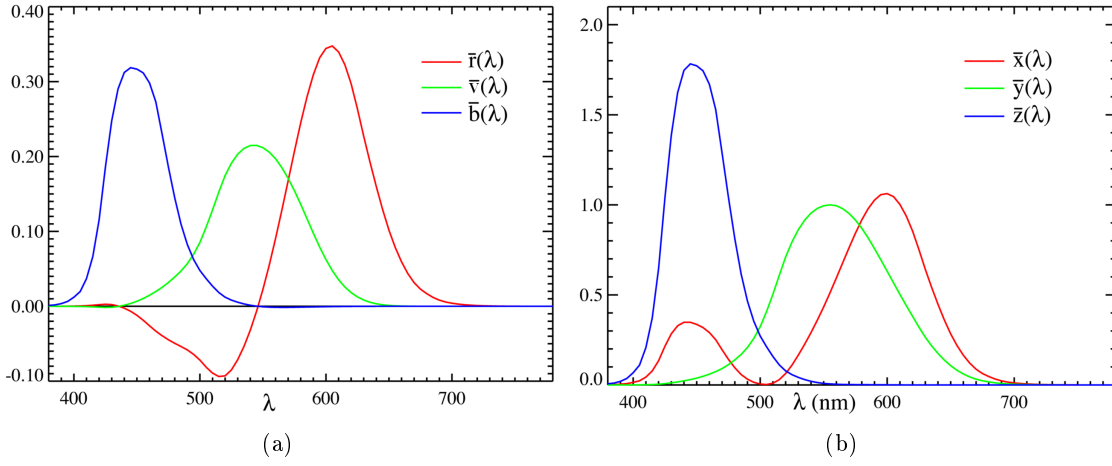
Le domaine $CIE\ XYZ$ de représentation de couleur a été défini par le CIE afin de pallier la limitation de la représentation dans le domaine RVB . En effet, les trois fonctions de couleurs du domaine RVB , $\bar{r}(\lambda)$, $\bar{v}(\lambda)$ et $\bar{b}(\lambda)$, déterminées à partir des trois couleurs primaires RVB présentent des valeurs négatives (Cf Fig. 3.7(a)) ce qui rend impossible la représentation d'un certain nombre de couleurs par superposition des trois spectres. *A contrario* les fonctions de couleurs du domaine XYZ définies par le CIE ne présentent que des valeurs positives permettant une représentation complète de la couleur perceptible par l'oeil humain (Cf. Fig. 3.7(b)). Finalement, le domaine $CIE\ XYZ$ est défini par une transformation linéaire du domaine RVB par :

$$\begin{cases} X = \frac{1}{0.17697}(0.49R + 0.31V + 0.20B) \\ Y = \frac{1}{0.17697}(0.17697R + 0.81240V + 0.01063B) \\ Z = \frac{1}{0.17697}(0.00R + 0.01V + 0.99B) \end{cases} \quad (3.9)$$

Le domaine RVB normalisé utilise également une représentation de la couleur dans un espace à deux dimensions en utilisant une équation similaire à (3.8). En effet, la représentation est obtenue à partir d'une normalisation des composantes RVB :

$$r = \frac{R}{R + V + B}, v = \frac{V}{R + V + B}, b = \frac{B}{R + V + B} \quad (3.10)$$

Il est à noter que la somme des trois composantes est égale à 1. Ainsi, seules deux composantes sont nécessaires pour décrire une couleur dans cet espace.


 FIGURE 3.7 – Les fonctions colorimétriques du domaine (a) RVB et (b) XYZ

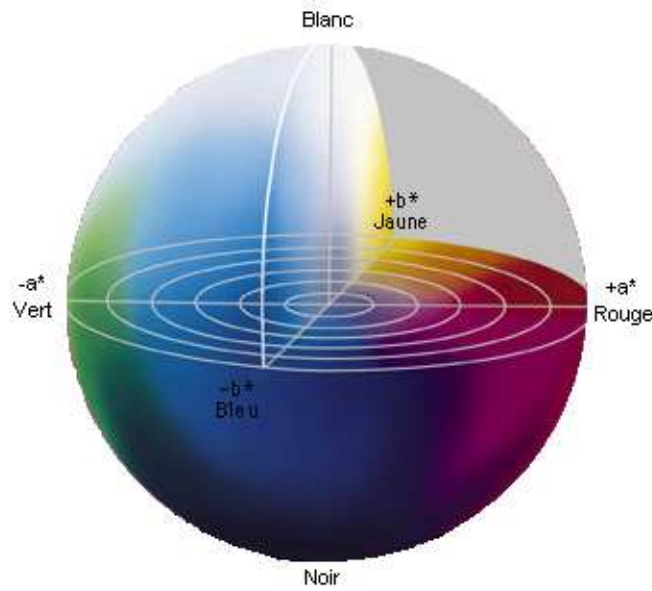
3.4.1.5 Espaces $CIE L^*a^*b^*$ et $CIE L^*u^*v^*$

Les espaces $CIE L^*a^*b^*$ et $CIE L^*u^*v^*$ définis en 1976 sont dérivés de l'espace $CIE XYZ$. L'objectif de ces espaces est de définir un domaine le plus perceptuel possible afin qu'un changement de couleur produise une modification visuelle en accord. Le passage des coordonnées du domaine $CIE XYZ$ vers les coordonnées du domaine $CIE L^*a^*b^*$ s'effectue par les transformations suivantes avec comme blanc de référence le point défini dans le domaine $CIE XYZ$ par les trois composantes chromatiques (X_0, Y_0, Z_0) :

$$\begin{cases} L^* &= 116 \times f\left(\frac{Y}{Y_0}\right) - 16 \\ a^* &= 500 \times \left(f\left(\frac{X}{X_0}\right) - f\left(\frac{Y}{Y_0}\right)\right) \\ b^* &= 500 \times \left(f\left(\frac{Y}{Y_0}\right) - f\left(\frac{Z}{Z_0}\right)\right) \end{cases} \quad \text{où } f(x) = \begin{cases} x^{\frac{1}{3}} & \text{si } x > 0.008856 \\ 7.787x + \frac{16}{116} & \text{sinon} \end{cases} \quad (3.11)$$

D'un point de vue colorimétrique le paramètre a^* représente l'opposition entre les couleurs vert et rouge tandis que b^* définit l'opposition des couleurs jaune et bleu. Une représentation sphérique de ce domaine est illustrée à la Fig. 3.8. De même, le passage des coordonnées $CIE XYZ$ vers les coordonnées du domaine $CIE L^*u^*v^*$ s'effectue en utilisant le blanc de référence (X_0, Y_0, Z_0) par :

$$\begin{cases} L^* &= 116 \times \left(\frac{Y}{Y_0}\right)^{1/3} - 16 \\ u^* &= 13 \times L^*(u' - u'_0) \\ v^* &= 13 \times L^*(v' - v'_0) \end{cases} \quad \text{où } \begin{cases} u' &= \frac{4X}{X+15Y+3Z} \\ v' &= \frac{9Y}{X+15Y+3Z} \\ u'_0 &= \frac{4X_0}{X_0+15Y_0+3Z_0} \\ v'_0 &= \frac{9Y_0}{X_0+15Y_0+3Z_0} \end{cases} \quad (3.12)$$

FIGURE 3.8 – Représentation sphérique du $CIE L^*a^*b^*$

3.4.2 Méthode de modélisation de la couleur de la peau

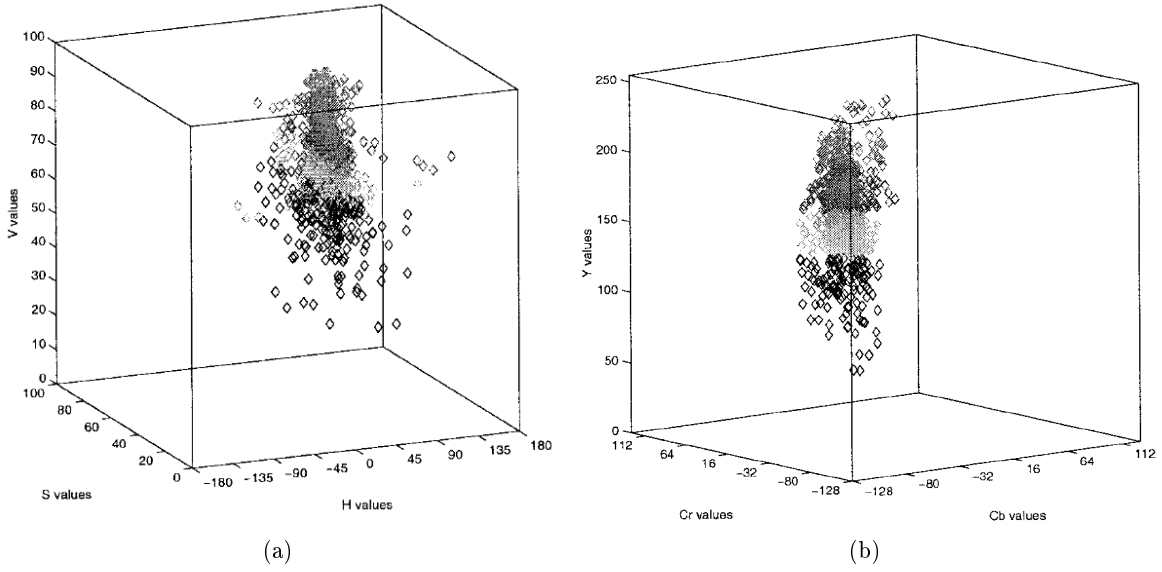
Des états de l'art de la détection de la couleur de la peau ont été récemment rédigés. Vezhnevets et al. [116] proposent de classer les méthodes de modélisation de la peau suivant quatre catégories :

1. définition explicite de la couleur de peau,
2. distribution non-paramétrique de la couleur de la peau ,
3. distribution paramétrique de la couleur de la peau,
4. modélisation dynamique de distribution de la couleur de la peau.

Similairement, Martinkauppi et al. [77] propose la même classification en séparant la dernière catégorie en deux sous-ensembles (paramétrique et non-paramétrique). Dans la thèse soutenue par Störking sur le traitement d'images et la couleur de peau [105], une subdivision initiale différencie les méthodes selon la source d'entrée. Ainsi les quatre catégories sont originellement divisées selon que les paramètres de la caméra sont connus ou non. Nous proposons dans ce paragraphe d'effectuer une description des méthodes de modélisation de la couleur de la peau décrites dans la littérature en les classant selon les catégories proposées par Vezhnevets et al. [116] qui sont plus générales et évitent la redondance dans les présentations.

3.4.2.1 Définition explicite de la couleur de la peau

La méthode la plus simple pour définir la couleur de la peau est d'effectuer un seuillage sur une ou plusieurs composantes couleurs avec un ou plusieurs seuils. Suivant le domaine couleur utilisé, différentes heuristiques ont été établies. La plupart des méthodes utilisent une segmentation dans le plan chromatique du domaine couleur. Sobottka et Pitas [101, 102] proposent une méthode de détection du visage dans le plan $H - S$ de l'espace HSV en définissant des seuils minimums et maximums pour les valeurs H et S . Deux autres méthodes proposées par Collobert et al. [17] et Hidai et al. [49] utilisent une méthode similaire en utilisant les composantes UV dans le domaine


 FIGURE 3.9 – Répartition des pixels de peau dans le domaine (a) HSV et (b) $YC_B C_R$

YUV et les composantes $C_B C_R$ du domaine $YC_B C_R$ respectivement. Garcia et Tziritas [43] ont proposé deux méthodes s'appuyant sur des heuristiques dans le domaine HSV et $YC_B C_R$ en étudiant la répartition des pixels de peau illustrée à la fig. 3.9 dans chacun des espaces. Dans ces deux espaces, un ensemble de plans est défini empiriquement afin de représenter la couleur de peau. Dans le domaine HSV , les plans sont définis par :

$$\begin{aligned} S \geq 10, V \geq 40, S \leq -H - 0.1V + 110, H \leq -0.4V + 75, \\ \text{si } (H \geq 0) S \leq 0.08H \times (100 - V) + 0.5V \text{ sinon } S \leq 0.5H + 35. \end{aligned} \quad (3.13)$$

De même, dans le domaine $YC_B C_R$ où la représentation de la peau est plus compacte les plans sont définis par :

$$\begin{aligned} C_R \geq -2(C_b + 24), C_R \geq -(C_B + 17), C_R \geq -4(C_B + 32), C_R \geq 2.5(C_B + \theta_1), \\ C_R \geq \theta_3, C_R \geq 0.5(\theta_4 - C_B), C_R \leq (220 - C_B)/6, C_R \leq 4/3(\theta_2 - C_B). \end{aligned} \quad (3.14)$$

Où les paramètres $\theta_{1,...,4}$ sont donnés par :

$$\text{si } (Y > 128) \begin{cases} \theta_1 = -2 + (256 - Y)/16 \\ \theta_2 = 20 + (256 - Y) \\ \theta_3 = 6 \\ \theta_4 = -8 \end{cases} \quad \text{sinon} \begin{cases} \theta_1 = 6 \\ \theta_2 = 12 \\ \theta_3 = 2 + Y/32 \\ \theta_4 = -16 + Y/16 \end{cases} \quad (3.15)$$

Les plans sont définis plus facilement dans le domaine HSV grâce à l'accès au paramètre H qui est le critère majeur de la définition de la peau. L'intersection de ces plans décrits par les équations (3.13) et (3.14) avec les plans HSV et $YC_B C_R$ pour $V = 70$ et $Y = 160$ respectivement est illustrée à la Fig. 3.10.

D'autres travaux utilisent directement le domaine RVB . Dans ce système de représentation, Peer et al. [86] propose une méthode utilisant de multiples règles :

$$\begin{aligned} R > 95, V > 40, B > 20, |R - V| > 15, \\ R > V, R > B \text{ et } \max(R, V, B) - \min(R, V, B) > 15. \end{aligned} \quad (3.16)$$

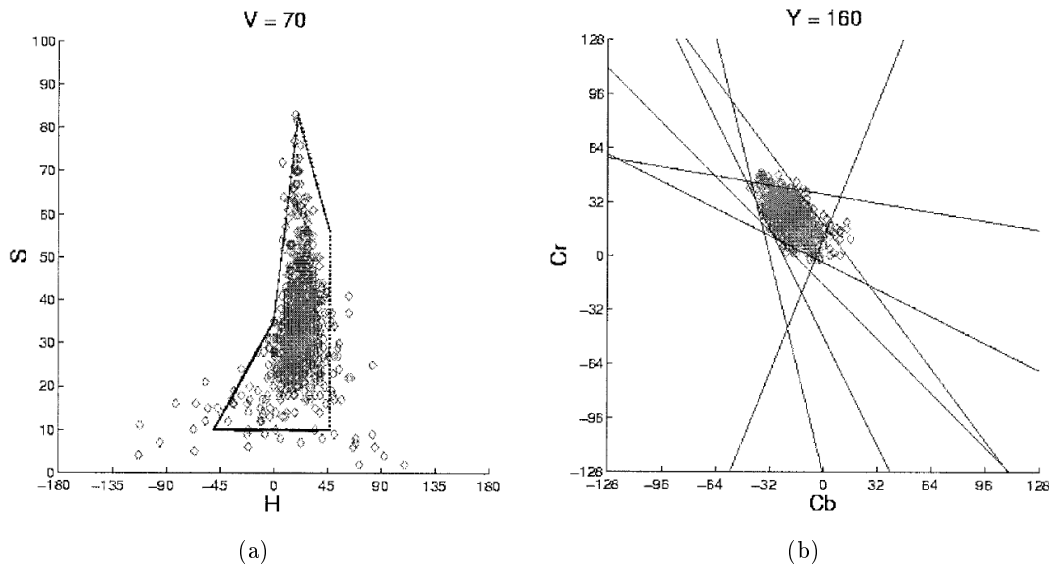


FIGURE 3.10 – Limitations de la surface représentant la couleur de peau dans le domaine (a) HSV et (b) $YC_B C_R$

Plusieurs propositions ont également été faites pour établir des seuils permettant une définition explicite de la couleur de la peau dans le domaine RVB normalisé. Par exemple, Chiang et al. [16] présentent une méthode permettant de détecter la couleur de la peau, des lèvres et des yeux. Dans ce système deux règles simples sont définies dans le domaine RVB normalisé pour décrire le ton de peau : $r > v > b$ et $r - v > 45$. Des méthodes similaires sont appliquées pour détecter les lèvres et les yeux.

Ces méthodes ont fait l'objet de nombreuses recherches puisqu'elles ont l'avantage d'établir des classifieurs très rapides. En contrepartie, la plus grande difficulté réside dans la recherche simultanée de l'espace couleur le plus représentatif et des règles de décision optimales qui sont difficiles à obtenir empiriquement pour des images avec une luminosité et un fond non fixé. Récemment, Gomez et al. ont proposé une méthode afin d'obtenir l'espace couleur optimal [45] et une simple règle de décision pour permettre un fort taux de détection [44]. Tout d'abord, de nouveaux espaces couleur sont déterminés à partir de l'espace RVB normalisé par un algorithme de construction d'inductions proche de l'algorithme C4.5. Ensuite, la meilleure règle de décision similaire à (3.16) est établie pour permettre le meilleur taux de classification de pixels de peau.

3.4.2.2 Modélisation non-paramétrique de la couleur de la peau

De nombreuses méthodes non-paramétriques ont été proposées pour modéliser la couleur de la peau. L'objectif de tous ces systèmes est d'estimer la distribution des pixels de peau à partir d'un ensemble de données sans déterminer explicitement un modèle caractérisant la peau. Le résultat de la détection de peau par ces méthodes est généralement nommé carte de probabilité de peau.

– **Table de consultation (Lookup Table) :**

Une méthode couramment utilisée pour classer les pixels entre pixels de peau et pixels de non-peau est nommée table de consultation (LUT) [15, 103]. Cette méthode s'appuie sur la distribution couleur des pixels de peau. Après la constitution d'une base de pixels de couleur de peau, le nombre de fois que chaque couleur apparaît dans la base est stocké. Ensuite, toutes ces valeurs sont normalisées. Plus formellement, la probabilité qu'un pixel de couleur c soit un pixel de peau est formulé par :

$$P_{\text{peau}}(c) = \text{peau}[c] / \text{Norm} \quad (3.17)$$

où $\text{peau}[c]$ est le nombre de fois que la couleur c est présente dans la base d'apprentissage de couleur de peau et Norm est un coefficient de normalisation (la somme de tous les coefficients [55] ou la valeur maximum [134]). Finalement, grâce à un seuil fixe Ω , un pixel est considéré comme un pixel de peau lorsque $P_{\text{peau}}(c) > \Omega$. Une autre méthode de décision définie par [91] utilise la carte de probabilité obtenue en envoyant celle-ci à un réseau de neurones qui détecte les visages et détermine leur orientation.

– Méthode de Bayes :

La probabilité $P_{\text{peau}}(c)$ déterminée par l'équation (3.17) correspond à la probabilité d'observer la couleur c sachant que cette couleur est une couleur de peau. Elle peut ainsi être formulée par la probabilité conditionnelle $P(c \mid \text{peau})$. La mesure recherchée est en fait plutôt la probabilité d'observer une couleur de peau sachant que le pixel est de couleur c soit $P(\text{peau} \mid c)$. Afin de déterminer cette probabilité, la règle de Bayes est utilisée :

$$P(\text{peau} \mid c) = \frac{P(c \mid \text{peau})P(\text{peau})}{P(c \mid \text{peau})P(\text{peau}) + P(c \mid \neg\text{peau})P(\neg\text{peau})} \quad (3.18)$$

où $P(c \mid \text{peau})$ et $P(c \mid \neg\text{peau})$ sont déterminés à partir de l'équation (3.17) avec une base d'apprentissage de pixels de peau et de non-peau respectivement. Les valeurs de probabilité $P(\text{peau})$ et $P(\neg\text{peau})$ peuvent être déterminées à partir du rapport entre le nombre de pixels de peau et de non-peau dans la base d'apprentissage [55, 11, 12]. Même si l'on peut directement utiliser la probabilité déterminée par (3.18), il est plus efficace d'utiliser la comparaison entre $P(\text{peau} \mid c)$ et $P(\neg\text{peau} \mid c)$. Ce rapport peut être écrit en s'appuyant sur l'équation (3.18) par :

$$\frac{P(\text{peau} \mid c)}{P(\neg\text{peau} \mid c)} = \frac{P(c \mid \text{peau})}{P(c \mid \neg\text{peau})} * \delta \text{ où } \delta = \frac{P(\text{peau})}{P(\neg\text{peau})} \quad (3.19)$$

Finalement un seuil de décision doit être défini afin de déterminer la frontière entre les pixels de peau et de non peau. Pour un seuil θ , cette règle de décision peut ainsi être définie par :

$$\begin{cases} c \text{ est un pixel de peau} & \text{si } (P(\text{peau} \mid c)/P(\neg\text{peau} \mid c) > \theta) \\ c \text{ est un pixel de fond} & \text{sinon} \end{cases} \quad (3.20)$$

Afin de ne pas définir explicitement les probabilités conditionnelles $P(\text{peau} \mid c)$ et $P(\neg\text{peau} \mid c)$, il est possible de s'appuyer directement sur $P(c \mid \text{peau})$ et $P(c \mid \neg\text{peau})$ grâce à l'équation (3.19). Ainsi l'équation (3.20) peut-être reformulée pour un seuil Θ par :

$$\begin{cases} c \text{ est un pixel de peau} & \text{si } (P(c \mid \text{peau})/P(c \mid \neg\text{peau}) > \Theta) \\ c \text{ est un pixel de fond} & \text{sinon} \end{cases} \quad (3.21)$$

où $\Theta = \theta/\delta$. Il est à noter que les probabilités $P(\text{peau})$ et $P(\neg\text{peau})$ n'ont en fait aucun impact sur la qualité de classification mais influent uniquement sur le choix du seuil Θ .

– **Intersection d'histogrammes :**

Comaniciu et Ramesh [18] utilisent les composantes chromatiques rv (du domaine RVB normalisé) afin de détecter les visages dans une séquence d'images. Une table de consultation est calculée à partir d'une fenêtre de l'image contenant le visage sous différentes conditions d'illumination pour initialiser le système. Ensuite, le visage est suivi en estimant la similarité entre la densité de l'histogramme appris et les densités d'histogrammes de régions de test dans l'image. Ces régions de test sont des fenêtres de l'image. Les régions ayant la plus haute similarité sont indiquées comme étant des régions de couleur de peau. Birchfield [2] présente une méthode proche en s'appuyant sur une intersection d'histogrammes dans un domaine RVB balancé ($B - V, V - R, R + V + B$).

– **Réseaux de neurones :**

Afin de modéliser la couleur de la peau, l'apport des algorithmes d'apprentissage a également été étudié. Les réseaux de neurones [1, 61] ont été très abondamment utilisés ces dernières années dans le traitement d'images [36, 9].

A l'aide de ce type de modèle, plusieurs travaux ont été menés pour décrire un modèle de peau. Chen et Chiang [14] ont élaboré différents réseaux classés suivant leurs structures (nombre de couches cachées et nombre de neurones par couches) et le domaine couleur d'entrée ($YUV, HSV, \dots$ sans utiliser la notion de luminance). Bousaïd et al. [4] et Seow et al. [95] ont proposé de modéliser la couleur de la peau par réseau de neurones dans le domaine couleur RVB .

3.4.2.3 Modélisation paramétrique

Plutôt que de définir la couleur de la peau par des règles simples ou par un ensemble de caractéristiques définissant la notion de la peau individuellement pour chaque couleur, les modèles paramétriques essayent de définir explicitement la couleur de peau par des ensembles.

– **Moyenne :**

La méthode la plus naturelle et la plus simple pour modéliser un ensemble de données est de caractériser cet ensemble par sa valeur moyenne. Ainsi un article s'appuyant sur cette idée a été proposé par Fieguth et Terzopoulos [39]. Lors de la phase d'initialisation, la valeur moyenne de chacune des composantes RVB est calculée. Ensuite, chacune de ces valeurs moyennes est comparée avec la valeur moyenne correspondante de la zone à tester. Si les valeurs moyennes des composantes initiales et courantes sont (relativement) similaires alors la zone à tester est déclarée comme étant une zone de pixels de peau.

– **Approximation par Gaussienne(s) :**

Les méthodes d'approximation par une ou plusieurs Gaussiennes utilisent, comme la technique évoquée précédemment, uniquement la distribution en couleur des pixels de peau.

Cette distribution est ensuite modélisée sous la forme d'un modèle Gaussien ce qui permet une représentation très compacte de la peau. Avec une approximation par une seule Gaussienne[131], la distribution de la couleur de peau c peut-être définie comme :

$$P(c | peau) = \frac{1}{\sqrt{2\pi} |\Sigma_p|} \exp \left\{ -\frac{1}{2} (c - \mu_p)^T \Sigma_p^{-1} (c - \mu_p) \right\}. \quad (3.22)$$

où les paramètres du modèle μ_p et Σ_p sont la moyenne de la distribution des couleurs des pixels de peau et la matrice de covariance respectivement. Ils peuvent être calculés à l'aide de la base d'apprentissage par :

$$\mu_p = \frac{1}{T_p} \sum_{j=1}^{T_p} c_j, \quad \Sigma_p = \frac{1}{T_p - 1} \sum_{j=1}^{T_p} (c_j - \mu_p)(c_j - \mu_p)^T \quad (3.23)$$

où T_p est le nombre total de pixels de peau dans la base d'apprentissage. La méthode de décision utilise comme précédemment un seuil et permet de déterminer si un pixel peut être considéré ou non comme un pixel de peau. Dans les modèles avec plusieurs Gaussiennes utilisés par [110, 130, 55] l'ensemble des pixels de peau de la base d'apprentissage est tout d'abord décomposé en différents ensembles par des méthodes de clustering. Ensuite chaque groupe de pixels est modélisé par une gaussienne ce qui permet de définir avantageusement la couleur de la peau.

– Modèle elliptique :

Afin de généraliser la méthode de modélisation par Gaussienne, Lee et Yoo [70] ont défini une alternative appelée modèle elliptique. En effet dans les systèmes utilisant une ou plusieurs Gaussiennes, les clusters sont représentés de façon symétrique par rapport à la moyenne. Or, dans la réalité Lee et Yoo ont constaté que les distributions de couleurs des pixels de fond et de peau ne sont pas symétriques. Le modèle de distribution elliptique qu'ils ont proposé est défini par :

$$\Phi = (c - \phi)^T \Lambda^{-1} (c - \phi) \quad (3.24)$$

Ce modèle est estimé en deux étapes. Premièrement, 5% des couleurs ayant la plus faible représentativité sont supprimés afin de limiter l'impact du bruit et les exemples négligeables. Ensuite les paramètres utilisés dans l'équation 3.24 sont définis :

$$\phi = \frac{1}{n} \sum_{i=1}^n c_i \quad (3.25)$$

$$\Lambda = \frac{1}{N} \sum_{i=1}^n peau[c_i] \cdot (c_i - \mu)(c_i - \mu)^T \quad (3.26)$$

Où

$$\mu = \frac{1}{N} \sum_{i=1}^n peau[c_i] \cdot c_i \quad \text{et} \quad N = \sum_{i=1}^n peau[c_i] \quad (3.27)$$

Et n est le nombre total des couleurs d'entraînement c_i de l'ensemble des pixels de peau d'entraînement et $peau[c_i]$ est le nombre d'exemples de peau de couleur c_i .

L'avantage de ce modèle est qu'il permet de modéliser la couleur de la peau de manière plus fine. En effet, il correspond visuellement plus précisément à la distribution de la couleur de la peau dans l'espace couleur choisi.

3.5 Détection de la main intégrant des connaissances à priori

En comparaison avec le nombre de méthodes s'appuyant sur la couleur de la peau, peu de méthodes ont été définies pour détecter une main en s'appuyant sur la notion de forme. La majorité des méthodes proposées sont, qui plus est, généralement une adaptation des systèmes de détection de visages. Les systèmes sont séparés en deux groupes généraux : les méthodes de modèles de forme et les approches par classification.

3.5.1 Modèle de forme

Les méthodes basées sur les modèles de forme ont été utilisées avec succès dans de nombreuses applications pour détecter des objets déformables. Ces méthodes se sont largement améliorées grâce à l'utilisation de distances pour les comparaisons entre un modèle et l'image telles que les distances de Hausdorff ou de chanfrein.

La comparaison par la distance de chanfrein est une technique pour trouver la meilleure correspondance entre deux ensembles de points de contours en minimisant la distance entre les ensembles. Plusieurs systèmes utilisent ce système afin de déterminer localement une comparaison optimale. Borgfors [3] propose une approche multi-résolution où une pyramide de contours à différentes résolutions est construite puis comparée à un modèle. Au niveau le plus bas, les paramètres de transformation sont optimisés en commençant par les points situés sur une grille prédéterminée. Ensuite à la résolution supérieure, la comparaison est effectuée avec ces paramètres optimaux et les localisations où le coût est trop important sont abandonnées. Breuel [6] propose une alternative à cette approche en subdivisant l'espace de transformation. La qualité de la comparaison est estimée par le nombre de points de similitude entre deux ensembles suivant un seuil d'erreur fixé. Cette méthode est appliquée à chaque sous-région de l'image et l'optimum global est déterminé en sub-divisant récursivement ces sous-régions.

D'autres approches se sont appuyées sur une structuration des modèles. Olson et Huttenlocher [82] ont proposé une méthode utilisant une structure en arbre. A chaque niveau de l'arbre, les deux modèles les plus proches sont regroupés. Chaque nœud de l'arbre possède les points que les modèles regroupés ont en commun. Lors du calcul de la distance entre les modèles et l'image, la distance entre la racine de l'arbre et les contours de l'image sont déterminés. Ce calcul est propagé aux nœuds fils si la distance déterminée n'est pas supérieure à un seuil fixe et ainsi de suite. Le gain principal de ce système est la rapidité d'exécution principalement lorsque les modèles ont un nombre important de points en commun. Un autre système présenté par Stenger [104] utilise une hiérarchie en arbre. Chaque nœud correspond à un modèle ayant une légère modification du modèle du niveau supérieur. Ainsi dans la recherche du modèle, seules les branches ayant un nœud avec une forte correspondance sont étudiées. Un exemple de la représentation hiérarchique de ce modèle est illustré à la figure 3.11.

3.5.2 Approches par classification de zones d'image

Les méthodes de classification utilisent l'image à différentes résolutions et sont utilisées avec succès pour la détection de visage. L'idée générale de ces méthodes est de glisser des fenêtres de taille fixe sur l'image à différentes résolutions et de classer chaque fenêtre comme visage ou non-visage. Ce principe de détection est illustré à la Fig. 3.12. La tâche est similaire à un problème de reconnaissance de modèle et toutes les méthodes proposées sont ainsi caractérisées suivant la méthode de classification utilisée. Pour chacun de ces systèmes, la gestion de différentes vues peut être considérée comme plusieurs classifieurs correspondant à chacune des vues à reconnaître. Les problèmes de luminosité et de contraste sont gérés par une étape de normalisation des images d'entrée ou lors de l'apprentissage en incluant un nombre important d'exemples dans différentes conditions d'éclairage.

L'idée de la classification hiérarchique a initialement été formulée avec l'algorithme de classification K-means par Sung et Poggio [107]. Dans cette approche, l'image d'entrée est normalisée afin d'être robuste au changement de luminosité. Ensuite, les images de la base d'apprentissage sont classifiées entre des groupes de visages et de non-visages. Lors de la phase de détection, un vecteur de distances entre la fenêtre de l'image à tester et le centroïde de chaque groupe (cluster) est envoyé à un réseau de neurones pour la classification.

Schneiderman [93] ont proposé une méthode basée sur une décomposition locale de l'image par ondelettes. La décision de classification s'appuie sur une règle bayésienne appliquée à des fonctions de densités sur l'histogramme des coefficients de l'ondelette. Une approche du même auteur, utilise une classification par l'algorithme d'Adaboost après décomposition en ondelettes [94].

Rowley et al. [90] utilisent une approche par réseau de neurones de haut niveau afin de modéliser un visage. Un visage est défini par des fenêtres de taille 20×20 pixels. Le réseau est décomposé en 26 unités cachées : 4 unités sous-régions de l'image de taille 10×10 , 16 sous-régions de taille 5×5 et 6 bandes horizontales de taille 20×5 . Les poids du réseau sont déterminés par l'algorithme de rétropropagation. Les images sont initialement pré-traitées de la même façon que dans les travaux de Sung et Poggio [108]. Plusieurs réseaux identiques sont entraînés pour détecter les visages, les résultats de ces réseaux sont fusionnés par un réseau de neurones permettant la

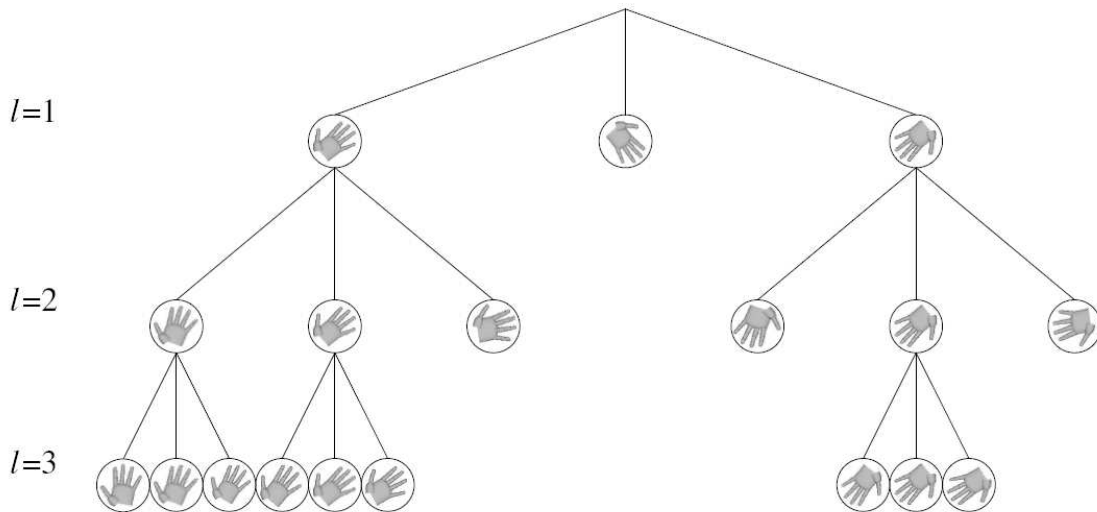


FIGURE 3.11 – Structuration hiérarchique de modèles de main

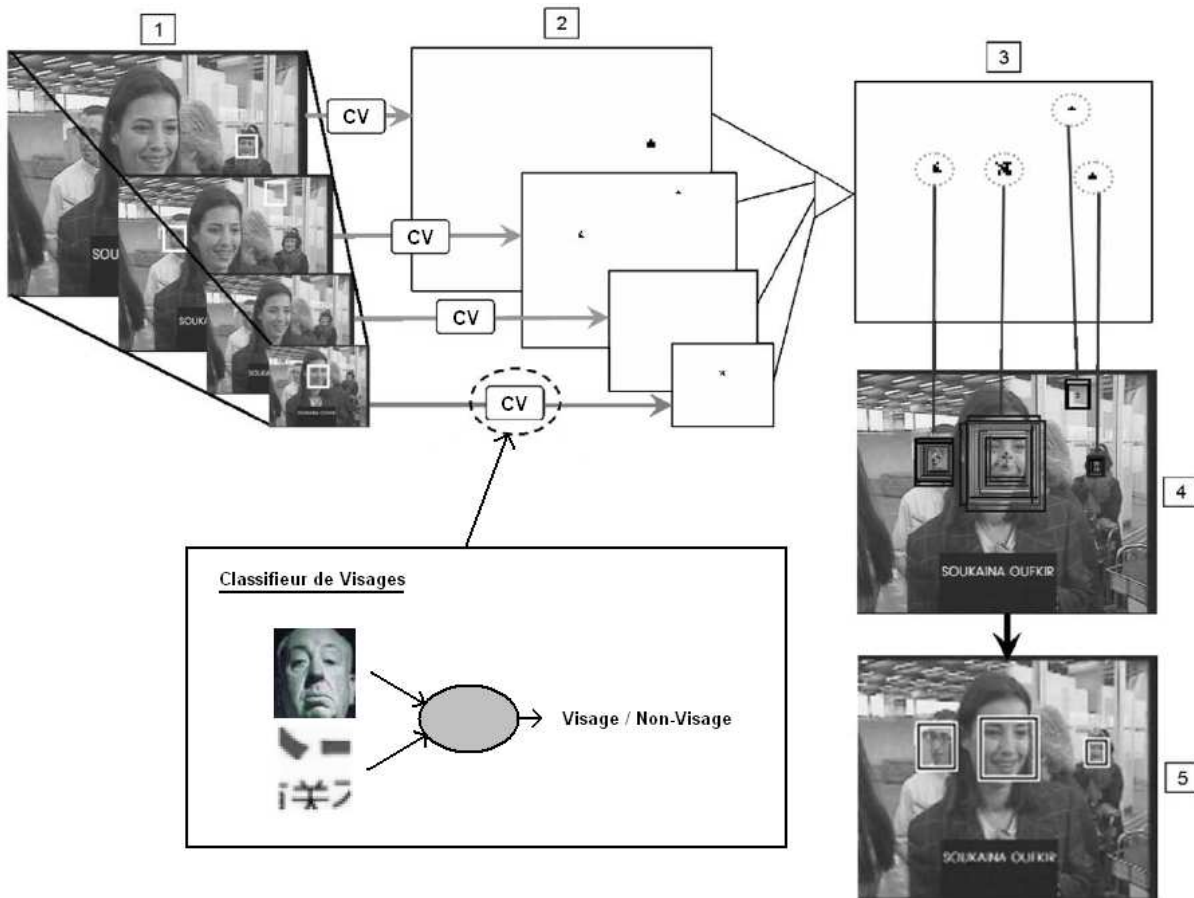


FIGURE 3.12 – Les différentes étapes du processus de localisation du visage (1) Création d'une pyramide multi-échelle de l'image d'entrée (2) Classification de chaque image de la pyramide par le classifieur de visages. (3) Projection des candidats sur l'image originale et fusion des candidats redondants (4) Application précise du classifieur dans les zones candidates (5) Classification de chaque visage suivant les réponses de classification précédente. (Garcia et Delakis [42])

décision.

Viola et Jones [117] ont défini un système basé sur une construction de classifieurs rectangulaires hiérarchiques. Ils présentent une méthode permettant de construire des classifieurs faibles correspondant à la différence entre les sommes de l'intensité des pixels entre des rectangles adjacents de l'image par l'algorithme d'Adaboost. Différentes caractéristiques sont obtenues suivant l'arrangement des rectangles. Une cascade de filtres est organisée suivant la qualité de discrimination et le coût de calcul. Cette structure hiérarchique permet une détection rapide et une bonne qualité de détection. Dans le contexte de la détection de la main, Kölsch et Turk [56] ont appliqué les détecteurs de Viola et Jones [117] afin de repérer des mains dans une image complexe. Dans cette approche une classification de postures de mains est également définie.

La méthode proposée par Garcia et Delakis [42] présente un système basé sur une architecture neuronale convolutionnelle s'appuyant sur les travaux de Lecun et al. [68] pour la reconnaissance de chiffres. Ce système de classification est une succession de convolutions et de sous-échantillonnages. Les coefficients de convolution sont déterminés par apprentissage neuronal à partir d'une base d'exemples et de contre-exemples. L'utilisation de sous-échantillonnages et de

convolutions permet de garder la cohérence spatiale. Cette méthode présente des résultats très performants même en cas d'occlusion partielle, de fonds très difficiles ou d'images bruitées. Pour la reconnaissance du visage, cette méthode présente les meilleures performances de détection sur les bases de références.

3.6 Comparaison des différentes méthodes de détection

Les méthodes basées sur la couleur de la peau sont très difficilement comparables quantitativement. En effet, la majorité des méthodes proposées ne présentent que des résultats qualitatifs sur quelques images tandis que d'autres systèmes emploient leurs propres bases de données. Cependant, des résultats sur des bases de données générales utilisées pour la détection du visage sont répertoriées à la table 3.1. Aucune méthode à notre connaissance n'a été évaluée sur des bases de données d'image de mains. Cependant, la couleur de la peau de l'intérieur de la main est plus stable que celle du visage. En effet, la peau possède une densité de mélanocytes peu variable entre les individus contrairement aux zones du corps. On peut donc supposer que la performance de classification sera meilleure sur une base d'images de paumes que sur les bases de visages testées. Les trois bases de visages servant à l'évaluation sont :

- la Compaq Database par Jones et Reg [55]
- la base de données de Chai et al. [11]
- la base de données ECU de l'Edith Cowan University [12]

Il est à noter que les résultats sont fournis tels que fournis dans les articles même si le nombre de pixels pour l'entraînement et pour l'évaluation sont différents pour chaque étude.

Les différents résultats présentés dans le tableau montrent que les méthodes de classification par Bayes ou réseau de neurones sont les plus performantes et indiquent des taux de classification similaires en utilisant les trois composantes couleurs du domaine *RVB* sur la base *ECU*. Les autres méthodes de classification ont des taux de classification relativement similaires et sensiblement moins élevés que les deux systèmes précédents. Il est à noter que la plus faible performance des systèmes paramétriques peut s'expliquer par le fait que le ton de couleur de peau ne peut pas être modélisé mathématiquement malgré le changement de représentation colorimétrique.

En terme de coût mémoire, la méthode de détermination Bayésienne est celle nécessitant le plus d'espace. En effet, en utilisant les trois composantes couleurs du domaine *RVB* codé sur 256 bits le nombre de probabilités à sauvegarder est égal à 256^3 soit 16777216 combinaisons. A contrario, les méthodes par réseaux de neurones et par Gaussiennes comportent peu de paramètres. Une gaussienne 2D contient par exemple 6 paramètres tandis qu'une gaussienne 3D en contient 12. Le coût de calcul peut également être un critère à prendre en compte. Les méthodes les plus simples et les plus rapides sont les méthodes basées sur des seuils. La méthode la plus coûteuse est la détection par réseau de neurones dont la vitesse dépend de l'architecture et de la complexité du réseau. Les autres méthodes sont très rapides et nécessitent peu de calculs.

La représentation couleur a également été largement étudiée pour déterminer le domaine le plus caractéristique du ton de peau. Les différents articles publiés comparant l'impact des différents domaines couleurs présentent des résultats variés, voire contradictoires suivant les implémentations et les bases de données. Cependant, dans tous ces articles les résultats montrent que la composante luminosité qui est parfois supprimée afin d'être robuste aux changements de luminosité a un impact important sur la qualité de la classification.

Méthode	Espace couleur	VP	FP	Référence
Compaq Database				
Seuillage	$I(YIQ)$	94.7%	30.2%	Brand et Mason [5]
Seuillage	RVB	94.7%	32.3%	Brand et Mason [5]
Bayes	RVB	90.9%	13.5%	Jones et Regh [55]
Bayes	RVB	93.4%	19.8%	Brand et Mason [5]
Réseau de neurones	TS	78%	32%	Brown et al. [7]
Ellipse	$CIE - xy$	90.0%	20.9%	Lee et Yoo [70]
Ellipse	$rv - chroma.$	90.0%	21.3%	Lee et Yoo [70]
Gaussienne	$C_B C_R$	90.0%	33.3%	Lee et Yoo [70]
Gaussienne	IQ	90.0%	33.3%	Lee et Yoo [70]
Ens. Gaussiennes	IQ	90.0%	30.0%	Lee et Yoo [70]
Ens. Gaussiennes	RVB	90.0%	15.5%	Lee et Yoo [70]
Chai et al. Database				
Bayes	RVB	90.0%	5%	Phung et al. [12]
Ens. Gaussiennes	$Y C_B C_R$	90.0%	12%	Phung et al. [12]
ECU Database				
Seuillage	HS	78.38%	19.48%	Sobottka et al. [101]
Seuillage	$Y C_B C_R$	82.00%	18.77%	Garcia et Tziritas [43]
Bayes	RVB	86.17%	15%	Jones et Regh [55]
Gaussienne	$C_B C_R$	81.07%	15%	Lee et Yoo [70]
Gaussienne	RVB	83.45%	15%	Jones et Regh [55]
Ens. Gaussiennes	RVB	83.63%	15%	Lee et Yoo [70]
Réseau de neurones	RVB	85.97%	15%	Brown et al. [7]

TABLE 3.1 – Comparaison des méthodes de détection de la couleur de la peau. VP désigne les Vrais Positif et FP les Faux Positifs

Les deux principaux défauts des méthodes de détection par les pixels de peau sont qu'ils ne sont pas robustes à des changements importants de luminosité et qu'il est difficile de différencier les objets ayant un ton de peau d'un visage ou d'une main. Par contre, la mise en place de ces méthodes est relativement aisée et toutes ces techniques sont robustes aux occlusions partielles et aux problèmes de hautes dimensions.

Les méthodes basées sur la forme de la main en se basant sur un modèle ou par classification hiérarchique présentent peu de résultats quantitatifs mais plutôt des résultats qualitatifs (contrairement aux systèmes de détection du visage). Certaines caractéristiques de ces méthodes

permettent d'émettre plusieurs conclusions. Ces systèmes sont beaucoup plus complexes à mettre en œuvre que les systèmes de classification des pixels de peau mais permettent des traitements en temps réel.

Au niveau de la détection, ces systèmes présentent de nombreux avantages. Toutes les méthodes basées sur les contours sont robustes aux changements de luminosité même importants. Ils permettent également de détecter une main (ou un visage) même dans un fond de ton de peau et sont robustes à des occlusions partielles. Le principal inconvénient de ces systèmes consiste à modéliser et à détecter des problèmes peu ou pas contraints tels que ceux présentant diverses orientations dans l'espace et les doigts pliés. Une autre limite est la précision de la détection et le manque d'information sur celle-ci dans le cas des classifications hiérarchiques. Par exemple avec ces détecteurs, l'orientation de l'objet détecté est inconnue ce qui nécessite un processus de recalage si l'on désire comparer l'élément extrait.

3.7 Extraction de la zone d'intérêt : la paume

La segmentation de la paume est le processus de correction des distortions et de la normalisation de toutes les paumes dans un même système de coordonnées. Cette section décrit deux méthodes d'extraction de la région d'intérêt (ROI). La première méthode extrait une zone carrée de la paume et la seconde effectue une segmentation circulaire.

3.7.1 Segmentation de la paume sous forme carrée

3.7.1.1 Extraction des points importants

La main doit être alignée de manière stable afin de permettre les comparaisons. Pour cet alignement, les axes la caractérisant doivent être définis par rapport à chaque saisie. Ces méthodes reposent sur la recherche des points T1 à T5 et B1 à B4 de la main (Figure 3.13) permettant de déterminer un repère. Ces points sont définis en utilisant les caractéristiques de la forme de la main : le contour de la main contient neuf points de références (les points au bout des doigts et les points entre deux doigts adjacents). Ces points sont extraits en déterminant les 9 plus fortes courbures du contour.

La paume est la surface interne à la main entre les doigts et le poignet. A partir des points détectés, deux méthodes sont principalement utilisées pour extraire la région d'intérêt.

3.7.1.2 Extraction de la paume

La première méthode d'extraction utilise directement les points importants pour extraire la paume [135]. La méthode est illustrée dans le Fig. 3.14a. Deux constantes sont définies pour cette méthode : la distance entre les points O et O' et la distance entre A et B. Cette méthode d'extraction a comme principale caractéristique d'obtenir une taille et une position fixe des points dans le système de coordonnées pour toutes les images de main. Cette segmentation a besoin d'une bonne détermination des points B2 et B4 sinon la détection de la partie centrale de la main devient très imprécise.

Dans la seconde méthode, deux points P1 et P2 sont déterminés comme indiqués dans la Fig. 3.14b. La paume est définie comme la région carrée ayant deux de ses coins placés sur les milieux

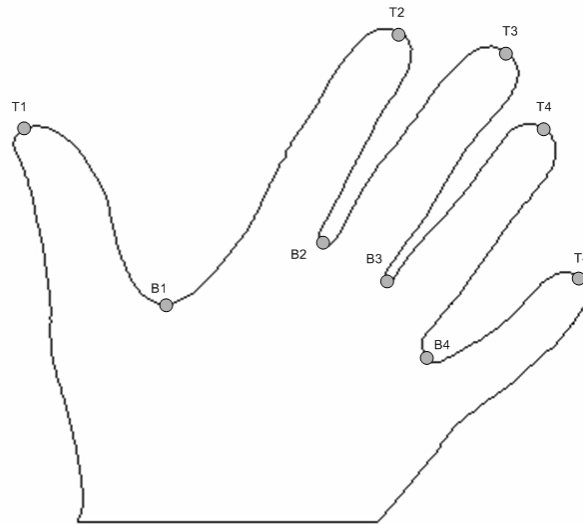


FIGURE 3.13 – Points importants d'une main

des segments $[P1, B2]$ et $[B4, P2]$. Avec ce principe, les deux points B2 et B4 sont fondamentaux pour une extraction précise de la paume. Contrairement à la méthode précédente, la taille de la région extraite est dépendante de la taille de la main. Pour la comparaison entre différentes paumes, celle-ci doit donc être normalisée après le processus d'extraction.

Dans les deux dernières méthodes d'extraction de la paume, la taille de la région extraite peut permettre un pré-filtrage suivant les utilisateurs avant même l'étude de la paume. En effet, pour un même utilisateur la taille de la région est sensiblement identique d'une acquisition à une autre. Ainsi si la dimension de la région de référence et la région extraite sont trop différentes, les deux zones extraites sont différentes et ne proviennent donc pas de la même main.

3.7.2 Segmentation de la paume de façon circulaire

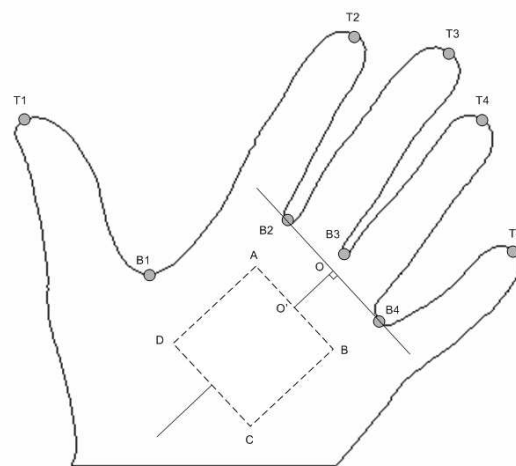
Une autre méthode d'extraction de la région d'intérêt effectue une segmentation circulaire de la paume (Fig. 3.14c). La paume est définie comme le plus grand cercle inscrit du contour de la main. Les mains étant de différentes tailles, le rayon obtenu dépend de chaque utilisateur et la région extraite doit donc être normalisée avant d'être comparée. Cette méthode ne présente que peu d'avantage. En plus d'avoir une extraction moins stable que les méthodes précédentes, la région est circulaire donc moins aisée à traiter numériquement.

3.8 Conclusion sur les approches de détection de la main

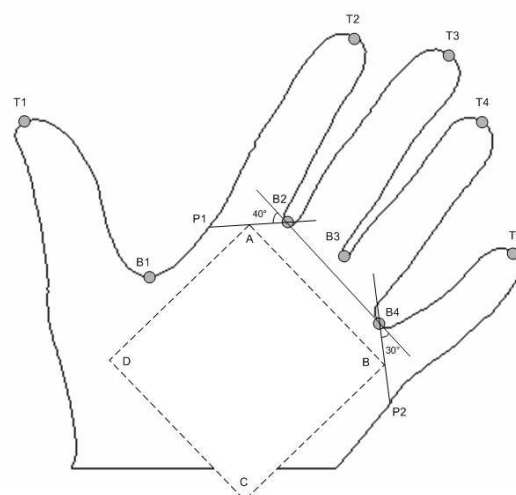
Dans ce chapitre, un état de l'art des méthodes de détection de mains et de paumes est présenté. Les différentes approches d'extraction de la main comparées dans le tableau 3.1 montrent que les méthodes d'extraction basées sur la couleur de la peau complètent les méthodes de détection intégrant la notion de forme. Les avantages et les inconvénients des deux approches

sont opposées. Les méthodes par couleur sont peu robustes aux changements de luminosité mais sont précis tandis que les algorithmes s'appuyant sur la forme sont robustes mais imprécis. Il est ainsi intéressant de combiner ces deux approches ce que nous présentons dans le chapitre suivant.

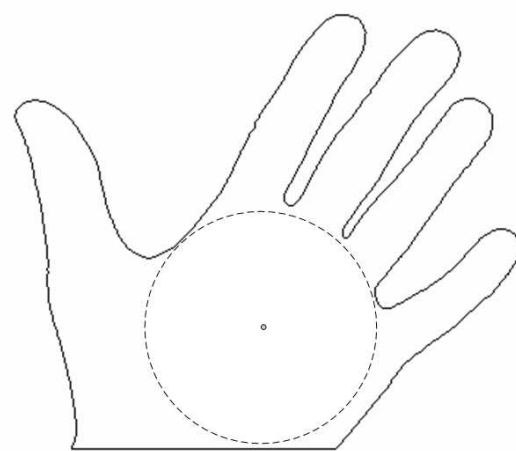
Les méthodes d'extraction de la paume sont peu nombreuses. Elles s'appuient sur des segmentations géométriques simples après une définition très précise du contour de la main. Bien que peu développée dans les articles biométriques, la phase de segmentation est une donnée fondamentale pour la reconnaissance. En cas de mauvaise définition de l'élément biométrique, il devient très difficile d'obtenir ensuite une caractérisation robuste et fiable.



(a)



(b)



(c)

FIGURE 3.14 – Extraction de la paume par les points importants (a), par les angles (b), ou de manière circulaire (c)

Chapitre 4

Fusion d'information pour la segmentation de la main dans une image complexe

Sommaire

4.1	La segmentation pour la reconnaissance biométrique sans contact	49
4.2	Segmentation couleur multi-résolution des zones de peau par apprentissage	51
4.2.1	Apprentissage de la couleur de la peau	51
4.2.1.1	Perceptron multi-couche	51
4.2.1.2	Machine à vecteurs de support	52
4.2.2	Détection multi-résolution	54
4.2.3	Post-traitement sur la carte de probabilité	55
4.3	Utilisation de connaissance de forme à priori	57
4.3.1	Apprentissage de la forme d'une main	58
4.3.2	Recherche d'une main à partir du modèle	61
4.3.2.1	Initialisation du modèle de forme	61
4.3.2.2	Évolution du contour du modèle	62
4.4	Évaluation de la méthode de segmentation	64
4.4.1	Base de données d'évaluation	64
4.4.2	Évaluation de la détection de peau	65
4.4.3	Évaluation de la segmentation par contour actif de forme	67
4.5	Conclusion sur la détection de la main dans une image complexe	68

4.1 La segmentation pour la reconnaissance biométrique sans contact

Dans notre approche, la segmentation habituelle utilisée pour la détection de la main en biométrie ne peut pas être appliquée. En effet, l'utilisateur doit pouvoir se faire reconnaître dans différentes conditions d'éclairage et de fond. Plusieurs contraintes doivent être définies lors de l'acquisition afin de permettre dans une étape postérieure l'identification d'un individu.

- L'utilisateur doit présenter sa main proche du plan perpendiculaire à l'axe du capteur. En effet sans cette contrainte, la détection de la paume peut s'avérer impossible, par exemple si l'utilisateur place sa main dans un plan parallèle à l'axe du capteur.
- L'utilisateur ne doit pas plier les doigts afin de supprimer les auto-occlusions et permettre une vue dégagée de la paume.
- L'utilisateur doit présenter sa main de façon complète afin de déterminer complètement la paume et les doigts.

En suivant ces contraintes, les degrés de libertés de détection de la main définis à la section 2.2.3 sont largement diminués. Finalement, le nombre de degrés de liberté autorisés dans notre approche est de 7 (5 pour la rotation des doigts dans le plan perpendiculaire à l'axe du capteur et 2 pour la rotation de la main également dans le plan perpendiculaire à l'axe du capteur). Des exemples d'acquisitions typiques avec différents capteurs sont présentés à la Fig. 4.1. Les images sont inversées dans l'axe horizontal pour une interaction naturelle avec les utilisateurs dans le logiciel d'acquisition.

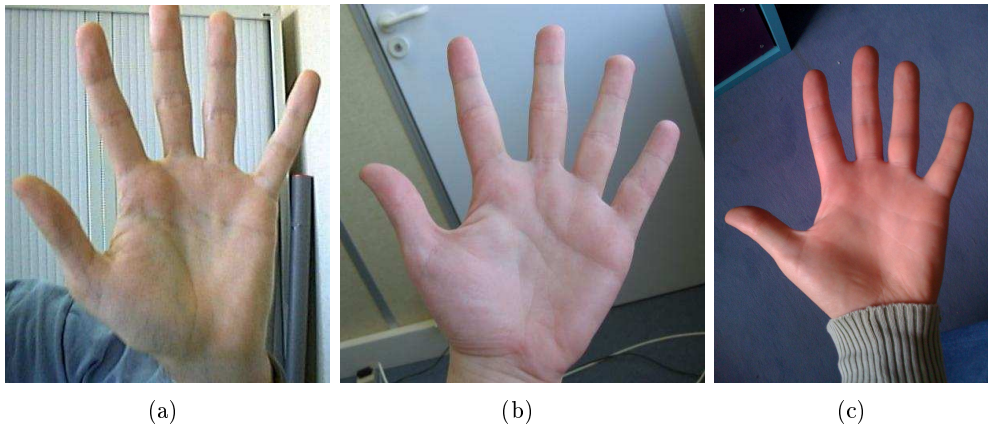


FIGURE 4.1 – Exemples d'acquisitions avec (a) une Webcam (b) un appareil photo numérique (c) un téléphone mobile

La problématique majeure de la détection dans les systèmes de reconnaissance biométrique réside dans la nécessité d'une segmentation de la modalité très précise. Une détection imprécise ne permettra pas une reconnaissance robuste à posteriori. Ainsi dans notre contexte, la précision doit être privilégiée vis-à-vis de la robustesse. Dans la majorité des systèmes définis s'appuyant sur la notion de forme, la main est détectée à partir d'un modèle générique ce qui implique une détection robuste mais imprécise dans le cas général puisque la méthode ne s'adapte pas à chaque individu. A contrario, les méthodes utilisant la détection par la couleur de la peau sont précises puisque s'adaptant localement à l'image (au pixel près) mais sont limitées au niveau de la robustesse à la luminosité par exemple.

Les images à traiter pour notre méthode de reconnaissance sont très variables au niveau du fond et de la luminosité ambiante. Les acquisitions envisagées sont effectuées dans des conditions non contrôlés. En effet, il n'est pas envisageable de contraindre l'utilisateur quand à son environnement. Ainsi, nous proposons une détection combinant à la fois des notions de structure et des notions de couleur [29] pour obtenir une segmentation robuste et précise. Dans une phase

d'initialisation, la main est recherchée en assumant qu'elle est la partie de peau la plus représentée dans l'image. Cette détection couleur détaillée dans le paragraphe 4.2 est déterminée en multi-résolution en utilisant un réseau de neurones multi-couche. Ensuite, un modèle de contour actif défini par Cootes [20] guide la recherche précise et locale de la main. Cette seconde méthode est décrite dans le paragraphe 4.3.

4.2 Segmentation couleur multi-résolution des zones de peau par apprentissage

Dans ce travail, l'objectif de la segmentation par la couleur de la peau est de classifier un pixel de peau sans utiliser les informations de voisinage. L'information de voisinage n'a pas été utilisée car cette phase n'est utilisée que comme pré-traitement avant la phase de segmentation de l'information de contour. Cette phase n'a donc pas besoin d'être extrêmement précise mais elle doit avant tout être très rapide. La segmentation de la peau par réseau de neurones a été utilisée comme décrit dans le chapitre précédent. Cependant, aucune étude comparative sur l'architecture du réseau et le domaine couleur n'a été menée. Elle est effectuée dans le paragraphe suivant. Dans un second temps, une architecture multi-résolution proche du temps réel est présentée. L'architecture du classifieur présente une amélioration significative des performances. A titre de comparaison, une approche basée sur les machines à vecteurs de support est également décrite.

4.2.1 Apprentissage de la couleur de la peau

Deux classifieurs sont présentés pour la caractérisation de la couleur de la peau, le perceptron multi-couche et les machines à vecteurs supports. Finalement, l'algorithme multi-résolution et les traitements après la classification des pixels sont détaillés.

4.2.1.1 Perceptron multi-couche

Le perceptron multi-couche est un classifieur bien connu. Une description complète de l'algorithme peut être trouvée en [61].

Plusieurs critères ont influé dans le choix d'une méthode d'apprentissage. Premièrement, les modèles représentant la couleur de peau sans généralisation, tels que les méthodes de table de consultation ou la méthode de Bayes, dépendent intégralement de la qualité de la base de données. Pour ces systèmes de segmentation, la base d'apprentissage doit complètement décrire tout les tons de peau et de fond. De plus dans le cas de la distribution de Bayes, elle doit également prendre en compte la représentativité des couleurs entre la peau et le fond. Ces méthodes possèdent également l'inconvénient de consommer une part mémoire très importante. Deuxièmement, les modèles paramétriques supposent que la distribution de la couleur de peau peut être modélisée par une fonction connue.

Les modèles d'apprentissage permettent de résoudre toutes ces contraintes. De par sa bonne faculté d'approximation et son exécution rapide, le réseau de neurones est l'une des méthodes les plus utilisées en apprentissage. Dans notre approche, trois types de réseaux ont été entraînés. Les deux premiers sont des perceptrons mono-couches possédant N neurones sur la couche cachée. Le premier a une fonction de transfert de type sigmoïde tandis que le second à une fonction de transfert de type tanh. La troisième structure est un réseau avec deux couches cachées. La première couche possède $N \in [1, 10]$ neurones avec la fonction de transfert sigmoïde, la seconde possède $N2 \in [1, N]$ neurones mais avec la fonction tangente hyperbolique (tanh). D'autres structures de

réseaux plus complexes n'ont pas été testées afin de limiter le temps de calcul lors de la classification. Chaque réseau est utilisé avec les composantes couleurs en entrée. La sortie est composée d'un seul neurone dont la valeur de sortie $s \in [0, 1]$ correspond à la probabilité que le pixel couleur d'entrée du réseau soit un pixel de peau. Deux exemples de réseaux sont représentés à la figure 4.2.

Dans la phase d'apprentissage, les valeurs de sorties sont fixées afin d'obtenir la probabilité qu'un pixel soit un pixel de peau. Ainsi, la valeur de sortie est égale à 1 lorsque un pixel couleur est un pixel de peau ou à 0 dans le cas où le pixel est un pixel de fond. Dans la phase d'évaluation, la probabilité que chaque pixel dans l'image soit un pixel de peau est déterminée par le réseau de neurone. Cette phase de détection est illustrée à la figure 4.3. Le processus complet calcule la carte de probabilité de l'image d'entrée.

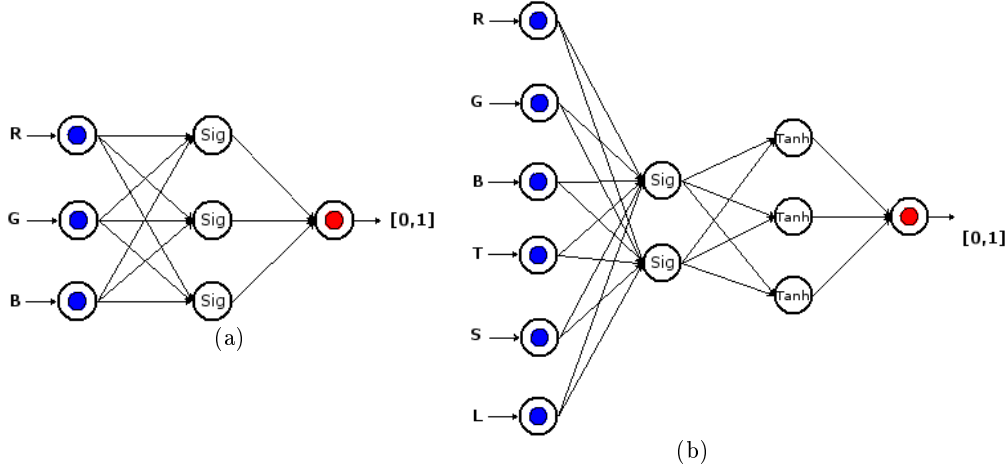


FIGURE 4.2 – Réseau de neurones (a) Réseau de type 1 avec $N=3$ dans le domaine RGB (b) Réseau de type 3 avec $N=2$, $N_2=3$ dans le domaine RGB+TSL

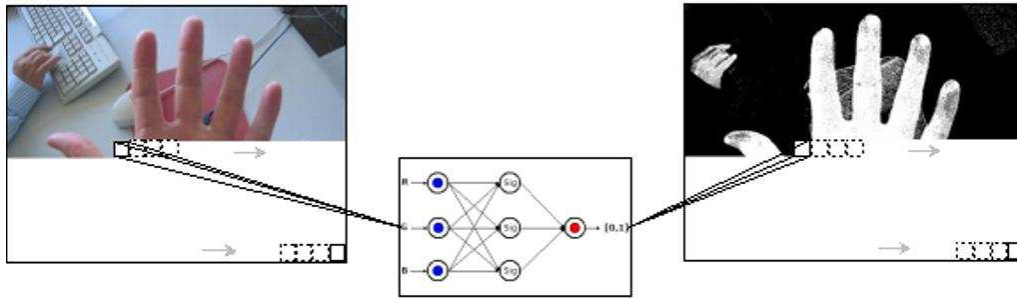


FIGURE 4.3 – Calcul de la carte de probabilité (pixels par pixels)

4.2.1.2 Machine à vecteurs de support

Une autre méthode d'apprentissage a été testée afin d'obtenir une classification optimale : Les machines à vecteurs de support (SVM) [115, 21, 10, 13]. Les SVM, introduites par Vapnik, permettent à l'aide d'un apprentissage de classifier des éléments dans différentes classes à l'aide de fonctions linéaires après projection non linéaire. Dans le cas de séparation entre deux classes, la fonction de discrimination correspond à un hyperplan maximisant la marge (la distance entre

les plus proches vecteurs de chaque classe). Cet hyperplan peut être défini, pour un ensemble de vecteurs $x \in \mathbb{R}^d$, dans l'espace à d dimensions, par $f(x) = (x.w) + b$ où w est le vecteur normal au plan et $|b| / \|w\|$ est la distance entre le plan et l'origine. Ceci peut-être représenté dans l'espace à 2 dimensions ($d=2$) par le schéma suivant :

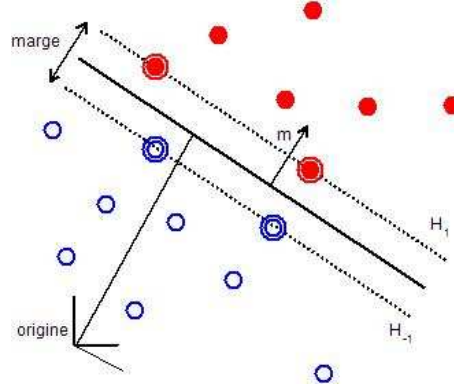


FIGURE 4.4 – Classification par les SVMs

Où H_1 (respectivement H_{-1}) représente la droite parallèle à $f(x)$ passant par les points de la classe 1 (respectivement -1) les plus proches de l'hyperplan. Ces points sont appelés les vecteurs supports (ils sont entourés sur le schéma).

La liberté de $f(x)$ (une modification de w et un changement cohérent de b laisseraient la fonction inchangée) nécessite une normalisation, celle-ci est effectuée en utilisant les vecteurs de support qui sont définis par $f(x_i) = y_i$ où les vecteurs (points en 2d) de l'ensemble d'apprentissage sont défini comme (x_i, y_i) , $y_i \in -1, +1$, où y_i représente la classe du vecteur x_i . La fonction normalisée, en utilisant le fait que les données sont linéairement séparables, permet de déterminer la classe de chaque vecteur. En effet :

$$\begin{cases} x_i.w + b \geq +1 & \Rightarrow y_i = +1 \\ x_i.w + b \leq -1 & \Rightarrow y_i = -1 \end{cases} \quad (4.1)$$

Le plan maximisant la marge peut être déterminé en minimisant $\|w\|^2/2$ sous les contraintes exprimées précédemment qui sont équivalentes à $y_i(x_i.w + b) - 1 \geq 0, \forall i$.

Dans le cas, le plus probable, où tous les vecteurs ne sont pas linéairement séparables une variable d'erreur ξ_i est introduite, l'équation précédente devient :

$$\begin{cases} x_i.w + b \geq +1 - \xi_i & \Rightarrow y_i = +1 \\ x_i.w + b \leq -1 + \xi_i & \Rightarrow y_i = -1 \\ \forall i \xi_i > 0 \end{cases} \quad (4.2)$$

L'erreur globale du système n'excède pas $\sum_i \xi_i$. La maximisation de la marge s'exprime donc par $\|w\|^2/2 + C * (\sum_i \xi_i)$ où C est un paramètre indiquant la pénalité assignée à une erreur (permet de régler la généralité de la solution trouvée). Pour les deux méthodes décrites, seul le cas de la séparation linéaire a été traité.

Tandis que les réseaux de neurones, dans le cas de problèmes non linéaires, déterminent une fonction non linéaire pour séparer les données, les SVMs déterminent un hyperplan en projetant

non-linéairement les données d'apprentissage (exemples et contre exemples) dans un autre espace et en les séparants linéairement dans celui-ci. Cette projection s'effectue en utilisant une fonction non linéaire $\Phi(x)$. Dans le but de supprimer des calculs dans des espaces à fortes dimensions, une fonction noyau $K(x, y)$ est utilisée correspondant à $K(x, y) = \Phi(x) \cdot \Phi(y)$. La fonction de décision $f(x)$ peut alors s'exprimer sous la forme :

$$f(x) = \sum_{i=1}^l \alpha_i y_i K(s_i, x) + b \text{ où } s_i \text{ sont les vecteurs supports.} \quad (4.3)$$

Tout comme la fonction d'activation pour les réseaux de neurones, de nombreux noyaux ont été spécifiés. Les plus courants sont définis par les équations suivantes :

$$K(x, y) = \gamma + (x \cdot y)^p \quad (4.4)$$

$$K(x, y) = \exp \left(-\frac{1}{2\sigma^2} \|x - y\|^2 \right). \quad (4.5)$$

$$K(x, y) = \tanh(x \cdot y - \Theta) \quad (4.6)$$

Pour la caractérisation de la couleur, les entrées et la sortie sont identiques à l'approche basée sur les réseaux de neurones avec le même pré-traitement en phase de détection et d'apprentissage.

4.2.2 Détection multi-résolution

Le coût des opérations de segmentation est inégal suivant que l'on utilise l'approche basée sur les SVMs ou l'approche basée sur les réseaux de neurones. Malgré l'amélioration algorithmique sensible des méthodes de classification principalement pour les méthodes basées sur les SVMs, ceux-ci restent trop coûteux pour une application temps réel.

Afin d'augmenter la vitesse de détection tout en gardant une bonne qualité de segmentation, nous avons utilisé une approche par décomposition pyramidale. L'approche se base sur un traitement itératif de l'image initiale suivant sa résolution. Elle est explicitée par l'algorithme 1 et illustrée à la Fig. 4.5. Dans cet algorithme, l'image initiale est compressée à une faible résolution. A cette résolution, la carte de probabilité est calculée. A partir de cette carte initiale, la résolution est augmentée et la probabilité des pixels de peau est recalculée près de la frontière d'une transition peau/non-peau. Dans les approches multi-résolution en cascade utilisées par exemple dans l'approche de Viola et Jones, les calculs sont effectués sur l'image complète à différentes résolutions puis fusionnés ensuite pour détecter l'élément recherché à différente taille. Ici, notre approche multi-résolution effectue d'abord des calculs à une faible résolution puis n'effectue des calculs à des résolutions supérieures que dans les zones d'incertitudes. Il n'est pas nécessaire de faire une recherche complète sur l'image à différentes résolutions car l'information recherchée est une information colorimétrique et non géométrique. Contrairement aux informations géométriques, la couleur ne dépend pas d'un facteur d'échelle.

Cette méthode permet un gain en vitesse très important puisque les éléments de peau ont généralement un rapport surface/contour très important ce qui implique que seulement une petite partie des pixels de peau sont recalculés à chaque résolution. L'algorithme multi-résolution permet également de garder les structures les plus fines de l'image tels que les doigts.

Bien que cette méthode présente une segmentation pertinente, un post-traitement est nécessaire afin d'améliorer la qualité de la détection. Pour une reconnaissance stable dans une seconde étape, la détection de la main doit être aussi précise que possible. Comme indiqué précédemment, la détection a un impact important sur la reconnaissance biométrique.

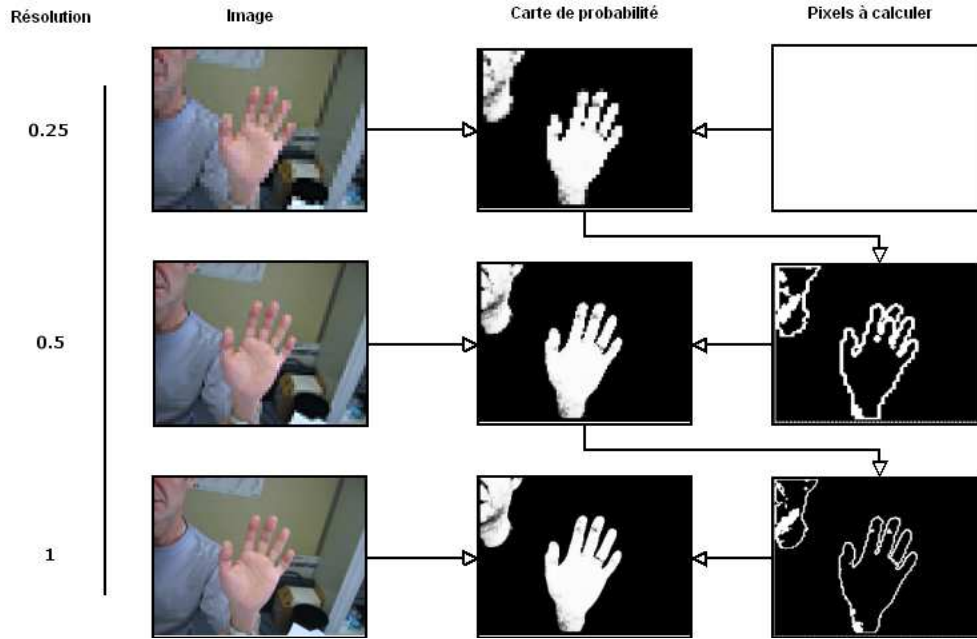


FIGURE 4.5 – Carte de probabilité pyramidale déterminé à différentes résolutions

4.2.3 Post-traitement sur la carte de probabilité

La carte de probabilité est une image en niveau de gris. Un traitement est ainsi nécessaire afin de binariser le ou les objets couleurs de peau [29].

De nombreuses méthodes ont été définies afin d'effectuer cette opération de binarisation. La méthode la plus utilisée consiste en une segmentation par seuillage de l'histogramme de la carte de probabilité. Le seuillage peut être fixe dans des conditions contrôlées avec un fond uniforme ou déterminé de manière optimale en maximisant la variance inter-classes [99]. D'autres approches plus complexes s'appuient sur un algorithme de ligne de partage des eaux dans le domaine chromatique [8], un filtre médian [75], une combinaison entre des informations de contours et de régions [87] ou un regroupement de pixels [112].

Notre approche s'appuie sur une succession de traitement permettant d'assurer une plus grande cohérence spatiale entre les pixels et de déterminer l'objet main dans l'image. Tout d'abord, une dilatation morphologique d'un élément de rayon 1 est appliquée afin de grouper les pixels voisins sans recoller les doigts entre eux. Ensuite, la carte est segmentée par un seuil fixé à 0.5 afin d'avoir une répartition ; chaque pixel est considéré comme un pixel de peau si la valeur de sortie de la carte de probabilité est supérieure à 0.5 sinon il est classé comme un pixel de fond.

A partir de cette carte binaire, différencier plusieurs objets de couleur de peau est une tâche

Algorithme 1 : Calcul multi-résolution de la carte de probabilité

Input :

I : Image d'entrée.

$Retaille(I, x)$: Fonction qui agrandit ou réduit l'image I d'un facteur x .

$Energie(I, x, y)$: Fonction qui retourne l'énergie moyenne autour du pixel (x, y) .

$f(C)$: Fonction qui indique la probabilité que la couleur C soit un pixel de peau.

N_k : Nombre de facteur de réduction de la carte de probabilité.

$\{K_k\}$: Ensemble des facteurs de résolution.

S_{peau} : Seuil qui indique que le pixel est un pixel de peau.

S_{fond} : Seuil qui indique que le pixel n'est pas un pixel de peau.

Output :

P_{N_k} : Carte de probabilité.

Initialisation

$I_0 = Retaille(I, K_0)$

for $i \leftarrow 1$ **to** $Hauteur(I_0)$ **do**

for $j \leftarrow 1$ **to** $Largeur(I_0)$ **do**

$P_0[i, j] = f(I_0[i, j])$

end

end

for $i \leftarrow 1$ **to** $Hauteur(I_0)$ **do**

for $j \leftarrow 1$ **to** $Largeur(I_0)$ **do**

$E = Energie(P_0, i, j)$

 si $E > S_{peau}$ ou $E < S_{fond}$ alors $M_0[i, j] = 0$

 sinon $M_0[i, j] = 1$

end

end

Calcul multi-résolution

for $k \leftarrow 1$ **to** $N_k - 1$ **do**

$P_k = Retaille(P_{k-1}, K_0/K_k)$

$M_k = Retaille(M_{k-1}, K_0/K_k)$

for $i \leftarrow 1$ **to** $Hauteur(P_k)$ **do**

for $j \leftarrow 1$ **to** $Largeur(P_k)$ **do**

 si $M_k[i, j] = 1$ alors $P_k[i, j] = f(I[i * K_k, j * K_k])$

end

end

for $i \leftarrow 1$ **to** $Hauteur(M_k)$ **do**

for $j \leftarrow 1$ **to** $Largeur(M_k)$ **do**

 si $M_k[i, j] = 1$ alors

$E = Energie(P_k, i, j)$

 si $E > S_{peau}$ ou $E < S_{fond}$ alors $M_k[i, j] = 0$

end

end

end

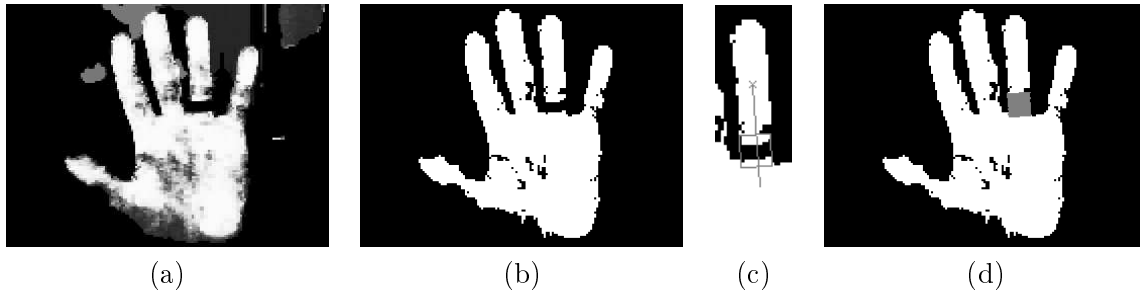


FIGURE 4.6 – Post-Traitement (a) Carte de probabilités (b) Carte binarisée (c) Ajustement du patch (d) Carte binarisée avec patch

complexe. Dans le but de les distinguer, le contour des objets est déterminé par l'algorithme de Di Zenzo [27]. Cet algorithme se base sur une recherche des composantes couleurs principales afin de calculer le gradient le plus pertinent dans une image multivaluée. En assumant que le gradient entre deux objets dans une image est supérieur à un seuil fixe δ , tous les pixels de peau pour lesquels le gradient est supérieur à ce seuil sont considérés comme des pixels de fond. Finalement, comme notre système de détection de main sera utilisé pour une reconnaissance biométrique, nous supposons que seulement une main est présentée au système, ainsi la région ayant la plus grande surface est préservée et considérée comme une main.

Une hypothèse n'est généralement pas formulée lors de l'utilisation de la segmentation par la couleur de la peau. Dans tous les cas de segmentation basée sur la couleur, la couleur est supposée homogène et représentative de l'objet à détecter. En fait, cette hypothèse n'est pas vérifiée dans le cas où l'utilisateur porte un artefact tel qu'une bague. Pour ce cas, un post-traitement supplémentaire a été défini. En plus de la région avec la plus grande surface, les régions ayant une surface et une elongation importantes sont conservées. Ces régions correspondent potentiellement à des doigts déconnectés de la main d'un utilisateur portant des bagues. Pour reconnecter ces régions à la main, chaque région est approximée par une ellipse et une recherche est effectuée dans la direction de l'axe majeur si la main est proche. Si la région est proche de la main, un patch ajusté est ajouté pour reconnecter le doigt à la main. Ce patch a une largeur égale à la largeur de l'axe mineur et une longueur dépendante de la distance entre le doigt et la main. Il est orienté suivant la direction de l'axe principal et centré entre la main et le doigt. Le processus complet de segmentation est illustré à la Fig.4.6.

4.3 Utilisation de connaissance de forme à priori

La segmentation par la couleur de la peau n'est pas suffisamment robuste (Cf. Paragraphe 3.6) pour détecter précisément une main dans une image présentant un fond non-uniforme. Deux difficultés majeures ne peuvent pas être complètement résolues : le changement de luminosité influence la qualité de la détection de la peau et il est très difficile de différencier directement une main d'un visage à partir de la carte de peau segmentée. Le cas d'une acquisition du visage et de la main en simultanée peut arriver dans les acquisitions à l'aide d'une Webcam. Les modèles de contours actifs introduits par Tim Cootes [20] permettent de résoudre ces deux problèmes. Ils sont en effet robustes aux changements de luminosité et segmentent des formes apprises.

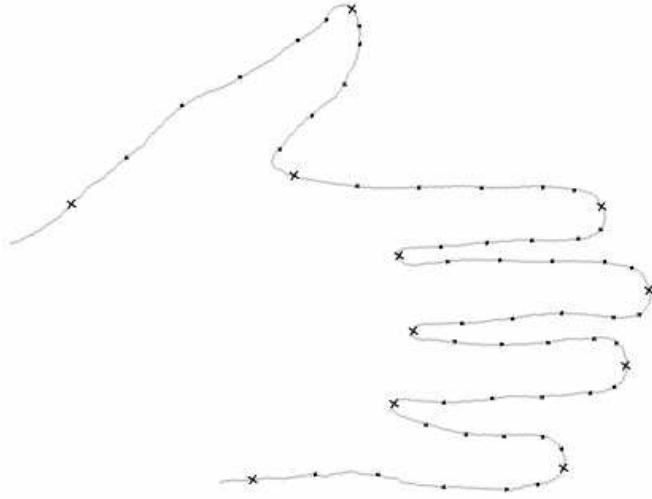


FIGURE 4.7 – Schéma d'une main annotée avec 60 points

4.3.1 Apprentissage de la forme d'une main

L'apprentissage de la forme de la main consiste à définir toutes les formes possibles que peuvent prendre les mains à partir d'une base. Afin de limiter la représentation de la forme d'une main, seuls quelques points caractéristiques sont utilisés pour représenter la forme de chaque main et non tous les points du contour. Ces points caractéristiques s'appuient sur les 9 points de références spécifiés dans le paragraphe 3.7.1.1.

La modélisation de la main nécessite cependant plus de précision. Tout d'abord deux points près du poignet sont ajoutés aux modèles. Le point du poignet près du pouce est défini par $B0$ et le point du poignet proche de l'auriculaire par le point $B5$. Suivant la formulation de la figure 3.13 du paragraphe 3.7.1.1, ces points sont définis par :

$$\begin{cases} \|B0T1\| &= \sigma_{B0}\|B1T1\| \\ \|B5T4\| &= \sigma_{B5}\|B4T4\| \end{cases} \quad (4.7)$$

Où $B0$ et $B5$ sont situés sur le contour de la main. σ_{B0} et σ_{B5} sont fixés empiriquement afin de modéliser au mieux les doigts de la main à 1.5 et 1.2 respectivement. Ensuite, M points intermédiaires sont disposés entre les points de référence sur le contour de la main afin de définir le modèle X d'une main. X est ainsi décrit par $[X[0], X[1], \dots, X[(M+1) \times 10]]$ où $X[i]$ est le i -ème landmark. Le schéma 4.7 illustre un exemple de marquage avec M fixé à 5.

A partir de la base de mains annotées, une transformation est nécessaire afin d'aligner tous les modèles en position, orientation et échelle. Pour ce faire, une analyse de Procruste généralisée [20] peut être employée. Cette approche itérative consiste en une minimisation de la somme des distances entre chaque forme et la forme moyenne. Elle est décrite par l'algorithme 2.

Cet algorithme consiste à modifier l'ensemble des points de chaque main apprise pour la faire coïncider avec une main de référence en terme de position, d'orientation et d'échelle. Dans un premier temps, une forme de référence correspondant à l'une des mains est choisie de façon arbi-

Algorithme 2 : Alignement des mains par analyse de Procruste généralisée

Input :

X : Ensemble d'images annotées.

e : Erreur d'alignement maximale autorisée

Output : L'ensemble X aligné.

begin

Initialisation

Translation de chaque exemple à son centre de gravité situé comme origine.

Choix du premier exemple comme référence et redimensionnement de chaque image afin que $|\bar{X}| = 1$.

Choix de $\tilde{M} = |X_0|$ comme estimation initiale de la forme moyenne.

$M = NULL$.

Alignement des mains :

while $|M - \tilde{M}| > e$ **do**

$M = \tilde{M}$

Alignement de toutes les formes suivant la moyenne M .

$\tilde{M} = |\bar{X}|$.

Changement de l'estimation en alignant la moyenne avec $|X_0|$ et redimensionnement afin que $|\bar{X}| = 1$.

end

end

traire. Puis, pour chacune des autres, les points pour l'aligner sont modifiés suivant la première. L'alignement se fait en quatre temps :

- Le calcul des centres de masse de la main de référence et des mains de la base de données (qui est en fait le barycentre des points caractéristiques). Les mains sont ensuite translatées au centre de masse du point de référence.
- Le calcul de la taille de chaque main. La norme utilisée est la norme L2. La distance pour chaque main correspondant à la somme des distances de chaque point par rapport au centre de masse. Une fois la distance des deux mains calculée, une homothétie est appliquée à la main à aligner avec un facteur (distance de la main de référence / distance de la main en cours).
- L'alignement en orientation : pour effectuer cette étape, la matrice de corrélation entre les deux mains est déterminée puis une décomposition en valeurs propres de celle-ci effectuée. Ceci permet d'obtenir la matrice de rotation permettant d'aligner la main concernée avec la référence en orientation.
- Le déplacement de la main au barycentre de la référence.

Une fois ces étapes terminées, la moyenne de toutes les mains est calculée. Le processus précédent s'applique avec la moyenne comme main de référence jusqu'à convergence. La figure 4.8 montre l'ensemble des contours de la base d'apprentissage avant et après alignement.

L'ensemble des listes de points est une donnée trop importante et trop difficile à exploiter telle quelle. L'analyse en composantes principales a pour but de fournir à partir de ces données et

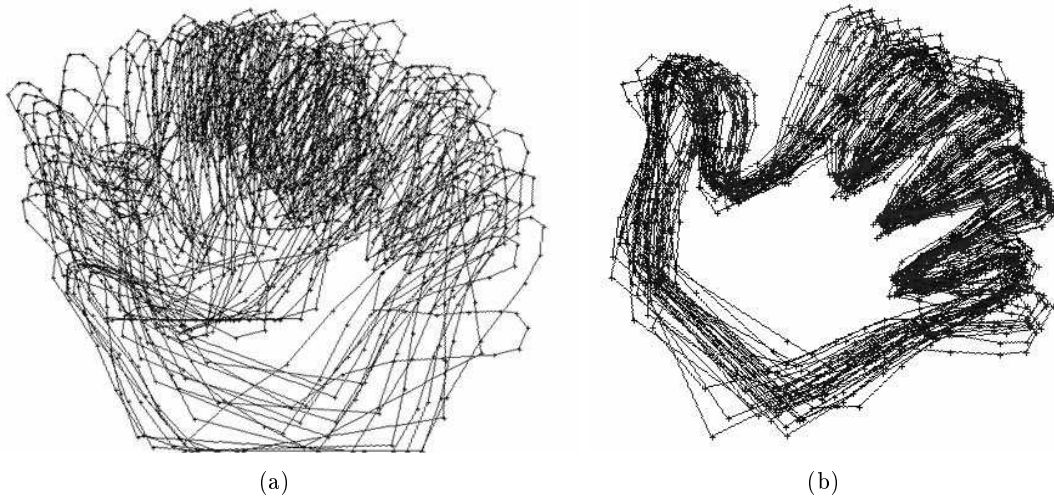


FIGURE 4.8 – Mains avant (a) et après alignement (b)

de la moyenne calculée précédemment un modèle qui contient les variations possibles d'une forme.

Le principe de cette méthode explicité à l'algorithme 3 est de calculer la matrice de covariance de l'ensemble de la base d'apprentissage, qui contient les informations sur les variations possibles, et de la diagonaliser. L'objectif de cette diagonalisation est d'obtenir d'une part des matrices de passage dans un espace où chaque variation se situe sur un axe, ce qui permet d'isoler les variations une à une, et d'autre part de pouvoir quantifier les variations en terme d'importance (une variation sera importante si elle apparaît souvent dans la base ; une variation faible sera probablement le fruit d'une exception ou d'une erreur), grâce aux valeurs propres. L'avantage est de pouvoir ne considérer que les variations dignes d'intérêt, c'est-à-dire celles dont les valeurs propres sont importantes. Les variations principales pour les 3 plus importantes valeurs propres sont illustrées à la Fig. 4.9. Souvent, quelques valeurs propres contiennent la plupart de l'information. Il est important pour améliorer la vitesse d'exécution de ne faire les calculs que sur les valeurs propres les plus importantes, les autres étant inutiles.

Algorithme 3 : Détermination du modèle de mains par PCA

Input :

X : Ensemble de N mains alignées et annotées.

p : Pourcentage d'informations conservées pour créer le modèle.

Output : Modèle de mains $(\bar{X}, [V_0, \dots, V_k], [v_0, \dots, v_k])$.

begin

$\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$ ou X_i est le i -ème modèle.

$C = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})(X_i - \bar{X})^T$

 Calcul des vecteurs propres $\{V_i\}$ et des valeurs propres correspondantes $\{v_i\}$ de C .

$k = \min_l \left(\sum_{i=1}^l v_i > p * \sum_{i=1}^N v_i \right)$.

end

Une fois les valeurs et vecteurs propres obtenus, on peut passer un vecteur de points dans l'espace propre en utilisant la matrice des vecteurs propres comme matrice de passage. On peut

ensuite voir grâce à la nouvelle expression du vecteur si celui-ci est proche de la moyenne ou pas. Ceci sera utilisé dans la phase de détection.

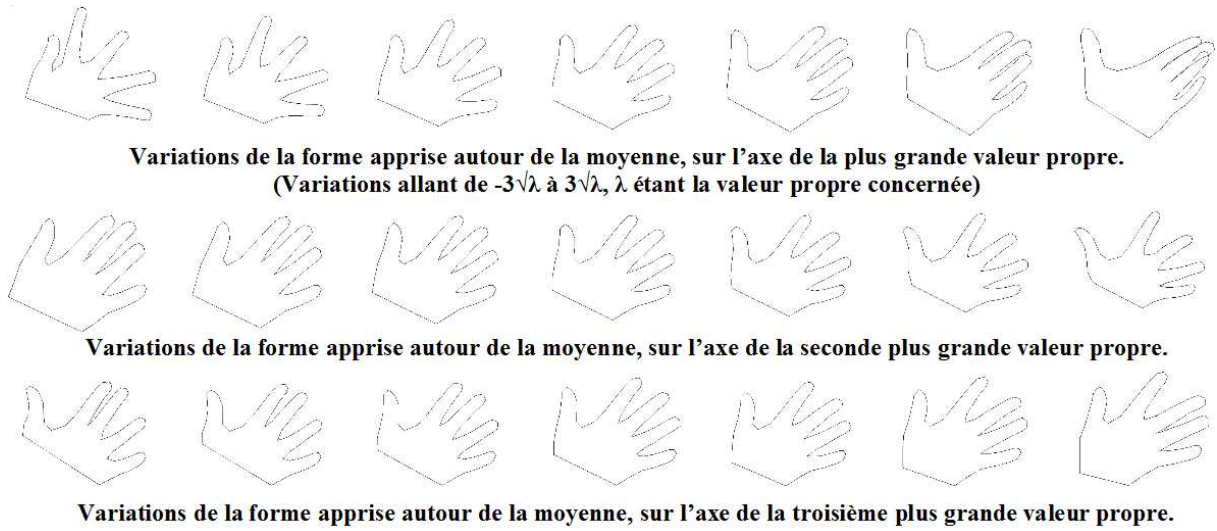


FIGURE 4.9 – Variations selon les trois principales valeurs propres

4.3.2 Recherche d'une main à partir du modèle

La phase de recherche est la phase principale de la segmentation de la main. Dans cette phase basée sur la méthode de Cootes et al. [20], en plus de l'image d'entrée et des données d'apprentissage (forme moyenne et déformations), une approximation de la segmentation est donnée. Afin de rendre ce système complètement automatique, une initialisation du modèle de forme est proposée dans le paragraphe suivant puis l'évolution de celui-ci est précisée.

4.3.2.1 Initialisation du modèle de forme

Le modèle de forme active est une méthode locale ; ainsi une mauvaise initialisation du modèle entraîne dans la plupart des cas une segmentation finale erronée ou imprécise. L'initialisation consiste à calculer les positions de départ des points caractéristiques du modèle de main sur l'image. Des techniques d'optimisation générales tels que les algorithmes génétiques peuvent être utilisées dans cet objectif. Une approche plus rapide et plus locale d'initialisation à partir de la segmentation de peau présentée dans le chapitre précédent est utilisée.

Afin d'aligner le modèle sur l'image de main segmentée, la première approche a consisté à rechercher les points de forte courbure dans l'objectif de détecter les neufs points importants utilisés pour aligner les mains dans les systèmes biométriques (Cf. section 3.7.1.1). Après expérimentations cette méthode s'est révélée très peu robuste. En effet, les contours de l'image de peau segmentée ne sont pas réguliers et ils entraînent donc de fortes courbures qui perturbent la recherche des points. Un exemple de forte courbure dû à une approximation de détection de peau est illustré à la Fig. 4.10.

Afin d'être robuste à des zones de forte courbure, un diagramme de distance entre les points du contour et un point fixe situé à la base du poignet est calculé. Ce point à la base du poignet

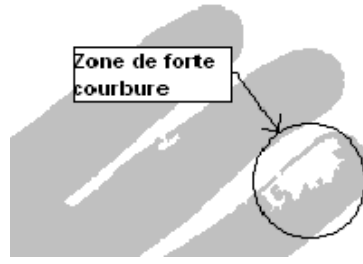


FIGURE 4.10 – Exemple de zone mal segmentée avec une forte courbure

est le barycentre des pixels de la main situé dans les N_c lignes en bas de l'images (N_c est fixé expérimentalement à 20). Les points particuliers de la main sont déterminés en détectant les cinq extréma et les quatre minima locaux de ce diagramme. Un algorithme spécifique est établi pour détecter ces points. Initialement, la dérivée pour chaque point est calculée pour un voisinage fixe. Si le nombre d'extrémas trouvés est trop important, le voisinage est augmenté afin de ne pas prendre en compte des zones mal segmentées. Si le nombre d'extrémas est trop faible, le voisinage est réduit pour détecter tous les points d'intérêt. L'algorithme s'arrête lorsque le nombre d'extrémas trouvés est égal à neuf.

Ces points sont définis pour une dérivée nulle en augmentant ou en réduisant le voisinage utilisé pour calculer la dérivée. Les points sont déterminés lorsque neuf extréma sont trouvés. Ce processus itératif permet de supprimer les dérivées nulles des petites zones qui correspondent à des zones mal segmentées.

A partir de ces points, deux points près du poignet sont ajoutés comme décrit précédemment puis les M inter-points sont fixés à une distance équidistante entre les landmarks pour représenter la main. Le processus complet de recherche est illustré à la Fig. 4.11.

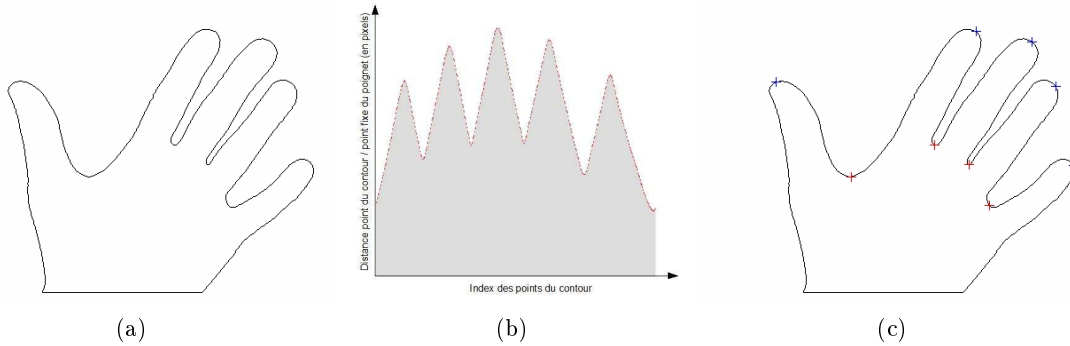


FIGURE 4.11 – Extraction automatique des landmarks (a) forme de main (b) diagramme de distance (c) points caractéristiques

4.3.2.2 Évolution du contour du modèle

A partir du modèle de distribution des points, l'évolution du modèle consiste à le déformer suivant les contours de l'image et ses contraintes. La recherche de la main est ainsi une succession d'évolutions suivant les gradients et d'applications de contraintes.

Pour détecter des contours, des méthodes répandues telles que l'algorithme de Deriche [26], le module de gradient de Sobel [100] ou l'algorithme de Di Zenzo sont utilisées. Dans le modèle de Cootes, en assumant que la forme à détecter correspond aux gradients importants de l'image, chaque point du modèle est déformé suivant la normale à la forme pour trouver les contours importants. Après cette déformation, les contraintes définies par le modèle sont appliquées afin de garder une forme correspondant à l'objet appris. On ne doit faire apparaître aucune contrainte de position, d'échelle ou de direction, c'est pourquoi, afin d'appliquer ces contraintes la forme courante est alignée sur la moyenne. Les opérations effectuées sont stockées pour pouvoir inverser le processus une fois les contraintes appliquées. Une fois l'alignement effectué, les contraintes sont appliquées sur la différence entre la moyenne et la forme. Ensuite, un vecteur dont les premières valeurs sont les projections sur les axes de variations principales et les autres sont nulles, est déterminé.

Les contraintes sont maintenant appliquées : pour chaque valeur hors de l'intervalle $[-3\sqrt{\lambda}, +3\sqrt{\lambda}]$, λ étant la valeur propre associée, la valeur est remplacée par $-3\sqrt{\lambda}$ si elle est négative, $+3\sqrt{\lambda}$ sinon. Ensuite ce vecteur est repassé dans l'espace initial, la moyenne est ajoutée et les transformations inverses de l'alignement sont effectuées. Les contraintes sont alors appliquées sur la forme. Cela revient à regarder si pour chaque variation principale, la distance entre la forme et la moyenne est bien inférieure à l'amplitude correspondant à la variation concernée. Dans le cas contraire, on modifie la forme pour qu'elle se situe dans cet intervalle et ainsi on garde un objet ressemblant au modèle appris.

Les deux processus de recherche de gradient sur la normale puis l'application des contraintes sont appliqués jusqu'à convergence, lorsque la déformation est minime. Une limitation sur le nombre d'itérations est appliquée afin de détecter une mauvaise détection ou de stopper le processus de recherche.

Dans cette évolution, le modèle présente cependant un inconvénient majeur. En effet, le modèle a tendance à diverger sans suivre le contour réel de l'objet à détecter. Cette erreur arrive lorsque l'initialisation n'est pas assez précise ou lorsque des gradients de fortes amplitudes se situent près de l'objet à rechercher. Le phénomène de divergence à cause d'une mauvaise initialisation est résolu par une définition locale du modèle initial défini dans la section précédente. Le phénomène de divergence dû à de forts gradients extérieurs est résolu en utilisant deux méthodes combinées.

Premièrement, le gradient $G(x, y)$ de l'image est déterminé en utilisant l'algorithme de Di Zenzo dans un espace de couleur de peau spécifique ce qui limite l'information de fond. L'espace couleur de peau est déterminé par ACP sur la base de pixels de peau utilisée pour l'apprentissage du modèle neuronal. Toujours afin de limiter les gradients de fond, un coefficient p correspondant à la probabilité qu'un pixel soit un pixel de peau ajuste ce gradient. Ainsi, le gradient ajusté s'exprime simplement par :

$$G_a x, y = G(x, y) \times p(x, y) \quad (4.8)$$

où $G(x, y)$ est le gradient de l'image I au point (x, y) déterminé par Di Zenzo dans l'espace couleur de peau et $p(x, y)$ est la probabilité que le pixel $I(x, y)$ soit un pixel de peau. Une illustration du calcul du gradient est présentée à la Fig. 4.12.

Pour limiter encore le phénomène d'expansion du modèle et pour que le modèle final corres-

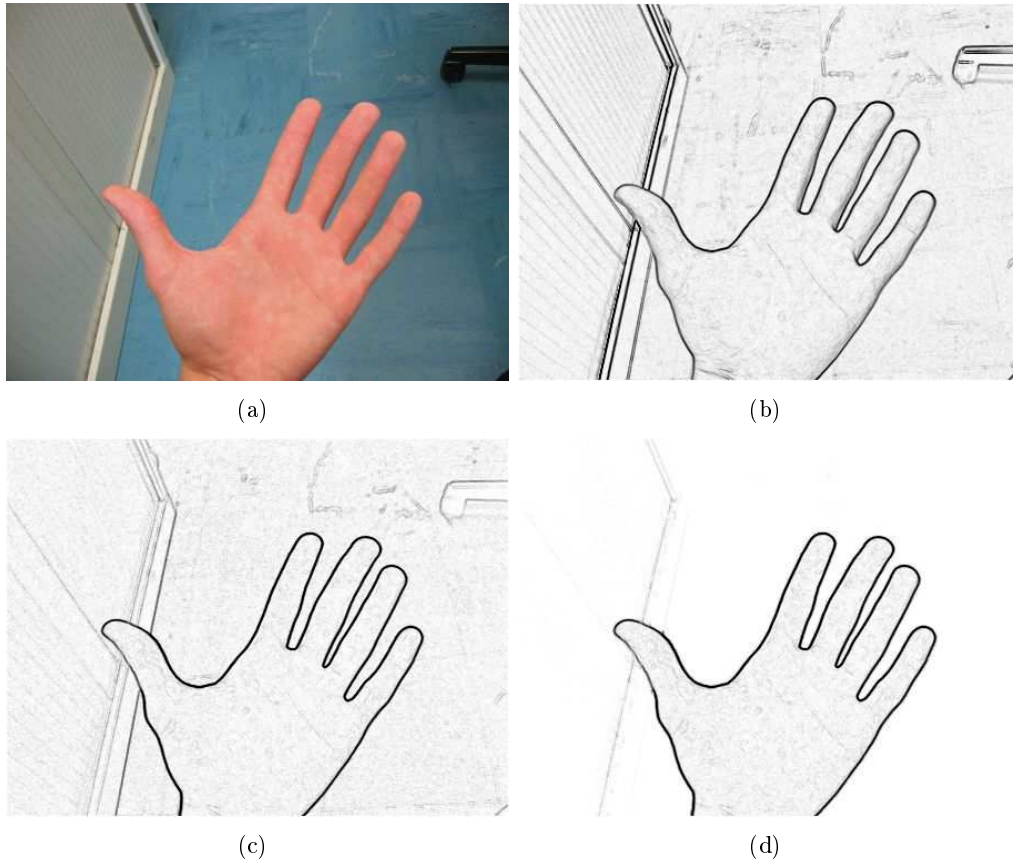


FIGURE 4.12 – Détermination du gradient (a) Image originale (b) Di Zeno dans l'espace RVB (c) Di Zeno dans l'espace de peau (d) Di Zeno ajusté

ponde au contour réel de la main, un poids $W \in [0, 1]$ limite le processus de déformation. Lorsque le poids est égal à 1 seules les contraintes du modèle sont utilisées (méthode de Cootes), lorsque le poids est égal à 0 les points sont déplacés sans contraintes suivant la normale au contour. Un poids intermédiaire permet d'obtenir une recherche proche des gradients tout en préservant un contour proche du modèle. Ce poids permet ainsi de détecter la forme précise de la main et non une forme trop dépendante du modèle. Une évolution du contour à différentes itérations du processus est présentée à la Fig. 4.13.

4.4 Évaluation de la méthode de segmentation

Dans cette section, des résultats expérimentaux sont présentés pour la détermination des différents paramètres et pour valider notre approche. Premièrement, la base d'expérimentation est décrite. Ensuite, les résultats sur la détection de la peau et l'évaluation de la qualité du processus de segmentation complet sont montrés.

4.4.1 Base de données d'évaluation

Toutes les images utilisées pour valider les différentes méthodes sont des images de mains provenant de nos acquisitions ou des images visages pour compléter notre base de données pour

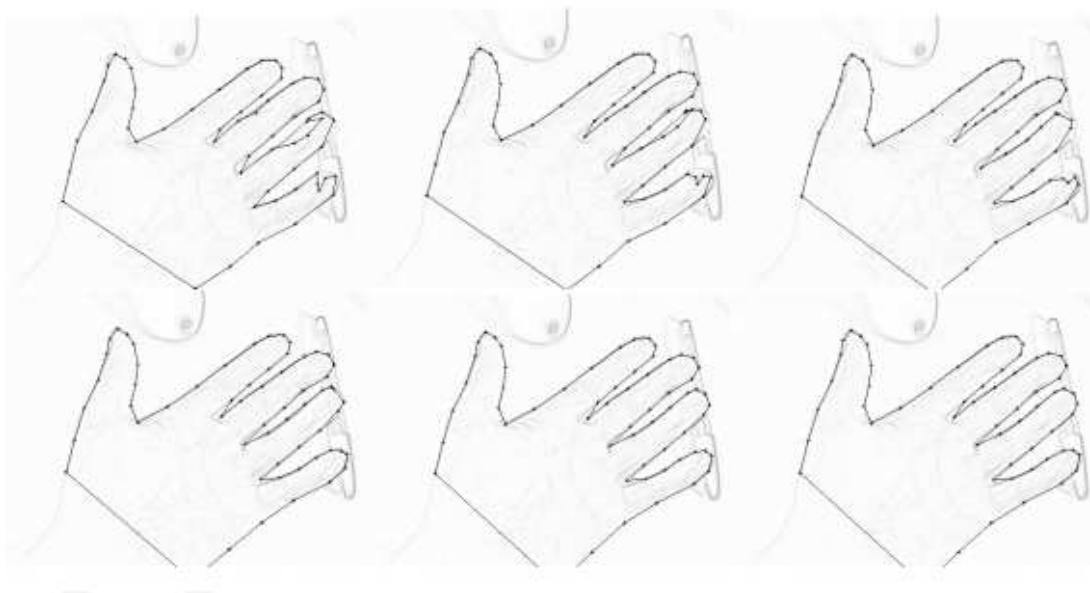


FIGURE 4.13 – Evolution du contour aux itérations 0, 5, 10, 20, 30 et 40

l'apprentissage de la couleur de la peau. Les images sont séparées en deux sous-ensembles : l'ensemble de détection de peau et l'ensemble pour détecter la main via l'algorithme.

La base de détection de peau contient des images de mains et de visages sous différentes luminosités, en intérieur, en extérieur et pour des individus de différentes couleurs de peau. Cette base est divisée en deux sous-ensembles. Le premier, correspondant à la base d'entraînement, contient 5789766 pixels dont 2189796 pixels de couleur de peau. Le second pour évaluer la détection contient 607800 pixels également distribués entre les pixels de peau et de fond.

L'ensemble ASM est composé de 260 images de taille 454*341. Il est divisée en trois sous-ensembles. Le premier contient 86 image segmentées de main pour lequel les landmarks sont précisés. Le second ensemble, permettant d'ajuster les paramètres de l'ASM, contient 87 images annotées. Le dernier ensemble contient également 87 images et permet de valider le processus complet de détection.

4.4.2 Évaluation de la détection de peau

L'estimation par Bayes, le modèle de mélange de gaussienne et l'approche par réseau de neurones sont comparés sur la base de données de pixels de peau. Les tableaux 4.1,4.2,4.3 présentent les taux d'erreurs de classification des pixels de peau sans post-traitement lorsque le taux de faux positifs et le taux de faux négatifs sont identiques. Les faux négatifs correspondent aux pixels de fond classés comme pixels de peau, les faux positifs correspondent aux pixels de peau classés comme fond.

Les méthodes par estimation de Bayes et par approximation par gaussiennes présentent des taux d'erreurs variant de 9.4% à 33.4% suivant la méthode et le domaine couleur employé. La méthode de segmentation par Bayes affiche un taux de détection optimale dans le domaine *YUV*.

Domaine	Bayes	Gaussiennes	UV	HS
<i>RVB</i>	13.6%	1	19.0%	33.4%
<i>YUV</i>	9.4%	2	19.2%	18.6%
<i>UV</i>	12.1%	3	21.1%	15.6%
<i>HSL</i>	12.9%	4	19.6%	15.7%
<i>HS</i>	11.3%	5	17.5%	15.4%
<i>TSL</i>	11.6%	6	17.0%	15.3%
<i>TS</i>	12.6%	7	16.7%	15.1%
<i>RVB_N</i>	11.9%	7	17.3%	15.6%

TABLE 4.1 – Taux d'erreur avec (a) l'estimation par Bayes et (b) la somme de gaussiennes pour la détection de peau

Pour l'approximation par gaussiennes, le plus bas taux d'erreur de 15.1% est obtenu dans le domaine *HSL* sans la composante de luminosité avec 7 gaussiennes. L'augmentation du nombre de gaussiennes est assez peu influant pour la modélisation de la peau à partir de 4 gaussiennes. On peut en déduire que la couleur de la peau peut se modéliser avec peu de paramètres. Cependant, les performances par Bayes sont bien supérieures à la modélisation par gaussiennes ce qui peut s'expliquer par le fait que l'estimation par une ou plusieurs gaussiennes n'est qu'une approximation de la vérité terrain.

Bien que de nombreuses études proposent une caractérisation de la couleur de la peau sans utiliser l'information de luminance afin d'être robuste aux changements de luminosité, nos expérimentations montrent une amélioration des performances en utilisant cette composante.

Le tableau 4.2 montre la performance de classification avec les réseaux de neurones suivant le domaine couleur et l'architecture optimale du réseau. Les résultats présentés sont les meilleures performances du système avec une couche cachée comprenant de 1 à 10 neurones avec les fonctions sigmoïdes et tangentes hyperboliques pour les colonnes 1 et 2 du tableau respectivement. La troisième colonne présente les résultats optimaux pour un réseau à deux couches cachées. La première comprend de 1 à 10 neurones sur la première couche cachée avec la fonction de transfert sigmoïde, la seconde de 1 à 10 neurones avec la fonction tangente hyperbolique. Avec les réseaux de neurones, le taux d'erreur est plus faible que par l'approximation par Bayes et le modèle par somme de Gaussiennes. La complexité de la représentation de la peau peut expliquer ce résultat. L'ajout d'une couche cachée ou de représentations couleurs à l'entrée du réseau ne change pas fondamentalement les performances de classification.

Pour compléter cette évaluation, l'approche par réseau de neurones a été testée dans l'espace couleur de peau et le classifieur SVM a été comparé aux performances du réseau de neurones. L'espace couleur de peau a été testé avec les deux premières composantes PCA_1 et les deux dernières PCA_2 . Les taux de détection sont décrits dans le tableau 4.3.

Le taux de classification dans le domaine *RVB* augmente de plus de 20% entre le réseau de neurones et les SVMs. Cependant, l'analyse en composantes principales de la couleur de peau donne un taux de reconnaissance similaire aux autres domaines sans utiliser l'information de luminosité. La classification par SVM est bien supérieure à la méthode de réseau de neurones du fait de la recherche d'optimalité.

Domaine	Sigmoïde	Tanh	Sigmoïde+Tanh
RGB	6.2%	7.8%	6.8%
YUV	6.7%	7.2%	6.4%
HSL	7.9%	8.1%	7.0%
TSL	6.8%	7.0%	6.2%
RGB+YUV	7.3%	7.3%	6.8%
RGB+TSL	6.2%	6.2%	6.3%
YUV+TSL	6.7%	6.2%	6.4%

TABLE 4.2 – Taux d’erreur de classification avec un réseau de neurones

Domaine	Sigmoïde	SVM
RVB	6.2%	4.9%
PCA_1	12.9%	12.0%
PCA_2	11.8%	11.0%

TABLE 4.3 – Taux d’erreur de classification dans l’espace RVB et l’espace couleur de peau avec un réseau de neurones et les SVMs.

Un facteur dominant dans le choix de la méthode de classification reste cependant le temps d’exécution. Le temps moyen d’exécution des deux approches principales est indiqué dans le tableau 4.4 où $P(X)$ est l’approche pyramidale appliquée à la méthode de détection X . A partir de ces résultats, l’approche par réseaux de neurones étant près de 30 fois plus rapide que l’approche par SVMs (155ms pour une image 454*341) a été utilisée pour une méthode proche du temps réel.

Domaine	RN	P(RN)	SVM	P(SVM)
RVB	2624ms	155ms	69097ms	4237ms

TABLE 4.4 – Temps de calcul moyen pour les images de taille 454 * 341.

4.4.3 Évaluation de la segmentation par contour actif de forme

Le modèle de forme est défini par le modèle de distribution de point déterminé par la première base d’ASM. Seules les 13 premières valeurs propres préservant 99% de l’information sur un total de 222 sont gardées pour modéliser la forme de la main. La figure 4.9 illustre les variations de la main dans les trois premiers axes avec b_i égal à trois fois l’écart type. Les points caractéristiques étant automatiquement situés, les points du poignet ne sont pas précisément détectés. Cependant, notre objectif étant de détecter la forme des doigts pour une reconnaissance biométrique, la précision de localisation des points du poignet n’est pas très importante. Pour une précision plus importante des points, un marquage à la main de l’ensemble d’apprentissage devrait être défini.

Toutes les constantes doivent être fixées pour évaluer la qualité de la détection du processus complet de segmentation. Afin d’évaluer la détection, la distance d’erreur est définie par l’écart

Composantes	1er	2ème	3ème	1ère et 2ème	1ère et 3ème	2ème et 3ème	Toutes
Ratio	3.46	4.22	2.41	3.59	3.33	3.55	3.49

TABLE 4.5 – Ratio entre la moyenne de l'intensité du gradient de la main et du fond

moyen de la distance minimale entre les points caractéristiques et la vérité terrain. Cette distance s'exprime par l'équation suivante :

$$e = \frac{1}{N} \sum_{i=1}^N \text{Min}(d(X_i, C)) \quad (4.9)$$

où N est le nombre total de points définissant la main. Le choix de la meilleure composante de couleur de peau est fixé par le meilleur ratio entre l'intensité du gradient déterminé par l'algorithme de Di Zenzo dans une zone proche du contour de la main et le reste de l'image. Le résultat de cette étude, décrit dans le Tableau 4.5, montre que le second plan est le plus discriminant entre toutes les combinaisons possibles.

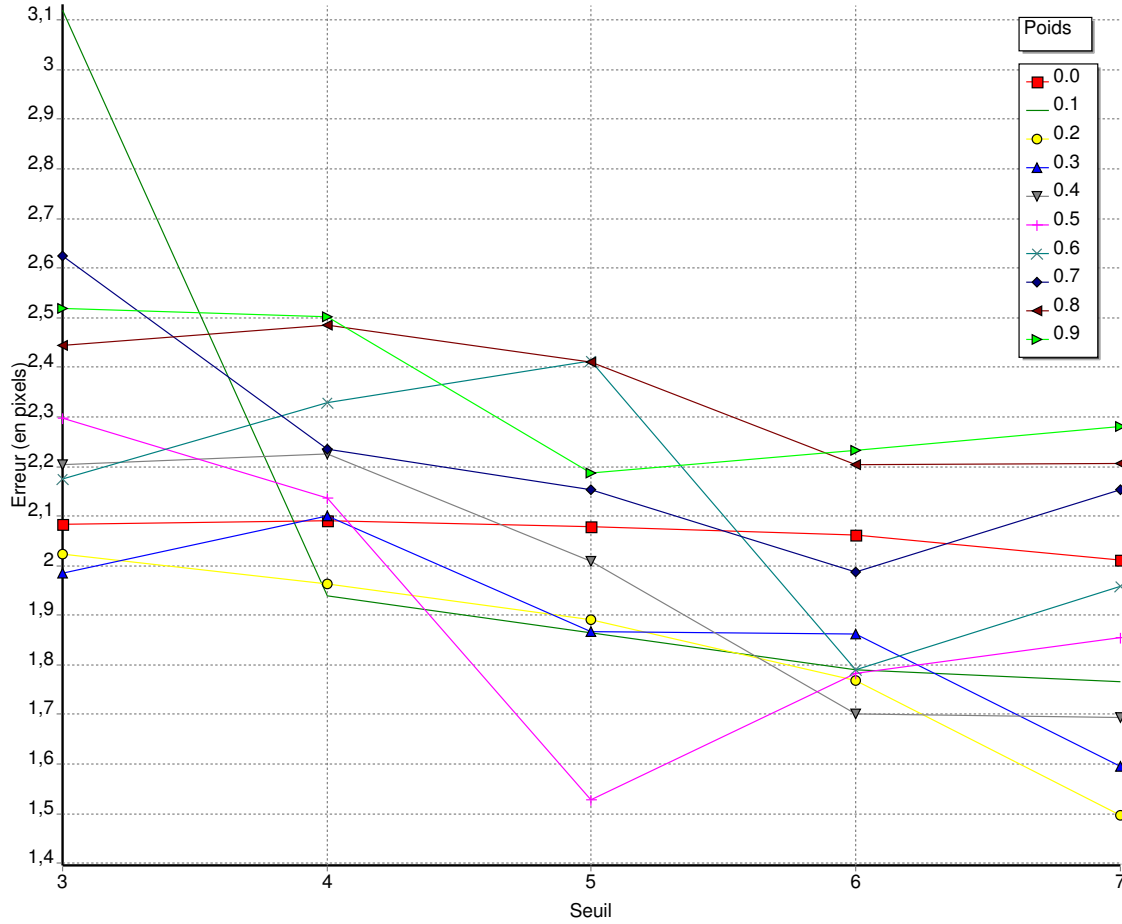
Les principaux paramètres d'évolution du contour sont le poids de déformations W du modèle limitant la déformation de la main et le seuil G définissant la valeur minimum définissant un contour *fort*. Les valeurs optimales de W et G sont définies par expérimentations avec le second plan de couleur de peau. Les résultats sont décrits suivant le taux d'erreur e et le nombre de mains non détectées dans les figures 4.14 et 4.15. Les expérimentations montrent que plus le seuil G augmente et plus la détection est précise en s'approchant des contours importants de la main mais plus il y a des échecs de détection. Un compromis doit ainsi être adopté entre W et G afin d'obtenir peu de fausses détections mais une grande précision. Les valeurs obtenues sont un seuil G de 6 pixels et un poids de déformation de 0.2. Avec ces paramètres, la distance d'erreur moyenne est de 1.8 pixels et le taux de non détection est de 3.8%.

Un autre paramètre important doit être défini pour compléter le processus : le nombre de points décrivant une main. Avec notre méthode, neuf points sont constants (les cinq extrémités de doigts et les quatre points à la vallée entre les doigts adjacents). Le nombre de points entre les points constants est fixé par expérimentations (Fig. 4.16) avec les paramètres optimaux (déterminé avec 8 points intermédiaires). Bien sûr, le taux d'erreur diminue tandis que le temps d'exécution augmente en ajoutant des points. Ainsi, le nombre de points est défini suivant les objectifs du système. Ici, le nombre de points entre les points fixes définissant le modèle est de 12.

La robustesse et la précision du modèle sont validées avec la troisième base d'évaluation qui est plus difficile que la seconde base. Le taux d'erreur obtenu est de 2.05 pixels avec un taux de mauvaises détections identiques. Ceci confirme la détermination optimale des paramètres et la robustesse de notre système de détection.

4.5 Conclusion sur la détection de la main dans une image complexe

Une nouvelle méthode d'extraction de main est présentée dans ce chapitre. La méthode combine les informations de couleur et de forme. La fusion des deux représentations implémentée

FIGURE 4.14 – Taux d’erreurs suivant les constantes W et G

permet une segmentation robuste, précise et rapide de la main.

En plus de la performance de vitesse, la méthode de segmentation de la peau par un réseau de neurones multi-échelle est comparée à d’autres approches de l’état de l’art. Cette comparaison démontre que cette approche est plus performante que les autres. Avec l’ajout de la méthode de modèle de forme actif ajusté suivant la couleur de la peau, une amélioration très sensible des performances est relevée principalement dans les cas difficiles. La figure 4.17 montre l’impact de la fusion des deux informations pour des images mal segmentées initialement.

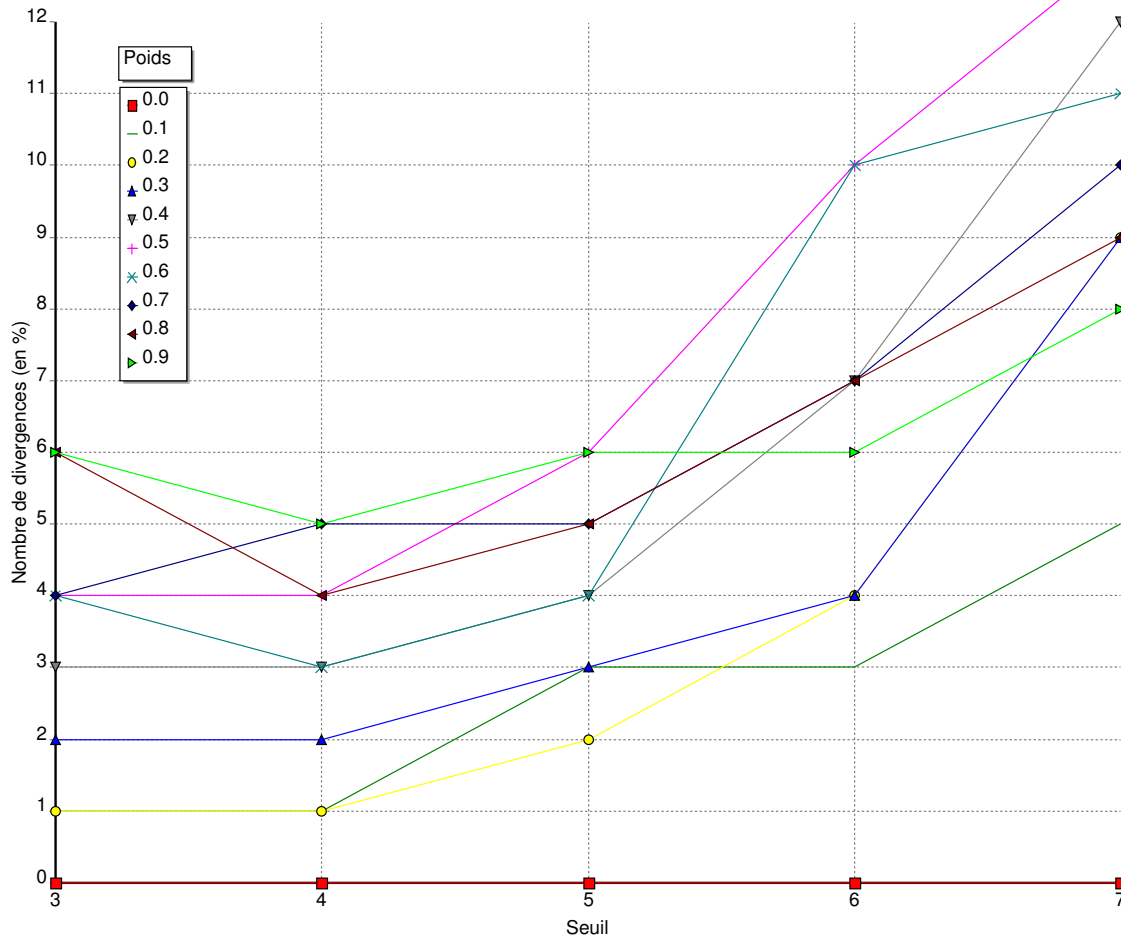


FIGURE 4.15 – Taux de divergences suivant les constantes W et G

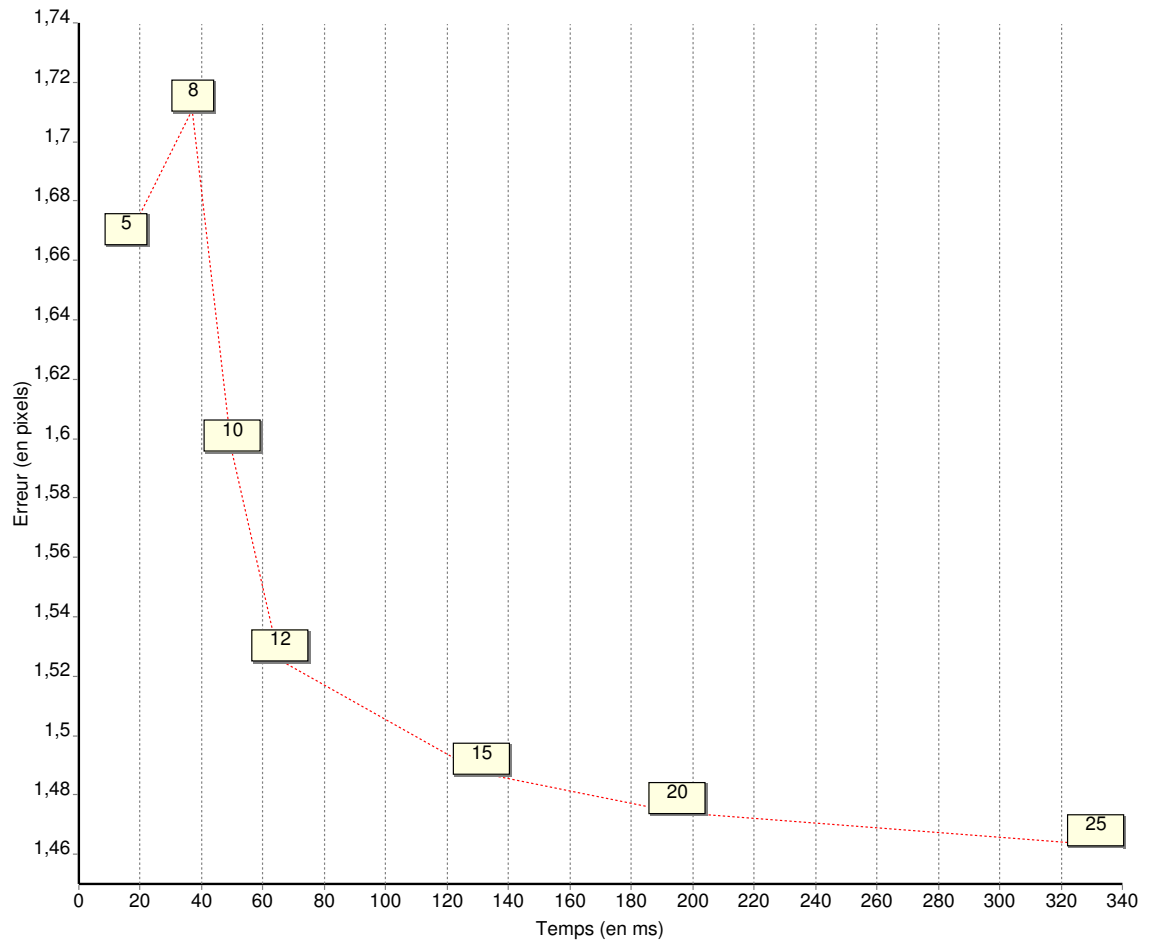


FIGURE 4.16 – Qualité de détection de main et vitesse d'exécution suivant le nombre de points caractéristiques

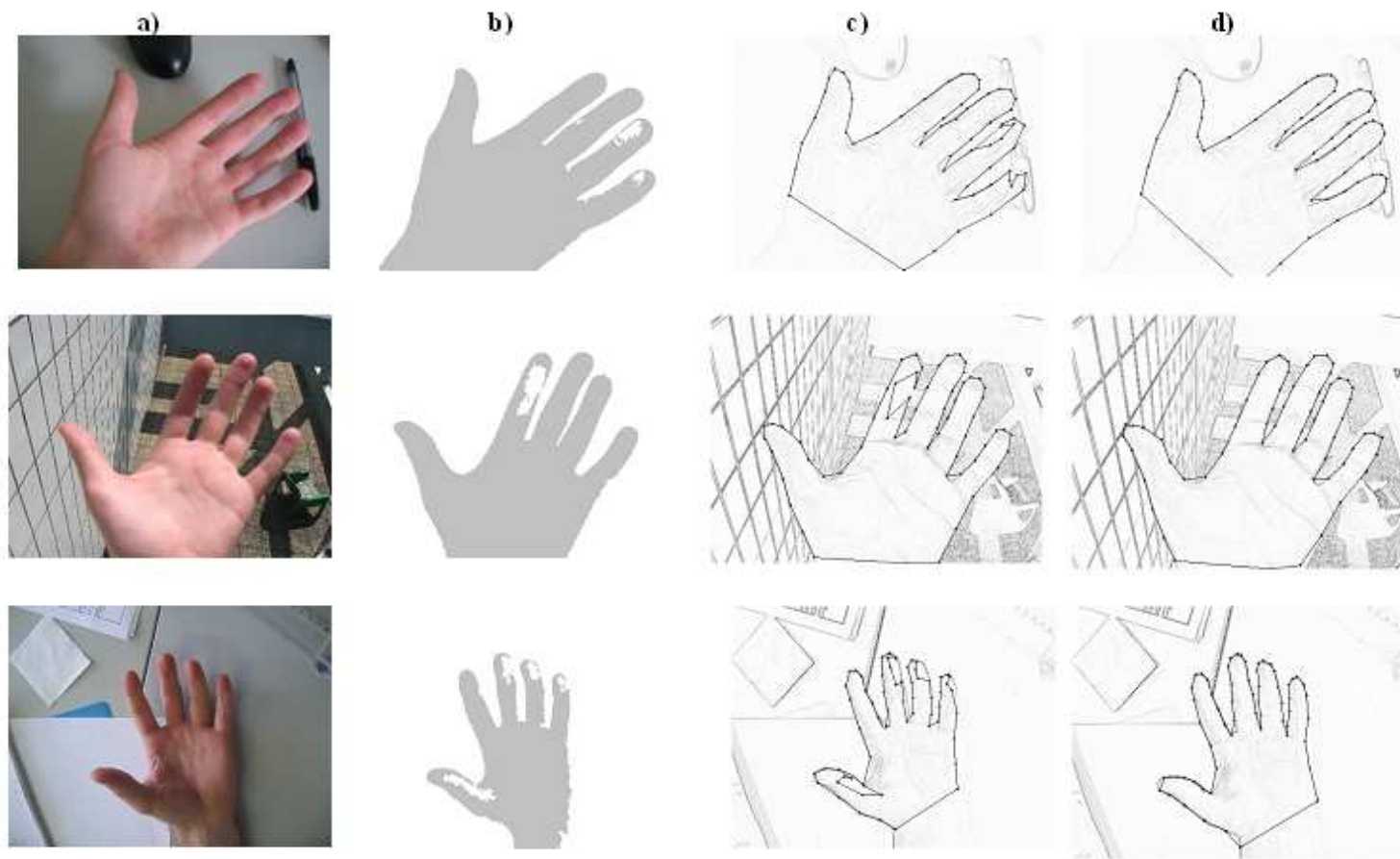


FIGURE 4.17 – Exemples de détection de la main. (a) image initiale, (b) image segmentée par la couleur, (c) initialisation du contour, (d) main segmentée.

Deuxième partie

Reconnaissance de la main

Chapitre 5

État de l'art des systèmes de reconnaissance biométrique par la main

Sommaire

5.1	Introduction à la reconnaissance de la main	75
5.2	Reconnaissance biométrique basée sur la forme de la main . . .	76
5.3	Authentification s'appuyant sur la paume de la main	78
5.3.1	Approche basée sur les lignes de la main	78
5.3.2	Approche par décomposition en sous-espaces	79
5.3.3	Approche par codage élémentaire	80
5.3.3.1	Analyse par transformée de Fourier	81
5.3.3.2	Analyse par ondelettes	83
5.3.3.3	Analyse par filtrage de Gabor	86
5.3.4	Approche par apprentissage	88
5.3.5	Autres approches	89
5.4	Multi-biométrie par fusion de méthodologies	90
5.5	Comparaison des différentes approches	91
5.6	Conclusion sur la reconnaissance par la main	93

5.1 Introduction à la reconnaissance de la main

La reconnaissance par la main est utilisée depuis plus d'une dizaine d'années au niveau commercial. La reconnaissance utilisée dans un premier temps s'appuie uniquement sur la reconnaissance de la forme de la main qui est supposée unique et stable pour chaque individu. Il y a deux types de système pour la reconnaissance de la forme, l'approche par composantes géométriques et l'approche par analyse en composantes principales.

Des méthodes plus robustes ont été élaborées ces dernières années. Elles reposent sur une analyse de la paume par extraction des lignes ou de la texture de la main. Contrairement à la forme de la main, les lignes de la paume ne sont pas dépendantes de la génétique. Ainsi, deux jumeaux homozygotes possèdent des paumes différentes [57].

De nombreuses méthodologies sont employées dans la littérature pour identifier une paume. Initialement, la recherche s'est portée sur des approches en haute résolution (supérieure à 500

dpi). A cette résolution, les lignes principales ainsi que les plis, les rides, les points singuliers et les *minutie* sont identifiés pour permettre la reconnaissance. Plus récemment, les recherches se sont focalisées sur les approches à faible résolution pour des applications civiles ou commerciales. Le chapitre présent expose un état de l'art des principales approches de biométrie de la main.

5.2 Reconnaissance biométrique basée sur la forme de la main

La reconnaissance par la forme de la main présente l'avantage d'utiliser un capteur de faible résolution et une très grande résistance aux facteurs extérieurs (coupures ou traces sur la main, humidité, luminosité...). Cette méthode est ainsi utilisée dans des conditions extérieures difficiles comme pour l'identification dans des zones portuaires et aéroportuaires.

Dans les systèmes commercialisés, l'image de la main est acquise à l'aide d'un capteur possédant des guides qui indiquent à l'utilisateur comment il doit positionner la main. Ce capteur est dans un espace semi clos avec une luminosité constante ce qui permet d'obtenir un fort contraste entre la main et le fond. Les guides étant fixes, 16 axes sont prédéterminés afin de mesurer les caractéristiques de la main dans le système de Jain et al.[53]. L'axe F1 permet par exemple de déterminer la largeur du pouce dans la figure 5.2 illustrant ce type d'acquisition.

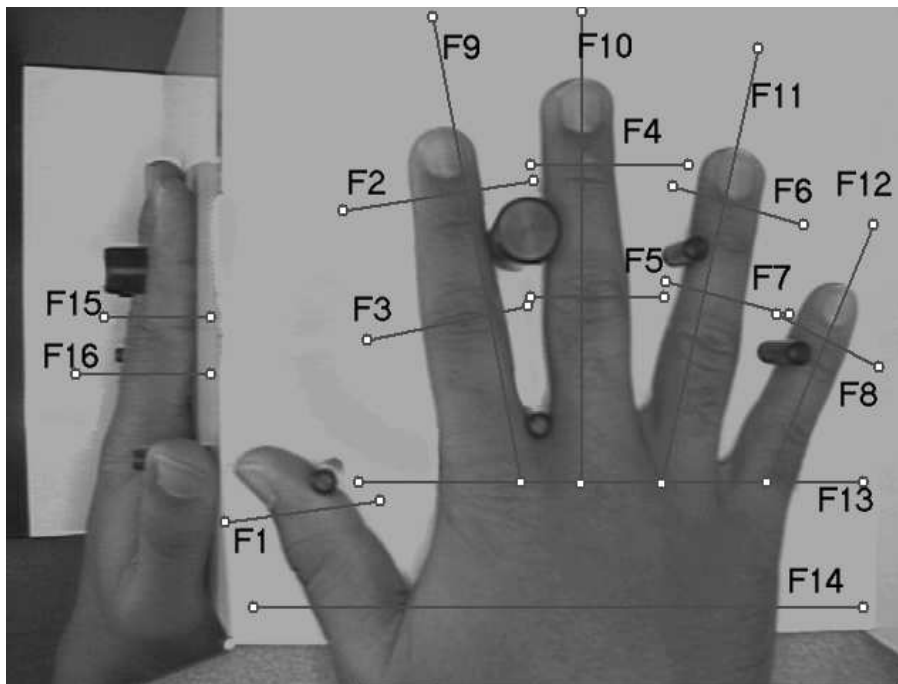


FIGURE 5.1 – Caractérisation de la main par axes dans le système de Jain et al. [53]

En ignorant l'effet de la lumière, de la couleur de la peau et du bruit, une séquence de pixels le long d'un axe a un profil idéal en forme de créneau. Soient P_s et P_e les points frontières entre les doigts ou la paume et le fond. Afin de déterminer P_s et P_e dans le profil réel deux suppositions sont effectuées sur la distribution. Tout d'abord, le modèle du profil observé est obtenu à partir du profil idéal par l'ajout d'un bruit Gaussien ; par exemple, le niveau de gris x d'un pixel entre P_s et P_e est supposé suivre la distribution :

$$G(x, B, \sigma_B^2) = \frac{1}{\sqrt{2\pi\sigma_B^2}} \exp\left(\frac{-1}{2\sigma_B^2}(x - B)^2\right) \quad (5.1)$$

Où σ_B^2 est la variance de x dans l'intervalle $R =]P_s, P_e]$, B le niveau de gris dans le même intervalle.

En plus, il est supposé que le niveau de gris d'un pixel est indépendant du niveau de gris des autres pixels de la ligne. Ces deux hypothèses permettent d'écrire la valeur de la distribution des niveaux de gris, le long d'un axe de longueur L pixels. Les paramètres ainsi obtenus permettent de décrire une main par 16 longueurs.

La comparaison entre deux vecteurs caractéristiques de la forme de la main $F = (f_1, f_2, \dots, f_{16})$ et $G = (g_1, g_2, \dots, g_{16})$ correspond à une distance euclidienne pondérée par la variance afin de rendre plus importante les valeurs stables. Deux mains sont ainsi déclarées comme provenant du même individu si :

$$\sqrt{\sum_{i=1}^{16} \frac{(g_i - f_i)^2}{\sigma_i^2}} < S \quad (5.2)$$

où σ_i^2 est la variance du i -ième élément et S un seuil prédéfini.

D'autres méthodes utilisant des guides s'appuient sur une recherche des contours de la main, à partir de l'image saisie, afin d'obtenir le profil idéal. La difficulté est ici d'effectuer une bonne segmentation de la main pour déterminer les points importants.

Les systèmes de capture opèrent dans un milieu semi fermé. Une segmentation basée sur la définition d'un seuil fixe sur l'histogramme bimodal des niveaux de gris est le plus souvent utilisée. Afin d'avoir un effet de contre-jour et faciliter la binarisation, certains systèmes effectuent une illumination sous la main. D'autres systèmes utilisent en plus un miroir lors de la capture de la main afin de déterminer les caractéristiques de hauteur de celle-ci à différents niveaux. De plus, la plupart du temps, des notions de surface pour les doigts et la main sont déterminées afin d'améliorer la représentation et donc les qualités du système en terme de reconnaissance.

De nouveaux prototypes ont été développés récemment pour supprimer les guides lors de la saisie de la main [65]. Sans guide, la main doit être alignée et les axes la caractérisant définis par rapport à l'acquisition. Ces systèmes reposent sur la recherche des points entre les doigts et des points au bout des doigts. Après détection de ces points, l'alignement de l'image s'effectue suivant l'axe du majeur ou la droite passant par les points entre l'index et le majeur et entre l'annulaire et l'auriculaire.

Contrairement aux méthodes avec guides, la position de la main n'est pas importante ce qui évite les erreurs de placement. Le système est par conséquent plus facile à utiliser. Les caractéristiques de la main s'appuient comme précédemment sur des calculs de largeurs et de longueurs des doigts. Dans la méthode de Kumar et al.[65], la reconnaissance est effectuée par une analyse hiérarchique des composantes de la main. Les caractéristiques de la main sont divisées en deux groupes. Le premier est constitué des longueurs et des largeurs de tous les doigts qui sont modélisés par un ensemble de Gaussiennes, le second groupe est constitué par l'ensemble des

régions au bout des doigts. Ces régions sont définies par les pixels du contour de chaque bout de doigt. La comparaison s'effectue en deux étapes. Dans un premier temps, le premier groupe de test est projeté par l'ensemble de Gaussiennes. Si ce groupe est proche de la référence, la seconde étape de comparaison est appliquée. La seconde étape est une comparaison par distance euclidienne entre chaque point du contour des bouts de doigt. Si trop de points des ensembles ne correspondent pas, l'image est rejetée sinon elle est reconnue comme valide.

D'autres approches modélisant la forme complète des doigts ont été présentées par Xiong et al. [129], Oden et al. [80] et Yörük et al. [132]. Les deux premières approches utilisent une modélisation des contours de la main par approximation polynomiale. Dans l'approche présentée par Yörük et al., la silhouette de la main est normalisée afin que les doigts soient toujours présentés selon la même pose. Après cette normalisation, une approche par distance de Hausdorff et une autre par analyse en composantes indépendantes sont proposées pour comparer les silhouettes.

Tous ces procédés se basant sur la forme de la main présentent l'avantage d'être, comme l'iris, très résistantes aux événements externes (coupure, problème de peau, ride, ...) contrairement aux empreintes digitales par exemple. Les performances de ces systèmes sont acceptables dans le cadre d'une recherche de sécurité moyenne et leur acceptabilité par le public en font le deuxième secteur biométrique en parts de marché. La forme de la main présente néanmoins l'inconvénient de ne pas être invariable au cours du temps (surtout à cause du vieillissement et de la perte ou prise de poids) ce qui oblige des mises à jour fréquentes des caractéristiques de chaque individu. Ceci peut également être un avantage pour des raisons de sécurité. La donnée biométrique se "périme" au cours du temps et elle devient inutilisable si les usagers ne souhaitent plus utiliser le système. Une autre méthode possédant sensiblement les mêmes qualités et défauts est la reconnaissance par la paume de la main. Les performances en terme de reconnaissance de ces systèmes sont cependant supérieures.

5.3 Authentification s'appuyant sur la paume de la main

La reconnaissance par la paume de la main est une méthode biométrique récente et n'est que très peu utilisée dans un contexte commercial pour le moment. Ces approches ont dans un premier temps utilisé les mêmes méthodes que la reconnaissance par empreintes digitales (extraction de *minutae*, de *cores* et de *deltas* [34]) dans un contexte *offline* avec des images à haute résolution. Le contexte *offline* est défini par une acquisition et une vérification sur deux systèmes différents. Pour les empreintes digitales, un système *offline* consiste en une acquisition des empreintes par encrage des doigts sur une feuille blanche puis une reconnaissance après numérisation de cette dernière.

Récemment, de nombreux systèmes proposent des méthodologies de reconnaissance sur des images à basse résolution et dans un contexte *online* (acquisition et reconnaissance à partir du même système) plus propice aux utilisations commerciales de la biométrie.

5.3.1 Approche basée sur les lignes de la main

Historiquement, les premiers systèmes emploient une technologie basée sur les empreintes digitales. Afin d'utiliser ces méthodes, une acquisition *offline* à haute résolution est nécessaire pour obtenir des images de qualité suffisante pour extraire les *minutae*, les rides et les lignes de

la paume de la main.

Pour des approches *online*, des approches à basse résolution sont préférées afin d'obtenir une reconnaissance fiable en un temps raisonnable. Le premier système décrit par Shu et al [98] extrait les lignes de la main par la méthode des filtres de Stack dans quatre directions (horizontale, verticale et les deux diagonales). Contrairement aux méthodes d'empreintes digitales qui sont basées sur une analyse par points, cette méthode effectue une vérification par comparaison de lignes. Les lignes une fois extraites sont normalisées dans un repère défini par les lignes principales. Elles sont ensuite comparées avec une méthode *ad-hoc* basée sur une comparaison des centres et des orientations de chaque segment.

Une autre approche proposée par Han et al. [47] s'appuie sur une extraction des lignes de la main par les masques de Sobel. L'image est segmentée par seuillage puis les lignes principales sont extraites. Les lignes sont caractérisées par leurs points d'intersection, leurs extrémités et leurs tangentes. Une autre approche s'appuyant sur une extraction analogue aux filtres de Sobel est présentée par Wu et al. [125]. Après cette extraction des lignes, une précision plus importante est obtenue par un algorithme de recherche de continuité des lignes. Cette recherche s'appuie classiquement sur le raccordement des segments proches dont l'inclinaison est similaire. Un an plus tard, Wu et al. [126] ont présenté une autre approche sur la caractérisation des lignes. Les lignes sont détectées par traitements morphologiques et filtrages orientés. Les lignes extraites selon l'orientation des filtres sont binarisées. La comparaison entre deux paumes consiste à comparer bit à bit les images contenant uniquement les lignes de la main dans chaque direction.

Li et al. [71] s'appuient sur une extraction des lignes par la méthode de Hough. Chaque ligne est ensuite représentée dans un espace paramétrique suivant une origine, un angle et une dimension. A partir de cette information, une comparaison globale est effectuée par distance Euclidienne entre les différents paramètres. Afin d'améliorer la reconnaissance, une seconde étape utilise une comparaison par distance de Hausdorff.

5.3.2 Approche par décomposition en sous-espaces

Les approches par décomposition en sous-espaces sont classiquement utilisées pour modéliser un objet. Plusieurs décompositions linéaires ou non linéaires d'un ensemble peuvent être appliquées.

L'analyse par composantes principales (*ACP*) utilisée dans notre approche par modélisation de la peau et de la main est la base de plusieurs articles sur la reconnaissance palmaire. Inspirés par une approche similaire pour la reconnaissance du visage [114], Lu et al. [73] décomposent une base d'empreintes palmaires en une petite base d'un ensemble caractéristique nommé *eigenpalms* (Cf. fig 5.2). Cette base réduite est l'ensemble des vecteurs propres de la base d'apprentissage et représente ainsi les composantes principales d'une empreinte palmaire. Les caractéristiques d'une paume sont extraites en projetant la paume dans l'espace *eigenpalms*. Ces données sont ensuite comparées par distance Euclidienne.

La même équipe provenant de l'école polytechnique de Hong-Kong menée par le professeur David Zhang¹ propose également une méthode s'appuyant sur la discrimination linéaire de Fi-

1. <http://www4.comp.polyu.edu.hk/~biometrics/>

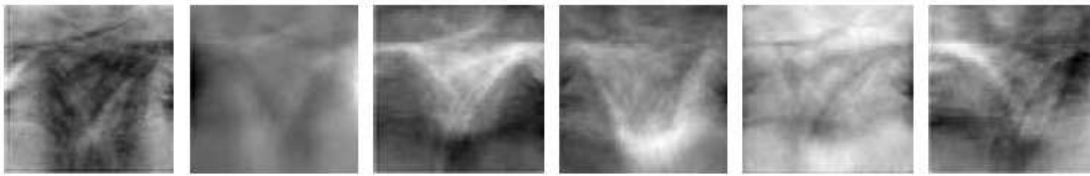


FIGURE 5.2 – *Eigenpalms* : Six vecteurs propres avec les six valeurs propres les plus importantes [19]

sher. Cette méthode de Wu et al. [123] maximise la séparation entre plusieurs classes en effectuant une projection linéaire des données. Dans le cas de la paume, l'espace de séparation est nommé *Fisherpalms*. Dans la phase de reconnaissance, la paume d'entrée est projetée dans le *Fisherpalms* puis comparée comme dans la méthode précédente par distance Euclidienne.

Durant l'année 2004, Connie et al. [19] ont comparé les approches par *ACP* et par analyse en composantes indépendantes (*ACI*). Cette analyse repose sur une décomposition d'un signal complexe en un ensemble indépendant de signaux linéaires (Cf. fig 5.3). Comme précédemment, les paumes sont projetées dans l'espace calculé puis plusieurs méthodes de comparaisons sont évaluées :

- distance Euclidienne,
- mesure Cosinus,
- réseau de neurones (Cf. 5.3.4).



FIGURE 5.3 – Six vecteurs de bases de l'*ACI* [19]

Plus récemment, Kumar et Negi [66] ont démontré qu'une utilisation de l'information de voisinage permet un gain important en reconnaissance. L'information locale n'est pas utilisée dans les méthodes précédentes puisque les images sont modélisées sous forme d'un vecteur. Afin de compléter l'approche, les caractéristiques des images de paume sont extraites par multi-résolution puis modélisée par une matrice. L'approche de l'article de Lu et al. [73] citée précédemment (extraction par *ACP* et comparaison par distance Euclidienne) est utilisée avec cette représentation en deux dimensions.

5.3.3 Approche par codage élémentaire

Dans ce paragraphe, les caractéristiques de la main sont extraites à partir de l'image de paume dans le domaine fréquentiel ou le domaine temps/fréquence. La première méthode d'extraction dans le domaine fréquentiel utilise la transformée de Fourier, la seconde une signature par ondelettes et la troisième approche le filtrage par Gabor.

5.3.3.1 Analyse par transformée de Fourier

La transformée de Fourier est l'une des plus utilisées dans le traitement d'image. Elle est principalement utilisée dans le rehaussement d'image et l'extraction de caractéristiques. La transformée de Fourier discrète permettant le passage du domaine spatial au domaine fréquentiel peut être définie comme :

$$F(u, v) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) \exp[-j2\pi(\frac{mu}{M} + \frac{nv}{N})] \quad (5.3)$$

où $f(m, n)$ est une image de taille $M \times N$, $j = \sqrt{-1}$, $u = 0, 1, \dots, M-1$, $v = 0, 1, \dots, N-1$. La transformée de Fourier inverse est définie par :

$$f(x, y) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} F(m, n) \exp[-j2\pi(\frac{mx}{M} + \frac{ny}{N})] \quad (5.4)$$

Pour l'identification de la paume, la transformée de Fourier peut être utilisée pour l'extraction des caractéristiques puisqu'il y a des correspondances entre les caractéristiques de la paume dans le domaine spatial et le domaine fréquentiel [120].

Plus les lignes principales sont importantes et plus l'information dans le domaine fréquentiel est diffuse. De plus, les principales lignes dans le domaine spatial sont marquées par une information importante dans la direction perpendiculaire aux lignes principales dans le domaine fréquentiel. A titre de comparaison, la figure 5.4 montre trois groupes de paume et leurs images fréquentielles correspondantes pour la même paume, des paumes similaires et des paumes différentes.

Afin de comparer différentes paumes, les images projetées dans le domaine fréquentiel sont exprimées en coordonnées polaires. L'image est convertie dans le système de coordonnées polaires par :

$$I'(r, \theta) = I(64 + r \cos \theta, 64 + r \sin \theta), \quad (0 \leq r \leq 64, 0 \leq \theta \leq \pi) \quad (5.5)$$

où I est l'image dans le système cartésien et I' est l'image dans le système polaire. Dans le but de représenter l'intensité des lignes principales, l'image dans le domaine fréquentiel est divisée en petites parties par une série d'anneaux ayant le même centre comme montré à la Fig. 5.5. L'énergie de chaque anneau est définie par :

$$R_i = \sum_{\theta=0}^{\pi} \sum_{r=8(i-1)}^{8i} I'(r, \theta), \quad (i = 1, 2, \dots, 8) \quad (5.6)$$

où I' est l'image dans le système de coordonnées polaires et R_i est la caractéristique du i -ème anneau.

Afin de représenter les lignes principales, l'image dans le domaine fréquentiel est divisée par une série de lignes partant du centre de l'image comme montré dans la Fig. 5.5. L'énergie de chaque éventail est défini par :

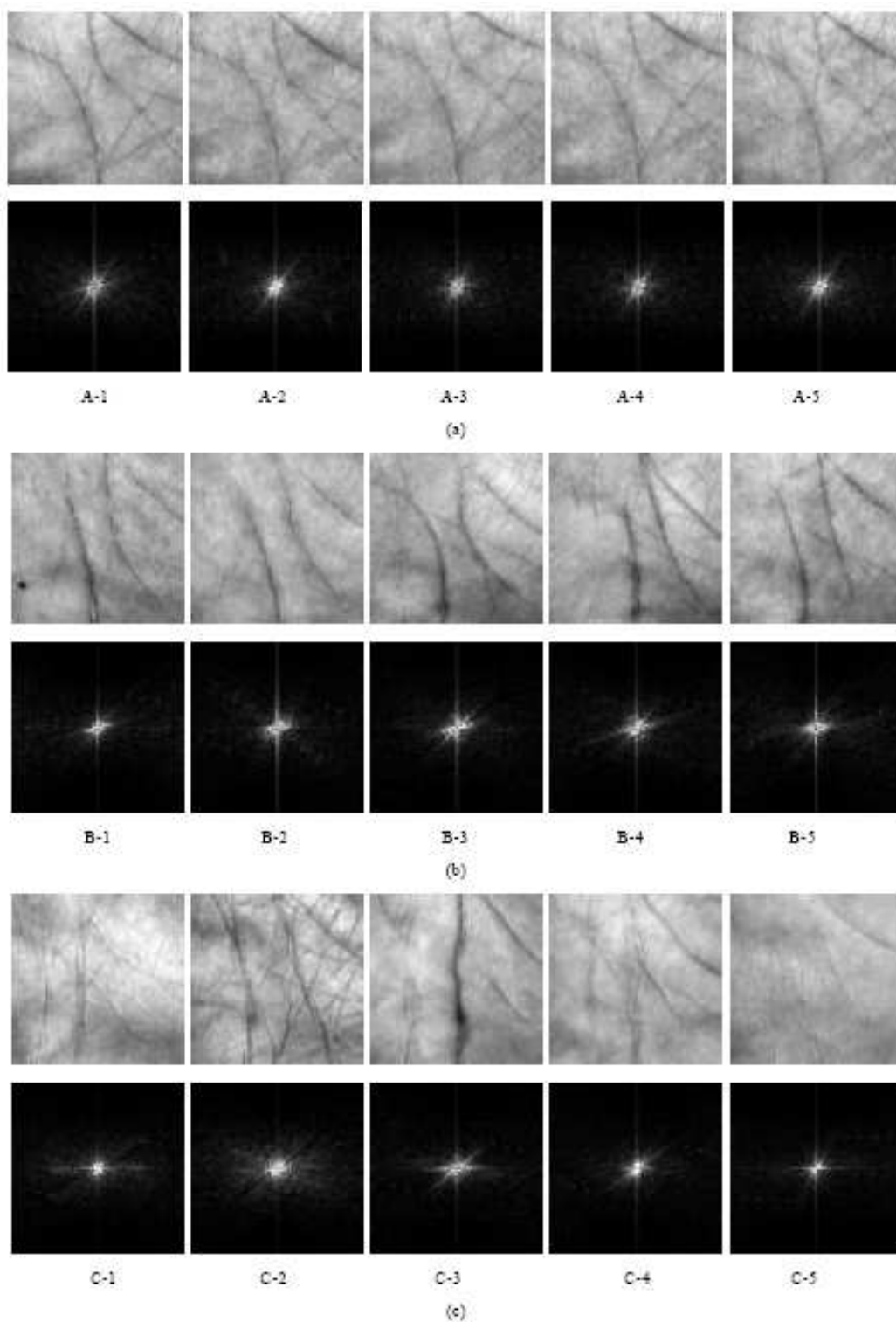


FIGURE 5.4 – Exemples de représentation dans le domaine fréquentiel : (a) Exemples provenant de la même paume, (b) Exemples provenant de paumes similaires, (c) Exemples provenant de paumes différentes.

$$\theta_i = \sum_{\theta=(i-1)}^i \sum_{r=0}^{64} I'(r, \theta\pi/8), \quad (i = 1, 2, \dots, 8) \quad (5.7)$$

où θ_i sont les caractéristiques en θ .

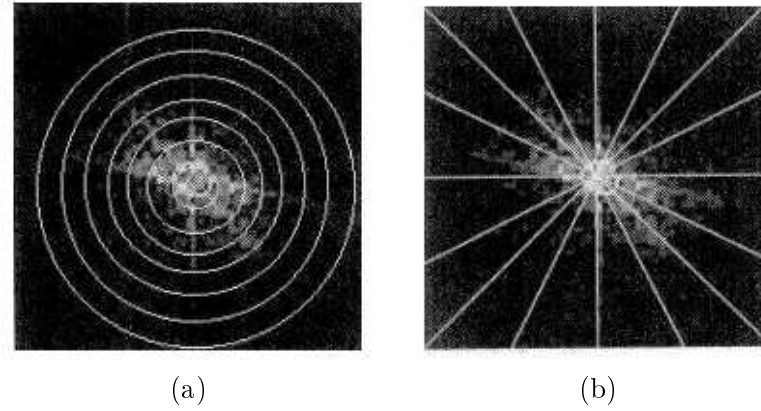


FIGURE 5.5 – Représentation des caractéristiques de la paume : (a) caractéristiques en R et (b) caractéristiques en θ .

Le calcul permettant de différencier deux paumes représentées par les caractéristiques angulaires et de rayons correspond à la combinaison des distances de ces caractéristiques.

Si $RX_i (i = 1, 2, \dots, 8)$ et $RY_i (i = 1, 2, \dots, 8)$ représentent deux caractéristiques R , la distance DR_{xy} entre RX_i et RY_i est définie par :

$$DR_{xy} = \frac{1}{8} \sum_{i=1}^8 |RX_i - RY_i| \quad (5.8)$$

De même si $\theta X_i (i = 1, 2, \dots, 8)$ et $\theta Y_i (i = 1, 2, \dots, 8)$ représentent deux caractéristiques θ , la distance $D\theta_{xy}$ est définie par :

$$D\theta_{xy} = (1 - \frac{l_{xy}l_{xy}}{l_{xx}l_{yy}}) \times 100 \quad (5.9)$$

$$\text{où} \quad l_{xx} = \sum_{i=1}^8 (\theta X_i - \frac{1}{8} \sum_{j=1}^8 \theta X_j)^2, \quad (5.10)$$

$$l_{yy} = \sum_{i=1}^8 (\theta Y_i - \frac{1}{8} \sum_{j=1}^8 \theta Y_j)^2, \quad (5.11)$$

$$\text{et} \quad l_{xy} = \sum_{i=1}^8 (\theta X_i - \frac{1}{8} \sum_{j=1}^8 \theta X_j)(\theta Y_i - \frac{1}{8} \sum_{j=1}^8 \theta Y_j). \quad (5.12)$$

A partir de ces deux distances, deux seuils δ_θ et δ_R correspondant aux seuils suivant θ et R respectivement, sont utilisés pour valider ou non l'identité d'un utilisateur.

5.3.3.2 Analyse par ondelettes

Les ondelettes sont des outils performant pour l'analyse multi-résolution. En deux dimensions, elles permettent de décomposer une image en plusieurs directions à différentes résolutions comme décrit en détail par Stéphane Mallat [76].

Pour une décomposition, deux ondelettes qui sont respectivement les parties dérivées le long de x et de y d'une fonction $\theta(x, y)$ sont définies par :

$$\psi^1(x, y) = \partial\theta(x, y)/\partial x \quad (5.13)$$

$$\psi^2(x, y) = \partial\theta(x, y)/\partial y \quad (5.14)$$

Ainsi, pour une fonction $f(x, y) : \mathbb{R}^2 \rightarrow \mathbb{R}$ la transformée en ondelettes a deux composantes $W^1 f(k, x, y)$ et $W^2 f(k, x, y)$ définies par :

$$\text{Soit } \psi_k^1(x, y) = (1/k^2)\psi^1(x/k, y/k) \quad (5.15)$$

$$\text{et } \psi_k^2(x, y) = (1/k^2)\psi^2(x/k, y/k) \quad (5.16)$$

$$W^1 f(k, x, y) = f(x, y) * \psi_k^1(x, y) \quad (5.17)$$

$$W^2 f(k, x, y) = f(x, y) * \psi_k^2(x, y) \quad (5.18)$$

Où $*$ est le symbole de convolution et k le niveau de décomposition. La définition du k -ième niveau de décomposition est illustrée au schéma de la Fig.5.6 où A_k est la composante basse fréquence de A_{k-1} , H_k , V_k et D_k sont les hautes fréquences horizontale, verticale et diagonale du k -ième niveau de décomposition respectivement. Le symbole $\downarrow 2$ indique que la résolution de l'image est divisée par deux horizontalement puis verticalement dans le sens gauche $\rightarrow$ droite du schéma. Initialement, A_0 est définie par l'image d'origine (la paume). Après décomposition au k -ième niveau l'image originale I est représentée par $3k + 1$ sous-images :

$$\left[A_k, \{H_i, V_i, D_i\}_{i=1, \dots, k} \right] \quad (5.19)$$

La figure 5.7 illustre la décomposition en ondelettes d'une paume au niveau 1 et au niveau 2.

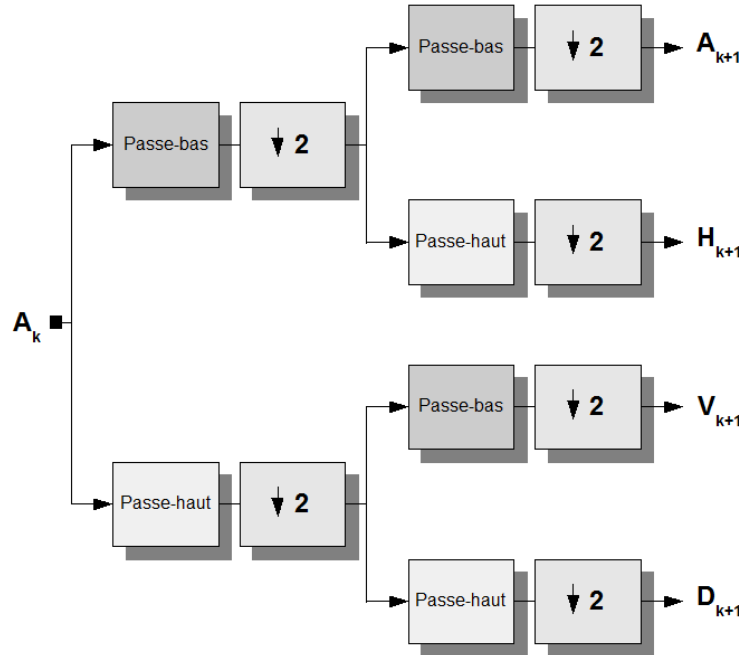


FIGURE 5.6 – Schéma de décomposition en ondelettes au niveau 1

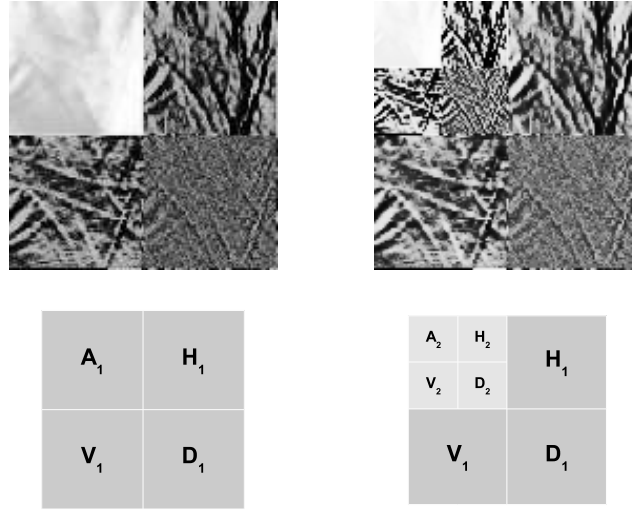


FIGURE 5.7 – Décomposition en ondelettes de Haar au premier et au second niveau

A partir de cette décomposition, différentes méthodes de comparaison sont abordées. Wu et al. [124] effectuent une comparaison par distance Euclidienne des énergies des imagerie transformées par ondelettes. Chaque imagerie est divisée en N blocs sans recouvrement et les énergies sont calculées sur chacun des blocs de chaque imagerie. La différence entre deux paumes correspond ainsi à la somme de la différence des énergies des blocs correspondants. Cette méthode est évaluée jusqu'à six niveaux de décomposition avec trois types d'ondelettes couramment utilisées :

- l'ondelette de Haar (Fig. 5.7)
- l'ondelette de Daubechies
- l'ondelette orthonormale Symmlet

Poon et al. [88] utilisent une approche similaire. Ils présentent une caractérisation de la paume par analyse des énergies après transformation par ondelettes. Les énergies sont calculées dans des zones ellipsoïdales sur la paume d'entrée. Lu et al. [72] utilisent comme décrit précédemment une reconnaissance par analyse en composantes indépendantes pour obtenir une référence puis effectuent une projection dans cet espace pour comparer une paume aux références.

Zhang et Zhang [138] proposent également une méthode s'appuyant sur une analyse par ondelettes. Pour la reconnaissance, une cascade hiérarchique de méthodes de comparaison est utilisée. Premièrement, les paumes sont différenciées en huit classes selon l'orientation des imagerie. Ensuite, quatre propriétés des imagerie sont extraites :

1. Centre de gravité des imagerie,
2. Densité des imagerie,
3. Disparité spatiale,
4. Energie.

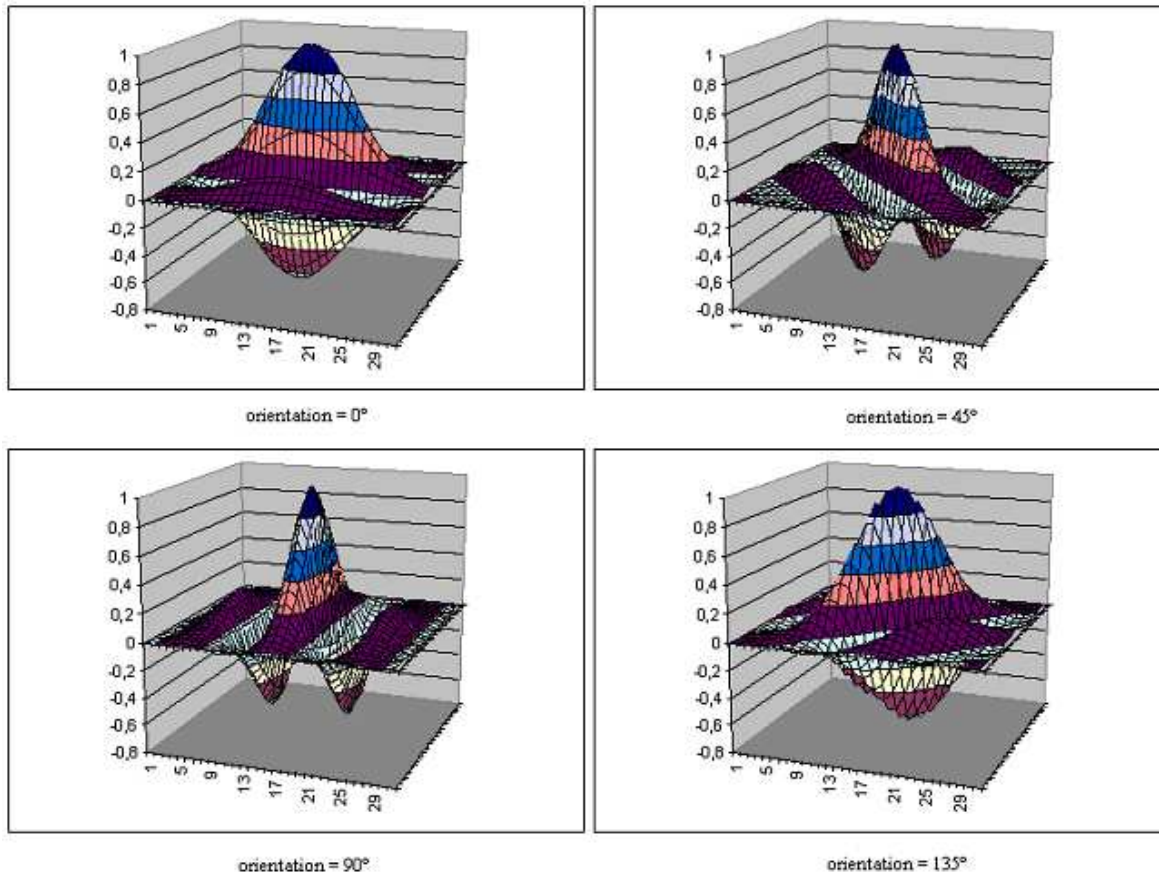


FIGURE 5.8 – Fonction de Gabor 2D

Lors de la phase de comparaison, deux tests sont effectués. Tout d'abord si deux paumes ne sont pas de la même catégorie elles sont déclarées différentes. Ensuite, l'erreur sur chaque propriété est calculée par distance de Manhattan. L'erreur finale est la somme des erreurs cumulées pondérées par un coefficient. Si l'erreur est inférieure à un seuil alors les paumes sont identiques.

5.3.3.3 Analyse par filtrage de Gabor

De nombreuses approches pour la caractérisation biométrique utilisent l'analyse par filtres de Gabor 2D. Initialement, introduite puis utilisée pour la reconnaissance de l'iris par Daugman [23, 24], la transformée de Gabor permet une location précise en échelle temps-fréquence. Un filtre de Gabor 2-D est une Gaussienne modulée par une onde (Fig. 5.8) et peut être formulé de la manière suivante :

$$G_{\sigma,\phi,\theta}(x, y) = g_{\sigma}(x, y) \cdot \exp^{2\pi i \phi(x \cos \theta + y \sin \theta)}$$

$$\text{où } g_{\sigma}(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (5.20)$$

Afin d'être résistant à la luminosité, le filtre de Gabor est centré au point (0,0). Cette normalisation pour un filtre de taille $(2k + 1)^2$ s'exprime par :

$$\tilde{G}_{\sigma,\phi,\theta}(x, y) = G_{\sigma,\phi,\theta}(x, y) - \frac{\sum_{i=-k}^k \sum_{j=-k}^k G_{\sigma,\phi,\theta}(i, j)}{(2k+1)^2} \quad (5.21)$$

Dans le domaine discret, le filtrage par Gabor d'une image $I(x, y)$ est obtenu par convolution avec la fonction de Gabor $\tilde{G}_{\sigma,\phi,\theta}(x, y)$. Pour un filtre de Gabor de taille $(2k+1)^2$, la convolution discrète de $I(x, y)$ par $\tilde{G}_{\sigma,\phi,\theta}(x, y)$ est :

$$C(x, y | \sigma, \phi, \theta) = \sum_{\lambda=-k}^k \sum_{m=-k}^k I(x + \lambda, y + m) * \tilde{G}_{\sigma,\phi,\theta}(\lambda, m) \quad (5.22)$$

Le temps de calcul et la taille des données stockées sont très importants pour une utilisation à grande échelle des systèmes biométriques. Pour cette raison, J.G. Daugman et Zhang et al. [136] binarisent le résultat de la convolution d'une image pour un filtre de Gabor codant chaque pixel sur deux bits, (b_r, b_i) , de la façon suivante :

$$\begin{cases} b_r = 1 & \text{si } \text{Re}(C) \geq 0 \\ b_r = 0 & \text{si } \text{Re}(C) < 0 \end{cases} \quad (5.23)$$

$$\begin{cases} b_i = 1 & \text{si } \text{Im}(C) \geq 0 \\ b_i = 0 & \text{si } \text{Im}(C) < 0 \end{cases} \quad (5.24)$$

En utilisant cette méthode de codage, la taille du vecteur caractéristique est de 256 octets (contre 2048 sans binarisation). Le résultat de ce codage correspond à deux matrices réelle et imaginaire pouvant être comparées grâce à la distance de Hamming normalisée. Soit P et Q deux vecteurs représentant la paume, la distance de Hamming normalisée est définie par :

$$D_0 = \frac{\sum_{i=1}^n \sum_{j=1}^n \text{Re}(P(i, j)) \otimes \text{Re}(Q(i, j)) + \text{Im}(P(i, j)) \otimes \text{Im}(Q(i, j))}{2n^2} \quad (5.25)$$

Où $\otimes$ est l'opérateur ou-exclusif. Cette distance est plus rapide à calculer que la distance Euclidienne et permet des comparaisons en un temps minimal. La distance de Hamming détermine la différence entre deux paumes : plus la distance est faible et plus les paumes sont proches. Afin d'être invariante en translation lors de la comparaison, l'équation précédente est complétée par :

$$D_{\min} = \min_{|s| < S, |t| < T} \frac{\sum_{i=\max(1, 1+s)}^{\min(N, N+s)} \sum_{j=\max(1, 1+t)}^{\min(N, N+t)} \text{Re}(P(i+s, j+t)) \otimes \text{Re}(Q(i, j)) + \text{Im}(P(i+s, j+t)) \otimes \text{Im}(Q(i, j))}{2H(s)H(t)} \quad (5.26)$$

où $S = 2$ et $T = 2$ contrôlent la variation maximum en translation horizontale et verticale du processus de comparaison et $H(s) = \min(N, N + s) - \max(1, 1 + s)$. La distance de Hamming ne peut cependant pas supporter les variations en rotation. Afin d'accroître la robustesse, lors de la phase d'enrôlement, une rotation de quelques degrés est appliquée à la référence. En combinant ces deux approches, les effets du pré-processus et les rotations, l'équation permet une correspondance invariante en rotation et en translation.

Kong et al. [57] proposent plusieurs approches s'appuyant sur des bancs de filtres de Gabor. Dans ces approches, un banc de filtres de Gabor est employé pour extraire l'information de phase d'une image de paume laquelle est regroupée selon une règle de fusion pour produire une signature unique. Dans le premier article de Kong et al. [58], l'information est fusionnée en gardant la rotation dont l'amplitude est la plus importante. Les angles sont finalement stockés selon quatre groupes décrivant l'ensemble des angles du banc de filtres. Les groupes sont codés sur 2 bits, deux paumes sont ainsi comparées par une distance de Hamming spécifique. Plus tard, Kong et al. [59] ont complété cette approche en utilisant un banc de filtres comprenant 6 orientations au lieu de 4 et une définition différente pour la fonction de Gabor. Très récemment, Kong et al. ont présenté un état de l'art de la reconnaissance de la paume [60].

Afin d'être plus robuste aux rotations, Kumar et Shen [63] ont présenté une méthode d'extraction circulaire. La paume est décomposée par une série de six anneaux concentriques, le vecteur caractéristique correspond à la moyenne et à la variance sur chacun des anneaux pour chaque direction de filtrage. Le score de comparaison est la somme des différences entre le vecteur caractéristique à tester et le vecteur de référence pour les anneaux de chaque direction.

Wu et al. [122, 121] fusionnent l'information de phase et d'orientation pour la reconnaissance biométrique. Comme dans l'article de Kong et al. [59], la rotation dont l'amplitude est la plus importante est conservée. En plus de l'information d'orientation, la paume filtrée par un filtre de Gabor d'orientation 0° puis binarisée complète la référence biométrique. La distance est calculée comme précédemment par une distance de Hamming spécifique. Cette méthode est nommée PPOC (palmpoint phase orientation code).

5.3.4 Approche par apprentissage

Afin de comparer deux paumes, plusieurs méthodes d'apprentissages sont proposées dans la littérature. Han et al. [46] propose une extraction de l'information par les filtres de Sobel puis une classification par réseau de neurones. Pour caractériser un utilisateur X , M exemples d'un individu X appelés exemples positifs et N exemples d'autres personnes appelés exemples négatifs sont utilisés. L'algorithme d'apprentissage est l'approche par retro-propagation du gradient. L'entrée du réseau est composée de plusieurs valeurs correspondant à la moyenne d'images de la paume :

- 64 moyennes d'images de taille $32 * 32$,
- 8 moyennes d'images de taille $256 * 32$,
- 8 moyennes d'images de taille $32 * 256$.

L'architecture du réseau est composée de 80 neurones en entrée, 40 neurones sur une couche cachée et 1 neurone de sortie comme décrit dans la Fig. 5.9.

Shang et al. [97] utilisent également un apprentissage par réseau de neurones. Différentes structures sont utilisées pour l'apprentissage qui est effectué après une décomposition de la paume par *ICA*. Wang et al. [118] extrait l'information de texture par *LBP* (*Local Binary Pattern*) [81]. Ce descripteur est invariant au niveau de gris. L'opérateur marque les pixels d'une image avec le résultat d'un filtrage $3 * 3$ de chaque pixel seuillé avec la valeur centrale. Les histogrammes des labels sont utilisés comme descripteur de texture. Ce processus de caractérisation est illustré à la figure 5.10.

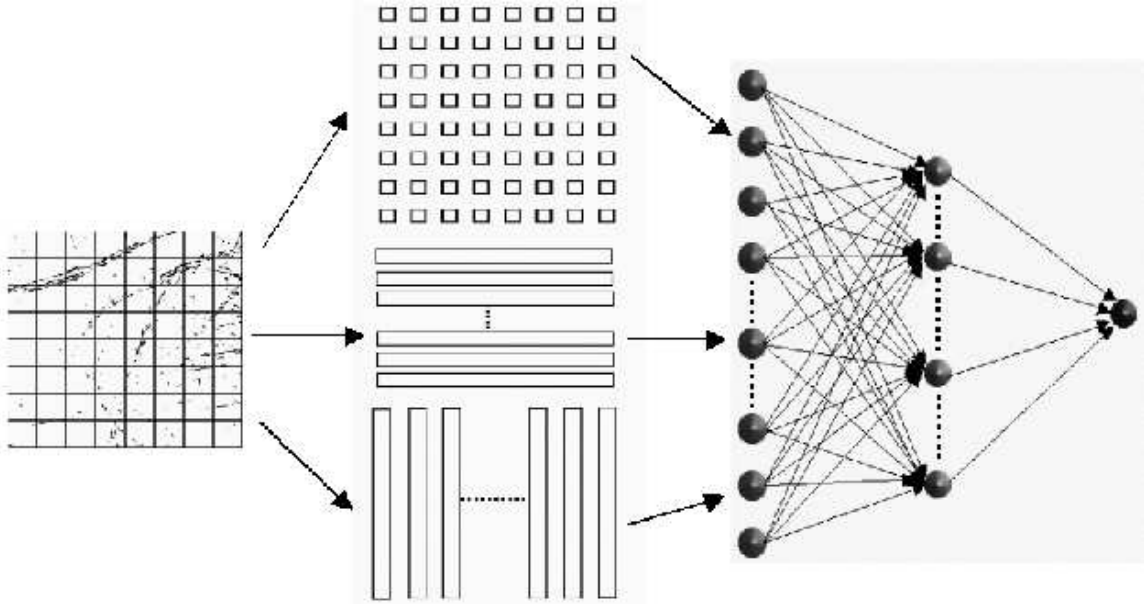
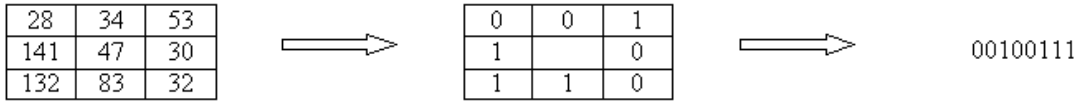


FIGURE 5.9 – Architecture dans le réseau de neurones de [46]

FIGURE 5.10 – Opérateur *LBP*

Afin de caractériser une paume, les caractéristiques *LBP* les plus significatives sont apprises par AdaBoost [40] à partir d'une large base de données. L'algorithme Adaboost est une méthode utilisant de nombreux classifieurs faibles qui sont combinés linéairement pour former un classifieur robuste.

5.3.5 Autres approches

D'autres approches spécifiques ont été proposées. Noh et al. [79] extraient les caractéristiques de paumes par une binarisation de Otsu [84] après égalisation d'histogramme. Finalement, les caractéristiques sont définies à l'aide des moments invariants de Hu puis comparées par distance Euclidienne.

You et al. [133] utilisent une méthode de caractérisation hiérarchique de texture pour obtenir une reconnaissance robuste et rapide. Deux types de caractéristiques sont ainsi définis :

1. 4 valeurs qui correspondent aux données d'énergies de la paume filtrée dans 4 directions (0°C , 45°C , 90°C , 135°C),
2. 64 valeurs qui sont les valeurs moyennes de 64 imageries divisant la paume.

Les caractéristiques de deux paumes sont comparées classiquement par distance Euclidienne.

5.4 Multi-biométrie par fusion de méthodologies

Afin de pallier les problèmes de reconnaissance par empreinte palmaire due à une faible résolution par exemple ou les limites de la biométrie de la main qui est moins discriminative, plusieurs approches par fusion sont proposées dans la littérature. Classiquement, trois niveaux de fusion sont suggérés :

- fusion à la représentation où les caractéristiques (features en anglais) de plusieurs biométries sont concaténées pour former une donnée fusionnée,
- fusion au niveau de décision où les scores de décision de plusieurs biométries sont fusionnés pour générer un score final,
- fusion au niveau d'abstraction où les décisions de plusieurs biométries sont consolidées.

Les méthodes de fusion biométrique et leurs applications sur différentes biométries sont présentées dans le rapport de thèse de Allano [67]. Dans ce document, les méthodes de normalisations et de combinaisons des scores sont évaluées sur plusieurs traits biométriques.

Plus particulièrement au trait biométrique de la main, Kumar et al. [64] ont présenté la première approche de fusion des paramètres de forme et de paume. Les caractéristiques des paumes sont les valeurs d'orientations principales des paumes. La forme est modélisée par 16 données géométriques : 4 longueurs de doigts, 8 largeurs de doigts (2 par doigt exceptés le pouce), largeur de la paume, hauteur de la paume, surface et hauteur de la main. Les données sont fusionnées à la représentation en concaténant les caractéristiques ou à la décision.

Goh Kah Ong et al. [83] utilisent une fusion des scores par machine à vecteurs supports. Les caractéristiques de paume sont extraites par analyse de Fisher tandis que les informations de main sont définies comme précédemment par un ensemble de longueurs et de largeurs des doigts et de la main.

Une autre approche plus complète de fusion de données de la main est proposée par Ribaric et Fratric [89]. Cette méthode utilise une caractérisation par analyse en composantes principales sur la paume (*eigenpalm*) mais également sur la surface des doigts (*eigenfinger*). Les zones de la paume et des doigts utilisées sont schématisées dans la Fig. 5.11. La décomposition en composantes principales de la paume et du petit doigt pour différentes valeurs propres est illustrée à la Fig. 5.12. A partir des décompositions, les scores de comparaisons définis par distance Euclidienne sont normalisés puis fusionnés par combinaison linéaire. La décision finale de classification est déterminée par l'algorithme des k plus proches voisins (KNN).

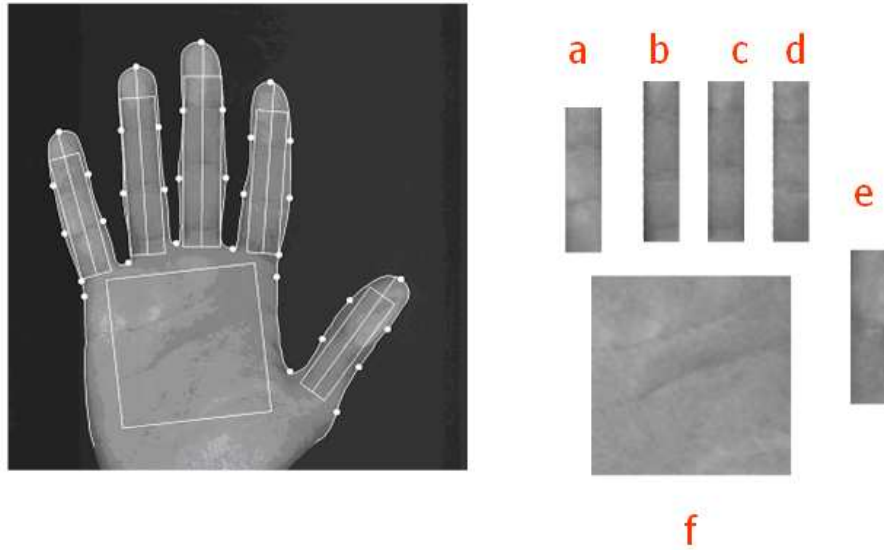


FIGURE 5.11 – Extraction des zones des doigts et de la paume : (a) auriculaire, (b) annulaire, (c) majeur, (d) index, (e) pouce et (f) paume

5.5 Comparaison des différentes approches

Contrairement aux reconnaissances biométriques par empreintes digitales ou par analyse du visage, chaque équipe de recherche possède sa propre base de travail. Cependant, récemment l'équipe BRC de Hong Kong du professeur Zhang propose en téléchargement sa base de données pour l'analyse de la paume dans le cadre de projets non-commerciaux².

Le tableau suivant présente les résultats de reconnaissance en mode authentification (A) ou identification (I) des systèmes de reconnaissance palmaire, sur la forme de la main ou en fusionnant les informations. Le nombre d'acquisition (A) par utilisateur (U) ainsi que le nombre d'images utilisées pour l'enrôlement (E) sont précisés pour chaque évaluation lorsque l'information est disponible. Les performances sont indiquées pour le taux d'égale d'erreur (TER). Dans tous les tests, les images contenues dans les bases de données sont acquises dans un environnement contrôlé : le fond est uni et la lumière est connue. Dans les approches de Zhang et Jain, l'emplacement de la main est fixé par une série de guides sur le plan de pose ce qui stabilise les acquisitions entre différentes sessions.

Les méthodes de reconnaissance par la paume affichent de très bonnes performances en terme de reconnaissance principalement pour les méthodes s'appuyant sur la caractérisation de la texture par les filtres de Gabor. Cette méthode de reconnaissance présente l'avantage d'être très rapide puisqu'elle utilise des images à faibles résolutions. Les approches basées sur l'extraction des lignes utilisent des caractéristiques plus difficiles à extraire automatiquement et nécessitent des images de bonne résolution (> 100 dpi) pour une caractérisation efficace. La décomposition en sous-espaces est une méthode très répandue pour la caractérisation du visage. Pour la définition de la paume, les performances de reconnaissance sont légèrement inférieures aux méthodes par Gabor. Dans ces approches la décomposition est très dépendante de la diversité et de la

2. PolyU Palmprint Database : <http://www.comp.polyu.edu.hk/~biometrics/>

Méthode	Comparaison	Mode	U	A	E	TER	Référence
Reconnaissance par la forme de la main							
16 longueurs	Distance Euclidienne	A	50	10	2	0.9%	Jain et al. [53]
ICA	Distance Euclidienne	I	756	-	-	0.45%	Yoruk et al. [132]
Reconnaissance palmaire							
Ondelette	Distance Euclidienne	A	200	5	1	0.5%	Wu et al. [124]
Gabor	Distance de Hamming	A	193	40	1	0.6%	Zhang et al. [136]
LDA	-	A	300	10	-	0.82%	Wu et al. [123]
PCA	Distance Euclidienne	A	382	8	4	0.85%	Lu et al. [73]
ICA	Réseau de neurones	A	100	6	-	1%	Conie et al. [19]
Gabor	Distance Euclidienne	A	80	10	1	2.5%	Kumar et al. [63]
Ondelette	Distance de Manhattan	I	50	4	1	2%	Zhang et Zhang [138]
Gabor	Distance Euclidienne	A	80	10	1	1.6%	Kong et al. [59]
Filtrage	Distance Euclidienne	A	320	10	6	2.08%	U et al. [125]
Morphologie	Distance de Hamming	A	100	6	1	1.96%	Wu et al. [126]
Moment de Hu	Distance Euclidienne	A	189	2	1	1%	Noh et al. [79]
Gabor	Distance de Hamming	A	392	20	1	0.31%	Wu et al. [121]
LBP	AdaBoost	A	50	10	2	2%	Wang et al. [118]
Hough	Distance de Hausdorff	A	100	6	3	1%	Li et al. [71]
Fusion biométrique							
PCA	K plus proches voisins	I	237	10	7	0.58%	Ribaric et al. [89]
Multiple	Distance Euclidienne	A	386	20	1	4.5%	You et al. [133]
PCA et 10 longueurs	Distance Euclidienne	A	50	30	10	4.5%	Han et al. [46]
Energie et 16 longueurs	Distance Euclidienne	A	100	10	5	1%	Kumar et al. [65]
LDA et 22 longueurs	SVM	A	100	6	4	1%	Ong et al. [83]

TABLE 5.1 – Comparaison des méthodes de reconnaissance de la main

généralité de la base de données d'apprentissage. Cette limite est également présente pour les méthodes basées sur l'apprentissage où la définition des contre-exemples dans une utilisation réelle est très difficile et non abordée dans les méthodes l'utilisant.

Les méthodes de reconnaissance par la forme de la main affichent de bonnes performances de reconnaissance également même si dans l'article de Jain et al.[53], un tri de la base de mains est effectué. Ces méthodes présentent cependant l'avantage d'être la modalité biométrique présentant la plus faible taille de référence (16 octets pour la méthode de Jain et al.) ainsi qu'une très grande vitesse d'exécution. Les performances en terme de reconnaissance sont cependant à relativiser. En effet dans les méthodes utilisant les informations de forme et de paume de la

main, le taux de reconnaissance de la paume est toujours nettement supérieur à celui de la forme.

Les différents articles montrent l'impact important en matière de performance de reconnaissance des méthodes utilisant la fusion d'information. De plus ces méthodes permettent de baisser de manière très importantes le taux de faux rejets pour des taux de fausses acceptations faibles. En contre-partie, le temps d'extraction des caractéristiques et la taille des références sont multipliés par le nombre de modalités employées.

L'analyse par les veines est une biométrie beaucoup plus récente. Peu d'articles sont parus pour le moment bien que la technologie soit suffisamment mûre pour que des modules de reconnaissance par les veines soient commercialisés. Ils semblent que les algorithmes les plus performants sont plutôt brevetés que publiés. Les principaux avantages de cette méthode résident dans la haute fiabilité de reconnaissance mais surtout dans la difficulté à frauder dans de tels systèmes qui reposent sur une technologie infra-rouge. Il est ainsi intéressant de noter que cette technologie est employée dans de nombreux systèmes de retrait automatique de billets au Japon. Un point important de cette technologie est qu'elle est sans contact ce qui crucial dans les pays Asiatiques où l'hygiène est un facteur prépondérant.

5.6 Conclusion sur la reconnaissance par la main

Ce chapitre liste les différentes approches utilisées pour la reconnaissance de la main. Les différentes méthodes présentées sont très variées. Le tableau 5.1 affiche les résultats de plusieurs caractérisations suivant les critères biométriques utilisés :

- Reconnaissance par la forme
- Reconnaissance palmaire
- Fusion biométrique

Clairement, les approches par analyse de la paume sont plus performantes que la reconnaissance par la forme de la main. Les méthodes de fusion permettent cependant une amélioration du taux de reconnaissance en cumulant les informations sur la forme ou la texture des doigts.

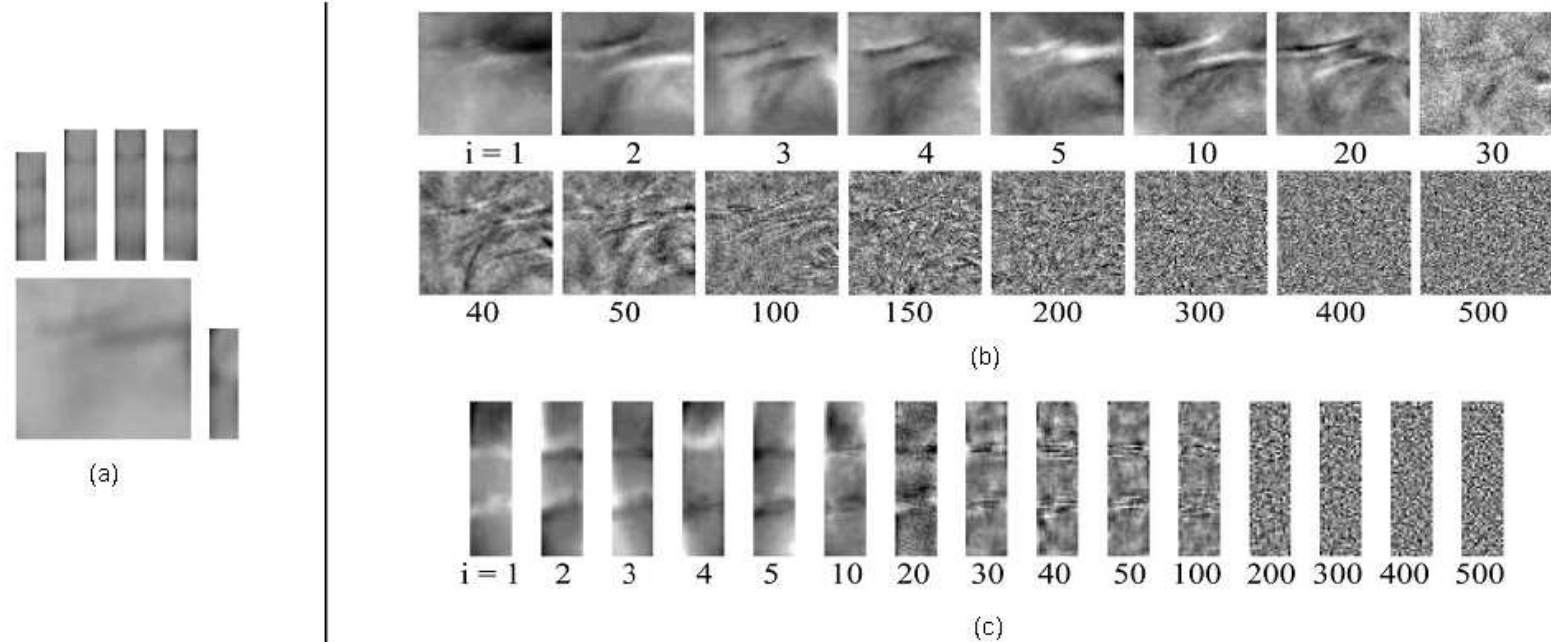


FIGURE 5.12 – Caractéristiques des doigts et de la paume : (a) valeurs moyennes de la paume et des doigts, (b) eigenpalms et (c) eigenfingers de l'auriculaire suivant différentes valeurs propres

Chapitre 6

Reconnaissance de la main sans contact

Sommaire

6.1	Extraction et normalisation de la paume	95
6.2	Combinaison des informations de forme et de texture	96
6.2.1	Extraction des caractéristiques	97
6.2.2	Evaluation des performances	99
6.3	Authentification par filtre de Gabor circulaire et distance d'édi- tion	102
6.3.1	Extraction des caractéristiques	102
6.3.2	Comparaison des données extraites	103
6.3.3	Evaluation des performances	105
6.4	Classification par estimation de distribution non linéaire	106
6.4.1	Extraction des caractéristiques	106
6.4.1.1	Extraction par banc de filtres de Gabor	106
6.4.1.2	Extraction par analyse robuste des niveaux de gris . . .	106
6.4.2	Comparaison par estimation de distribution non linéaire	109
6.4.2.1	Estimation de la distribution	110
6.4.2.2	Fonction de reconnaissance individuelle	111
6.4.3	Évaluation des performances	112
6.5	Etude sur la robustesse aux paramètres extérieurs	113
6.5.1	Description des filtres log-normaux	114
6.5.2	Evaluation des performances suivant les perturbations extérieures	115
6.5.2.1	Les changements de luminosité et de contraste	116
6.5.2.2	Le flou Gaussien et le bruit blanc	116
6.5.2.3	Les translations, rotations et zooms	116
6.5.2.4	Résumé des performances	119
6.6	Conclusion sur la reconnaissance de la main sans contact . . .	119

6.1 Extraction et normalisation de la paume

Après la détection de la main dans une image complexe décrite dans le chapitre 4, il est nécessaire d'extraire la paume indépendamment de la distance entre la main et le capteur d'acquisition.

La méthode de définition de la paume s'appuie sur les dimensions de la main et l'extraction de paume comme décrit dans l'article de Zhang et al. [136] et expliqué au paragraphe 3.7.1. Dans la littérature, deux valeurs illustrées à la Fig. 6.1 sont fixées : la distance entre les points $O1$ et $O2$ et la largeur de zone extraite $\|A1A2\|$. Dans ces méthodes, la zone extraite est carrée. Dans notre approche la paume est définie avant normalisation de façon rectangulaire en ajoutant une notion de longueur de dimension $\|A2A3\|$.

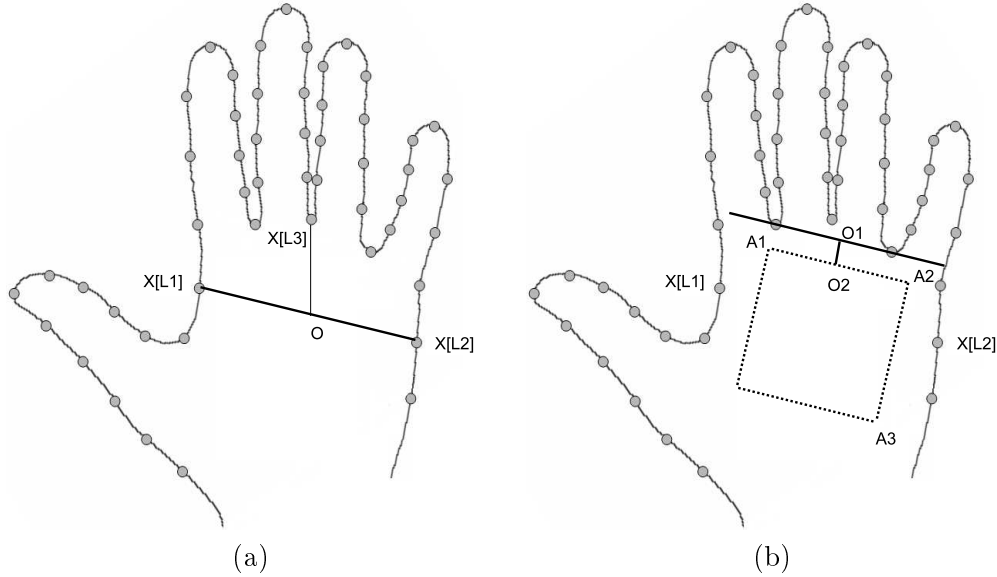


FIGURE 6.1 – Extraction de la paume (a) distances de référence (b) paume extraite

Ces valeurs, constantes dans les systèmes de reconnaissance de paume, sont définies ici suivant la taille de la main. Elles sont déterminées suivant la largeur et la longueur de la paume. La largeur de la main est fixée par la distance Euclidienne entre les points $X[L1]$ et $X[L2]$ où X est l'ensemble des points du contour de la main utilisée par l'ASM, $L1$ et $L2$ sont les index fixés après expérimentations à 30 et 125 avec 12 points intermédiaires entre chaque extrémité des doigts. La longueur de la paume est, elle, approximée à un coefficient près par la distance entre O et $X[L3]$ où O est le centre du segment $[X[L1], X[L2]]$. L'utilisation de la longueur de la paume et non de la longueur de la main permet d'être robuste aux doigts pliés.

Finalement, $\|A1A2\|$, $\|A2A3\|$ et $\|O1O2\|$ sont définis par :

$$\begin{cases} \|O1O2\| &= \alpha \|X[L1]X[L2]\| \\ \|A1A2\| &= \beta \|X[L1]X[L2]\| \\ \|A2A3\| &= \gamma \|OX[L3]\| \end{cases} \quad (6.1)$$

où α , β et γ sont les coefficients de dimension choisis à $1/10$, $2/3$ et $3/2$, respectivement. La paume est ensuite retaillée à une taille fixe $N \times N$ pour être indépendante de la distance entre la main et le capteur d'acquisition et pour pouvoir comparer efficacement les paumes entre elles.

6.2 Combinaison des informations de forme et de texture

De nombreuses méthodes de fusion d'information pour la reconnaissance sont proposées dans la littérature (Cf. chapitre précédent). Dans cette approche, le processus d'authentification fu-

sionne par convolution les caractéristiques géométriques des doigts et la texture de la paume.

6.2.1 Extraction des caractéristiques

Etant données la qualité des images extraites, les lignes de la paume ne peuvent pas être déterminées de manière robuste, une caractérisation globale est ainsi préférable. Différentes méthodes permettent d'obtenir les caractéristiques de la paume : l'analyse en ondelettes, la transformée de Fourier, l'Analyse en Composantes Principales, le filtre de Gabor, etc. Grâce à ses bonnes performances en reconnaissance de l'iris et de la paume et à ses qualités propres : localisation précise en temps/fréquence et robustesse aux variations de contraste et de luminosité, nous avons utilisé un filtre de Gabor. Différentes implémentations de ce filtre existent. Dans les approches basées sur l'étude de Daugman, un filtre de Gabor 2D dans le domaine complexe est utilisé. Pour limiter le temps de calcul et la taille des caractéristiques, le filtre dans le domaine réel décrit dans Kruinzinga et Pektov [62] est employé :

$$G(x, y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \times \cos\left(2\pi \frac{x'}{\lambda} + c\right) \quad (6.2)$$

$$\begin{cases} x' = (x - \xi) \cos(\theta) - (y - \eta) \sin(\theta) \\ y' = (x - \xi) \sin(\theta) + (y - \eta) \cos(\theta) \end{cases}$$

Où le couple (ξ, η) définit le centre de la fonction, θ contrôle l'orientation de la fonction, σ est la déviation standard de l'enveloppe gaussienne, γ est le ratio de l'aspect visuel fixé à 0.5, λ est la période de l'onde et φ est la phase. σ est défini par le rapport constant $\sigma / \lambda = 0.56$ [62]. Pour plus de robustesse à la luminosité, le filtre est centré au point $(0, 0)$ en utilisant pour un filtre de taille $(2k + 1)^2$ la formule :

$$\Omega(x, y) = G(x, y) - \frac{\sum_{i=0}^k \sum_{j=0}^k G(i, j)}{(2k + 1)^2} \quad (6.3)$$

Ainsi les caractéristiques de la paume de la main sont obtenues par le résultat de la convolution de l'image de la paume I avec ce filtre de Gabor robuste par :

$$C(x, y) = I(x, y) * \Omega(x, y) \quad (6.4)$$

Où $*$ est l'opérateur de convolution.

Pour compléter la représentation, les caractéristiques de la forme de la main sont extraites. Il s'agit des largeurs et des longueurs des doigts. Les longueurs et les largeurs sont approximées par la distance entre les points du modèle de main et sont définies par :

$$L(i) = d(m(X(H_{2i}), X(H_{2i+2})), X(2i + 1)) \quad (6.5)$$

$$l(i, j) = d(X(H_{2i+1} - j), X(H_{2i+1} + j)) \quad (6.6)$$

Où $d(a, b)$ est la distance Euclidienne entre les points a et b , $m(a, b)$ retourne le milieu du segment $||ab||$, $L(i)$ est la longueur du i ème doigt, $l(i, j)$ est la j ème largeur du i ème doigt et $H_i = i(N + 1)$ est l'indice du i ème point caractéristique de la main (Cf. Figure 6.2).

Dans les systèmes biométriques, trois méthodes de combinaison de données sont utilisées. La combinaison peut être effectuée à la représentation en regroupant les caractéristiques extraites,

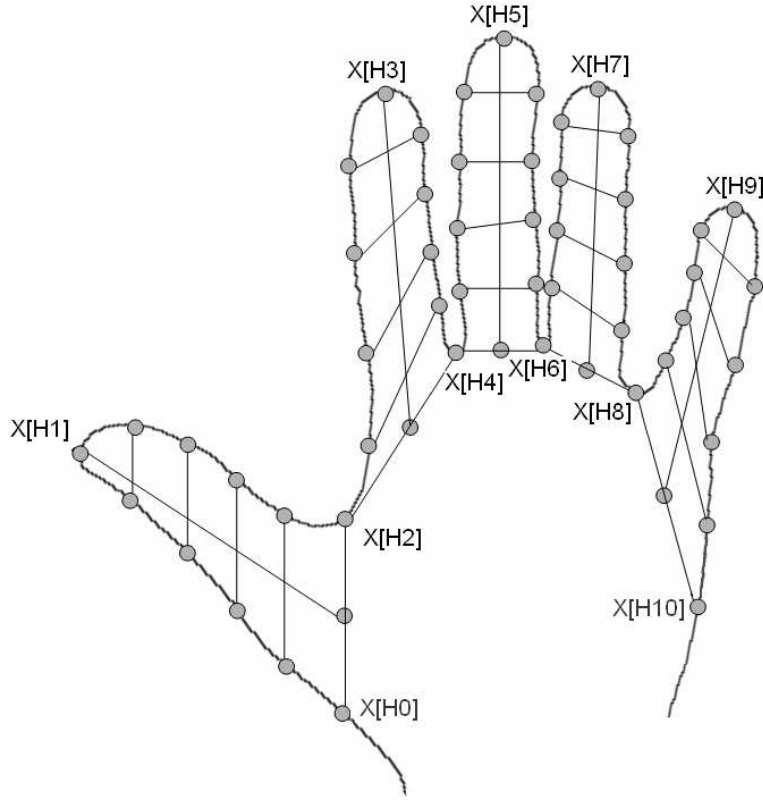


FIGURE 6.2 – Schéma d'extraction des dimensions de la main

au niveau de la comparaison ou au niveau des décisions. Notre fusion est basée sur la convolution de la texture de la paume avec les caractéristiques de la forme de la main afin d'ajouter un facteur géométrique à la texture. Cette fusion permet également de contenir les deux informations dans une référence de taille identique à la caractéristique de la paume seule. Le résultat de cette convolution est binarisé pour limiter la taille des caractéristiques et les temps de calcul dans la phase de comparaison. Le processus complet est défini par :

$$S(x, y) = b(C(x, y) * H) \quad (6.7)$$

Où $b(x) = 0$ si $x < 0$ et $b(x) = 1$ sinon., $*$ est l'opérateur de convolution et H est un filtre de taille 5×5 correspondant aux caractéristiques de formes. Il peut être explicité par :

$$H(x, y) = H'(x, y) \times \frac{1}{\sum_{i=0}^4 \sum_{j=0}^4 H'(i, j)} \quad (6.8)$$

$$\text{Où } H'(x, y) = \begin{pmatrix} L(0) & l(0, 1) & l(0, 2) & l(0, 3) & l(0, 4) \\ L(1) & l(1, 1) & l(1, 2) & l(1, 3) & l(1, 4) \\ L(2) & l(2, 1) & l(2, 2) & l(2, 3) & l(2, 4) \\ L(3) & l(3, 1) & l(3, 2) & l(3, 3) & l(3, 4) \\ L(4) & l(4, 1) & l(4, 2) & l(4, 3) & l(4, 4) \end{pmatrix} \quad (6.9)$$

Le coefficient de normalisation de H permet d'être robuste à la distance entre la main et le système d'acquisition. La fusion par convolution des données caractéristiques augmente l'unicité

de la main. Elle permet en effet de distinguer des personnes ayant des formes de main ou des paumes très similaires. L'extraction de la texture seule et la combinaison de la texture et de la géométrie de la main sur une paume sont illustrées à la Fig. 6.3.

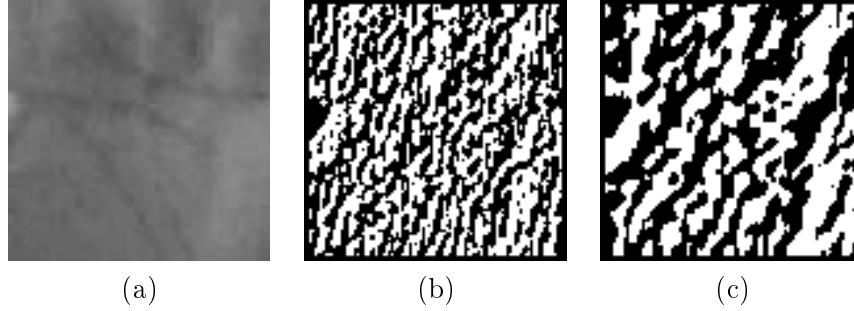


FIGURE 6.3 – Extraction des caractéristiques (a) paume extraite, (b) texture de la paume, (c) convolution de la texture et de la forme

Pour la comparaison, les caractéristiques de la main S sont assimilées à une matrice. Une méthode classique de comparaison de matrices binaires est appliquée : la distance de Hamming normalisée. Cette distance est une comparaison pixel par pixel et elle donne une réponse normalisée entre 0 et 1, 0 étant la correspondance parfaite. Elle est définie pour deux caractéristiques de mains X et Y par :

$$D_0(X, Y) = \frac{\sum_{i=1}^N \sum_{j=1}^N X(i, j) \otimes Y(i, j)}{N^2} \quad (6.10)$$

Où $\otimes$ est l'opérateur ou-exclusif. La segmentation de la main ainsi que la détection de la paume n'étant pas parfaites, une tolérance en translation et en rotation est appliquée au processus de comparaison. Cette comparaison souple s'exprime pour deux caractéristiques de mains X et Y par :

$$D_r(X, Y) = \min_{|s| < S, |t| < T, |a| < A} D_0(Tr(R(X, a), s, t), Y) \quad (6.11)$$

Où $Tr(X, s, t)$ est la translation de l'image X horizontalement par s et verticalement par t et $R(X, a)$ est la rotation de l'image X par un angle a . Les coefficients S et T sont fixés à 2 pixels et A est limité à 2° afin de limiter le temps d'exécution.

Ainsi deux mains X et Y sont considérées comme provenant de la même personne si $D_r(X, Y) \leq \delta$ sinon elles sont déclarées d'utilisateurs différents avec δ un seuil fixé.

6.2.2 Evaluation des performances

Une base de données spécifique a été élaborée pour valider notre approche. Toutes les images de la base sont acquises à partir d'une webcam *Philips ToUcam Pro 740K* avec une taille de 640×480 . La base contient 180 images de mains, certaines avec des mains possédant des bagues, provenant de 18 personnes. 10 images de la main droite sont acquises pour chaque individu de la base. Les images sont inversées pour une interaction naturelle de l'utilisateur avec le logiciel d'acquisition. Les contraintes d'acquisitions de la main sont les mêmes que celles précisés dans le chapitre de détection de mains :

- La main doit être présentée face à la caméra,

- la paume est dans l'axe perpendiculaire au capteur,
- les doigts sont écartés pour obtenir une plus grande précision de détection.

Les utilisateurs n'étant pas habitués au système, en pratique, une mauvaise acquisition est automatiquement rejetée. Une comparaison est considérée comme une comparaison d'authentique si deux images de paumes proviennent de la même main, sinon elle est comptée comme une comparaison d'imposteur.

Pour obtenir les coefficients optimaux et le meilleur filtre représentant la forme, un ensemble de paramètres est utilisé pour valider notre approche. Premièrement, un banc de filtres est créé pour déterminer le meilleur score de reconnaissance pour la paume seule. Chaque paume de la base de mains est comparée avec toutes les autres afin de déterminer ce score.

Les paramètres pour le filtre de Gabor sont inspirés de [136], l'orientation du filtre est testée suivant huit valeurs $\theta = 22.5^\circ C, \theta = 45^\circ C, \dots, \theta = 180^\circ C$, trois valeurs sont utilisées pour la fréquence spatiale $\lambda = 5.47, \lambda = 8.20$ et $\lambda = 10.93$, le couple (ξ, η) et la phase φ_0 sont fixés à $(0, 0)$ et 0 respectivement, tandis que la déviation standard σ est déterminée suivant le rapport $\sigma/\lambda = 0.56$. Le meilleur taux de comparaison présenté à la courbe de la figure 6.4 est obtenu avec les paramètres $\theta = 157.5^\circ C, \lambda = 10.93$ et $\sigma = 6.12$.

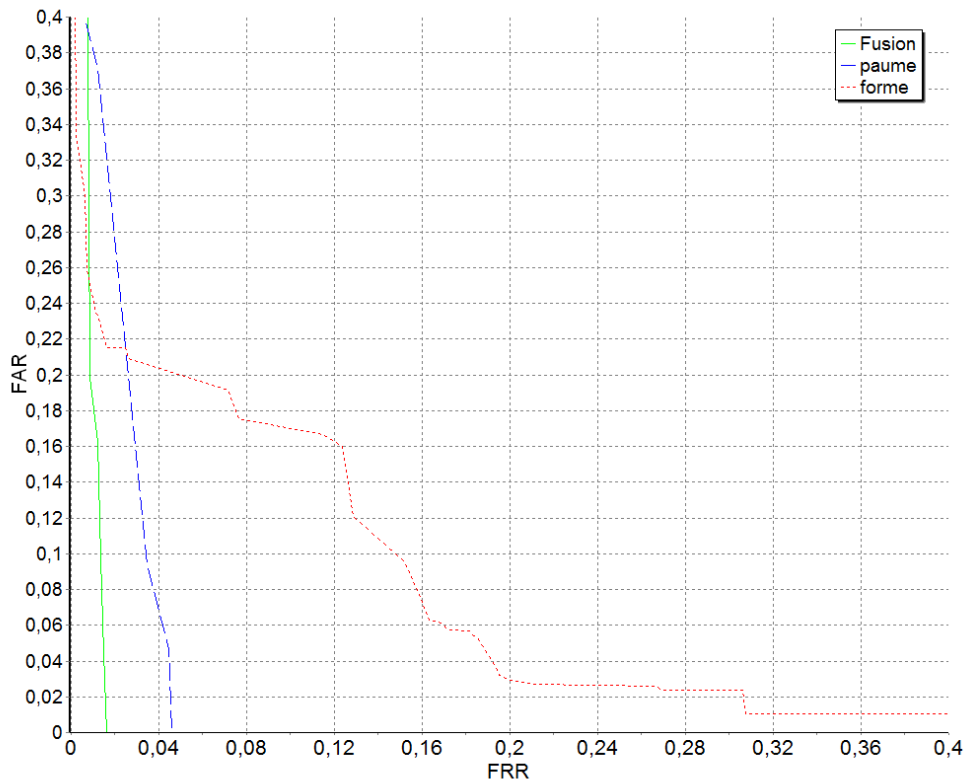


FIGURE 6.4 – Courbe ROC comparant les performances avec et sans fusion par convolution

Avec les coefficients optimaux, le TER (Taux d'égale erreur) correspondant à $FAR=FRR$ indique une erreur de 4.25%. Nous avons également testé le filtre de Gabor complexe présenté

en [136] et les résultats obtenus sont similaires sur notre base de données.

Pour diminuer les erreurs de reconnaissance, nous avons proposé une nouvelle méthode de fusion des caractéristiques de forme et de texture de la paume. Après expérimentations, le filtre optimal définissant la forme est seulement composé des largeurs des doigts. Ainsi, le filtre optimal H est défini par :

$$H(X, Y) = l(x, y + 1) \times \frac{1}{\sum_{i=0}^4 \sum_{j=0}^4 l(i, j + 1)} \quad (6.12)$$

Ceci peut être expliqué par le fait que dans certaines acquisitions de mains, celle-ci n'est pas complètement présente dans l'image. Les bouts des doigts sont interpolés lors de la recherche par l'*ASM* ajusté ce qui entraîne une approximation de la caractérisation et une baisse des performances de reconnaissance.

Les performances du système de reconnaissance globale sont bien augmentées entre la reconnaissance par la paume seule et la reconnaissance globale de la main (Fig. 6.4). En effet, le *TER* est égal à 1.85%. Les courbes de distribution de distances entre les imposteurs et les authentiques avec ces paramètres sont exposées à la figure 6.5.

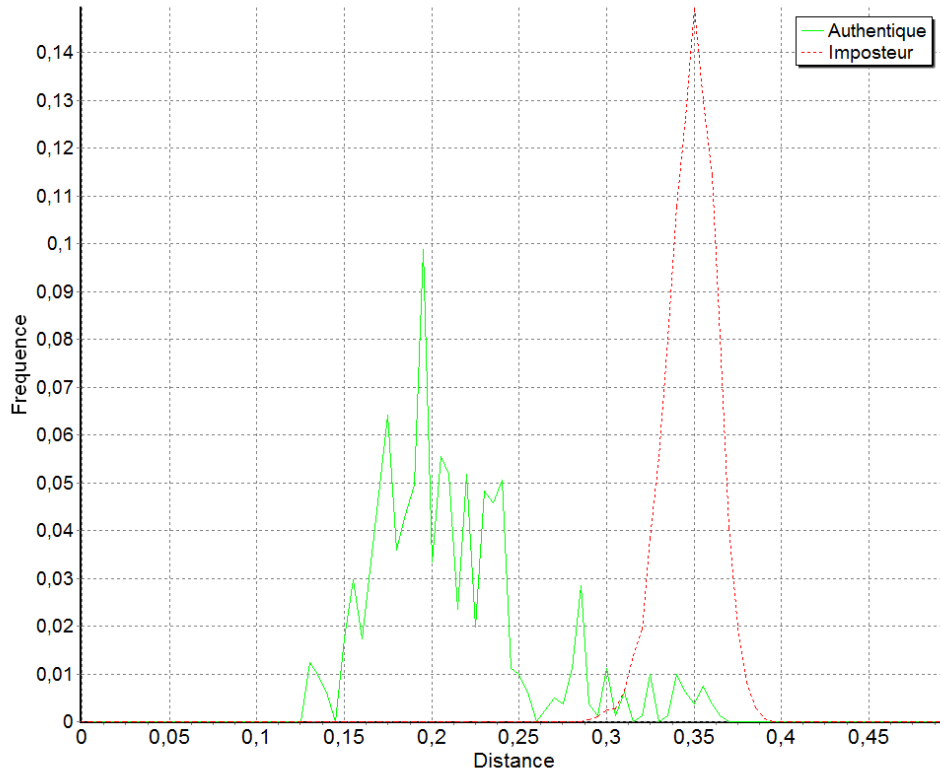


FIGURE 6.5 – Courbes de distributions des comparaisons

Le processus de reconnaissance complet (segmentation, extraction des caractéristiques et comparaison) est effectué en moins de 1 seconde sur un *Pentium M* à 1.60GHz.

6.3 Authentification par filtre de Gabor circulaire et distance d'édition

La méthode de reconnaissance présentée précédemment s'appuie sur une combinaison des informations de formes et de textures. Les informations de formes permettent d'améliorer la reconnaissance mais ne sont pas suffisamment stables pour obtenir un niveau de performances plus élevé. Particulièrement, les caractéristiques de formes sont très dépendantes de l'orientation de la main face au capteur. Ainsi, dans une nouvelle approche utilisant l'information de paume seule, une caractérisation plus robuste est formulée. Elle repose sur une extraction des caractéristiques par un filtrage de Gabor circulaire et une comparaison avec une distance d'édition normalisée.

6.3.1 Extraction des caractéristiques

Comme indiquée dans le paragraphe 5.3.3.3, les filtres de Gabor orientés ont été largement utilisés pour la reconnaissance de texture et en particulier dans des applications de reconnaissances biométriques. Les filtres de Gabor sont résistants aux changements de luminosité et de contraste mais ne sont pas robustes aux translations et aux rotations. Dans cette approche, un filtre de Gabor circulaire indépendant de l'orientation de la texture est utilisé. Ce filtre est une Gaussienne modulée par une onde circulaire. Il peut être défini comme dans l'article de Zhang et al. [137] par :

$$Gc_{\sigma,\phi,\theta}(x,y) = g_{\sigma}(x,y) \cdot \exp\left(2\pi i\phi\sqrt{x^2+y^2}\right) \\ \text{où } g_{\sigma}(x,y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \quad (6.13)$$

Où ϕ est la fréquence du filtre de Gabor circulaire. Les parties réelles du filtre de Gabor orienté défini au paragraphe 5.3.3.3 et du filtre de Gabor circulaire avec $\phi = 0.3666$, $\sigma = 1.4045$ et $\theta = 0^\circ C$ (pour le filtre orienté) sont montrées à la Fig. 6.6.

Comme précédemment pour plus de robustesse à la luminosité, le filtre de Gabor est centré au point (0,0). Un filtre circulaire robuste peut être explicité pour un filtre de taille $(2k+1)^2$ par :

$$\widetilde{Gc}_{\sigma,\phi,\theta}(x,y) = Gc_{\sigma,\phi,\theta}(x,y) - \frac{\sum_{i=-k}^k \sum_{j=-k}^k Gc_{\sigma,\phi,\theta}(i,j)}{(2k+1)^2} \quad (6.14)$$

Dans le domaine discret, les caractéristiques d'une image $I(x,y)$ sont obtenues par convolution avec la fonction de Gabor circulaire $Gc_{\sigma,\phi,\theta}(x,y)$. Pour un voisinage donné de taille $W * W$ avec $W = 2k+1$, la convolution discrète de $I(x,y)$ par la fonction de Gabor est :

$$C(x,y|\sigma,\phi,\theta) = \sum_{\lambda=-k}^k \sum_{m=-k}^k I(x+\lambda,y+m) \times \widetilde{Gc}_{\sigma,\phi,\theta}(\lambda,m) \quad (6.15)$$

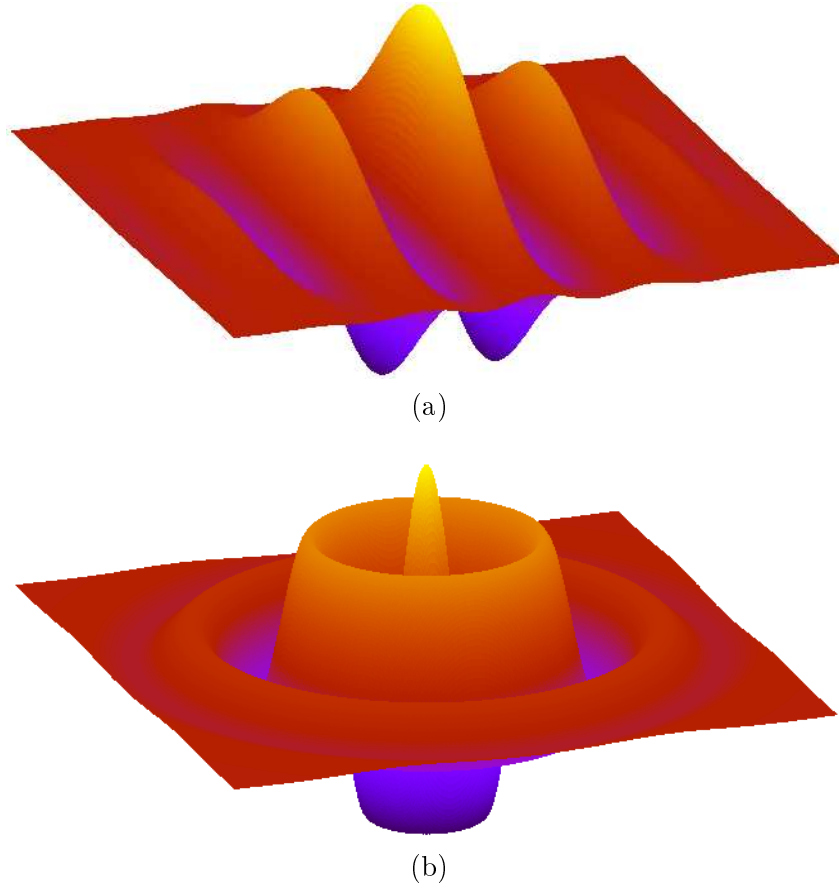


FIGURE 6.6 – Fonctions de Gabor (a) orientée et (b) circulaire

Il est à noter que contrairement aux méthodes de l'équipe de Zhang, l'image convoluée n'est pas binarisée afin de garder le maximum de précision. La taille de la référence est ainsi multipliée par un facteur important (8 dans ce cas) mais permet une caractérisation plus fine de la paume. Bien sûr, dans un système plus contraint en mémoire ou en puissance, la référence peut être binarisée pour gagner en taille de référence mais également en vitesse de comparaison.

6.3.2 Comparaison des données extraites

A partir des caractéristiques calculées, de nombreuses méthodes comme le calcul d'énergie, la distance Euclidienne ou la distance de Hamming pour les références binaires sont employées. Dans notre méthode, un algorithme de comparaison de chaîne de caractères approchée (ACCA) décrit par [51] est préférée. Cette méthode permet de prendre en compte les informations de continuité dans la référence biométrique contrairement à la distance de Hamming qui agit de façon bit à bit. L'algorithme s'appuie sur la distance d'édition qui est une généralisation de la distance de Hamming. Initialement, l'ACCA permet de comparer deux vecteurs en calculant une matrice de comparaison à partir d'une fonction de score. La fonction détermine la similarité entre deux éléments. La matrice de score est remplie suivant une comparaison ou une non-comparaison entre les éléments des vecteurs. Le score de comparaison final est la plus haute valeur dans la matrice de comparaison. Cette méthode de comparaison est robuste aux translations, aux suppressions,

aux substitutions et aux additions d'éléments dans les vecteurs. L'algorithme 4 présente un remplissage d'une matrice de comparaison où T est le seuil définissant une comparaison ou une non-comparaison, $f_1(d)$ est la fonction de score de comparaison et $f_2(d)$ est la fonction de score de non-comparaison. L'algorithme prend en entrée deux vecteurs V_1 et V_2 de taille s_1 et s_2 respectivement. Dans le tableau 6.1, un remplissage de matrice à partir de deux vecteurs $V_1 = \{8, 3, 11, 32, 21, 7, 6, 15, 20, 7\}$ et $V_2 = \{11, 31, 20, 7, 1, 15\}$ avec les paramètres $T = 1$, $f_1(d) = 4$ et $f_2(d) = 6$ est présenté comme exemple.

Algorithme 4 : Algorithme de distance d'édition : Calcul de la matrice de score

ACCA ($V_1[1 \dots s_1], V_2[1 \dots s_2]$)

Declaration of the scoring matrix $M[0 \dots s_1, 0 \dots s_2]$

for $i = 0$ to s_1

$M(0, i) = 0$

for $j = 0$ to s_2

$M(j, 0) = 0$

for $i = 1$ to s_1

for $j = 1$ to s_2

$d = |V_1(i) - V_2(j)|$

$$M(i, j) = \begin{cases} M(i-1, j-1) + f_1(d) & \text{si } d \leq T \\ \text{Max}(0, M(i, j-1) - f_2(d), M(i-1, j) - f_2(d)) & \text{si } d > T \end{cases} \quad (6.16)$$

	V1	8	3	11	32	21	7	6	15	20	7
V2	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	4	0	0	0	0	0	0	0
31	0	0	0	0	8	2	0	0	0	0	0
20	0	0	0	0	2	12	6	0	0	4	0
7	0	4	0	0	0	6	16	10	4	0	8
1	0	0	0	0	0	0	10	4	0	0	0
15	0	0	0	0	0	0	4	0	8	2	0

TABLE 6.1 – Exemple de matrice de score

Dans cet exemple, le score de comparaison entre V_1 et V_2 est 16. Les fonctions f_1 et f_2 peuvent être plus complexes et dépendre de la différence d dans l'algorithme 4. Il est cependant nécessaire que les fonctions f_1 et f_2 soient strictement croissantes et que $f_1/f_2 > 1$ afin que le score de comparaison décroisse moins vite qu'il n'augmente durant des comparaisons positives.

La procédure de comparaison est appliquée sur des vecteurs. Les caractéristiques de paumes P formées par les parties réelle et imaginaire de la convolution de la paume avec le filtre de Gabor circulaire sont représentées par un vecteur de taille $S = N \times N \times 2$. Afin d'augmenter le poids d'une longue séquence de bonnes comparaisons, l'équation 6.16 est modifiée. Ainsi, la définition de l'équation est modifiée par une fonction utilisant le dernier score de comparaison.

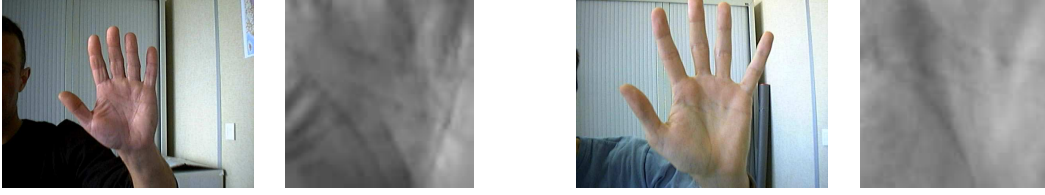


FIGURE 6.7 – Deux exemples de notre base de données et les paumes correspondantes

Pour comparer deux caractéristiques de paumes V_1 et V_2 , nous utilisons l'algorithme 4 avec l'équation suivante :

$$M(i, j) = \begin{cases} f_1(M(i-1, j-1), d) & \text{si } d \leq T \\ f_2(\text{Max}(M(i, j-1), M(i-1, j)), d) & \text{si } d > T \end{cases} \quad (6.17)$$

où T est fixé à 0.1, $f_1(x, d) = x \times a + b$ est la fonction de coïncidence et $f_2(x, d) = x/a$ est la fonction de non-coïncidence. Sans contrainte de temps, Les fonctions peuvent être optimisés en utilisant la différence d et des fonctions plus complexes. La valeur maximale du score de comparaison est définie par $M_{Max} = -q \frac{1-a^S}{1-a} + S * q$ où $q = \frac{b}{1-a}$. La valeur b est fixée empiriquement à 0.01. Avec cette valeur, la valeur a est définie de façon optimale sur notre base par $a = 0.01$. Finalement, la distance entre deux caractéristiques de paumes V_1 et V_2 où la comparaison parfaite est 0 et la plus mauvaise comparaison est 1, est définie par

$$D(V_1, V_2) = 1 - \text{Max}_{0 \leq i < S, 0 \leq j < S} (M(i, j)) / M_{Max} \quad (6.18)$$

6.3.3 Evaluation des performances

Afin d'évaluer le taux de reconnaissance de cette méthode, la base de données contenant 180 images décrites précédemment est utilisée. En plus, les "mauvaise acquisitions" supprimées précédemment sont rajoutées à la base. Ces acquisitions sont principalement des images où la main est présentée de façon orientée face à la caméra. Sur cette base, la méthode est comparée à deux systèmes de l'équipe de Zhang et al. : filtre de Gabor 2D [136] et PPOC [121].

Pour comparer notre système à ces méthodes, nous utilisons une taille de paume de 128×128 . La base considérée est appelée base DB_1 . La base nommée DB_2 contient des paumes de taille 32×32 ce qui limite le temps de calcul et la taille de la référence. Les figures 6.7 montrent deux exemples d'images de main avec les paumes extraites pour la DB_1 . Chaque paume de la base est comparée avec toutes les autres images de paume. Pour augmenter la robustesse du système et limiter l'impact d'une détection imprécise de la main, cinq paumes sont utilisées comme référence. Pour comparer deux images de paumes V_1 et V_2 , quatre paumes de l'utilisateur V_2 sont aléatoirement ajoutées pour former un ensemble de références $V_2^i, i \in [0, 5[$. La distance de comparaison est définie par la distance minimale entre V_1 et les éléments de l'ensemble V_2 :

$$D_f(V_1, V_2) = \text{Min}_{0 \leq i < 5} D(V_1, V_2^i) \quad (6.19)$$

Pour limiter l'impact de l'aléa dans la définition des ensembles de tests, le processus complet d'évaluation est effectué 10 fois.

La méthode proposée est évaluée avec les paramètres optimaux $f = 0.0916$ et $\sigma = 5.6179$ de la base de Zhang. Le taux de comparaison sur la base DB_2 est présenté à la Fig. 6.8. Par

Méthode	Filtres de Gabor 2D	PPOC	Méthode proposée
EER	5.8%	1.7%	1.2%
FAR (quand FRR=1%)	19.8%	13.4%	12.7%
FRR (quand FAR=1%)	6.0%	3.4%	1.5%

TABLE 6.2 – Comparaison du système proposé avec les méthodes de Zhang et al.

comparaison, les résultats des méthodes de Zhang et al. avec les paramètres optimaux et une comparaison un contre cinq sont présentés à la table 6.2. Les résultats montrent que notre système améliore les performances des méthodes de Zhang et al. sur notre base de données. Le processus complet de reconnaissance est effectué en moins d'une seconde sur un Pentium M à 1.60GHz.

6.4 Classification par estimation de distribution non linéaire

Les algorithmes de classifications présentées précédemment s'appuient sur des heuristiques liées à la reconnaissance de la paume. Dans l'approche par estimation de distribution non linéaire, nous avons voulu définir une méthode générique pour classifier les utilisateurs.

Ainsi dans ce paragraphe, une classification par apprentissage sans contre-exemples est proposée (méthodes génératrices). L'apprentissage définit un classifieur spécifique aux données de références et la non utilisation de contre-exemples permet à la fois de n'exclure aucun individu et d'être indépendant du reste de la base.

6.4.1 Extraction des caractéristiques

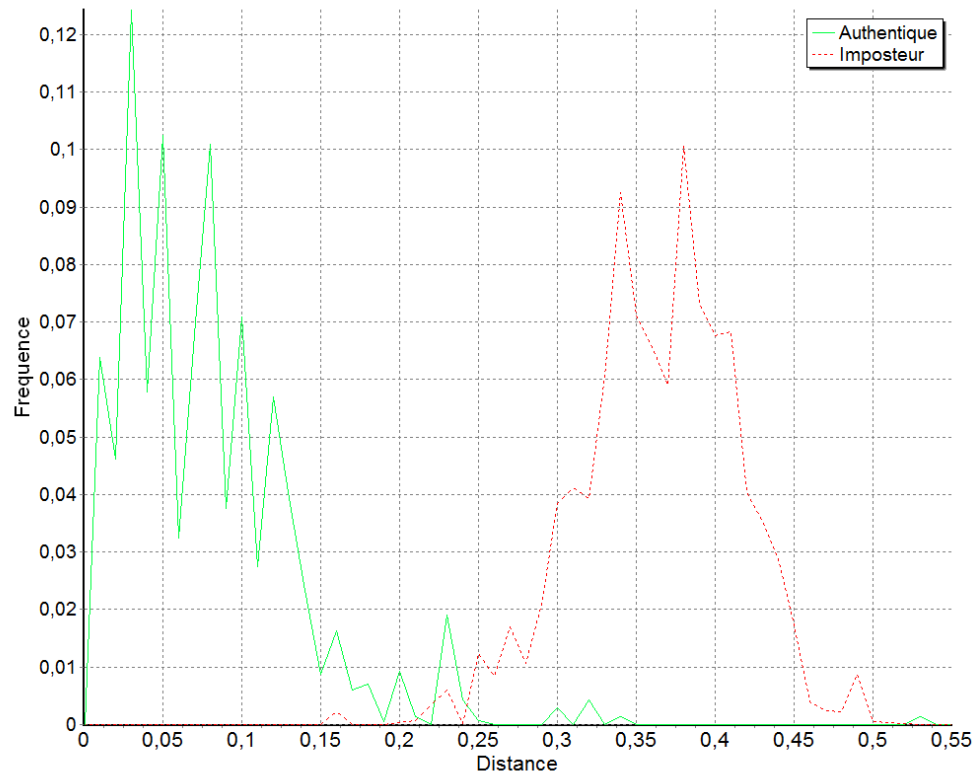
Pour valider cette approche par classification, deux caractéristiques sont évaluées. La première se base sur la texture par un banc de filtres de Gabor. La seconde utilise l'information d'énergie des images de paume comme référence.

6.4.1.1 Extraction par banc de filtres de Gabor

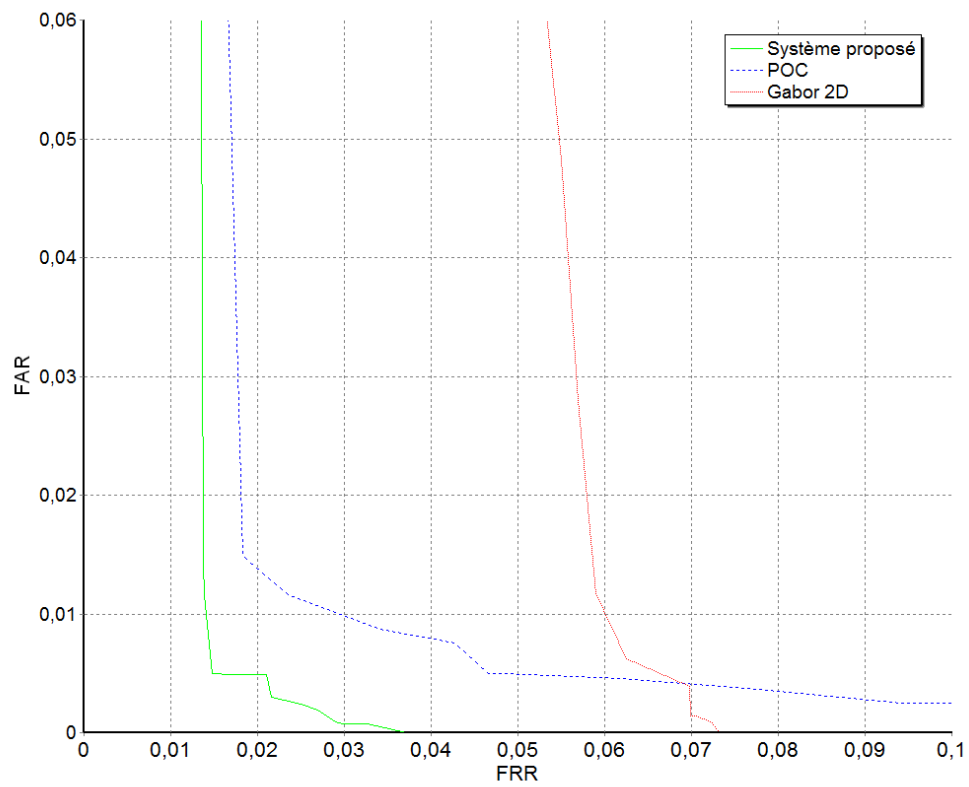
Les filtres de Gabor utilisés pour la définition de la texture sont les mêmes que ceux présentés dans la section 6.2.1. Ils sont également centrés en zéro. Dans les méthodes utilisant ce filtrage, un seul angle est utilisé et l'information filtrée est binarisée afin de limiter la taille de la référence. Afin de garder toute l'information disponible, la texture de la paume est extraite pour différentes orientations (0° , 45° , 90° , 135°) et sans binarisation. Les filtres de Gabor sont respectivement notés suivant ces orientations par Ω_j où $j \in \{0, 1, 2, 3\}$. Soit une paume P , P est convoluée avec chaque filtre Ω_j donnant la référence $R' = \{P_{\Omega_j}\}$ où $j \in \{0, 1, 2, 3\}$ et $P_{\Omega_j} = P * \Omega_j$. Les paramètres des filtres n, φ et σ sont fixés pour limiter le temps de calcul et pour obtenir des résultats optimaux à $n = 9, \varphi = 1.4045$ et $\sigma = 0.3666$.

6.4.1.2 Extraction par analyse robuste des niveaux de gris

Une analyse sur l'énergie des pixels nécessite une normalisation de l'image de référence. Dans la majorité des méthodes le bruit n'est pas pris en considération car les images d'entrées sont prises avec des capteurs adéquats dans un environnement contraint. Les images acquises avec



(a)



(b)

FIGURE 6.8 – Résultat des tests d'authentification : distribution des distances de comparaison (a) et courbe ROC (b)

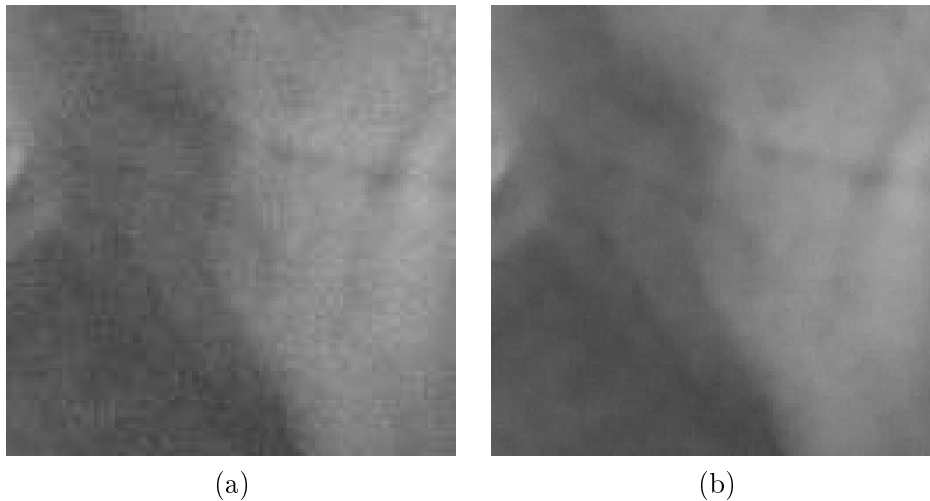


FIGURE 6.9 – Un exemple de restauration d’une image bruitée : (a) image bruitée (b) paume restaurée

un capteur de faible qualité (une webcam ou un capteur de téléphone pour notre système) présentent un bruit important et des artefacts de compression spécialement dans un environnement avec des conditions de luminosité difficiles. Pour les approches utilisant une extraction des informations de texture, ces défauts se trouvent limités par un filtrage de l’image d’entrée. Dans cette méthode, le niveau de gris des pixels sera utilisé. L’impact du bruit devient ainsi beaucoup plus important. Une régularisation par filtrage anisotrope est ainsi proposée [113] afin de lisser l’image en gardant les structures principales de la paume.

La régularisation par diffusion anisotrope est un lissage local. Le lissage est dépendant de l’information de gradient :

Avec un gradient important : le lissage est effectué le long de la direction du gradient,

Dans une région homogène : le lissage est effectué dans toutes les directions.

Cette méthode donne ainsi une régularisation anisotropique de l’image sans détruire les contours. Récemment, Tschumperlé a proposé un lissage anisotropique des images utilisant des équations aux dérivées partielles rapides et préservant les détails fins de l’image. En plus, l’auteur définit une régularisation prenant en considération les courbures des contours de l’image. Cette méthode est basée sur une séparation des processus de géométrie et de lissage respectant ainsi la géométrie locale de lissage. La description complète du processus est décrite dans [113]. Pour illustrer la régularisation, un exemple appliqué sur une image de paume de mauvaise qualité est présenté à la Fig. 6.9.

Il est important de noter que cette restauration est limitée à la paume pour éviter un temps de calcul trop important. En fait, le calcul multi-résolution dans le processus de détection de peau et l’apprentissage de la forme de la main permet de limiter l’impact du bruit lors de la phase de détection.

Dans les approches par Gabor, les filtres de Gabor sont centrés en 0 afin d’être robustes aux changements de luminosité. Le filtrage étant effectué localement la convolution permet d’être

robuste par la régularisation anisotropique, relativement à la taille du filtre, aux changements de luminosité non-homogène sur l'ensemble de l'image. A partir de la paume restaurée, nous proposons de faire une normalisation locale pour être robuste au changement de luminosité non-homogène. Soit I_l une image I avec un impact de luminosité additionnelle et $I(x, y)$ le niveau de gris au point (x, y) . $I_l(x, y)$ est défini par :

$$I_l(x, y) = I(x, y) + L(x, y) \quad (6.20)$$

Où $L(x, y)$ définit l'impact de luminosité au point (x, y) . Une approximation consiste à estimer que deux points proches dans une image ont un impact de luminosité identique. La définition de l'équation 6.21 est redéfinie par :

$$I_l(x, y) = I(x, y) + L_r(x, y) \quad (6.21)$$

Où $L_r(x, y)$ est l'impact de luminosité dans une région proche du point (x, y) . Utilisant cette définition, nous proposons de normaliser l'image de paume par une décomposition et une standardisation de cette image en un ensemble adjacent de régions connexes lesquelles sont supposées avoir le même impact de luminosité.

Premièrement une image de paume est décomposée en Λ régions constituant un ensemble de sous-images $\{R_1, R_2, \dots, R_\Lambda\}$. Le nombre de régions dépend de la taille des sous-images $S \times S$ et de la largeur de l'intersection entre chaque région W . Ensuite, chaque région R_i est normalisée à une moyenne et une variance prédéfinie. La moyenne ϕ_i et la variance ρ_i de chaque région sont calculées avant standardisation. Avec ces données, chaque région normalisée R'_i est déterminée par :

$$R'_i = \begin{cases} \phi_i^d + \kappa(x, y) & \text{si } R_i(x, y) > \phi_i \\ \phi_i^d - \kappa(x, y) & \text{sinon} \end{cases} \quad (6.22)$$

Où

$$\kappa(x, y) = \sqrt{\frac{\rho_i^d (R_i(x, y) - \phi_i)^2}{\rho_i}} \quad (6.23)$$

Et ϕ_i^d et ρ_i^d sont respectivement la moyenne et la variance désirées de la région R_i . Finalement, la caractéristique de paume R' est un ensemble de sous-images normalisées $\{R'_1, R'_2, \dots, R'_\Lambda\}$.

Comme chaque utilisateur a des paumes différentes, les données désirées (moyenne et variance) sont définies individuellement pour chaque utilisateur pour garder le plus d'information possible. Dans la phase d'apprentissage, ces valeurs sont fixées par la valeur médiane de la moyenne et de la variance de chaque région de paume pour chaque utilisateur. La figure 6.10 illustre la décomposition et la normalisation de la paume restaurée à la figure 6.9 avec les paramètres $S = N/2$ et $W = N/4$.

6.4.2 Comparaison par estimation de distribution non linéaire

A partir des caractéristiques de paume, une fonction de décision doit être définie pour vérifier si deux références sont similaires. Dans cette approche, une méthode de classification par estimation du support d'une distribution de haute dimension [92] est utilisée. De par leurs bonnes performances de classification les machines à vecteurs supports (SVM) introduites par Vapnik

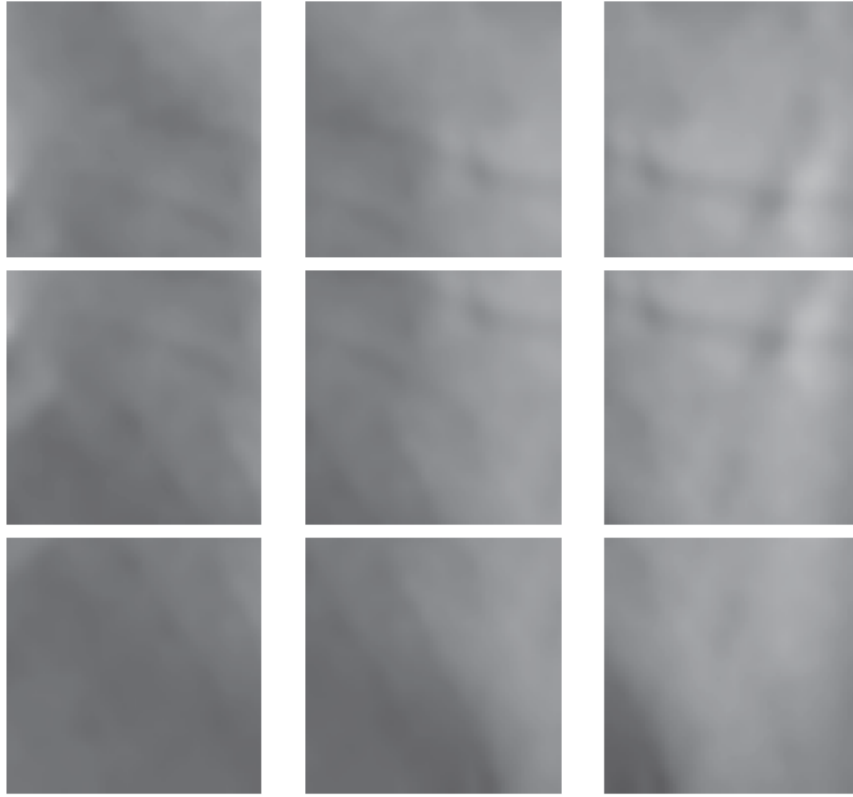


FIGURE 6.10 – Décomposition et normalisation de la paume restaurée de la figure 6.9 en neuf sous-images

[115] sont évaluées sur notre problème.

6.4.2.1 Estimation de la distribution

Initialement, les SVM permettent de classifier des éléments dans différentes classes avec des fonctions linéaires en projetant les données dans un espace de grandes dimensions. Dans [92], Scholkopf et al. proposent une extension de l'algorithme de classifications dans le cas de données non labélisées en estimant le support d'une distribution de haute dimension.

Soit un ensemble de données $\{x_1, \dots, x_l\}, i = 1, \dots, l$ où $l \in \mathbb{N}$ est le nombre d'exemples et $x_i \in \mathbb{R}^d$ où $d \in \mathbb{N}$ est la dimension de la donnée. Soit Φ la fonction de projection $\mathbb{R}^d \rightarrow F$, une fonction noyau $K(x, y) = \langle \Phi(x), \Phi(y) \rangle$ est définie. Des fonctions noyaux typiques sont décrites dans le paragraphe 4.2.1.2. Le but de la méthode est de déterminer une fonction f qui retourne $+1$ dans une région proche de l'ensemble de données and -1 sinon. L'objectif est de projeter les données dans un espace caractéristique and de les séparer avec une marge maximum. L'ensemble de données peut être séparé à partir de l'origine en résolvant le problème suivant :

$$\min_{w, \xi, \rho} \frac{1}{2} w^T w + \frac{1}{vl} \sum_{i=1}^l \xi_i - \rho \quad (6.24)$$

$$\text{avec } w^T \Phi(x_i) \geq \rho - \xi_i, \xi_i \geq 0. \quad (6.25)$$

où $\xi \in \mathbb{R}^l$, $\rho \in \mathbb{R}$, $w \in F$, $v \in (0, 1]$ contrôlent le nombre de vecteurs supports et les erreurs. v est une limite supérieure des erreurs (points d'entraînement en dehors de la région définie) et une limite inférieure des vecteurs supports. Le problème dual peut être obtenu en introduisant un Lagrangien. Il est ainsi exprimé par :

$$\min_{\alpha} \frac{1}{2} \sum_{ij} \alpha_i \alpha_j K(x_i, x_j) \quad (6.26)$$

$$\text{avec } 0 \leq \alpha_i \leq \frac{1}{vl}, \sum_{i=1}^l \alpha_i = 1. \quad (6.27)$$

où α_i est un multiplicateur de Lagrange et $\alpha_i \geq 0$. En utilisant le Lagrangien, la fonction de décision f peut être exprimée avec la fonction noyau par :

$$f(x) = \text{sign}\left(\sum_{i=1}^l \alpha_i K(x_i, x) - \rho\right) \quad (6.28)$$

$$\text{où } \text{sign}(a) = \begin{cases} 1 & \text{si } a \geq t \\ -1 & \text{sinon} \end{cases} \quad (6.29)$$

La démonstration complète de la définition de cette fonction est présentée dans l'article de Scholkopf et al. [92]. Classiquement, le seuil t est fixé par la valeur 0.

6.4.2.2 Fonction de reconnaissance individuelle

Pour la comparaison de paumes, nous proposons de fixer une fonction de décision pour chaque individu. Ainsi, dans la phase d'entraînement une fonction g_{I_d} est estimée pour chaque individu I_d . Les caractéristiques des paumes de chaque individu R'_{I_d} définies précédemment (extraction par banc de filtres de Gabor ou normalisation de l'énergie) sont transformées en un vecteur $1D$ X_{I_d} par concaténation de chaque élément de la référence R'_{I_d} ligne par ligne.

Γ références $X_{I_d}^j$, $j = 1, \dots, \Gamma$ définies de classe +1 sont utilisées comme données d'entraînement. Comme l'extraction de la paume est un processus imparfait, chaque caractéristique de paume utilisée pour définir la fonction de décision est translatée (verticalement et horizontalement) et subit une rotation pour augmenter la taille de l'ensemble d'entraînement. Les valeurs de translations et de rotations sont définies de -2 à $+2$ pixels et de -2° à $+2^\circ$ respectivement.

La phase d'entraînement est très rapide avec si peu d'exemples mais il n'est pas possible de déterminer un seuil fixe t pour tous les utilisateurs pour séparer les imposteurs et les véritables utilisateurs. Ce facteur est lié au fait que la fonction de décision est seulement dépendante d'un ensemble d'exemples positifs sans être contre-balancé par un ensemble de contre-exemples fixes. Pour chaque utilisateur I_d un seuil t_{I_d} doit ainsi être fixé. Ce seuil t_{I_d} est calculé en projetant les

références $X_{I_d}^j$, $j = 1, \dots, \Gamma$ dans la fonction g_{I_d} . Ensuite, la valeur médiane de ces projections est extraite et ajustée avec une valeur τ pour définir le seuil de décision. Cette définition peut être explicitée par :

$$t_{I_d} = \text{medium}_j(g_{I_d}(X_{I_d}^j)) - \tau \quad (6.30)$$

Le paramètre τ limite l'impact d'une mauvaise extraction de la main et du bruit dans les images. Il indique également la disparité maximale acceptée dans l'ensemble d'apprentissage. La valeur médiane est utilisée pour introduire une notion de variabilité entre les paumes d'apprentissage et la fonction de décision. Elle permet contrairement à une valeur moyenne d'être plus robuste et de moins prendre en compte des images de mauvaises qualités.

6.4.3 Évaluation des performances

Comme dans les études précédentes, une base d'évaluation est constituée. La base précédente contenant 180 images est complétée pour finalement contenir 49 personnes. 10 images sont acquises pour chaque individu dans le même contexte que précédemment.

Soit $X_{I_d}^j$ la caractéristique de la paume j d'un utilisateur I_d . Chaque utilisateur I_d est décrit par une fonction de décision g_{I_d} déterminée par l'estimation de la distribution. Les données de chaque utilisateur pour permettre la définition de cette fonction sont définies par l'ensemble $\{X_{I_d}^j\}(j = 1, \dots, \Gamma)$ de classe +1 où Γ indique le nombre d'images utilisées pour définir la fonction de décision. Dans la phase d'évaluation, une base de tests est également définie pour chaque utilisateur. Cette base s'exprime pour un identifiant I_d par l'ensemble de références $\{X_{I_d}^j\}(j = \Gamma + 1, \dots, 10)$ de classe +1 et la base d'imposteurs $\{X_{I_e}^j\}(e = \{1, \dots, 49\} - \{d\}, j = 1, \dots, 10)$ de classe -1. Pour augmenter le nombre de tests et limiter l'impact du choix des références, le processus complet d'expérimentation (apprentissage et évaluation) est effectué 10 fois en changeant aléatoirement l'index j de chaque référence. Le taux d'erreur indiqué est le taux d'égale erreur moyen sur les 10 expérimentations.

Dans les phases d'évaluation, les taux de classification sont déterminés. Pour les caractéristiques extraites par analyse robuste des niveaux de gris, différentes valeurs de paramètres Γ , K et τ sont testés pour définir le meilleur taux de comparaison. Le tableau 6.3 montre les performances du système suivant les variables Γ et τ . Cette évaluation montre comme attendu que plus le nombre de données d'apprentissage est important et plus le taux de reconnaissance augmente. Avec un τ fixé optimalement à 0.05, le taux de reconnaissance est de 98.5% avec sept images d'entraînement. Le processus d'authentification est calculé avec un noyau K de type *RBF* (fonctions à base radiale). Cependant, la sélection de la fonction du noyau est un problème ouvert et difficile. Les résultats de reconnaissance avec les fonctions linéaires, polynomiales et *RBF* sont montrés à la Fig. 6.11 et indique que le *RBF* est optimal pour notre système.

La méthode utilisant l'extraction par un banc de filtres de Gabor a été évaluée avec les paramètres optimaux définis précédemment. Le taux d'égale erreur optimal obtenu avec ces paramètres est de 1.7%.

Le temps d'enrôlement est dépendant du nombre d'exemples d'entraînement. Avec Γ égal à 5, le calcul de la fonction de décision est effectué en 4.7 secondes. Le processus complet de reconnaissance est toujours effectué en moins de 1 seconde.

τ	$\Gamma = 3$	$\Gamma = 4$	$\Gamma = 5$	$\Gamma = 6$	$\Gamma = 7$
0.02	81.5%	83.5%	84.7%	85.8%	86.1%
0.03	85.2%	88.1%	92.5%	93.4%	93.6%
0.04	92.5%	94.9%	96.2%	97.0%	97.3%
0.05	93.7%	96.0%	97.3%	98.4%	98.5%
0.06	88.1%	92.5%	92.7%	93.3%	93.6%
0.07	83.2%	84.2%	85.9%	86.3%	86.4%

TABLE 6.3 – Comparaison des performances du système suivant le nombre d'exemples d'entraînement et le seuil d'ajustement τ

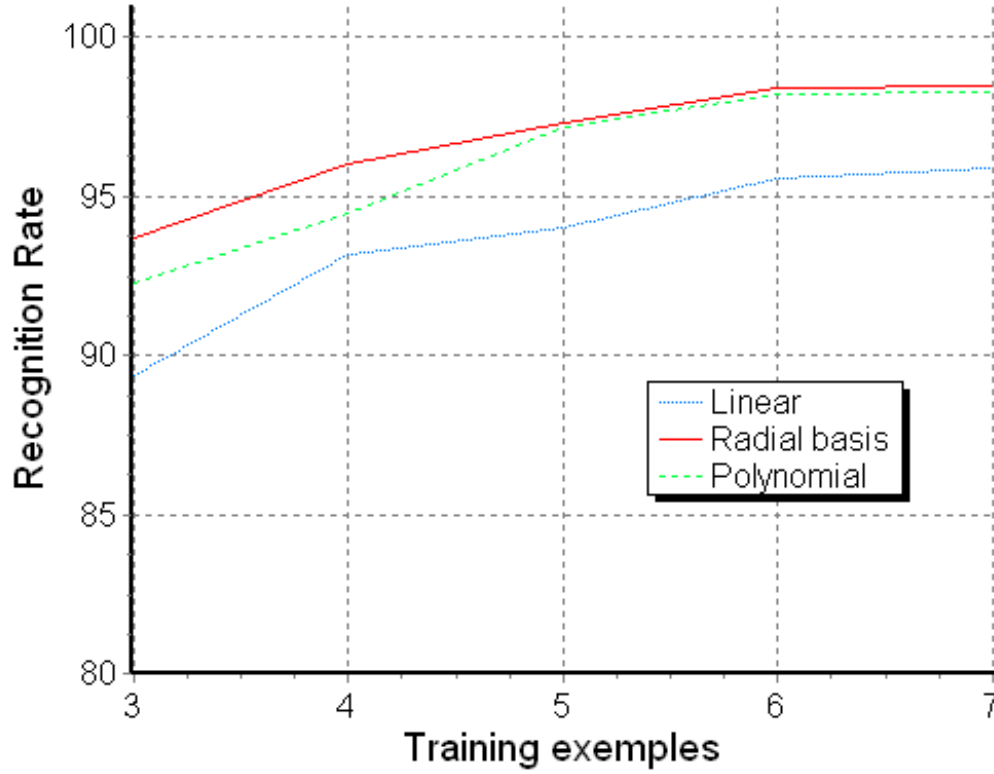


FIGURE 6.11 – Taux de reconnaissance avec différents noyaux

6.5 Etude sur la robustesse aux paramètres extérieurs

Dans les trois méthodes présentées précédemment, l'impact des paramètres extérieurs sur la reconnaissance n'est pas abordé. Dans cette section, nous avons évalué la robustesse de la méthode de Zhang et al. basée sur la reconnaissance par filtre de Gabor à nos conditions d'utilisation. A titre de comparaison, les filtres log-normaux sont également testés.

6.5.1 Description des filtres log-normaux

Les filtres log-normaux sont utilisés pour réaliser l'échantillonnage du spectre photométrique, ils permettent de développer une technique pour estimer la fréquence locale qui tire avantage de l'ensemble des échelles disponibles après filtrage. Ils sont à variables séparables en fréquence et en orientation, ce qui permet d'obtenir une méthode robuste de l'estimation de la fréquence locale de l'image.

La conception de l'opérateur de filtrage log-normal étudié est tiré de l'article de Massot et Herault [78]. L'équation du filtre log-normal est exprimé par :

$$|G_{i,j}(f, \theta)|^2 = A \times \frac{1}{f^2} \times \exp \left[-\frac{1}{2} \left(\frac{\ln(f/f_i)}{\sigma_r} \right)^2 \right] \times \cos^{2n} \left[\frac{\theta - \theta_j}{2} \right] \quad (6.31)$$

Avec la normalisation de la fonction A définie par :

$$A = \frac{f_i^2}{e^{2\sigma^2}} \quad (6.32)$$

Où f_i est la fréquence centrale du filtre, θ_j est l'orientation centrale, σ_r est la bande passante suivant les fréquences et n contrôle la bande passante suivant les orientations.

On remarque que dans la définition les composantes radiales et fréquentielles sont séparées et exprimées en coordonnées polaires. Pour la caractérisation de paume, le filtre est centré au point $(0,0)$ afin d'être robuste aux changements de luminosité. Les figures 6.12 et 6.13 présentent un filtre log-normal et son application sur une image de paume vis-à-vis d'un filtrage par un filtre de Gabor.

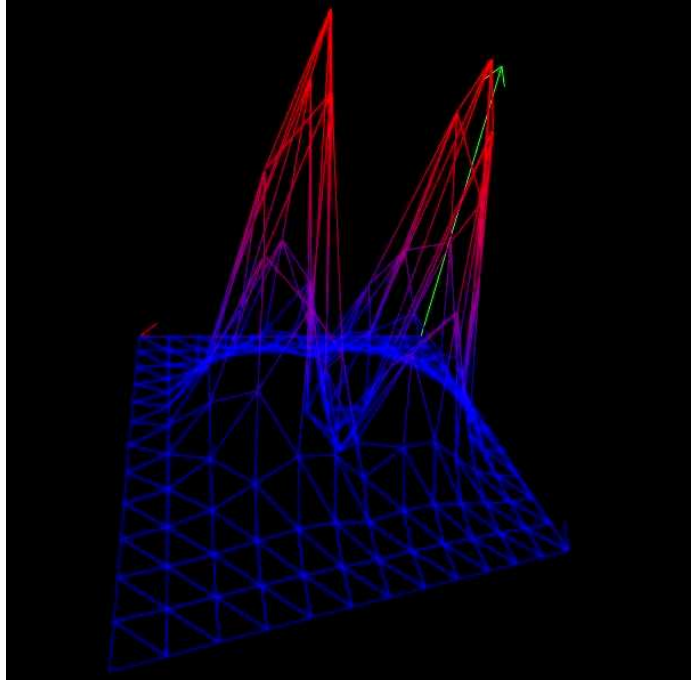


FIGURE 6.12 – Filtre bidimensionnel log-normal en coordonnées log-polaires

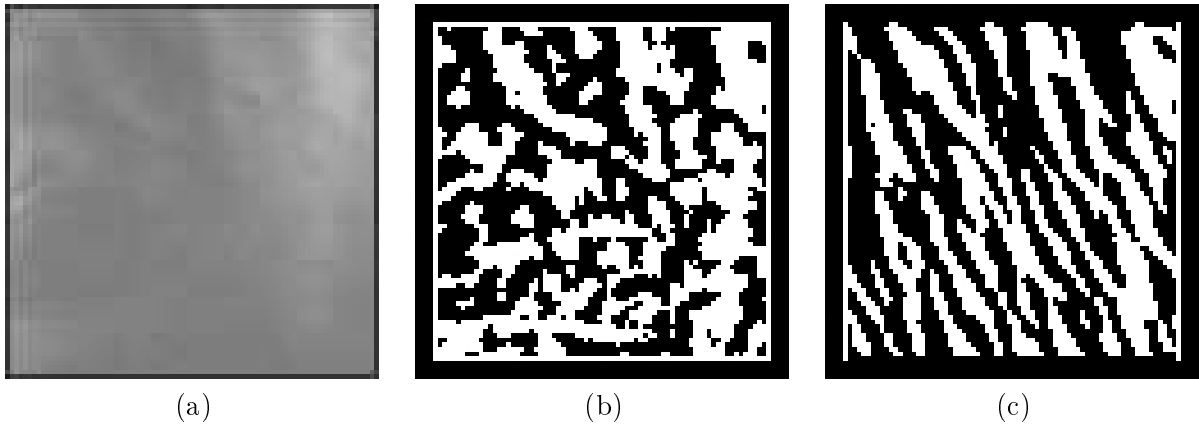


FIGURE 6.13 – Exemple de filtrages : (a) image originale, (b) image filtrée par filtre log-normal et (c) image filtrée par filtre de Gabor

6.5.2 Evaluation des performances suivant les perturbations extérieures

Les deux filtres sont comparés suivant l'effet des sept perturbations suivantes :

- les changements de luminosité,
- les changements de contraste,
- le flou (Gaussien),
- le bruit blanc,
- les translations,
- les rotations 2D et 3D,
- les zooms.

Pour l'évaluation, la distance de comparaison utilisée est la distance de Hamming normalisée. Les perturbations sont évaluées les unes après les autres indépendamment. Pour cette étude, chaque image de la base de paumes de main subit la perturbation étudiée. Ensuite, les images obtenues sont ajoutées à la base originale afin d'obtenir les erreurs de reconnaissance induite par la perturbation. Du fait de l'ajout des données de déformation, le nombre d'échantillons se trouve augmenté par rapport à la base initiale. Les scores de performance dans les tests de robustesse sont élevés car les images sont comparées à elles-mêmes avec quelques perturbations.

A titre de référence, la base est évaluée sans pré- ou post- traitement avec une seule image de référence. Toutes les valeurs indiquées dans le tableau 6.4 sont présentées pour un seuil fixe de 0.3 qui correspond à un TFA faible.

	FAR	FRR
Log-normal	0.512	16.780
Gabor	1.025	16.180

TABLE 6.4 – Référence pour l'évaluation de la robustesse aux variations

6.5.2.1 Les changements de luminosité et de contraste

Du fait de la normalisation au point $(0, 0)$, les systèmes de reconnaissance par filtres de Gabor ou par filtres log-normaux sont très robustes aux changements de luminosité et de contraste. Ainsi, aucune variation dans les performances n'est observée suivant ces deux variations. Un exemple de mesure de distances avec une image dans la base montre ce constat dans le tableau 6.5.

Luminosité (en%)	Log-normal	Gabor
-60	0.0311	0.0277
-40	0.0294	0.0263
-20	0.0204	0.0209
+20	0.0238	0.0195
+40	0.0296	0.0263
+60	0.0311	0.0277

Contraste (en%)	Log-normal	Gabor
-60	0.0401	0.0345
-40	0.0278	0.0300
-20	0.0312	0.0271
+20	0.0264	0.0205
+40	0.0174	0.0184
+60	0.0147	0.0162

TABLE 6.5 – Exemple de distance selon la luminosité et le contraste entre une image de référence et l'image ayant subi la perturbation

6.5.2.2 Le flou Gaussien et le bruit blanc

Comme précédemment, les performances sont très peu dégradées avec un ajout de flou Gaussien ou de bruit blanc sur les images de la base de données. Une illustration d'un ajout de bruit ou de flou est affichée à la Fig. 6.14. Un exemple de mesure avec une image de la base illustre la robustesse du système tableau 6.6.

6.5.2.3 Les translations, rotations et zooms

Comme indiqué précédemment, le taux d'erreur est plus faible que dans la base originale car les déformations sont comparées aux images originales ce qui augmente la qualité de reconnaissance. Cependant, cette étude a pour objectif de tester la robustesse du système et non d'établir des performances réelles suivant des déformations.

Les tableaux 6.7,6.8,6.10 présentent les résultats de reconnaissance de la paume de la main suivant les déformations de translation, rotation et zoom. A titre de comparaison en testant la base d'entrée en ajoutant les images sans déformations (ce qui revient à avoir des doublons d'images identiques), le taux de reconnaissance devient égal à 98.8%. Les perturbations dans les tableaux sont en valeur absolue. Par exemple, pour un angle de $|5^\circ C|$, la base est complétée par

Bruit (en dispersion)	Log-normal	Gabor
5	0.1529	0.1466
10	0.2394	0.2323
20	0.3035	0.3063

Flou Gaussien (en écart type)	Log-normal	Gabor
1.0	0.0614	0.0486
3.0	0.1184	0.0953
9.0	0.2634	0.2403

TABLE 6.6 – Exemple de distance selon le bruit et le flou entre une image de référence et l'image ayant subi la perturbation

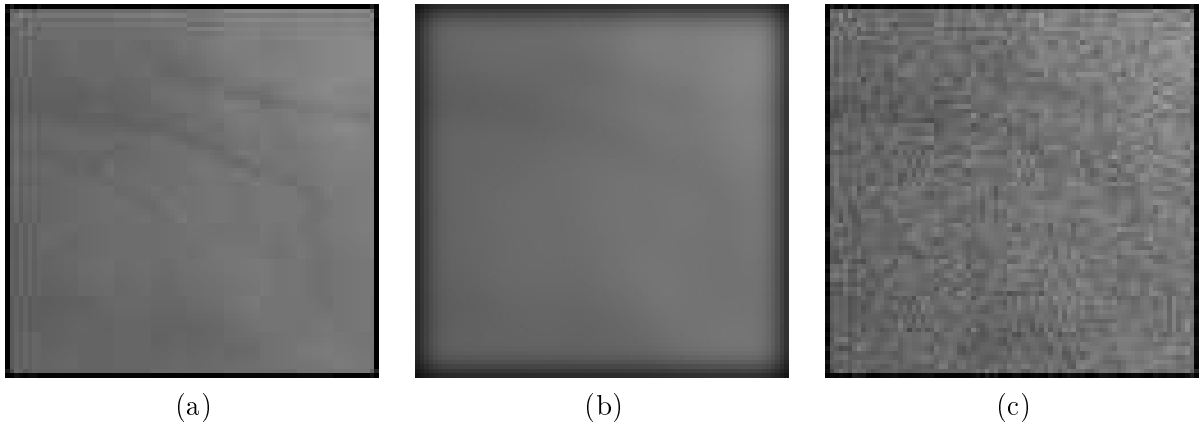


FIGURE 6.14 – Exemple de bruits et de flous : (a) image originale, (b) image ayant subi un flou Gaussien d'écart type 9 et (c) image bruitée avec une dispersion de 20

des images ayant subi des perturbations de $5^{\circ}C$ et de $-5^{\circ}C$.

Angle (en $^{\circ}C$)	Log-Normal		Gabor	
	FAR	FRR	FAR	FRR
5	0.279	3.348	0.1223	3.5385
20	0.061	4.873	0.099	4.847
45	0.048	5.277	0.031	5.314
90	0.039	5.593	0.026	5.610

TABLE 6.7 – Robustesse à la rotation

Les résultats selon ces différentes contraintes sont presque similaires entre l'extraction des textures par un filtre de Gabor et un filtre log-normal. On remarque également que logiquement plus la perturbation est importante et plus le taux de reconnaissance décroît. Cependant malgré de très fortes perturbations, le taux de reconnaissance baisse au maximum de 4% par rapport à la référence ce qui montre une relative bonne robustesse à ces types de perturbation dans le plan.

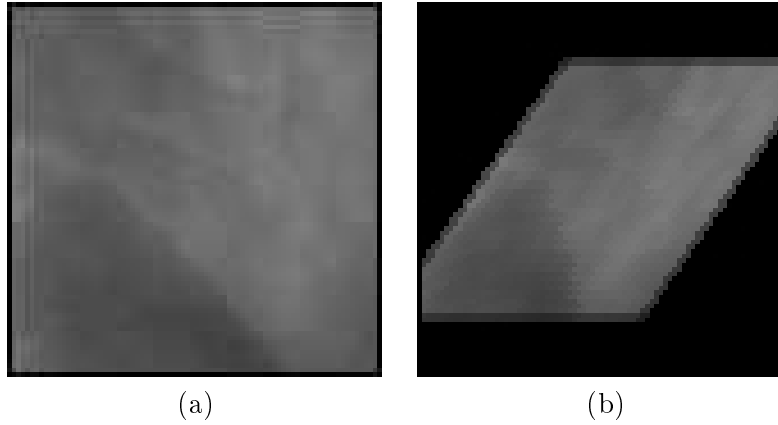
Zoom (en %)	Log-Normal		Gabor	
	FAR	FRR	FAR	FRR
150	0.066	3.815	0.171	3.732
300	0.047	4.993	0.024	5.083
500	0.044	5.287	0.022	5.371

TABLE 6.8 – Robustesse au zoom

Translation (en pixels)	Log-Normal		Gabor	
	FAR	FRR	FAR	FRR
5	0.053	5.030	0.066	4.887
10	0.036	5.569	0.050	5.473
25	0.013	5.826	0.016	5.820

TABLE 6.9 – Robustesse au translation

Comme le système de reconnaissance subit en réalité des perturbations dans l'espace, celles-ci sont également étudiées. Pour cela, les images de paumes subissent des rotations suivant le *tilt* et le *slant*, c'est-à-dire selon l'axe médian vertical et horizontal. Un exemple de rotation dans l'espace d'une image de paume est illustrée à la Fig. 6.15.


 FIGURE 6.15 – Exemple de rotation 3D : (a) image originale et (b) image tournée de $45^\circ C$ en slant et en tilt

Tilt (en $^\circ C$)	Slant (en $^\circ C$)	Log-Normal		Gabor	
		FAR	FRR	FAR	FRR
45	0	0.109	3.188	0.127	2.886
70	0	0.248	4.531	0.250	3.967
0	45	0.185	3.309	0.220	3.716
0	70	0.297	4.475	0.142	4.812
45	45	0.296	3.869	0.284	3.813
70	70	0.214	5.433	0.167	5.524

TABLE 6.10 – Robustesse au translation

Le tableau 6.10 récapitule les résultats de comparaison pour différentes rotations dans l'espace. En ce qui concerne le tilt qui est la rotation la plus courante dans notre base de données, le filtre de Gabor présente de bien meilleurs résultats que le filtre log-normal (près de 1%). Il est intéressant de remarquer que pour des rotations inférieures à $45^\circ C$ qui correspondent à l'erreur maximale avec un utilisateur coopératif (qui tente de placer sa main face à la caméra), les résultats restent stables vis-à-vis des autres perturbations spatiales testées.

6.5.2.4 Résumé des performances

Les différentes analyses présentent les performances du système de reconnaissance par la main selon plusieurs perturbations.

Tout d'abord, la binarisation des filtres autour du point $(0, 0)$ permet une robustesse très forte aux changements de contraste et de luminosité. Pour les perturbations spatiales dans le plan de l'image, les différences de performance entre les deux filtres testés sont très faibles. Le taux de reconnaissance déterminé indique une robustesse intéressante à ces perturbations. Il est à noter que la distance de Hamming robuste aux faibles translations présentée précédemment permet un gain de performance très significatif puisque le taux de reconnaissance en mode authentification devient égal à 98.061%.

Les tests sur la rotation 3D permettent quand à eux de tirer certaines conclusions et de mettre en avant les différences entre ces deux filtres : le filtre de Gabor est plus robuste aux rotations de tilt que le filtre log-normal. Les performances selon ces perturbations indiquent cependant une perte de performance du système face à une rotation non prise en compte dans l'extraction de la paume. Cette perte de performances est de l'ordre de 2% pour une rotation maximale de $45^\circ C$.

6.6 Conclusion sur la reconnaissance de la main sans contact

Dans ce chapitre, plusieurs approches pour la reconnaissance de la main sans contact sont présentées. La première méthode combine les informations de forme et de texture par convolution. Ensuite, une approche plus robuste généralisant la méthode de Zhang et al. [136] est décrite. La méthode de comparaison est la distance d'édition qui permet une bonne robustesse aux bruits. Le taux d'égale erreur avec cette méthode, de 1.2%, améliore sur notre base de données les approches antérieures.

Une autre approche par classification d'estimation de distribution non linéaire spécifique est présentée. Les caractéristiques par filtrage de Gabor et par analyse de l'énergie sont évaluées sur une base de données de 49 personnes. Le taux d'égale erreur pour cette approche est de 1.5% dans le meilleur des cas. Dans toutes les approches présentées, le processus complet de reconnaissance est effectué en moins de 1 seconde sur un Pentium M à 1.60GHz.

De plus, une étude montre la robustesse du filtrage par Gabor aux paramètres extérieurs et aux imprécisions de calculs. Cette étude précise cependant que la rotation dans l'espace limite la reconnaissance avec des acquisitions de la paume dans un plan non-orthogonal à l'axe de la caméra.

Chapitre 7

Conclusion et perspectives

Notre travail s'est effectué dans le domaine de la reconnaissance biométrique. Dans le cadre des télécoms, nous avons proposé le premier système biométrique de reconnaissance de la main sans contact. La thèse s'est décomposée en deux parties principales :

Segmentation de la main : La segmentation de la main sans contact est la première phase de l'application de reconnaissance biométrique. Elle a consisté à extraire la main dans une image complexe puis d'en définir la zone centrale représentant la paume.

Reconnaissance biométrique : A partir de la segmentation, des algorithmes de reconnaissance ont été définis. Ces algorithmes ont permis la caractérisation de la paume à analyser dans l'objectif d'une reconnaissance biométrique.

Résumé

Dans la première partie de la thèse, une nouvelle approche fusionnant les informations de forme et de couleur pour la segmentation de la main dans un fond complexe a été développée. Les résultats expérimentaux ont montré que cette méthode a permis un gain important en performance en terme de précision et de robustesse. Le principe de la détection s'appuie sur une pré-segmentation à basse résolution par un réseau de neurones multi-échelle. Cette pré-segmentation initialise le module de recherche basé sur les contours et limite les déformations de celui-ci. Bien que définie dans le cadre de la segmentation de main, cette approche peut être généralisée pour la segmentation d'objets de formes et de couleurs connues.

La seconde étape de la thèse a permis de déterminer plusieurs méthodes de caractérisation de la paume et d'en illustrer également leurs applications pour la reconnaissance des veines.

Tout d'abord, une approche combinant les informations de forme et de texture a été proposée. Ensuite, une généralisation de la méthode de référence de Zhang et al.[136] utilisant la distance d'édition est définie. Cette méthode a été évaluée puis comparée aux principales approches de la littérature. Cette comparaison montre que nous obtenons de meilleures performances que les autres approches de reconnaissance palmaire sur notre base de données. Nous avons également mené une étude sur la robustesse du filtre de Gabor aux paramètres externes. Celle-ci indique

que le filtre de Gabor est bien robuste à la majorité des perturbations extérieures.

Finalement, une méthode de reconnaissance de la paume a été définie en analysant la distribution non-linéaire des paumes de la main. Cette approche générique permet à l'aide d'une fonction de décision définie spécifiquement de déterminer de façon robuste et élégante l'identité d'un individu dans un temps de reconnaissance raisonnable.

Perspectives

Cette thèse a couvert plusieurs aspects de la reconnaissance biométrique de la main incluant la segmentation, la vitesse d'exécution, la robustesse au contexte ou la caractérisation de paume. Sur la base de ce travail, plusieurs axes de recherche restent ouverts afin d'améliorer les performances du système :

Performance : La performance est l'objectif numéro 1 en reconnaissance biométrique. Plusieurs méthodes sont présentées dans le document. L'objectif comme dans toutes les méthodes en biométrie est d'augmenter la précision de la reconnaissance tout en maintenant une vitesse de traitement acceptable.

Base de données : Dans cette thèse, les performances sont évaluées sur une base de données créée en une seule session. Il serait intéressant d'examiner les performances des méthodes dans la durée en effectuant plusieurs sessions. Afin d'obtenir des performances plus précises, il conviendrait également de compléter cette base.

Segmentation : La méthode de segmentation est robuste aux changements de luminosité raisonnables. Cependant dans des conditions extérieures très défavorables, une initialisation du contour par la couleur de la peau n'est pas suffisante. Des approches par réseau de neurones convolutionnels semblent intéressantes pour résoudre cette difficulté.

Extraction de la paume : La caractérisation de la paume dépend de la qualité de la segmentation de la main et de l'extraction de la paume. L'étude sur la caractérisation de la paume présentée dans le document a montré que la rotation dans l'espace influe fortement dans les performances. Une étude sur une meilleure extraction de la paume pour ce contexte pourrait être envisagée.

Attaque : En plus des ces éléments, une analyse et une évaluation sur les vulnérabilités du système biométrique sont nécessaires. Il est par exemple important de noter que pour le moment, le système ne fait aucune différence entre une photo et une acquisition réelle d'une main.

Publications

1. J. Doublet, O. Lepetit et M. Revenu. Hand Detection for Contactless Biometrics Identification. Proc. Conf on COgnitive Systems with Interactive Sensors (COGIS 2006), Paris, France, 2006.

-
2. J. Doublet, O. Lepetit et M. Revenu. Reconnaissance Biométrique sans contact de la main intégrant des informations de formes et de textures. COMpression et REprésentation des Signaux Audiovisuels (CORESA 2006), Caen, France, 2006.
 3. J. Doublet, O. Lepetit et M. Revenu. Contact less Hand Recognition Using Shape and Texture Features. Proc. Conf. on 8th International Conference on Signal Processing (ICSP 2006), Guilin, China, 2006.
 4. J. Doublet, O. Lepetit et M. Revenu. Contact less Palmprint Authentication Using Circular Gabor Filter and Approximated String Matching. Signal and Image Processing (SIP 2007) Honolulu, USA, 2007.
 5. J. Doublet, O. Lepetit et M. Revenu. Contact less Hand Recognition Based on Distribution Estimation. Biometrics Symposium, Baltimore, USA, pages 1-6, 2007.
 6. J. Doublet, O. Lepetit et M. Revenu. Robust GrayScale Distribution Estimation for Contact-less Palmprint Recognition. IEEE Conference on biometrics : theory, applications and systems, Washington ,USA, pages 1-6, 2007.

Annexe A

Physiologie de la main

A.1 Anatomie de la main

La main est la partie du corps située à la fin de chaque bras de tous les primates. La main humaine est constituée d'une paume avec cinq doigts attachés au bout du bras par le poignet. Ces cinq doigts peuvent être séparés en un groupe de quatre doigts au bout de la main (l'index, le majeur, l'annulaire et l'auriculaire) et le pouce situé sur un côté. Les structures de la main déterminant les capacités physiques de mobilité peuvent quant à elles être séparées en cinq catégories [35] :

- **Les os** : une main humaine comporte 27 os (Fig. A.1(a)). Le carpe (poignet) est composé de 8 os arrangés en deux lignes de 4. Le métacarpe (la paume) comporte 5 os, un pour chaque doigt. Les phalanges (doigts) contiennent 14 os, 2 pour le pouce et 3 pour chaque doigt.
- **Les articulations** : Les principales jointures des articulations des doigts sont formées par les connexions des phalanges au métacarpe, elles sont appelées les articulations Méta-carpoPhalangiennes (MCP). Les trois phalanges de chaque doigt sont séparées par deux articulations nommées articulations InterPhalangiennes. Les jointures les plus proches du bout du doigt sont appelées articulations InterPhalangiennes Distales (IPD) tandis que les jointures les plus proches du MCP sont nommées articulations InterPhalangiennes Proximales (IPP). Le pouce a seulement une jointure nommée articulation Interphalangienne du Pouce (IP). Finalement, les articulations Carpo-MétaCarpiennes (CMC) connectent le carpe au métacarpe. Le CMC du pouce est une articulation particulière et complexe spécifiquement nommée Trapézo-MétaCarpienne (TMC).
- **Les ligaments** : Les ligaments sont les larges bandes de tissus qui connectent les os entre eux (Fig. A.1(b)). Deux de ces structures nommées ligaments collatéraux sont situées de chaque côté des jointures des doigts. Le rôle de ces ligaments est de limiter des déformations anormales de courbures aux jointures des phalanges afin que les phalanges de chaque doigt restent alignées. D'autres ligaments connectent la phalange proximale (qui relie le doigt au métacarpe) et la phalange moyenne sur la jointure du côté de la paume. Ces ligaments permettent de garder ces jointures rectilignes limitant les hyper-extensions.
- **Les tendons** : Les tendons (Fig. A.1(b)) permettant à chaque jointure de se redresser

s'appellent les tendons extenseurs. Ces tendons sont reliés aux muscles de l'avant bras. Ainsi, lorsque les muscles extenseurs se contractent, ils tirent sur les tendons extenseurs redressant le doigt. Il est à noter que la flexion d'un doigt peut entraîner naturellement la flexion d'autres doigts. Cette remarque est cependant limitée aux personnes n'étant pas musiciens ; en effet ces derniers s'entraînent à rechercher la parfaite autonomie de chacun de leurs doigts pour jouer de leur instrument.

- **Les muscles :** La plupart des muscles qui contrôlent la main proviennent du coude ou de l'avant bras (Fig. A.1(b)). Certains muscles contrôlent les orientations du poignet tandis que d'autres influencent le mouvement des doigts ou du pouce. De même, des muscles aident à placer et tenir le poignet et la main lorsque le pouce et les doigts saisissent ou effectuent des actions précises. Deux autres muscles permettent au pouce de se déplacer de l'autre côté de la paume de la main. Finalement, les plus petits muscles appelés muscles intrinsèques guident les mouvements fins des doigts afin qu'ils soient bien positionnés et stables durant l'activité de la main.

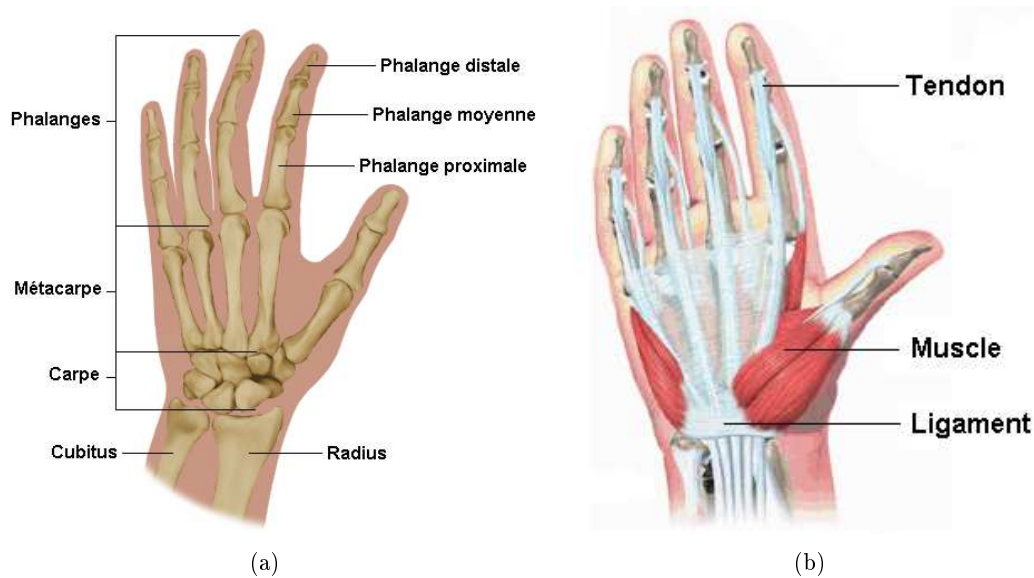


FIGURE A.1 – Anatomie de la main : (a) squelette et (b) structure interne

A.2 Physiologie de la peau

En plus de son anatomie, la main peut être caractérisée par sa couleur. Le teint de la peau humaine schématisée à la Fig. A.2(a) présente une forte variabilité de couleur. Il est bien connu que la génétique joue un rôle majeur dans la détermination de la couleur de la peau mais d'autres facteurs tels que l'exposition au soleil ou certaines pathologies influent également.

Sur le plan anatomique [41], la peau est composée de trois couches successives. La partie supérieure, la plus mince, s'appelle l'épiderme. La partie intermédiaire nommée le derme est disposée sur l'hypoderme qui est la couche la plus profonde. Le teint de la peau est déterminé par les mélanines qui sont situées dans la couche épaisse de l'épiderme de la peau Fig. A.2(b).

Elles sont créées par les cellules appelées mélanocytes qui ne sont pas réparties de façon homogène sur toute la surface de la peau. Par exemple, les paumes de la main ont un nombre de mélanocytes par surface de peau moins important que le reste du corps et sont ainsi d'une couleur plus claire que les autres parties du corps. Cependant, leur nombre moyen par surface de peau est le même chez tous les êtres humains mais la quantité, la nature et la répartition des mélanines produites sont très variables. Ce sont ces trois critères qui définissent complètement la couleur de la peau d'un être humain. Il est aussi intéressant de noter que l'intérieur de la main (paume et face interne des doigts) est l'une des rares parties du corps humain avec la plante des pieds ou les lèvres qui est dépourvue de poils.

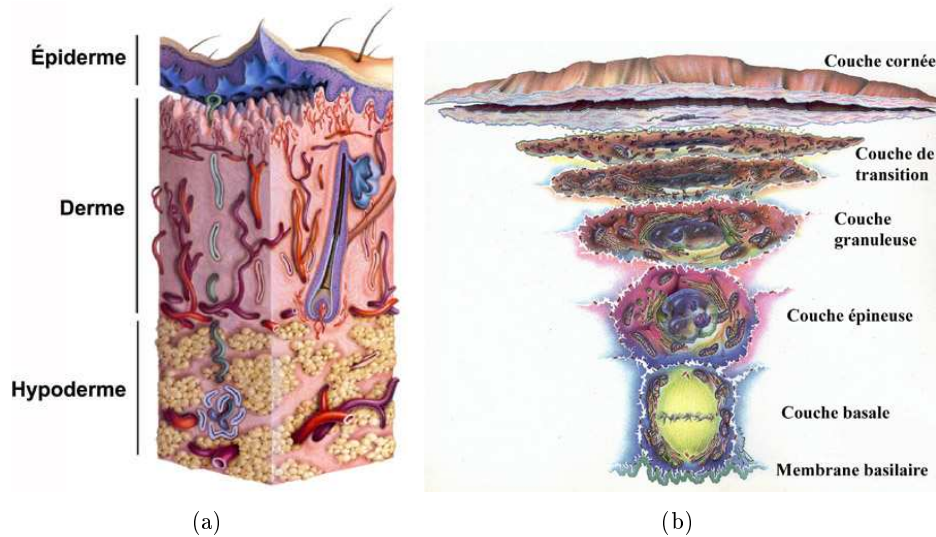


FIGURE A.2 – Schémas de coupe de (a) la peau et de (b) l'épiderme

Bibliographie

- [1] H. K. D. H Bhadesia. Neural networks in materials science. *ISIJ International*, 39(10) :966–979, 1999.
- [2] S. Birchfield. Elliptical head tracking using intensity gradients and color histograms. *In Proceedings of CVPR*, pages 232–237, 1998.
- [3] G. Borgefors. Distance transformations in digital images. *Comput. Vision Graph. Image Process.*, 34(3) :344–371, 1986.
- [4] F. Boussaid, D. Chai, and A. Bouzerdoun. On-chip skin color detection using a triple-well CMOS process. *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 5274 :206–214, 2004.
- [5] J. Brand and J. Mason. A comparative assessment of three approaches to pixellevel human skin-detection. *In Proc of the International Conference on Pattern Recognition*, 1 :1056–1059, 2000.
- [6] T.M. Breuel. Fast recognition using adaptive subdivisions of transformation space. *Computer Vision and Pattern Recognition*, pages 445–451, 1992.
- [7] D. Brown, I. Craw, and J. Lewthwaite. A som based approach to skin detection with application in real time systems. *in Proc. of the British Machine Vision Conference*, 2001.
- [8] L. Brèthes, P. Menezes, F. Lerasle, and J. Hayet. Face tracking and hand gesture recognition for human-robot interaction. *In IEEE 2004 International Conference on Robotics and Automation*, 2 :1901–1906, 2004.
- [9] G. Burel. *Réseaux de neurones en traitement d’images - Des Modèles théoriques aux Applications Industrielles*. PhD thesis, Université de Bretagne occidentale, Brest, 1991.
- [10] C.J.C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2 :121–167, 1998.
- [11] D. Chai and A. Bouzerdoun. A bayesian approach to skin color classification in ycbcr color space. *IEEE TENCON*, 2 :421–424, 2000.
- [12] D. Chai, S. L. Phung, and A. Bouzerdoun. A bayesian skin/non-skin color classifier using non-parametric density estimation. *IEEE International Symposium on Circuits and Systems (ISCAS’2003)*, 2 :464–467, 2003.
- [13] C.C. Chang and C.J. Lin. *LIBSVM : a library for support vector machines*, 2001.
- [14] C. Chen and S.P. Chiang. Detection of human faces in color images. *VISP*, 144(6) :384–388, December 1997.
- [15] Q. Chen, H. Wu, and M. Yachida. Face detection by fuzzy pattern matching. *ICCV ’95 : Proceedings of the Fifth International Conference on Computer Vision*, pages 591–596, 1995.

- [16] C.C. Chiang, W.K. Tai, M.T. Yang, Y.T. Huang, and C.J. Huang. A novel method for detecting lips, eyes and faces in real time. *Real-Time Imaging*, 9(4) :277–287, 2003.
- [17] M. Collobert, R. Feraud, and G.L. Tourneur. Listen : A system for locating and tracking individual speakers. *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 283–286, 1996.
- [18] D. Comaniciu and V. Ramesh. Robust detection and tracking of human faces with an active camera. *Proc. Third IEEE International Workshop on Visual Surveillance*, pages 11–18, 2000.
- [19] T. Connie, A. Teoh, M. Goh, and D. Ngo. Palmprint recognition with pca and ica. *Proceedings of Image and Vision Computing New Zealand*, pages 227–232, 2004.
- [20] T.F. Cootes, D. Cooper, C.J. Taylor, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1) :38–59, 1995.
- [21] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, pages 273–297, 1995.
- [22] J.M. Cross and C.L. Smith. Thermographic imaging of the subcutaneous vascular network of the back of the hand for biometric identification. *Proceedings of IEEE 29th Annual International Carnahan Conference on Security Technology*, pages 20–35, 1995.
- [23] J. G. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(11) :1148–1161, 1993.
- [24] J. G. Daugman. How iris recognition works. *IEEE transactions on circuits and systems for video technology*, 14 :21–30, 2004.
- [25] K. Delac and M. Grgic. A survey of biometric recognition methods. *46th International Symposium Electronics in Marine, Zadar, Croatia*, pages 184–193, 2004.
- [26] R. Deriche. Using canny’s criteria to derive an optimal edge detector recursively implemented. *Int. J. Computer Vision*, 1 :167–187, 1987.
- [27] S. Di Zenzo. A note on the gradient of a multi-image. *Comput. Vision Graph. Image Process.*, 33(1) :116–125, 1986.
- [28] J. Doublet, O. Lepetit, and M. Revenu. Contact less hand recognition using shape and texture features. *Proc. Conf. on 8th International Conference on Signal Processing (ICSP 2006), Guilin, China*, 2006.
- [29] J. Doublet, O. Lepetit, and M. Revenu. Hand detection for contactless biometrics identification. *Proc. Conf on COgnitive Systems with Interactive Sensors (COGIS 2006), Paris, France*, 2006.
- [30] J. Doublet, O. Lepetit, and M. Revenu. Reconnaissance biométrique sans contact de la main intégrant des informations de formes et de textures. *COmpression et REprésentation des Signaux Audiovisuels (CORESA 2006), Caen, France*, 2006.
- [31] J. Doublet, O. Lepetit, and M. Revenu. Contact less hand recognition based on distribution estimation. *Biometrics Symposium, Baltimore, USA*, pages 1–6, 2007.
- [32] J. Doublet, O. Lepetit, and M. Revenu. Contact less palmprint authentication using circular gabor filter and approximated string matching. *Signal and Image Processing (SIP 2007) Honolulu, USA*, 2007.
- [33] J. Doublet, O. Lepetit, and M. Revenu. Robust grayscale distribution estimation for contactless palmprint recognition. *IEEE Conference on biometrics : theory, applications and systems, Washington ,USA*, pages 1–6, 2007.

-
- [34] N. Duta, A.K. Jain, and K.V. Mardia. Matching of palmprints. *Pattern Recognition Letters*, 23(4) :477–485, 2002.
 - [35] e orthopod. A patient’s guide to hand anatomy. *Elsevier Science*, 2006. http://www.eorthopod.com/public/files/Hand_Anatomy.pdf.
 - [36] M. Egmont-Petersen, D. de Ridder, and H. Handels. Image processing with neural networks – a review. *Pattern Recognition*, 35(10) :2279–2301, 2002.
 - [37] A. Erol, G. Bebis, M. Nicolescu R.D. Boyle, and X. Twombly. Vision-based hand pose estimation : A review. *Elsevier Science*, 108 :52–73, 2006.
 - [38] K.C. Fan and C.L. Lin. The using of thermal images of palm-dorsa vein-patterns for biometric verification. *ICPR04*, 4 :450–453, 2004.
 - [39] P. Fieguth and D. Terzopoulos. Color-based tracking of heads and other mobile objects at video frame rates. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 21–27, 1997.
 - [40] Y. Freund. Boosting a weak learning algorithm by majority. *Information and Computation*, pages 256–285, 1995.
 - [41] V. Gagnon. *Etude des Interactions Entre les Nerfs Sensoriels et les Follicules Pileux Dans un Modèle in Vitro De Peau Reconstituée par Génie Tissulaire*. PhD thesis, Faculté de médecine, université Laval, Québec, Canada, 2005.
 - [42] C. Garcia and M. Delakis. Convolutional face finder : A neural architecture for fast and robust face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11) :1408–1423, 2004.
 - [43] C. Garcia and G. Tziritas. Face detection using quantized skin color regions merging and wavelet packet analysis. *IEEE Transactions on Multimedia*, 1(3) :264–277, 1999.
 - [44] G. Gomez and E. Morales. Automatic feature construction and a simple rule induction algorithm for skin detection. *Proc. of the ICML Workshop on Machine Learning in Computer Vision*, pages 31–38, 2002.
 - [45] G. Gomez, M. Sanchez, and Luis Enrique Sucar. On selecting an appropriate colour space for skin detection. *MICAI*, pages 69–78, 2002.
 - [46] C. Han. A hand-based personal authentication using a coarse-to-fine strategy. *IVC*, 22(11) :909–918, 2004.
 - [47] C. Han, H. Chen, C. Lin, and K. Fan. Personal authentication using palmprint features. *Pattern Recognition*, 36(2) :371–381, 2003.
 - [48] R.M. Haralick and L.G. Shapiro. Image segmentation techniques. *Computer Vision, Graphics, and Image Processing*, 29(1) :100–132, 1985.
 - [49] K. Hidai, H. Mizoguchi, and K. Hiraoka. Robust face detection against brightness fluctuation and size variation. *International Conference on Intelligent Robots and Systems*, 1 :127–130, 2000.
 - [50] E. Hjelmas and B.K. Low. Face detection : A survey. *CV&IU*, (83) :236–274, 2001.
 - [51] T. Hoad. Video representations for effective retrieval from large collections. *PhD thesis, Department of Computer Science and Information Technology, RMIT University*, 2004.
 - [52] S.K. Im, H.M. Park, S.W. Kim, C.K. Chung, and H.S. Choi. Improved vein pattern extracting algorithm and its implementation. *Consumer Electronics*, pages 2–3, 2000.

- [53] A. K. Jain, A. Ross, and S. Pankanti. A prototype hand geometry-based verification system. *Proc. of 2nd Int'l Conference on Audio and Video-based Biometric Person Authentication(AVBPA)*, pages 166–171, 1999.
- [54] A.K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Trans. on Circuits and Systems for Video Technology*, 14 :4–20, 2004.
- [55] M.J. Jones and J.M. Rehg. Statistical color models with application to skin detection. *International Journal of Computer Vision*, 46(1) :81–96, 2002.
- [56] M. Kölsch and M. Turk. Robust hand detection. *In International Conference on Automatic Face and Gesture Recognition, Seoul, Korea*, pages 614–619, 2004.
- [57] A.W.K. Kong. *Palmprint Identification Based on Generalization of Iris Code*. PhD thesis, University of Waterloo, Ontario, Canada, 2007.
- [58] A.W.K. Kong and D. Zhang. Palmprint texture analysis based on low-resolution images for personal authentication. *ICPR02*, 3 :807–810, 2002.
- [59] A.W.K. Kong and D. Zhang. Competitive coding scheme for palmprint verification. *ICPR04*, 1 :520–523, 2004.
- [60] Zhang D.; Kamel M. Kong A. A survey of palmprint recognition. *Pattern recognition*, 42 :1408–1418, 2009.
- [61] Ben Krose and Patrick Van Der Smagt. An introduction to neural networks. *CiteSeerX - Scientific Literature Digital Library and Search Engine [http://citeseerx.ist.psu.edu/oai2/ (United States)]*, 1993.
- [62] P. Kruizinga and N. Petkov. Nonlinear operator for oriented texture. *IEEE Trans. on Image Processing*, 8(10) :1395–1407, 1999.
- [63] A. Kumar and H.C. Shen. Palmprint identification using palmcodes. *ICIG '04 : Proceedings of the Third International Conference on Image and Graphics*, pages 258–261, 2004.
- [64] A. Kumar, D.C. M. Wong, H.C. Shen, and A.K. Jain. Personal verification using palmprint and hand geometry biometric. *AVBPA 2003 : audio- and video-based biometric person authentication*, pages 668–678, 2003.
- [65] A. Kumar, D.C.M. Wong, H.C. Shen, and A.K. Jain. Personal authentication using hand images. *Pattern Recogn. Lett.*, 27(13) :1478–1486, 2006.
- [66] K. Kumar and A. Negi. A feature partitioning approach to subspace classification. *TEN-CON*, pages 1–4, 2007.
- [67] Allano L. *La Biométrie multimodale : stratégies de fusion de scores et mesures de dépendance appliquées aux bases de personnes virtuelles*. PhD thesis, Institut national des télécommunications, Evry, France, 2009.
- [68] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Intelligent Signal Processing*, pages 306–351, 2001.
- [69] J. Lee and T. Kunii. Constraint-based hand animation. *Models and Techniques in Computer Animation*, pages 110–127, 1993.
- [70] J.Y. Lee and S.I. Yoo. An elliptical boundary model for skin color detection. *International Conference on Imaging Science, Systems and Technology*, 2002.
- [71] F. Li and M.K.H. Leung. Hierarchical identification of palmprint using line-based hough transform. *Pattern Recognition ICPR 2006*, 4 :149–152, 2006.

-
- [72] G. Lu, K. Wang, and D. Zhang. Wavelet based feature extraction for palmprint identification. *Proceeding of Second International Conference on Image and Graphics*, pages 780–784, 2002.
 - [73] G.M. Lu, D. Zhang, and K.Q. Wang. Palmprint recognition using eigenpalms features. *Pattern recognition letters*, 24(9-10) :1463–1467, 2003.
 - [74] L. Lucchese and S.K. Mitra. Color image segmentation : A state-of-the-art survey. *Proceedings of the Indian National Science Academy*, 67 :207–221, 2000.
 - [75] T. Lui and E. Izquierdo. Automatic detection of human faces in natural scene images by use of skin color and edge distribution. *5th Int. Workshop on Image Analysis for Multimedia Interactive Services, Lisbon, Portugal*, 2004.
 - [76] S. Mallat. A theory for multiresolution signal decomposition : the wavelet representation. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 11 :674–693, 1989.
 - [77] B. Martinkauppi. Detection of skin color under changing illumination : A comparative study. *IEEE 12th International Conference on Image Analysis and Processing*, pages 652–657, 2003.
 - [78] C. Massot and J. Herault. Extraction de la forme et de la perspective dans des textures artificielles et naturelles par modeles corticaux. *GRETSI*, 2006.
 - [79] J.S. Noh and K.H. Rhee. Palmprint identification algorithm using hu invariant moments and otsu binarization. *ICIS '05 : Proceedings of the Fourth Annual ACIS International Conference on Computer and Information Science*, pages 94–99, 2005.
 - [80] C. Oden, A. Ercil, V.T. Yldz, H. Krmzta, and B. Buke. Hand recognition using implicit polynomials and geometric features. *Proceedings of the Third International Conference on Audio- and Video-Based Biometric Person Authentication*, pages 336–341, 2001.
 - [81] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(7) :971–987, 2002.
 - [82] C.F. Olson and D.P. Huttenlocher. Automatic target recognition by matching oriented edge pixels. *IEEE Transactions on Image Processing*, 6 :103–113, 1997.
 - [83] M.G.K. Ong, T. Connie, A.T.B Jin, and D.N.C Ling. A single-sensor hand geometry and palmprint verification system. *WBMA '03 : Proceedings of the 2003 ACM SIGMM workshop on Biometrics methods and applications*, pages 100–106, 2003.
 - [84] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Trans. System Man Cybernetics*, 9(1) :62–69, 1979.
 - [85] V.I. Pavlovich, R. Sharma, and T.S. Huang. Visual interpretation of hand gestures for human computer interaction : A review. *IEEE Transactions and Pattern Analysis and Machine Intelligence*, 19(7) :677–695, 1997.
 - [86] P. Peer, J. Kovac, and F. Solina. Human skin colour clustering for face detection. *EURO-CON 2003. Computer as a Tool*, 2 :144–148, 2003.
 - [87] S.L. Phung, A. Bouzerdoum, and D. Chai. Skin segmentation using color and edge information. *This paper appears in : Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium*, 1(1) :525–528, 2003.
 - [88] C. Poon, D. C. M. Wong, and H. C. Shen. A new method in locating and segmenting palmprint into region-of-interest. *ICPR '04 : Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 4*, pages 533–536, 2004.

- [89] S. Ribaric and I. Fratric. A biometric identification system based on eigenpalm and eigenfinger features. *PAMI*, 27(11) :1698–1709, November 2005.
- [90] H.A. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Transactions On Pattern Analysis and Machine intelligence*, 20 :23–38, 1998.
- [91] B. Schiele and A. Waibel. Gaze tracking based on face-color. In *Proc. Int. Workshop on Auto. Face and Gesture Recog.*, pages 344–349, 1995.
- [92] B. Schölkopf, J.C. Platt, J. Shawe-taylor, A.J. Smola, and R.C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13 :1443–1471, 2001.
- [93] H. Schneiderman. A statistical approach to 3d object detection applied to faces and cars. *Ph.D.*, 2000.
- [94] H. Schneiderman and T. Kanade. Object detection using the statistics of parts. *Int. J. Comput. Vision*, 56(3) :151–177, 2004.
- [95] M.J. Seow, D. Valaparla, and V.K. Asari. Neural network-based skin color model for face detection. *Proceedings of the 32nd Workshop on Applied Imagery Pattern Recognition*, pages 141–145, 2003.
- [96] M. Sezgin and B. Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1) :146–165, 2004.
- [97] L. Shang, D.S. Huang, J.X. Du, and Z.K. Huang. Palmprint recognition using ica based on winner-take-all network and radial basis probabilistic neural network. *Lecture Notes in Computer Science*, 3972 :216–221, 2006.
- [98] W. Shu, G. Rong, Z.Q. Bian, and D. Zhang. Automatic palmprint verification. 1(1) :135–151, 2001.
- [99] S.K. Singh, D.S. Chauhan, M. Vatsa, and R. Singh. A robust skin color based face detection algorithm. *Tamkang Journal of Science and Engineering*, 6(4) :227–234, 2003.
- [100] I. Sobel and G. Feldman. A 3x3 isotropic gradient operator for image processing. *Pattern Classification and Scene Analysis*, 1968.
- [101] J. Sobottka and I. Pittas. Segmentation and tracking of faces in color images. *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition*, pages 236–241, 1996.
- [102] K. Sobottka and I. Pitas. Extraction of facial regions and features using color and shape information. In *Int. Conf. on Pattern Recognition (ICIP)*, 1996.
- [103] M. Soriano, S. Huovinen, B. Martinkauppi, and M. Laaksonen. Skin detection in video under changing illumination conditions. *ICPR, Proceedings of the International Conference on Pattern Recognition*, 1 :839–842.
- [104] B. Stenger, A. Thayananthan, P.H.S. Torr, and R. Cipolla. Model-based hand tracking using a hierarchical bayesian filter. *Pattern Analysis and Machine Intelligence*, 6(9) :1372–1384, 2006.
- [105] M. Störting. *Computer Vision and Human Skin Colour*. PhD thesis, Faculty of Engineering and Science, Aalborg University, Denmark, 2004.
- [106] D.J. Sturman and D. Zeltzer. A survey of glove-based input. *Computer Graphics and Applications*, 14(1) :30–39, 1994.
- [107] K.K. Sung and T. Poggio. Example based learning for view-based human face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20 :39–51, 1995.

-
- [108] K.K. Sung and T. Poggio. Example-based learning for view-based human face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20 :39–51, 1998.
 - [109] D. Taubman. Jpeg2000 - compression standard for interactive imaging.
 - [110] J. Terrillon and S. Akamatsu. Comparative performance of different chrominance spaces for color segmentation and detection of human faces in complex scene images. *International Conf on Face and Gesture Recognition*, pages 54–61, 2000.
 - [111] J. Terrillon, M. David, and S. Akamatsu. Automatic detection of human faces in natural scene images by use of a skin color model and of invariant moments. *In Proc. of the Third International Conference on Automatic Face and Gesture Recognition*, pages 112–117, 1998.
 - [112] F. Tomaz, T. Candeias, and H. Shahbazkia. Fast and accurate skin segmentation in color images. *Computer and Robot Vision, 2004. Proceedings. First Canadian Conference (CRV04)*, pages 180–187, 2004.
 - [113] D. Tschumperlé. Fast anisotropic smoothing of multi-valued images using curvature-preserving pde’s. *Int. J. Comput. Vision*, 68(1) :65–82, 2006.
 - [114] M. Turk and A. Pentland. Eigenfaces for recognition. *J. Cognitive Neuroscience*, 3(1) :71–86, 1991.
 - [115] V. Vapnik. The nature of statistical learning theory. *Springer-Verlag*, 1995.
 - [116] V. Vezhnevets, V. Sazonov, and A. Andreeva. A survey on pixel-based skin color detection techniques. *in Proc. Graphicon*, 2003.
 - [117] P. Viola and M. Jones. Robust real-time object detection. *International Journal of Computer Vision*, 2001.
 - [118] X. Wang, H. Gong, H. Zhang, B. Li, and Z. Zhuang. Palmprint identification using boosting local binary pattern. *ICPR ’06 : Proceedings of the 18th International Conference on Pattern Recognition*, pages 503–506, 2006.
 - [119] R. Watson. *A survey of gesture recognition techniques*, 1993.
 - [120] L. Wenwin, D. Zhang, and Zhuoqun X. Palmprint identification by fourier transform. *International journal of pattern recognition and artificial intelligence*, 16 :417–432, 2002.
 - [121] X. Wu, W. Kuanquan, D. Zhang, and K. Wang. Fusion of phase and orientation information for palmprint authentication. *Pattern Anal. Appl.*, 9(2) :103–111, 2006.
 - [122] X. Wu, K. Wang, and D. Zhang. Palmprint authentication based on orientation code matching. *Fifth International Conference on Audio- and Video-based Biometric Person Authentication*, pages 555–562, 2005.
 - [123] X. Wu, D. Zhang, and K. Wang. Fisherpalms based palmprint recognition. *Pattern Recogn. Lett.*, 24(15) :2829–2838, 2003.
 - [124] X. Q. Wu, K. Q. Wang, and D. Zhang. Wavelet based palmprint recognition. *Machine Learning and Cybernetics*, 3 :1253–1257, 2002.
 - [125] X.Q. Wu, K.Q. Wang, and D. Zhang. Palmprint recognition using directional line energy feature. 4 :475–478, 2004.
 - [126] X.Q. Wu, K.Q. Wang, and D. Zhang. Palmprint recognition using valley features. *Machine Learning and Cybernetics*, 8 :4881–4885, 2005.
 - [127] Y. Wu and T.S. Huang. Vision-based gesture recognition : A review. *Proceedings of the International Gesture Workshop on Gesture-Based Communication in Human-Computer Interaction*, 1739 :103–115, 1999.

- [128] Y. Wu and T.S. Huang. Hand modelling, analysis and recognition. *IEEE Signal Processing Magazine*, 18(3) :51–60, 2001.
- [129] W. Xiong, K.A. Toh, W.Y. Yau, and X.D. Jiang. Model-guided deformable hand shape recognition without positioning aids. *Pattern Recognition Letters*, 38(10) :1651–1664, 2005.
- [130] M. Yang and N. Ahuja. Gaussian mixture model for human skin color and its application in image and video databases. *Conf. on Storage and Retrieval for Image and Video Databases*, 3656 :458–466, 1999.
- [131] M.H. Yang and N. Ahuja. Detecting human faces in color images. *International Conference on Image Processing (ICIP)*, 1 :127–130, 1998.
- [132] E. Yoruk, E. Konukoglu, B. Sankur, and J. Darbon. Shape-based hand recognition. *IEEE Transactions on Image Processing*, 15(7) :1803–1815, 2006.
- [133] J. You, W.X. Li, and D. Zhang. Hierarchical palmprint identification via multiple feature extraction. *Pattern Recognition*, 35(4) :847–859, 2002.
- [134] B.D. Zarit, B.J. Super, and F.K.H. Quek. Comparison of five color models in skin pixel classification. *RATFG-RTS '99 : Proceedings of the International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, page 58, 1999.
- [135] D. Zhang. *Palmprint Authentication*. Kluwer Academic Publishers, 2004.
- [136] D. Zhang, W.K. Kong, J. You, and M. Wong. Online palmprint identification. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(9) :1041–1050, 2003.
- [137] J.G. Zhang, T.N. Tan, and L. Ma. Invariant texture segmentation via circular gabor filters. *ICPR02*, 2 :901–904, 2002.
- [138] L. Zhang and D. Zhang. Characterization of palmprints by wavelet signatures via directional context modeling. *Systems, Man, and Cybernetics, Part B : Cybernetics*, 34 :1335–1347, 2004.