

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'enseignement Supérieur et de la Recherche scientifique



Université Mohamed Khider Biskra
Faculté des Sciences et de la Technologie
Département de Génie Electrique
Filière :Électronique

Option :Électronique Télécommunication

Réf:.....

Mémoire de Fin d'Etudes
En vue de l'obtention du diplôme:

MASTER

Thème

***FUSION DE FACE 3D COULEUR,
PROFONDEUR ET PROFIL POUR SRV3D***

Présenté par :
NECIB Soufiane
Soutenu le :juin 2013

Devant le jury composé de :

Mme A.BARKAT

MAA

Président

Mme M.BELAHCENE

MCB

Encadreur

Mlle M.FEDIAS

MCB

Examineur

Année universitaire : 2012 / 2013

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'enseignement Supérieur et de la recherche scientifique



Université Mohamed Khider Biskra
Faculté des Sciences et de la Technologie
Département de Génie Electrique
Filière :Électronique
Option: Électronique Télécommunication

Mémoire de Fin d'Etudes
En vue de l'obtention du diplôme:

MASTER

Thème

**FUSION DE FACE 3D, COULEUR ET
PROFONDEUR POUR SRV 3D**

Présenté par :

NECIB Soufiane

Avis favorable de l'encadreur :

M.BELAHCENE

signature

Avis favorable du Président du Jury

Signature

Cachet et signature

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'enseignement Supérieur et de la Recherche scientifique



Université Mohamed Khider Biskra
Faculté des Sciences et de la Technologie
Département de Génie Electrique
Filière :Électronique
Option :Électronique Télécommunication

Thème :

FUSION DE FACE 3D, COULEUR ET PROFONDEUR POUR SRV 3D

Proposé par : NECIB Soufiane
Dirigée par : M.BELAHCENE

RESUMES (Français et Arabe)

Reconnaissance des visages traditionnels d'images de couleur ou l'intensité présente de nombreux défis, tels que les variations de l'éclairage, la pose et l'expression. En fait, le visage humain non seulement génère des informations de texture 2D, mais aussi des informations de forme 3D. Dans ce travail, nous examinons quelles sont les contributions des informations de profondeur, de couleur et de profil pour faire la reconnaissance faciale lorsque la variation de l'éclairage et de poses sont pris en compte. Nous présentons une méthode d'extraction de caractéristiques basées sur la réduction de l'espace unidimensionnel: l'analyse en composantes principales en amont à l'Enhanced Fisher Modèle linéaire discriminante (EFM). Une présentation théorique de ces approches et leurs applications sur les images profondeur, la couleur et profil est faite. Il y'a aussi une étude comparative sur la fusion de l'information de profondeur, couleur et profil pour les deux niveaux: les caractéristiques et les scores pour sélectionner les fonctions les plus efficaces et robustes et de construire ainsi une forte classification. La concaténation des vecteurs de caractéristiques et de la fusion des pixels des différentes modalités retenues : la moyenne, le produit, le minimum et le maximum sont utilisés dans le cas de fusion des scores par combinaison. Une normalisation par min_max est réalisée au préalable. La classification est réalisée par le Support Vector Machines (SVM). Les expériences sont effectuées sur la base de données CASIA visage 3D V4, des ensembles de données complexes avec des variations, y compris les variations de l'éclairage, de poses et les échecs de longue date entre les deux scans. Les résultats expérimentaux montrent la performance prometteuse du système proposé. Notez que dans notre système, tous les processus sont exécutés automatiquement.

Mots-clés : Profondeur, EFM, Fusion, Profil, ClassificationSVM, SRV3D.

ملخص

اعتراف من الصور التقليدية من الوجوه لون أو كثافة بطرح تحديات كثيرة مثل الاختلافات في الإضاءة، وتشكل والتعبير. في الواقع، وجه الإنسان لا يولد سوى معلومات الملمس D2 ولكن أيضا معلومات شكل D3. في هذه الورقة، ونحن نتفحص ما عمق المعلومات الاشتراكات واللون والتشكيل الجانبي للتعرف على الوجوه عندما تؤخذ تباين الإضاءة وي طرح في الاعتبار. نقدم استرجاع استنادا إلى الخفض من خصائص الأسلوب الأبعاد الفضاء: تحليل المكون الرئيسي لالمنبع المحسن فيشر الخطية التمايز نموذج (EFM). ويرصد العرض النظري لهذه المناهج وتطبيقاتها على الصور عمق واللون والشخصية. وهناك أيضا دراسة مقارنة عن الانصهار من عمق المعلومات واللون والتشكيل الجانبي للمستويين: خصائص وعشرات لتحديد وظائف الأكثر كفاءة وقوية، وبالتالي بناء تصنيف قوي. يتم استخدام متوسط المنتج، والحد الأدنى والحد الأقصى في حالة من الاندماج من خلال الجمع بين عشرات: في سلسلة من نواقل ميزة بكسل وأساليب مختلفة مزج المستخدمة. يتم إجراء تطبيع من قبل min_max مسبقا. يتم تنفيذ تصنيف من قبل أجهزة المتجهات الإعتدالية (SVM). وتجري التجارب على أساس D3 وجه كاسيا V4، ومجموعات البيانات المعقدة مع وجود اختلافات بما في ذلك التغييرات في الإضاءة، ويشكل والفشل طويلة بينهما هو إجراء الفحوصات. تظهر النتائج التجريبية أداء واعدة للنظام المقترح. لاحظ أنه في نظامنا، يتم تنفيذ جميع العمليات تلقائيا.

كلمات البحث: العمق، هذه الآلية، فيوجن، الشخصي، SRV3D تصنيف SVM.

Dédicaces

Je dédie ce modeste travail

À mes chers parents qui m'ont
toujours soutenu et encouragé,

À ma femme, mes sœurs,

frères :Walid, Mourad,

Bilal ,Hichem,Mohcen,Farid,Mohame

À tous mes amis.

REMERCIEMENTS

MES REMERCIEMENTS VONT EN PREMIER

**À MON ENCADREUR: DR ELKOURD KAOUTHER POUR SON
MERVEILLEUX SOUTIEN ET SES PRÉCIEUX CONSEILS.**

**À MES ÉDUCATEURS QUI ONT MÉNAGÉS LEURS
EFFORTS POUR MA**

RÉUSSITE.

AUX MEMBRES DU JURY, POUR LEUR COMPRÉHENSION.

Liste des tableaux

Table 1.1 – Récapitulatif des différentes techniques de réduction de dimension.....	14
Table 1.2 – Comparaison des différentes architectures présentées dans la figure 1.7 :.....	20
Table 2.1 - DISTRIBUTION DES 7317 SCANS PAR POSE :.....	34
Tableau 4.1 - Répartition des photos dans les différents ensembles pour les images 3D et 2D de visage :.....	61
Tableau 4.2 - Répartition des photos dans les différents ensembles pour les images 2D de profile :.....	62
Tableau 4.3 - Comparaison de TEE en fonction de nombre des caractéristiques pour les trois systèmes d'authentification de visage 3D, visage 2D et profile 2D :.....	70
Tableau 4.4 - Comparaison de TEE pour la fusion des scores :.....	72
Tableau 4.5 - Comparaison de TEE pour les trois systèmes d'authentification monomodaux et les quatre combinaisons de fusion de scores :.....	74

Liste des figures

Figure 1.1 . Processus d'estimation de la pose par une méthode globale:.....	4
Figure 1.2 . Les techniques de réduction de dimension cherchent un sous-espace de projection de faible dimension qui représente au mieux les variations de pose du visage. :.....	7
Figure 1.3 . Représentation des images d'un visage en rotation horizontale dans un espace formé par les 3 premiers vecteurs propres de l'ACP (tiré de . [32]). :.....	8
Figure 1.4 . Illustration de la méthode Isomap sur le jeu de données « brioche suisse » (tiré de [39]). :.....	10
Figure 1.5 . Les méthodes par comparaison avec des prototypes. L'orientation d'un nouveau visage correspond à la pose du visage de la base de prototypes le plus ressemblant :....	15
Figure 1.6 . Les méthodes par classification : un détecteur est appris par classe d'orientation et la pose d'un nouveau visage est donnée par le détecteur ayant obtenu le meilleur score. :.....	17
Figure 1.7 . Différentes architectures pour l'estimation de la pose par classification. (a) En parallèle, (b) multi-classes, (c) avec routeur, (d) descendante. :.....	18
Figure 1.8 . Architecture par routeur proposée par Rowley . :.....	20
Figure 1.9 . Les méthodes par régression non-linéaire visent à apprendre une relation fonctionnelle entre l'apparence d'un visage et sa pose. :.....	21
Figure 1.10 . Architecture d'un réseau LLM tirée de Rae et Ritter. :.....	22
Figure 1.11 . Architecture de notre réseau de neurones à convolution :.....	24
Figure 1.12 . Architecture du réseau MLP proposée par [60]. :.....	25
Figure 1.13 . Détection et normalisation des visages proposées par Gourier:.....	26
Figure 1.14 . Construction du modèle tensoriel. (a) Organisation matricielle des vignettes.	

(b) Modèle tensoriel proposé par Vasilescu et Terzopoulos . :.....	27
Figure 1.15. Illustration de l'estimation de la pose par projection sur le modèle tensoriel tirée de [59] :.....	27
Figure 2.1. Système d'acquisition de l'image :.....	31
Figure 2.2 - a profil pointe l'image b sommets Silhouette avant et après ajustements, c modèle 3D final ajusté d projetés modèle de c sur l'image :.....	32
Figure 2.3 . Détection de nez utilisant des images frontales et non frontal (Sous réserve 04385). :.....	33
Figure 2.4 . Exemple varie images correspondant à la figure 2.4 (Objet05078). :.....	35
Figure 2.5 . Les composants du système intégré de reconnaissance profil 3D assistée par ordinateur (URxD-PV). :.....	35
Figure 2.6 . Un exemple de mauvaise correspondance 2D-3D :.....	36
Figure 2.7 . P et son point voisin dans les deux sphères. :.....	37
Figure 2.8 . L'approche proposée. :.....	38
Figure 2.9 . Tête en temps réel pose exemple d'estimation. :.....	39
Figure 2.10. Exemples de vue latérale images où la texture des couleurs fiables ne sont pas disponibles (de gauche à droite) :.....	40
Figure 2.11. D'inscription et d'identification des modes du système de reconnaissance faciale à base :.....	41
Figure 2.12. Représentation du procédé de montage d'un modèle de visage générique pour la première 3D scan: (a) les données brutes, et (b) équipée AFM. :.....	42
Figure 2.13. Exemples d'images (avec des échelles de visage originales conservées)	

issus de sa propre base de données sans contrainte du visage de l'environnement.	45
Figure 2.14. L'utilisation de régions anatomiques pour la tête pose.	46
Figure 3.1. Schéma de principe de l'approche SRV3D :.....	49
Figure 3.2: (a) Image texture (b) Image 2.5D (c) Image 3D [GUE 08]. :.....	50
Figure 3.3 : Taxonomie des techniques optiques d'acquisition de formes 3D [AMO 08]. :.....	51
Figure 3.4 Exemples de vue de côté l'image où la texture des couleurs fiables ne sont pas disponibles (de gauche à droite):.....	53
Figure 3.5. Modèle 3D de visage humain constitué de triangles :.....	54
Figure 3.6. Illustration du principe de séparation optimale des classes par le LDA. :.....	54
Figure 3.7. Limite optimale recherchée par la SVM.	58
Figure 3.8 . L'hyperplan optimal H et la marge :.....	59
Figure 3.9. Images caractéristiques de visages :.....	59
Figure 4.1. Illumination variations of the CASIA 3D Face Database :.....	61
Figure 4.2 . Expression variations of the CASIA 3D Face Database :.....	61
Figure 4.3 . Exemple d'image de profile 2D, (a) image 19, (b) image 30. :.....	61
Figure 4.4 . Illustration du TFR et du TFA.:.....	63
Figure 4.5 . Courbe caractéristique de fonctionnement (ROC) d'un système de Vérification Biométrique. :.....	63
Figure 4.6 . Schéma de l'architecture structurelle du système de conception. :.....	64
Figure 4.7 . Prétraitement d'image de profondeur:.....	66
Figure 4.8 . Prétraitement d'image de couleur:.....	66
Figure 4.9 . Prétraitement d'image de profile 2D :.....	66
Figure 4.10 . Schéma de la fusion de scores :.....	68

Figure 4.11. Comparaison de TEE en fonction de nombre des caractéristiques pour les trois systèmes d'authentification de visage 3D, visage 2D et profile 2D :.....	70
Figure 4.12. Courbes ROC pour les trois systèmes d'authentification de visage 3D, visage 2D et profile 2D :.....	71
Figure 4.13. Comparaison de TEE pour la fusion des scores :.....	72
Figure 4.14. Courbes ROC pour les quatre combinaisons de fusion de scores :.....	73
Figure 4.15. Comparaison de TEE pour les trois systèmes d'authentification monomodaux et les quatre combinaisons de fusion de scores :.....	73
Figure 4.16. Courbes ROC pour les trois systèmes d'authentification monomodaux et les quatre combinaisons de fusion de scores :.....	74

Table des matières

Chapitre 1 : Reconnaissance de visage 3D

1.1 Introduction :	3
1.2 Extraction de caractéristiques :	4
1.2.1 Descripteurs globaux :	4
1.2.2 Descripteurs locaux :	5
1.3 Réduction de dimension :	6
1.3.1 Méthodes non supervisées :	7
1.3.2 Méthodes supervisées:	12
1.3.3 Synthèse :	14
1.4 Estimation de la pose :	15
1.4.1 Méthodes par comparaison avec des prototypes :	15
1.4.2 Méthodes par classification :	17
1.4.3 Méthodes par régression :	21
1.5 Méthodes de comparaison :	23
1.5.1 Réseau de Neurones à Convolution :	23
1.5.2 Méthodes de l'évaluation CLEAR 2007 :	24
1.5.2.1 Approche par un perceptron multicouches (MLP, Multi Layer Perceptron):	24
1.5.2.2 Approche par des mémoires auto-associatives linéaires (MAAL) :	25
1.5.2.3 Approche par un modèle tensoriel:	26
1.6 Conclusion :	27

Chapitre 2 : Etat de l'art des techniques de reconnaissance faciale à base de profil

2.1 Introduction :.....	29
2.2 Etats de l'art :.....	30
2.3 Conclusion :.....	47

Chapitre 3 : Etude de l'approche de fusion basée sur le profil

3.1 Introduction :.....	48
3.2 Présentation de l'approche SRV3D étudiée :.....	49
3.3 Système d'acquisition 3D :.....	50
3.4 Prétraitement de l'image 3D :.....	52
3.5 Descripteurs du visage 3D :.....	52
3.5.1 Utilisation du profil dans la reconnaissance du visage 3D :.....	52
3.5.2 Utilisation de la couleur et profondeur de visages pour la reconnaissance faciale :.....	53
3.6 Extraction des caractéristiques :.....	54
3.7 La fusion des scores :.....	57
3.8 Conclusion :.....	59

Chapitre 4 : Conception et résultats

4.1 Introduction :.....	60
4.2 La base de données CASIA 3D :.....	60
4.3 Mesure de la performance des systèmes d'authentification biométriques :.....	62
4.4 Architecture structurelle du système de conception :.....	64
4.4.1 Prétraitement d'image :.....	64
4.4.2 L'analyse en composantes principales(PCA) :.....	66
4.4.3 Le Modèle Discriminant Linéaire Amélioré de Fisher (EFM) :.....	67
4.4.4 Distance :.....	67

4.4.5 Normalisation par la méthode Min-Max :.....	68
4.4.6 Classification des scores par SVM :.....	68
4.5 Résultats obtenus :.....	69
4.5.1 Résultats obtenus pour l'authentification monomodaux :.....	69
4.5.2 Résultats obtenus pour la fusion :.....	71
4.6 Conclusion :.....	75

Introduction général

L'accroissement international des communications, implique le besoin de s'assurer de l'identité des individus. Les technologies biométriques de reconnaissance apportent la simplicité et le confort aux utilisateurs. Elles sont une brique dans tout système de sécurité actuel et futur. Cette technologie est applicable à un large champ d'applications.

De nombreuses technologies biométriques ont été développées, toutes basées sur les identificateurs biométriques (iris, voix, empreintes digitales, face, signature...). En effet, la biométrie est l'usage de différentes caractéristiques physiologiques et comportementales afin de réaliser une reconnaissance automatique d'un individu.

Notre travail consiste à étudier et évaluer par des expériences la fusion (image de profil & image 2D & image 2.5D profondeur). Ainsi, notre but c'est d'avoir un système biométrique de reconnaissance de visages (identification et vérification) basé sur la fusion de scores.

Nous utiliserons plusieurs méthodes de combinaison des scores (**moyenne**, **max** et **min**) obtenus de l'**ACP** pour essayer d'exploiter la complémentarité entre ces méthodes.

Les résultats des expériences sur la base de données, et celle-ci montre le développement de l'**ACP** dans cette fusion et comment choisir les paramètres pour avoir des meilleures performances. Donc, en plus de voir ce que la fusion apportera à la performance, on comparera les résultats obtenus avant et après la fusion.

Le présent mémoire est organisé comme suit :

Le **premier chapitre** : il représente les méthodes qui utilisent globalement l'image d'un visage pour en déduire sa pose. Nous les regroupons sous le terme de méthodes globales, par opposition aux méthodes par alignement d'un modèle déformable ou aux approches géométriques qui infèrent la pose à partir des positions relatives de certains éléments du visage tels que les yeux et la bouche.

Le **deuxième chapitre** commence par un aperçu sur l'état de l'art de système de reconnaissance de visage 3D SRV. Et décrit aussi les différents technique qu'on applique sur le profil et un peu plus en détail les méthodes d'extractions de visage.

Le **troisième chapitre** conception les organigrammes de notre système et leur fonctionnement avec une explication de chaque étape

Le **quatrième chapitre** Jusqu'ici nous avons approché la reconnaissance automatique des visages d'un point de vue théorique. Les résultats expérimentaux que nous allons présenter sont divisés en deux parties. Nous donnerons tout d'abord les résultats des comparaisons des systèmes monomodaux d'authentification de visage 3D, visage 2D et profile 2D. Nous présentons ensuite la fusion de scores par SVM des quatre combinaisons des systèmes (fusion 3D visage et 2D visage, fusion 3D visage et 2D profile, fusion 2D visage et 2D profile, fusion 3d et 2D visage et 2D profile).

1.1 Introduction

Dans ce chapitre nous présentons les méthodes qui utilisent globalement l'image d'un visage pour en déduire sa pose. Nous les regroupons sous le terme de méthodes globales, par opposition aux méthodes par alignement d'un modèle déformable ou aux approches géométriques qui infèrent la pose à partir des positions relatives de certains éléments du visage tels que les yeux et la bouche.

Les approches par alignement d'un modèle seront traitées dans la seconde partie du chapitre. Le processus d'estimation par les méthodes globales peut se décomposer en quatre étapes successives (*cf.* figure 1.1) :

- **Prétraitements** : le signal capturé par la caméra est prétraité pour simplifier les étapes suivantes, sans perdre l'information pertinente. Par exemple, une égalisation d'histogramme ajustera automatiquement les niveaux de gris ou les pixels de l'image seront seuillés afin d'éliminer l'arrière-plan de l'image.

- **Localisation du visage** : Cette étape extrait la zone de l'image contenant le visage dont on souhaite connaître la pose. Dans certaines méthodes, elle est intrinsèquement liée à l'étape d'estimation de pose

- **Représentation des images** : L'objectif est d'extraire les informations pertinentes et discriminantes contenues dans l'image en mesurant certaines propriétés (extraction de caractéristiques, section 1.2) ou en projetant les données dans un espace où la distribution des visages en fonction de leur pose est plus régulière. (section 1.3).

- **Décision**: un algorithme d'apprentissage supervisé est utilisé pour estimer les paramètres de pose du visage.

Nous avons différencié trois grandes approches. Les méthodes par comparaison avec des prototypes (section 1.4.1) déterminent la pose d'un visage en fonction du label des exemples (les prototypes) les plus ressemblants de la base d'apprentissage. Les méthodes par classification (section 1.4.2) regroupent les visages dont la pose est semblable et apprennent à déterminer le groupe auquel appartient un visage inconnu. Les méthodes par régression (section 1.4.3) apprennent la relation continue qui associe une pose à chaque image de visage.

Une étape facultative de post-traitements valide ou rejette la décision prise à l'étape précédente à partir d'informations supplémentaires liées au contexte. Dans une application de suivi par exemple, l'estimation de la pose pourra être filtrée en fonction des estimations précédentes. Nous nous intéresserons plus particulièrement à l'étape de représentation des images (section 1.2 et 1.3) et d'estimation de la pose (section 1.4).

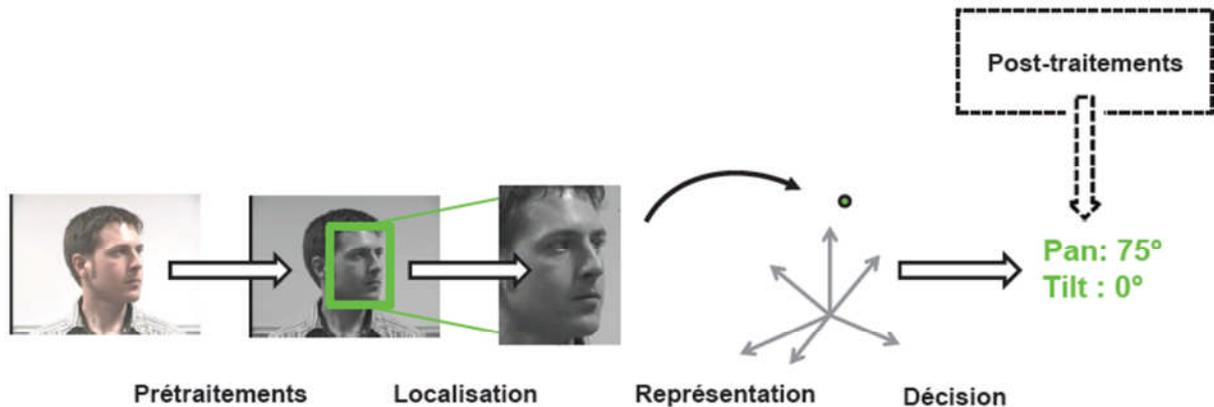


Figure 1.1 Processus d'estimation de la pose par une méthode globale:4

1.2 Extraction des caractéristiques

L'objectif de cette étape est d'utiliser un ou plusieurs descripteurs pour représenter au mieux l'image en fonction de la tâche à réaliser.

On cherchera généralement des descripteurs qui présentent des propriétés d'invariance par rapport aux caractéristiques non pertinentes.

Pour un détecteur de véhicule par exemple, on s'intéressera à des descripteurs invariants aux variations colorimétriques puisque une voiture peut être de n'importe quelle couleur.

Il existe de très nombreux descripteurs dans la littérature et l'objectif de cette section n'est pas de tous les énumérer. Nous présenterons les grandes catégories ainsi que les descripteurs les plus souvent utilisés.

1.2.1 Descripteurs globaux

Ils permettent d'extraire un ensemble d'attributs calculés sur toute l'image.

Filtrage de l'image Son but est d'extraire des informations pertinentes dans l'image.

Les filtres de Sobel ou Canny, par exemple, sont utilisés pour extraire les contours qui présentent l'avantage d'être plus robustes vis-à-vis des changements d'illumination que les niveaux de gris.

Les filtres de Gabor sont très largement utilisés en analyse d'image en général [1] et en

estimation de la pose de la tête en particulier [2];[3];[4]. Des opérateurs non linéaires tels que les Local Binary Patterns (LBP, [5]) ont également donné de bons résultats en analyse de visages [6]; [7]. [8] combinent les ondelettes de Gabor et les LBP pour estimer la pose de la tête.

- **Représentation par silhouette** Elle consiste à binariser l'image de manière à séparer l'arrière-plan de la forme à analyser. La silhouette est ensuite caractérisée par un ensemble de descripteurs de forme tels que les descripteurs de Fourier ou les moments géométriques [9]. [10] propose un panorama détaillé des descripteurs de forme. Les silhouettes issues de vues de différentes caméras peuvent également servir à reconstruire l'enveloppe visuelle de l'objet considéré.

On ne cherche alors plus à caractériser les silhouettes directement mais l'objet 3D qui en résulte [11].

- **Représentation par histogramme** La représentation par histogramme est un outil simple et puissant qui a été utilisé dans de nombreux domaines. Les histogrammes de couleur sont invariants en translation et en rotation dans le plan image et varient lentement lors des rotations hors-plan, des changements d'échelles et des occultations. Les histogrammes donnent une signature compacte, stable et discriminante bien adaptée pour l'indexation de grandes bases de données [12]. Les histogrammes de couleur sont sensibles à l'intensité et à la couleur de la source lumineuse ainsi qu'à la couleur de l'objet à détecter.

Pour parer ce problème, [13] proposent par exemple de construire des histogrammes à champs récepteurs gaussiens multidimensionnels. L'histogramme de l'orientation des gradients pondérés par leur module (HoG de l'anglais Histograms of Oriented Gradient) est aussi très utilisé [14]. Pour conserver en partie l'information spatiale, on peut concaténer les histogrammes calculés dans des fenêtres glissantes ou disjointes de l'image [15]; [16].

1.2.2 Descripteurs locaux

Par opposition aux descripteurs globaux, les descripteurs locaux ne caractérisent qu'une zone restreinte de l'image.

Chaque descripteur extrait une information partielle et doit, par conséquent, être combiné à d'autres descripteurs pour fournir une représentation complète de l'image à analyser.[17] proposent un tour d'horizon des descripteurs locaux tandis que [18] évaluent les performances de différents descripteurs tels que les *Shape Context* [19], les filtres orientables [20] ou SIFT [15].

On peut distinguer deux manières de les utiliser :

1. On définit un ensemble de descripteurs locaux calculables en tout point de l'image avec différents paramètres. Cet ensemble dense et redondant doit être couplé avec une méthode de sélection de descripteurs.

[21] combinent par exemple les descripteurs de [22] inspirés des ondelettes de Haar (plus de 45'000 descripteurs au total) avec l'algorithme AdaBoost [23] qui sélectionne itérativement les meilleurs descripteurs.

2. On identifie des points d'intérêt dans l'image à l'aide de détecteurs spécifiques tels que [24] ou [25], puis on utilise des descripteurs locaux (SIFT, [15] ou SURF [26], par exemple) pour caractériser le voisinage de ces points.

Cette méthode offre une représentation compacte de l'image à analyser et ne nécessite pas d'étape de sélection des descripteurs. Il faut toutefois que la détection des points d'intérêt soit répétable, c'est-à-dire que les mêmes points soient détectés au même endroit quelles que soient les conditions de prise de vue.

1.3 Réduction de dimension

Les images des visages sont des données de grande dimension. Il est toutefois raisonnable de supposer que ces données sont sur ou à proximité d'une variété de faible dimension immergée dans l'espace ambiant R_N où N est le nombre de pixels (*cf.* figure 1.2).

Les techniques qui cherchent une relation entre l'espace d'origine et un espace de faible dimension sont regroupées sous le terme de « réduction de dimension » ou plus récemment d'apprentissage de variétés (manifold learning). Dans notre cas, l'objectif est de trouver un espace de représentation sensible aux changements de poses et invariant aux autres sources de variation de l'image. Nous présentons dans cette partie un aperçu des méthodes de réduction de dimension, ainsi que leurs applications pour l'estimation de la pose de la tête.

Pour une description plus exhaustive, le lecteur pourra se reporter à [27]; [28]; [29]

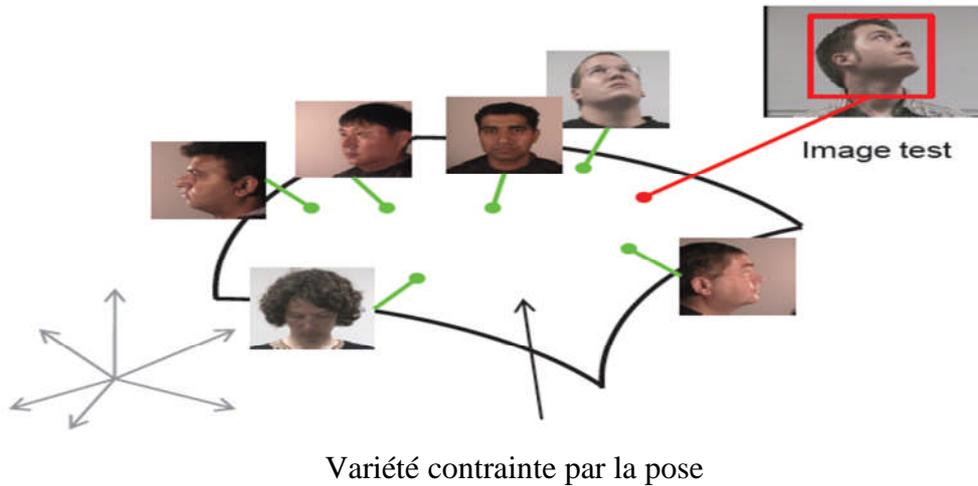


Figure 1.2 – Les techniques de réduction de dimension cherchent un sous-espace de projection de faible dimension qui représente au mieux les variations de pose du visage. :7

Un visage dont la pose est inconnue sera projeté sur cet espace pour simplifier la prise de décision.

1.3.1 Méthodes non supervisées

Nous traitons dans un premier temps les méthodes dites non supervisées car elles ne prennent pas en considération la grandeur que l'on souhaite estimer (les paramètres de pose dans notre cas).

- **Méthodes linéaires** : L'Analyse en Composantes Principales (ACP) est une méthode de réduction de dimension linéaire introduite par [30]. L'idée principale de l'ACP est de trouver une projection des données $\{x_i\}_i^n = 1 \in R^d$ qui maximise la variance.

La base de ce nouveau sous-espace est formée des m premiers vecteurs propres de la matrice de covariance estimée,

$$c = \frac{1}{n} \sum_i (X_i - \bar{X})(X_i - \bar{X})^T \quad (1.1)$$

avec \bar{X} , le vecteur moyenne estimée. Si on range les vecteurs propres dans une matrice \mathbf{W} de taille $d \times m$, la projection U_i de chaque donnée X_i est calculée par

$$u_i = W^T (X_i - \bar{X}) \quad (1.2)$$

Cette méthode à été utilisée dans de nombreux domaines, en particulier en analyse faciale [31]. [32] ont montré qu'il suffisait de trois axes principaux pour représenter la distribution des visages en fonction de leur pose (cf. figure 1.3).

[33] construisent un sous-espace propre par pose discrète.

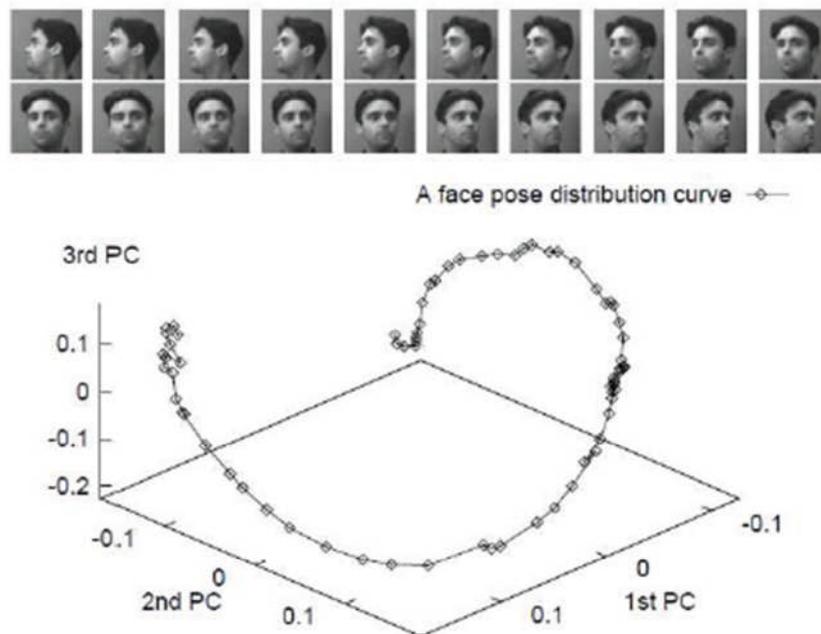


Figure 1.3 – Représentation des images d'un visage en rotation horizontale dans un espace formé par les 3 premiers vecteurs propres de l'ACP (tiré de [32]). :8

Le Multi dimensional Scaling (MDS) est une autre méthode très connue de réduction de dimension linéaire [34].

Au lieu de conserver la variance des données au cours de la projection, elle s'efforce de préserver toutes les distances entre chaque paire d'exemples $dist(X_i, X_j)$ en cherchant une transformation linéaire qui minimise l'énergie :

$$\varepsilon_{m ds} = \sum_{i,j=1}^n \left(\text{dist}(X_i, X_j) - \|U_i - U_j\| \right)^2 \quad (1.3)$$

Ce problème de minimisation peut être résolu par une décomposition en valeurs propres [35]; [36]. Lorsque la fonction de distance entre les données est la distance euclidienne, les sorties $u_i \in R^m$ de la MDS sont les mêmes que les sorties de l'ACP. Elles sont obtenues par une rotation suivie d'une projection sur les axes de plus grande variance.

- **Méthodes par noyaux :** Les méthodes précédentes trouvent un espace de représentation fidèle aux données lorsque la structure de ces données est linéaire ce qui n'est généralement pas le cas. L'idée est d'utiliser une fonction noyau pour construire un espace de grande dimension dans lequel le problème devient linéaire. On peut ainsi appliquer des méthodes linéaires de réduction de dimension lorsque la structure intrinsèque des données n'est pas linéaire. Ces méthodes utilisent généralement « l'astuce du noyau » (de l'anglais kernel trick) qui établit que tout algorithme formulé avec une fonction noyau peut être reformulé avec une autre fonction noyau. Une démarche courante est d'exprimer la méthode avec un produit scalaire, puis de le remplacer par une fonction noyau. Le kernel trick permet alors de travailler dans l'espace transformé sans avoir à calculer explicitement l'image de chaque donnée. [3] utilisent par exemple une Analyse en Composante Principale avec Noyau (KPCA de l'anglais Kernel Principal Component Analysis, [37]) pour représenter des visages dans différentes poses. Dans la KPCA, le choix de la fonction noyau n'est pas clairement défini. La MVU (de l'anglais Maximum Variance Unfolding) est une technique qui apprend conjointement la transformation et la fonction noyau [38].

- **Méthodes par graphe :** PCA et MDS tentent de conserver les distances euclidiennes entre chaque paire d'exemples sans prendre en compte la distribution du voisinage de chaque exemple. Si les données reposent sur une variété très incurvée, deux points très proches au sens de la distance euclidienne peuvent être très éloignés si on considère la distance le long de la variété. Le jeu de données « brioche suisse » est couramment utilisé pour illustrer ce phénomène (cf. figure 1.4).

Depuis les années 2000 de nombreuses solutions ont été proposées pour réduire la dimension de l'espace en prenant en compte la topologie de l'ensemble de données. La méthode Isomap (pour isometric feature mapping, de [39]) par exemple, cherche une transformation qui préserve la distance géodésique entre les données. La distance géodésique est la distance la plus courte entre deux points, mesurée le long de la variété.

L'algorithme se décompose en trois étapes :

1. Création d'un graphe d'adjacence G dans lequel chaque point X_i est connecté à ses k plus proches voisins (k -Isomap) ou aux données dont la distance est inférieure à ε (ε -Isomap).
2. Calcul du plus court chemin $d_G(X_i, X_j)$ entre chaque paire de nœuds du graphe à l'aide d'un algorithme du plus court chemin tel que [40] ou [41]. La matrice $D_G = d_G(X_i, X_j)$ contient alors une estimation de la distance géodésique entre chaque point.
3. Construction d'une représentation de faible dimension par MDS appliquée à la matrice des distances D_G .

[42] utilisent Isomap pour découvrir la variété des visages dans différentes poses. La méthode Isomap présente toutefois quelques inconvénients. Elle est topologiquement instable [43] car elle peut créer des connexions

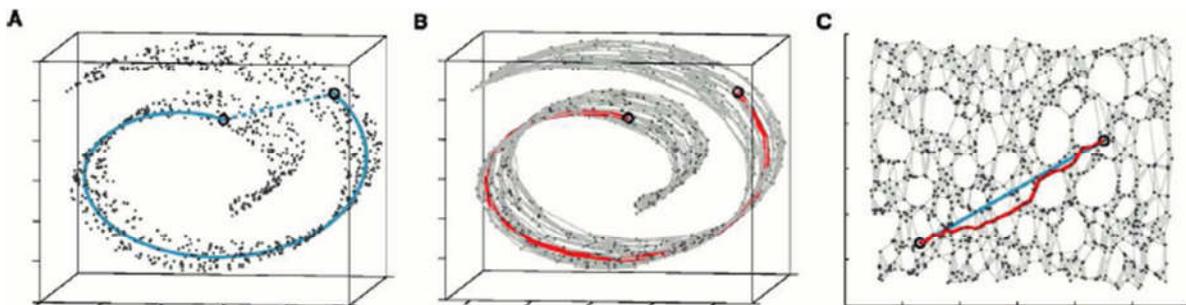


Figure 1.4 – Illustration de la méthode Isomap sur le jeu de données « brioche suisse » (tiré de [39]). :10

(A) La distance entre deux exemples du jeu de données (pointillés) ne reflète pas leur ressemblance intrinsèque mesurée par la distance géodésique (trait plein). (B) Le graphe d'adjacence permet d'approximer (courbe rouge) la distance géodésique. (C) Projection des données dans l'espace 2D trouvée par Isomap.

Fausse dans le graphe G ou mal approximer la distance géodésique lorsque la variété est trouée [44].

LLE (pour Locally Linear Embedding) est une technique de représentation de données de grande dimension dans un espace de faible dimension préservant localement la structure linéaire des exemples à proximité [45]. En ne conservant que la structure locale des données, le problème se rapporte à la décomposition en valeurs propres d'une matrice creuse (par opposition aux méthodes précédentes qui devaient décomposer une matrice pleine). La première étape est semblable à Isomap puisqu'elle consiste à construire un graphe des k -plus proches voisins de \mathbf{x}_i .

Chaque exemple est ensuite décrit par une combinaison linéaire de poids w_{ij} des k -plus proches voisins de X_i . Les poids de reconstruction représentent la contribution du j^{eme} point dans la reconstruction du i^{eme} exemple. La dernière étape revient à trouver la représentation qui conserve ces relations entre exemples en minimisant la fonction de coût :

$$\mathcal{E}_{lle} = \sum_i \left\| u_i - \sum_j w_{ij} u_j \right\| \quad (1.4)$$

La matrice de covariance des sorties y_i doit être unitaire pour éviter les solutions triviales.

[45] ont montré que l'espace réduit qui minimise l'équation (1.4) est formé des m vecteurs propres associés aux m plus faibles valeurs propres non nulles de la matrice $(I - W)^T (I - W) \cdot I$ est la matrice identité $n \times n$ et \mathbf{W} est la matrice creuse $n \times n$ dont les composantes sont égales aux poids de reconstruction si x_i est connecté à x_j et nulles sinon.

L'aspect local de LLE la rend moins sensible aux court-circuit que Isomap car seule une partie restreinte de la matrice \mathbf{W} est affectée.

Il existe de nombreuses autres techniques par graphe. Certaines sont locales telles que les Laplacian Eigenmaps [46] ou les Local Tangent Space Alignment [47]. D'autres sont globales telles que les diffusion maps [48] ou le Maximum Variance Unfolding [38].

Le but de ces méthodes est de représenter chaque noeud du graphe par un vecteur de faible dimension qui préserve les similarités entre les paires de vecteurs. Toutefois ces techniques ne fournissent pas de projection explicite sur l'espace de faible dimension et il n'est par conséquent pas possible de projeter un nouvel exemple qui n'était pas présent dans l'ensemble de départ (out of sample problem). Plusieurs solutions ont été proposées pour résoudre ce problème [49]; [50]; [51], par exemple, utilisent les Locality Preserving Projections [52] qui définissent une relation pour n'importe quel point de l'espace.

L'autre inconvénient majeur des méthodes par graphe est que l'information de pose des visages n'est pas exploitée. Nous allons donc aborder les méthodes supervisées qui intègrent explicitement cette information dans le choix du sous-espace.

1.3.2 Méthodes supervisées

La méthode de réduction de dimension la plus connue et la plus souvent utilisée est certainement l'Analyse Discriminante Linéaire (LDA de l'anglais Linear Discriminant Analysis).

L'objectif de la LDA est de trouver un sous-espace de projection qui maximise le ratio entre la matrice de covariance interclasse C_{inter} et la matrice de covariance intra-classe C_{intra} (critère de Fisher). Cette optimisation se rapporte à la résolution d'un problème aux valeurs propres généralisé :

$$C_{inter} W_k = \lambda_K C_{intra} W_k \quad (1.5)$$

Les vecteurs propres associés aux plus grandes valeurs propres forment le sous-espace de projection. La LDA est une technique très efficace de réduction de dimension, mais elle se limite toutefois aux problèmes de classification [53] proposent une extension de cette méthode aux problèmes de régression. L'idée est de considérer que les exemples x_i qui ont de faibles différences de label y_i appartiennent à la même classe. Les matrices de covariance sont remplacées par :

$$S_{intra} = \frac{1}{n_{intra}} \sum_{i,j \in A_{intra}} f(y_i, y_j) (X_i - X_j)(X_i - X_j)^T \quad (1.6)$$

$$S_{inter} = \frac{1}{n_{inter}} \sum_{i,j \in A_{inter}} f(y_i, y_j) (X_i - X_j)(X_i - X_j)^T \quad (1.7)$$

Avec $A_{intra} = \{(i, j) : |y_i - y_j| < \tau, i, j \in \{1, \dots, n\}, i \neq j\},$

$$A_{inter} = \{(i, j) : |y_i - y_j| \geq \tau, i, j \in \{1, \dots, n\}, i \neq j\},$$

$$n_{\text{intra}} = \text{card} (A_{\text{intra}}) \text{ et } n_{\text{inter}} = \text{card} (A_{\text{inter}}).$$

La fonction f est une fonction de pondération positive qui décroît avec l'amplitude de la différence. [53] ont montré que cette méthode donnait de bons résultats pour l'estimation de la pose de la tête en comparaison avec des méthodes classiques de réduction de dimension telles que la Sliced Inverse Regression (SIR, [54]), ou les Principal Hessian Directions (PHD, [54]). Plusieurs solutions ont été proposées pour intégrer l'information des labels dans les techniques d'apprentissage de la variété. SE-Isomap [55] construit tout d'abord une matrice des distances géodésiques pour chaque classe. Les différentes matrices sont ensuite regroupées au sein de la matrice des distances globales discriminantes. Pour finir, cette matrice est utilisée dans l'étape de réduction de dimension par MDS. Dans [51], le graphe intrinsèque caractérise la compacité intra-classe et connecte les échantillons de même classe, alors que le graphe de pénalité connecte les points à la marge et caractérise la séparabilité inter-classe. La Marginal Fisher Analysis cherche une transformation qui maximise la compacité intra-classe et la séparabilité inter-classe utilisent une métrique qui prend en compte les labels associés aux données pour mesurer les distances entre échantillons. Cette métrique, qui augmente les distances inter-classes et diminue les distances intra-classes, remplace la distance euclidienne lors de la construction du graphe d'adjacence de la LLE. La méthode WheightedIso de Vlachos joue également sur la métrique et multiplie par un facteur constant $Y < 1$ la distance entre deux échantillons d'une même classe. [56] ont montré que la qualité des résultats était très dépendante du choix de Y . Ils définissent une métrique telle que la distance entre deux échantillons de la même classe soit toujours inférieure à 1 et que la distance entre deux échantillons de classes différentes soit toujours supérieure à $1 - \alpha$. Le coefficient α , défini empiriquement, a été introduit pour permettre à deux échantillons de classes différentes de se « ressembler » plus que certains échantillons de la même classe et pour limiter les risques de déconnexion dans le graphe d'adjacence. De même que la LDA, ces méthodes concernent les problèmes de classification. Le Biased Manifold Embedding [43] s'adresse principalement aux problèmes de régression en pondérant chaque distance inter-exemples, par un facteur qui dépend de la différence des labels. Cette méthode a donné de bons résultats pour l'estimation de la pose de la tête.

Une autre approche supervisée de réduction de dimension consiste à définir la structure idéale de la variété et d'apprendre la transformation vers cette structure topologique. Lee et Elgammal modélisent la marche d'un humain par un tore. Une dimension est utilisée pour décrire la configuration du corps et une autre dimension représente le point de vue de la caméra. La

relation entre une silhouette et sa position sur le tore est apprise à l'aide d'un réseau de neurones à base radiale.

1.3.3 Synthèse

Le tableau 1.1 résume, par ordre chronologique, les différentes techniques de réduction de dimension et d'apprentissage de variétés. Il spécifie si la projection est linéaire ou nonlinéaire, supervisée ou non-supervisée et s'il existe une manière de projeter sur l'espace réduit un exemple qui n'appartient pas à l'ensemble d'apprentissage (out of sample). La dernière colonne précise si la méthode cherche à préserver des propriétés géométriques locales ou globales lors de la projection. Les méthodes globales reflètent plus fidèlement la structure globale des données, tandis que les méthodes locales sont plus faciles à calculer (matrice creuse) et plus à même de modéliser les variétés dont la géométrie est localement proche de la géométrie euclidienne, mais dont la géométrie globale ne l'est pas [de Silva et Tenenbaum].

Table 1.1 – Récapitulatif des différentes techniques de réduction de dimension.

Méthodes Non Supervisée			Méthodes Supervisée
M.Lineaire	M .Anoyeux	graphe	
ACP (Analyse en composantes)	KPCA(<i>Kernel Principal Component Analysis</i>)	Locales iso map	LDA(<i>Analyse discriminante linéaire</i>) MFA(<i>Marginal Fisher Analysis</i>). BME(<i>Biased Manifold Embedding</i>)
MDS		Globales DM(<i>Diffusion map</i>) MVU(<i>Maximum Variance Unfolding</i>)	

1.4 Estimation de la pose

1.4.1 Méthodes par comparaison avec des prototypes

Cette catégorie de méthodes est certainement la plus simple et la plus intuitive. Elle consiste à comparer l'apparence du visage dans une image à un ensemble de visages d'une base de données étiquetée en pose. On attribue au visage de l'image l'orientation du visage le plus ressemblant dans la base de données (cf. figure 1.5). On utilise couramment l'erreur quadratique moyenne Niyogi et Freeman, La Cascia, Sherrah et Gong, la corrélation croisée normalisée Beymer ou l'information mutuelle Goudelis comme mesure de ressemblance.

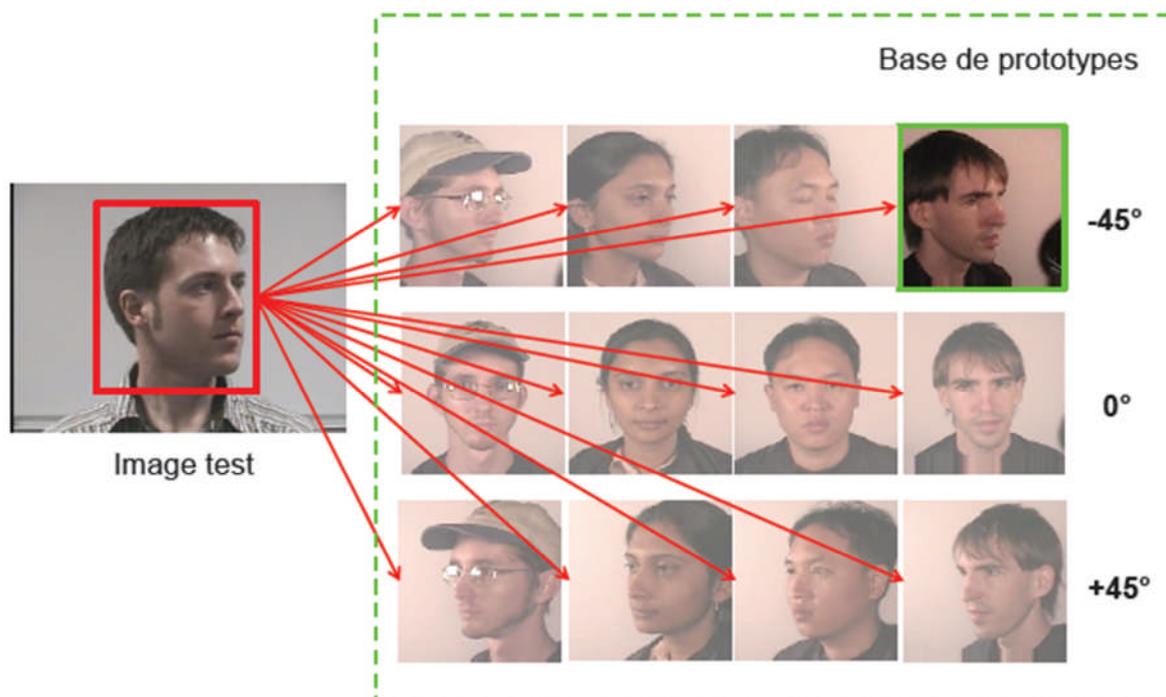


Figure 1.5 – Les méthodes par comparaison avec des prototypes. L'orientation d'un nouveau visage correspond à la pose du visage de la base de prototypes le plus ressemblant :15

Ces méthodes n'ont pas besoin d'une phase d'apprentissage et sont incrémentales en données (on peut augmenter sans effort l'ensemble des prototypes). Toutefois, des problèmes d'efficacité surviennent lorsque le nombre de prototypes est trop important puisque la complexité de ces algorithmes est linéairement dépendante du nombre d'éléments dans la base de données. Une solution à ce problème consiste à organiser les données de la base sous forme d'un arbre Niyogi et Freeman, Sengupta.

Dans Niyogi et Freeman, par exemple, l'arbre est construit en séparant récursivement les exemples en deux groupes. A chaque noeud de l'arbre, une analyse en composante principale

(ACP) détermine la direction du maximum de variance du groupe. On projette alors les exemples sur cet axe et on les sépare en deux classes en fonction du signe de leur coordonnée sur cet axe. Cette opération est répétée jusqu'à ce qu'il n'y ait plus assez d'exemples dans le groupe pour faire une ACP. Dans le cas d'un arbre binaire, le coût de la recherche diminue d'un facteur M à un facteur $\log(M)$, avec M le nombre d'éléments dans la base de données.

Dans les méthodes par comparaison avec des prototypes, le visage doit être préalablement détecté et une erreur de localisation peut fortement dégrader la précision du système. Niyogi et Freeman proposent de générer des prototypes présentant des décalages en translation et en changement d'échelle. Si l'image la plus ressemblante correspond à un prototype avec décalage, on peut facilement ajuster le cadrage de manière à avoir la ressemblance la plus forte avec un prototype centré.

Ce type d'approche s'apparente à des méthodes par suivi ; l'objectif est de maximiser la ressemblance entre l'apparence d'un modèle pour différents paramètres (translation, rotation, changement d'échelle par exemple) et l'image du visage dont on cherche à connaître la pose. Le modèle correspond le plus souvent à une image frontale du visage que l'on plaque sur un modèle géométrique simple tel qu'une ellipsoïde Choi et Kim, un cylindre La Cascia et Xiao ou un modèle générique de visage Malciu et Prêteux. La texture du modèle est parfois mise à jour au cours de la séquence afin de prendre en compte des changements d'illumination par exemple Xiao.

Par ailleurs, les méthodes par comparaison à des prototypes supposent que des visages différents dans une même pose sont toujours plus semblables qu'un même visage dans deux poses différentes. En d'autres termes, on souhaite que la mesure de similarité dans l'espace des images soit représentative de la distance dans l'espace des poses. Il est possible d'agir à deux niveaux :

- Appliquer une transformation de l'image qui accentue les caractéristiques liées à la pose et qui atténue les autres sources de variation telles que l'illumination, les expressions faciales et l'identité (*cf.* section 1.1 et 1.2).
- Choisir une mesure de ressemblance adaptée. La Cascia . Introduisent par exemple un terme qui modélise l'illumination dans la fonction de coût à optimiser La Cascia. Everingham et Zisserman utilisent une distance de Chamfer sur les contours orientés.

1.4.2 Méthodes par classification

L'idée principale est de regrouper par classe les visages de la base d'apprentissage qui ont des orientations proches et d'utiliser une méthode d'apprentissage pour déterminer si un visage inconnu appartient à cette classe d'orientation. Cette approche, illustrée par la figure 1.6, présente deux avantages par rapport aux méthodes par comparaison à des prototypes.

D'une part, le visage n'est pas comparé à tous les exemples de la base d'apprentissage mais il est uniquement traité par les détecteurs spécifiques à chaque classe d'orientation. D'autre part, le problème de ressemblance équivalente dans l'espace image et l'espace des poses évoqué précédemment est pris en charge par l'algorithme d'apprentissage. La plupart des algorithmes proposés dans cette partie sont capables de traiter conjointement les problèmes de détection du visage et d'estimation de la pose, ce qui constitue un avantage certain sur les autres approches.

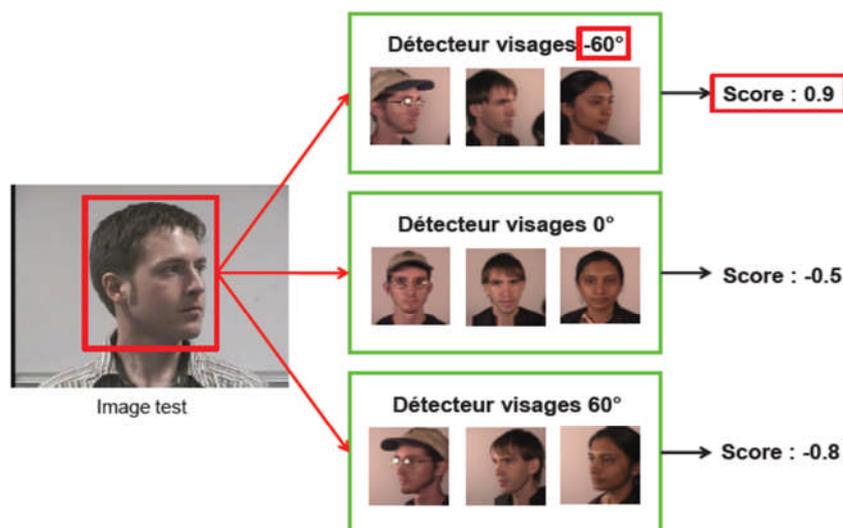


Figure 1.6 – Les méthodes par classification : un détecteur est appris par classe d'orientation et la pose d'un nouveau visage est donnée par le détecteur ayant obtenu le meilleur score. :17

Différentes techniques d'apprentissage ont été utilisées. Dans [13] et plus récemment [Voit et al.] chaque sortie d'un Perceptron Multicouche (MLP de l'anglais Multi Layer Perceptron) correspond à une classe d'orientation. La sortie est active lorsque le visage présenté en entrée correspond à cette classe. [58] entraînent autant de réseaux de neurones qu'il y a de classes. Chaque réseau est une Mémoire Auto- Associative Linéaire qui apprend à synthétiser en sortie, une image ressemblant à celle en entrée. Si le visage en entrée a une orientation proche de celle des visages utilisés pour l'apprentissage, les images en entrée et en sortie seront ressemblantes, sinon elles seront différentes. La pose est alors donnée par le réseau qui a obtenu l'erreur de reconstruction la plus faible. Dans Féraud, le procédé est semblable mais chaque classe

d'orientation est modélisée par un réseau auto-associatif non linéaire et la décision finale est donnée par un MLP en fonction des sorties de chaque réseau. Sauquet et Rowley font également appel à des réseaux MLP. D'autres techniques d'apprentissage sont utilisées telles que la classification bayésienne [3] ou les Séparateurs à Vaste Marge (SVM) [51], [55] [8].

Depuis le succès du détecteur de visages frontaux de [21], de nombreux travaux ont cherché à étendre cette technique à la détection de visages multi-orientations [3],[55]. La solution la plus directe consiste à faire fonctionner plusieurs détecteurs en parallèle (cf. figure 1.7(a)), chacun étant spécialisé pour une classe d'orientation Baluja, [3]. Cette architecture présente trois inconvénients majeurs. Le premier

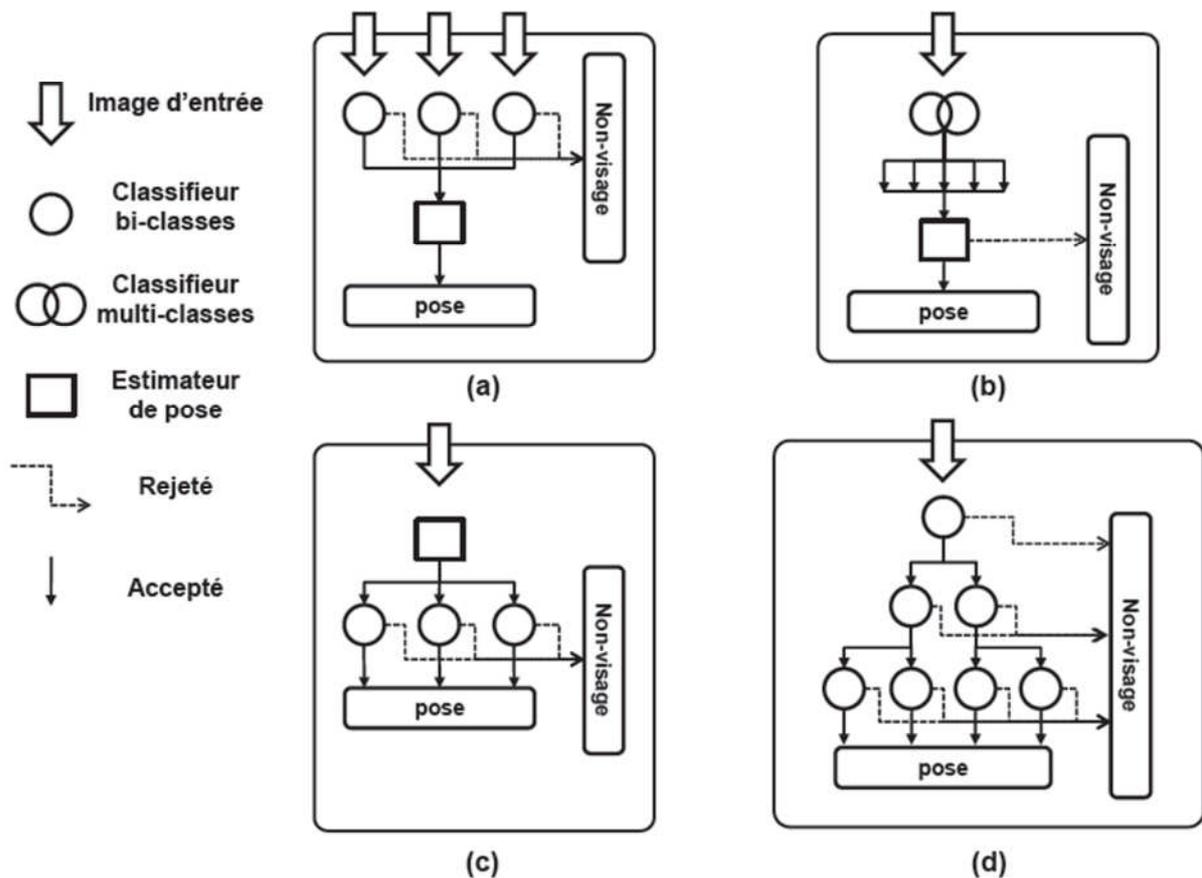


Figure 1.7 – Différentes architectures pour l'estimation de la pose par classification. (a) En parallèle, (b) multi-classes, (c) avec routeur, (d) descendante. :18

Concerne la prise de décision finale puisque différents classifieurs peuvent répondre positivement. Lorsque l'on peut associer une mesure de confiance à la sortie de chaque classifieur, il est possible d'appliquer des règles simples de décision telles qu'une moyenne pondérée ou la règle du « winner takes all » (la pose est attribuée en fonction du classifieur ayant

répondu avec le plus de certitudes) [3] Gourier. Toutefois, trouver la relation entre la sortie des classifieurs et la pose n'est pas toujours si évidente et certains ont recours à un MLP Féraud ou un SVM [51]. Cet expert est également utilisé pour décider si l'image correspond ou non à un visage. Le deuxième problème a trait à son efficacité puisque à chaque classification, il faut faire appel à tous les classifieurs. Le partitionnement de l'ensemble d'apprentissage constitue un autre inconvénient de cette architecture. En effet, il n'est pas évident de déterminer les frontières des différentes classes puisque les changements de pose ne sont pas discrets mais continus. De plus, le nombre d'exemples disponibles pour chaque classifieur est inversement proportionnel au nombre de poses que l'on pourra discriminer.

Dans une architecture multi classes Schiele et Waibel, Zhao, tous les exemples sont utilisés pour apprendre simultanément toutes les classes d'orientation (*cf.* figure 1.7(b)). Toutefois le problème de la prise de décision reste ouvert lorsque plusieurs sorties du classifieur répondent positivement. Schiele et Zhao utilisent une représentation gaussienne en sortie ; non seulement la sortie désirée est active, mais les sorties correspondant à des orientations proches sont également actives avec une intensité moindre. Cela permet de prendre partiellement en compte la continuité des valeurs de pose.

Une autre stratégie est d'inverser les étapes de détection de visage et d'estimation de pose [55]. Considérant que l'image en entrée est un visage, l'algorithme estime sa pose et l'envoie au détecteur concerné qui décide si l'image correspond effectivement à un visage (*cf.* figure 1.7(c)). Ainsi, un seul détecteur est requis à chaque estimation. Rowley estiment la rotation dans le plan et appliquent la rotation inverse à l'image afin que le visage se retrouve dans une position canonique. Le même détecteur de visages frontaux peut être utilisé pour n'importe quelle pose (*cf.* figure 1.8).

Le principal problème est que la fiabilité de ces méthodes repose sur les performances de l'estimateur de pose en amont. De plus, l'approche de Rowley ne peut pas s'étendre aux rotations hors-plan du visage.

[21] proposent de diviser cette tâche en une succession de décisions simples binaires structurée en arbre. Cette solution appartient aux approches descendantes (*coarse to fine* en anglais) illustrée sur la figure 1.7(d). Il existe différentes stratégies descendantes dans la littérature. Par exemple, l'espace des visages est subdivisé en sous-espaces de plus en plus réduits. Dans les premiers niveaux de la pyramide, les visages d'un sous-espace présentent de grandes variations de pose mais sont traités comme un ensemble afin d'extraire leurs caractéristiques communes et de les différencier des non-visages. Dans les étages inférieurs, les caractéristiques sélectionnées

sont discriminantes par rapport à la pose du visage. A chaque étage de la pyramide, les visages sont soit considérés comme des non-visages, soit transmis à tous les classifieurs du niveau

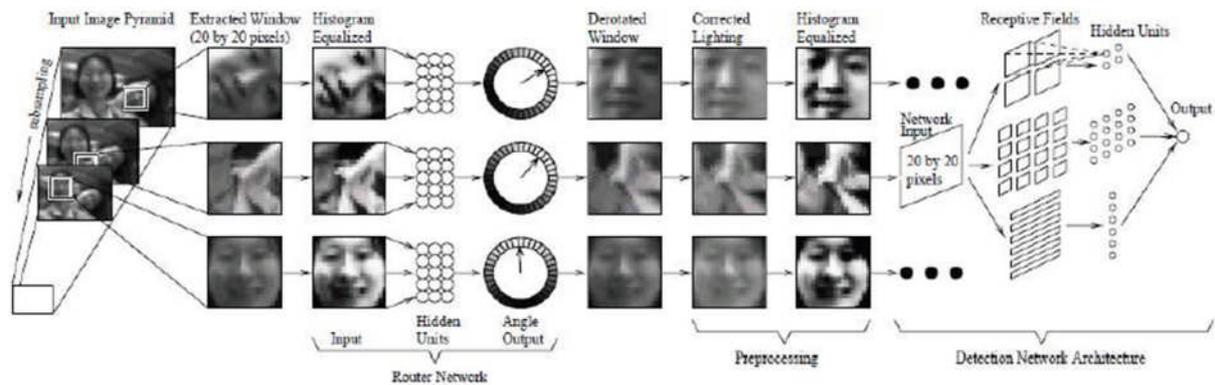


Figure 1.8 – Architecture par routeur proposée par Rowley . :20

suivant. Huang proposent une solution intermédiaire dans laquelle le visage à analyser est transmis à un nombre variable de détecteurs de l'étage suivant. La figure 1.7(d) présente une arborescence générique pour les stratégies « approche-précision » qui ne tient pas compte de toutes les subtilités structurelles exposées dans Huang.

Table 1.2 – Comparaison des différentes architectures présentées dans la figure 1.7

	Parallèle	Multi-classes	Routeur	approche-précision
Partitionnement des données	--	++	--	+
Temps de calcul	--	++	++	+
Simplicité de mise en oeuvre	++	++	++	--
Précision	--	-	--	-
Détection de la tête	+	--	+	++

Le tableau 1.2 synthétise les points forts et les points faibles de chaque architecture.

Les méthodes par classification sont performantes car elles s'appuient sur des méthodes d'apprentissage robustes et éprouvées telles qu'Adaboost ou les SVM. De plus, elles combinent, pour la plupart, deux tâches distinctes : la détection de visages et l'estimation de la pose.

Cette dernière propriété peut toutefois être un problème car les deux tâches s'opposent ; la détection cherche à séparer les visages des non-visages en s'appuyant sur des caractéristiques

partagées par tous les visages quelle que soit l'orientation alors que l'estimation de pose cherche justement à les différencier. Osadchy ont proposé une méthode d'unification des deux tâches. Un réseau de neurones à convolution apprend à projeter un visage sur une variété paramétrée par la pose et à garder les non-visages le plus éloigné possible de cette variété.

1.4.3 Méthodes par régression

Nous avons vu dans la partie précédente que le principal problème était le partitionnement des données. L'objectif des approches par régression est d'apprendre une relation fonctionnelle entre l'apparence d'un visage et sa pose. Ce principe est illustré par la figure 1.9. La relation est modélisée à partir d'un ensemble d'apprentissage étiqueté et fournit une estimation continue de la pose pour n'importe quel nouveau visage. Cette tâche est complexe car il s'agit d'approximer une fonction fortement non-linéaire dans un espace de grande dimension.

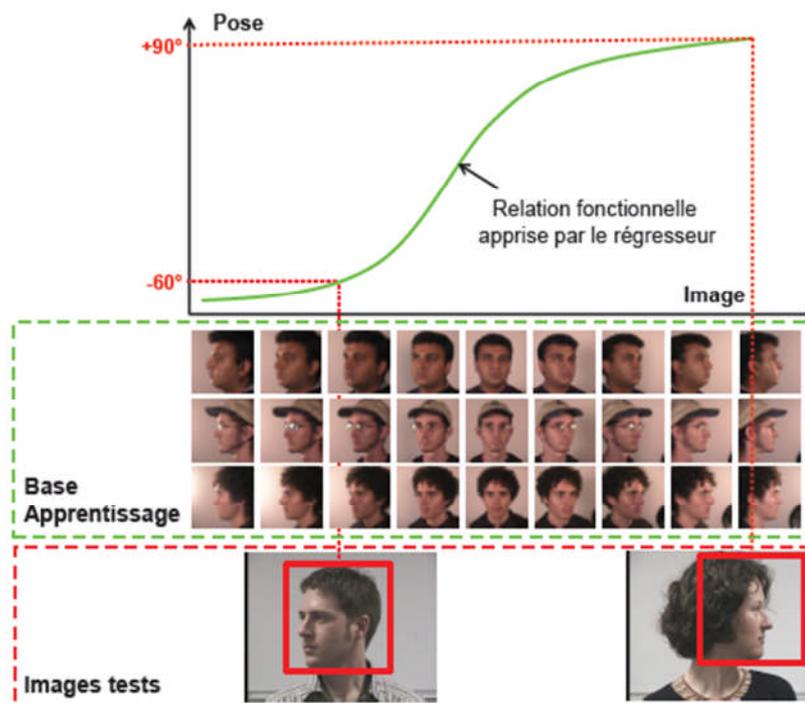


Figure 1.9 – Les méthodes par régression non-linéaire visent à apprendre une relation fonctionnelle entre l'apparence d'un visage et sa pose. :21

[20], [16] utilisent avec succès une Support Vector Regression (SVR) pour apprendre cette fonction. [20] filtrent au préalable l'image par un opérateur de Sobel et réduisent la dimension des images filtrées par une ACP. [16] utilisent la même méthode de représentation que SIFT

[15], une concaténation d’histogrammes des orientations de contours pour différentes zones de l’image.

La régression est le plus souvent apprise par un réseau de neurones. [60] entraînent un MLP (cf. section 1.5.2) ; l’activation d’une sortie est proportionnelle à l’amplitude de la rotation sur un des degrés de liberté de la tête.

Rae et Ritter utilisent une LLM (Locally Linear Map) pour apprendre la relation entre l’espace image et l’espace des poses. L’idée est de subdiviser l’espace d’entrée en régions représentées par des prototypes et d’apprendre une relation linéaire entre la distance d’un exemple à ce prototype et les paramètres de pose (figure 1.10). Le processus d’estimation de la pose d’un visage inconnu consiste à chercher le prototype le plus proche et d’appliquer la régression linéaire correspondante. Cette approche a été étendue à des exemples dont la dimension a été réduite par une ACP Bruske et à des visages représentés par des ondelettes de Gabor Krüger et Sommer.

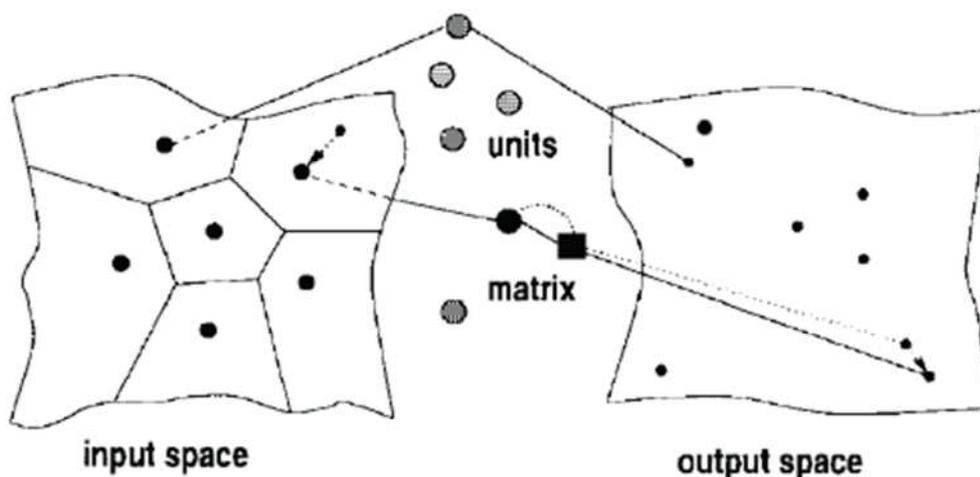


Figure 1.10 – Architecture d’un réseau LLM tirée de Rae et Ritter. :22

Certaines méthodes n’apprennent pas la relation entre l’image et l’espace des poses Mais une projection entre l’espace image et un sous-espace de faible dimension. Dans [42], un réseau RBF (Radial Basis Function) apprend la projection sur une variété à deux dimensions trouvée par Isomap. De même, [43] utilisent un GRNN (Generalized Regression Neural Network) pour apprendre la projection vers un espace réduit.

Les méthodes par régression sont rapides à calculer, n’ont besoin que d’images de visages cadrés (par opposition aux méthodes de classification qui nécessitent également des contreexemples) et sont les méthodes globales les plus précises [16]. Toutefois, ces méthodes réalisent une estimation de la pose indépendamment de l’étape de détection du visage et sont dépendantes de la qualité de ce cadrage.

1.5 Méthodes de comparaison

Dans la section 1.5.2, nous présentons les méthodes proposées lors de la campagne d'évaluation CLEAR 2006 que nous prendrons comme références pour situer notre approche. Nous nous comparons également à un réseau de neurones à convolution dont l'architecture est présentée à la section 1.5.1.

1.5.1 Réseau de Neurones à Convolution

Nous avons entraîné un réseau de neurones à convolution (CNN de l'anglais Convolutional Neural Network) proposé initialement par Le Cun. Ce choix a été motivé principalement par deux raisons : il intègre implicitement une phase d'extraction de caractéristiques et il a été utilisé avec succès dans de nombreuses applications Garcia et Delakis. Le CNN est un réseau multicouche bioinspiré Hubel et Wiesel qui combine trois idées principales : les champs récepteurs locaux, les poids partagés et le sous-échantillonnage. Dans une architecture classique, chaque cellule des couches de convolution est connectée à un ensemble de cellules regroupées dans un voisinage rectangulaire sur la couche précédente. Les champs récepteurs locaux permettent d'extraire des caractéristiques basiques tels que des contours par exemple. Les couches sont dites « à convolution » car les poids sont partagés et chaque cellule de la couche réalise la même combinaison linéaire (avant d'appliquer la fonction sigmoïde) qui peut être vue comme une simple convolution. Ces caractéristiques sont alors combinées à la couche suivante afin de détecter des caractéristiques de plus haut niveau. Entre deux phases d'extraction de caractéristiques, le réseau réduit la résolution de la carte des caractéristiques par un moyennage et un sous-échantillonnage. Cette réduction se justifie à deux titres : diminuer la taille de la couche et apporter de la robustesse par rapport aux faibles distorsions. La figure 1.11 décrit l'architecture que nous avons développée en suivant les recommandations de Simard.

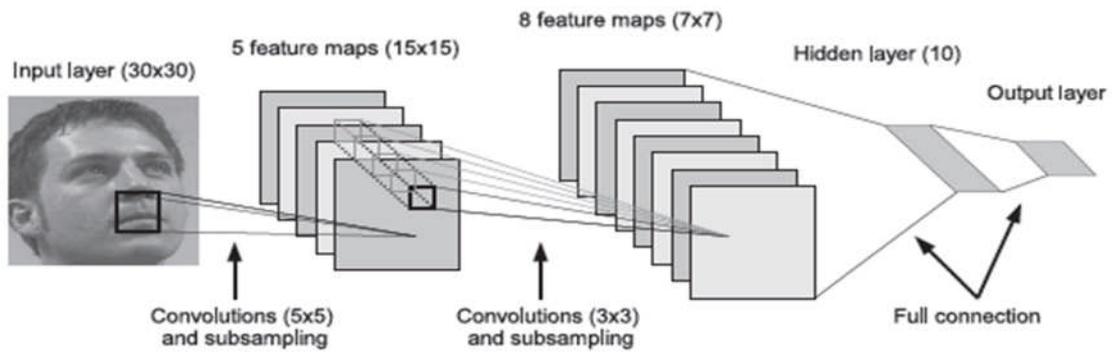


Figure 1.11 – Architecture de notre réseau de neurones à convolution

Le réseau est entraîné avec un algorithme classique de rétro propagation. Nous avons utilisé un pas adaptatif et une régularisation du gradient par un moment Um.

1.5.2 Méthodes de l'évaluation CLEAR 2007

Nous présentons dans cette partie, les méthodes proposées lors de la campagne d'évaluation CLEAR'07 Stiefelhagen et Garofolo. Trois méthodes ont obtenu des résultats sur la base Pointing 04 et serviront de références pour comparer nos algorithmes.

1.5.2.1 Approche par un perceptron multicouches (MLP, Multi Layer Perceptron)

[60] proposent une approche en deux étapes. Dans un premier temps, un classifieur détecte les pixels de teinte chair. Il s'agit d'un classifieur dont la frontière de décision est linéaire dans l'espace colorimétrique Teinte Saturation Valeur (HSV, de l'anglais Hue Saturation Value). La zone du visage correspond à la boîte englobante de la plus grande composante connexe. Dans un second temps, la pose de la tête est estimée par un réseau MLP. L'image du visage est redimensionnée en 64×64 pixels, convertie en niveaux de gris et l'histogramme est égalisé. On extrait une carte des modules des gradients de l'image.

L'entrée du réseau de neurones correspond à la concaténation de l'image du visage et de la carte des modules. Le réseau a une sortie par angle à estimer. La figure 1.12 illustre l'architecture du MLP.

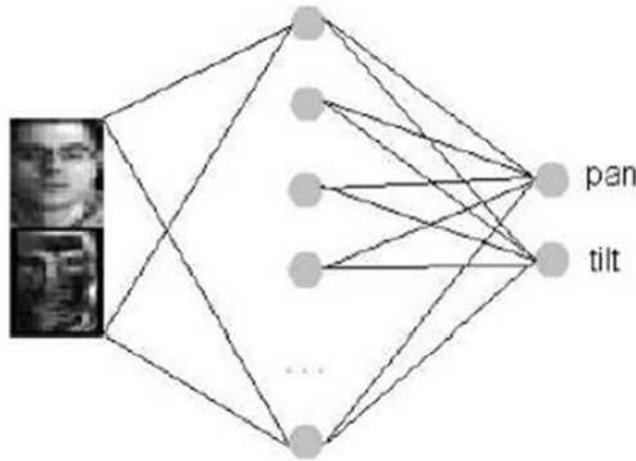


Figure 1.12 – Architecture du réseau MLP proposée par [60].

1.5.2.2 Approche par des mémoires auto-associatives linéaires (MAAL)

Gourier sépare également l'étape de détection du visage de celle d'estimation de la pose. La première étape, décrite en détail dans Gourier consiste à détecter les pixels de teinte chair par une décision bayésienne. La densité de probabilité conditionnelle d'un pixel d'appartenir à une région de peau peut être estimée en utilisant un histogramme de la chrominance. Le calcul de la chrominance (r, g) d'un pixel (x, y) est effectué en normalisant les composantes rouge et verte du vecteur de couleur $(R + G + B)$ par son intensité lumineuse $R + G + B$. La probabilité $p(x, y) = \text{Peau} | (r, g)$ pour un pixel (x, y) de chrominance (r, g) d'être de teinte chair est donnée par:

$$p(x, y) = \frac{p(r, g)}{p(r, g) + p(\text{non-peau})} \quad (1.8)$$

Les probabilités sont calculées pour tous les pixels de l'image. Les premiers et deuxièmes moments de ces cartes de probabilités donnent la position et l'orientation et le facteur d'échelle du visage dans le plan image. Le visage est alors normalisé en taille et en inclinaison dans une vignette de 23×30 pixels en niveaux de gris (cf. figure 1.13)



Figure 1.13 – Détection et normalisation des visages proposées par Gourier

Dans une deuxième étape, une MAAL est apprise par orientation. Ainsi, lorsqu'on présente en entrée une image \mathbf{x} dont la pose p correspond à celle apprise par la MAAL, l'image reconstruite \mathbf{x}_p en sortie sera très semblable à \mathbf{x} . Le processus d'estimation de pose consiste donc à sélectionner le réseau qui maximise la ressemblance entre l'image source et l'image reconstruite :

$$pose = \arg \max_p (\cos(X, X_p)) \quad (1.9)$$

1.5.2.3 Approche par un modèle tensoriel

Cette méthode a été proposée par [59]. Pendant la phase d'apprentissage, le bout du nez de chaque visage est localisé manuellement et une vignette de taille 18×18 pixels est extraite autour de cette position. Les vignettes de la base d'apprentissage sont organisées dans un tableau de dimension $324 \times 15 \times 13 \times 7$ correspondant respectivement au nombre de pixels, d'identités, de variations de l'angle pan, et de variation de l'angle tilt (*cf.* figure 1.14(a)). Le modèle tensoriel, illustré sur la figure 1.14(b) est créé à l'aide d'une analyse en composantes indépendantes multilinéaire, comme proposé par Vasilescu et Terzopoulos.

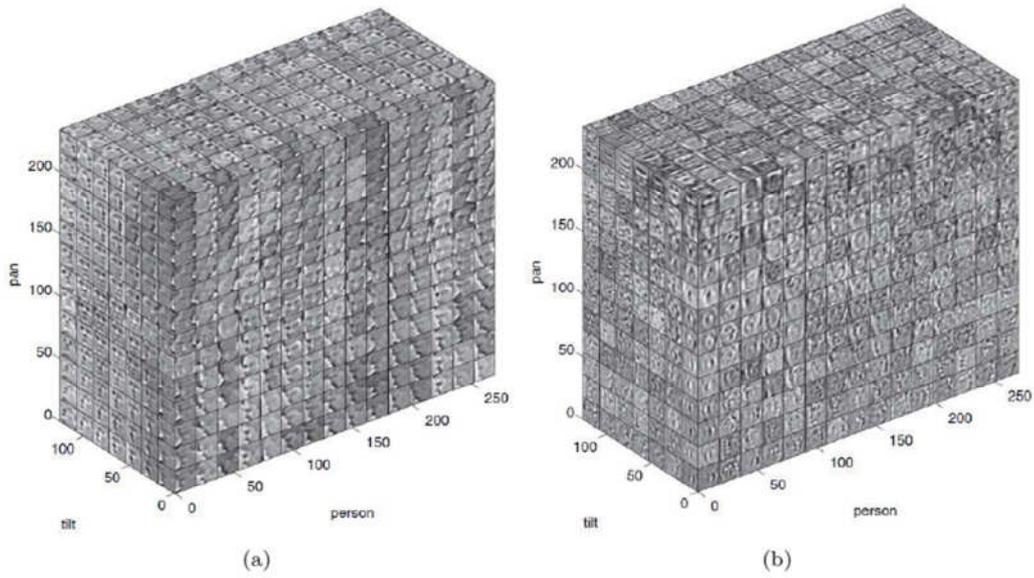


Figure 1.14 – Construction du modèle tensoriel. (a) Organisation matricielle des vignettes. (b) Modèle tensoriel proposé par Vasilescu et Terzopoulos .

En test, le visage est grossièrement localisé à l’aide du modèle de teinte chair proposé par Jones et Rehg et la pose est inférée en projetant l’image sur le modèle tensoriel (*cf.* figure 1.15). La localisation est affinée en maximisant la corrélation entre l’image extraite et l’image reconstruite à l’aide du modèle tensoriel. Pour finir, la pose est affinée par un algorithme du plus proche voisin dans un espace réduit par une analyse en composante principale (ACP).

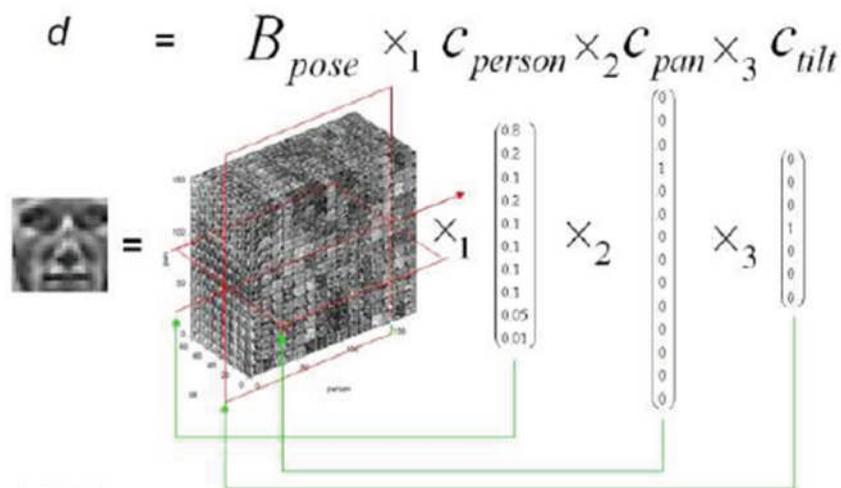


Figure 1.15 – Illustration de l’estimation de la pose par projection sur le modèle tensoriel tirée de [59]

Il existe d'autres méthodes qui ont été testées sur la base pointing 04 [57], Stiefelhagen mais nous ne les prenons pas en compte dans nos comparaisons car elles ne suivent pas le protocole défini pour l'évaluation CLEAR. Dans Stiefelhagen par exemple, 80% des exemples sont utilisés en apprentissages, 10% en validation et seulement 10% en test. [57] utilisent toute l'image et ne présélectionnent pas la zone du visage. Par ailleurs leurs résultats sont moins bons quelle que soit la méthode de réduction de dimension envisagée.

1.6 Conclusion

Dans ce chapitre, nous avons décrit la chaîne de traitement des méthodes globales d'estimation de la pose de la tête. On a focalisé notre analyse sur l'étape de représentation des images et l'étape de décision que nous l'avons divisée en trois catégories : les approches par comparaison avec des prototypes, par classification et par régression.

Ce chapitre a révélé la simplicité de mise en oeuvre et le caractère intuitif des méthodes par comparaison avec des prototypes. Nous nous sommes également intéressé aux méthodes par régression car elles sont à la fois rapides et plus précises que les autres approches globales.

2.1 Introduction

Certaines voix critiquent la reconnaissance du visage comme biométrie pour les systèmes en raison pour sa faible précision. Mais dans la dernière décennie, les recherches sur l'intelligence ambiante (AMI) a apporté une identification en évidence non critique de personnes [67]. Alors que la précision de la reconnaissance du visage est encore inférieure à d'autres données biométriques comme l'iris ou des empreintes digitales, c'est une approche très appropriée pour éviter toute intrusion, soit de façon autonome procédée ou être condensée à un système multimodal. Mais non intrusif n'est pas gratuit !!! Cela signifie généralement que images de visages disponibles (acquises) vont souffrir de mauvaise qualité (basse définition, le manque de fidélité des couleurs et autres) et la résolution variable, un problème analysé dans [62]. Dans les cas où l'ennemi à battre est le manque de résolution, la reconstruction... certaines techniques ont démontré que ça peut être une approche appropriée. Par exemple le cas de visage hallucination [61] qui génère des images de super-résolution à partir d'une petite face d'entrée, en utilisant un système spécialement formé pour préserver (ou recréer) les traits du visage. Plusieurs algorithmes d'hallucination [64, 68] ont été développés depuis que le travail de pionnier a été publié. Les résultats empiriques montrent que les images générées améliorent la précision de la reconnaissance des visages. L'hallucination recrée les composants perdus à haute fréquence en exploitant la connaissance sur la structure du visage humain acquis pendant la formation. Cela pourrait être un problème, parce que l'information ajoutée est inventée en fonction de ce que nous avons appris à partir d'un ensemble de données réduites: par rapport aux images d'entrée de faible résolution, l'apparition de visages hallucinés est plus semblable à ceux de l'ensemble de données de formation, augmentant ainsi la probabilité de faux positifs. Un autre aspect négatif de cette technique est que, dans certains cas, les images reconstruites montrent des artefacts et des déformations qui conduisent à des erreurs de classification pendant la reconnaissance. Une autre technique liée à la reconstruction consiste à synthétiser les vues 3D des visages ensemble de données de formation, d'un ou plusieurs images 2D [63]. Le taux d'erreurs de reconnaissance pour la classification de profil (pose) est d'environ la moitié de celui obtenu sans l'aide de cette technique [66].

Revenir à notre scénario d'environnement intelligent, l'obstacle majeur pour la reconnaissance de visage n'est pas la résolution, mais plutôt les conditions de capture. Ils ne sont pas contrôlés : variabilité des illuminations de visage, pose et taille, bruit introductif dû à la poussière dans

l'objectif de la caméra (images floues et hétérogène en raison de la combinaison de caractéristiques des appareils dans des environnements coopératifs). En outre, le matériel d'acquisition et de traitement des données peut probablement limiter la puissance de calcul [67], l'analyse et l'évaluation des performances de reconnaissance de visage des différents appareils portables, y compris les téléphones mobiles. La reconnaissance faciale par ordinateur a atteint des résultats très satisfaisants dans les dernières années grâce aux efforts d'une communauté très active, dépassant même les capacités humaines avec des visages non familiers [65]. La mise en revue de la littérature existante, montre qu'il n'est pas difficile de trouver un certain nombre d'algorithmes appropriés pour la construction de système performant. Néanmoins, la mise en œuvre d'un système de reconnaissance de visage haut de gamme satisfaisant à toutes les exigences est un projet coûteux et nécessite des améliorations. Maintenant, nous allons soulever une question: compte tenu de la marge de gain de performance et les conditions de ce scénario en béton, comment construire le meilleur système possible en valeur de coûts associés? Notre hypothèse est que nous pouvons utiliser l'un parmi ceux qui sont largement étudiés à base d'approches simples, et encore obtenir de bons résultats tout en affinant l'algorithme. Pour faire un bon choix un état de l'art sur les techniques nouvelles s'impose. C'est ce que nous tentons d'exposer dans ce présent chapitre.

2.2 Etats de l'art

Les auteurs **A-Nasser Ansari et al** [69] ont travaillé en 2008 sur une approche multimodale pour la modélisation 3D du visage et la reconnaissance en utilisant le masque facial 3D déformable. Il s'agit d'une approche multimodale pour la modélisation 3D du visage et la reconnaissance en utilisant les informations 2D et 3D. La nécessité d'une modélisation précise de la tête pour l'animation faciale réaliste, la compression et la reconnaissance a conduit les chercheurs à explorer des algorithmes basés sur des modèles 3D. L'objectif dans cette recherche est d'utiliser une approche multimodale pour représenter les données faciales 3D, capturées à partir d'images stéréo d'un sujet donné, par un modèle de masque difforme générique mettant l'accent sur la zone la plus distinctive du visage: la région centrale qui comprend les yeux, la bouche et le nez. Le processus de déformation capture essentiellement les caractéristiques de surface codées dans les données faciales et représente le visage avec un nombre réduit de points de données discriminatoires, à savoir les sommets du modèle de masque mesh. Cette approche multimodale pour la modélisation de visage et reconnaissance utilise trois caméras pour capturer des images stéréo, les deux frontales gauche-droite et une de profil voir figure 2.1.

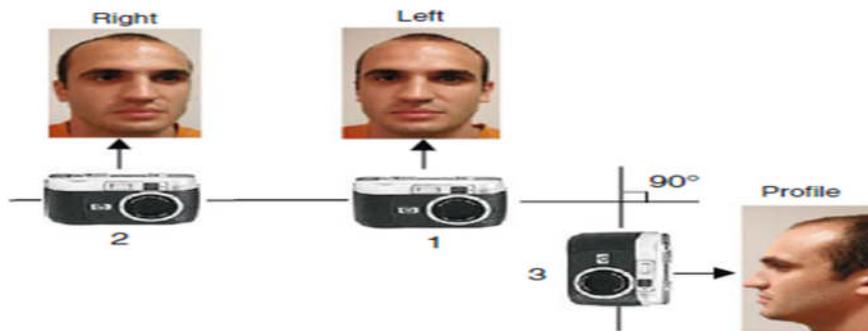
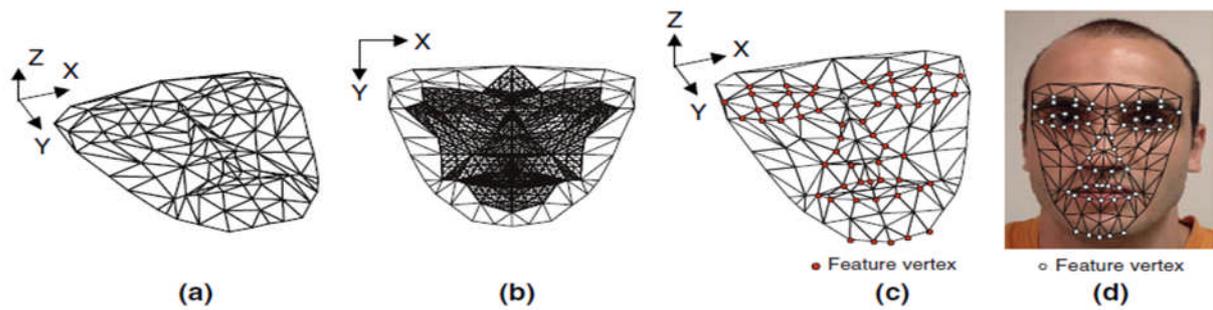


Figure 2.1 Système d'acquisition de l'image

Les traits du visage en 2D sont extraits de l'image de face et une carte de disparité dense est calculée à partir des deux images frontales. L'utilisation des fonctions 2D extraites et leurs disparités correspondantes permet le calcul de leurs coordonnées. La carte des écarts est calculée en utilisant les images frontales des caméras 1 et 2 et la caméra 3 de profil, complète l'information des deux caméras frontales et fournit des informations de fonctionnalité qui sont difficiles à obtenir avec précision par des caméras frontales. Par exemple, la pointe du nez et de l'ensemble de la silhouette de profil sont extraites de l'image de la caméra 3. On voit que l'addition de l'image de la vue de profil améliore la modélisation de la surface et le taux de reconnaissance. L'alignement 3D est réalisé d'où l'obtention de 3D « mesh model » de basses résolutions aux paramètres 3D (figure 2.2). Ensuite on projète ses sommets sur l'image frontale 2D et on ajuste ses sommets profil en utilisant l'image de la vue de profil. Pour atteindre cet objectif, le modèle doit être tourné, traduit et mis à l'échelle. L'équation 4 donne la somme de l'erreur quadratique pour les points caractéristiques N en termes d'échelle S , rotation R , et translation T .

$$\text{Min} \quad E(S, R, T) = \sum_{j=1}^N \|P_i - V_i\|^2 \quad (2.1)$$



a modèle 3D, **b** masque de modèle 2D, **c** suite après l'alignement du modèle 3D aux caractéristiques extraites, **d** projetée modèle de **c** sur l'image

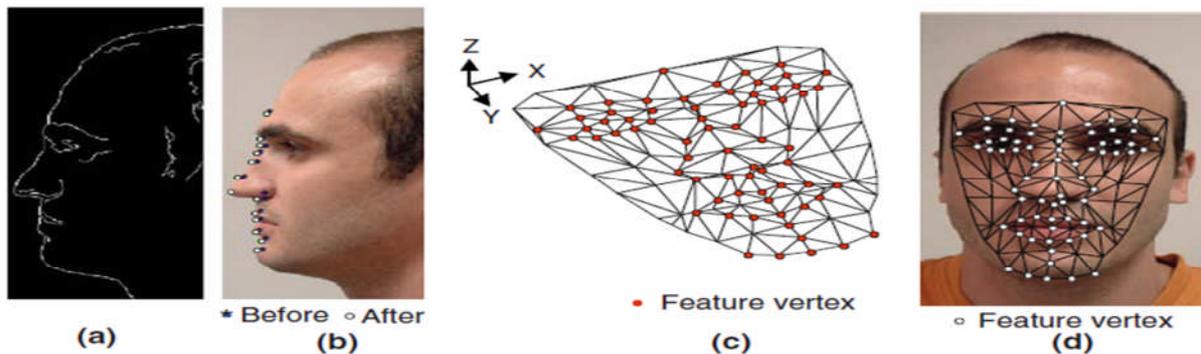


Figure 2.2 - **a** profil pointe l'image **b** sommets Silhouette avant et après ajustements, **c** modèle 3D final ajusté **d** projetés modèle de **c** sur l'image

La résolution du modèle 2D obtenue la région centrale est augmentée pour obtenir un modèle de masque facial couvrant les traits distinctifs du visage. Pour le visage d'un sujet donné suite à un modèle 3D de masque déformé spécifique on extrait les coordonnées 2D des sommets, ainsi que leurs écarts. Figure 2 **c**, **d** illustrent le résultat après avoir aligné le modèle 3D pour les fonctions 3D calculées, et ses sommets projetés sur l'image.

La méthode intègre des informations à partir des traits du visage extraite de la modalité d'image 2D avec des informations à partir de la modalité 3D obtenus à partir des images stéréo. Application des modèles dans la reconnaissance de visage en 3D, pour 112 sujets, valide un taux d'identification de 95% et le taux de vérification de 92% à 0,1% de taux de fausse acceptation. La stéréo a montré l'augmentation de la résolution des modèles de maille pour fournir un taux de reconnaissance élevé. Les résultats sont encourageants pour la modélisation 3D du visage et de la reconnaissance. Des méthodes fiables sont étudiées pour incorporer des expressions faciales dans ces modèles et de démontrer des résultats de reconnaissance sur un grand nombre de sujets.

Une approbation juridique de tous les sujets dans la base de données est également en cours afin de fournir aux chercheurs la première base de données accessible au public des images faciales stéréo.

Les auteurs **Timothy C et al** [70] se sont intéressés en **2008** sur les signatures de profil pour la détection robuste d'entité 3D. On introduit un nouveau système de reconnaissance faciale 3D qui est robuste aux variations de poses. De grands degrés de variation de poses du visage peuvent conduire à une fraction significative d'occlusions des caractéristiques visibles dans les images frontales. Une fonction automatique de haute précision et détection de pose est réalisée par une nouvelle technique appelée Rotated Signatures de Profil (RPS). Les expériences sont effectuées sur la plus grande base de données disponible de visages 3D acquises en vertu de divers pose. Cette base de données contient plus de 7300 images total de 406 sujets uniques se sont réunis à l'Université de Notre Dame. Les résultats expérimentaux montrent que l'algorithme de détection RPS est capable d'effectuer la détection de nez avec une précision supérieure à 96,5% dans la variation poser représentées dans les données utilisées.

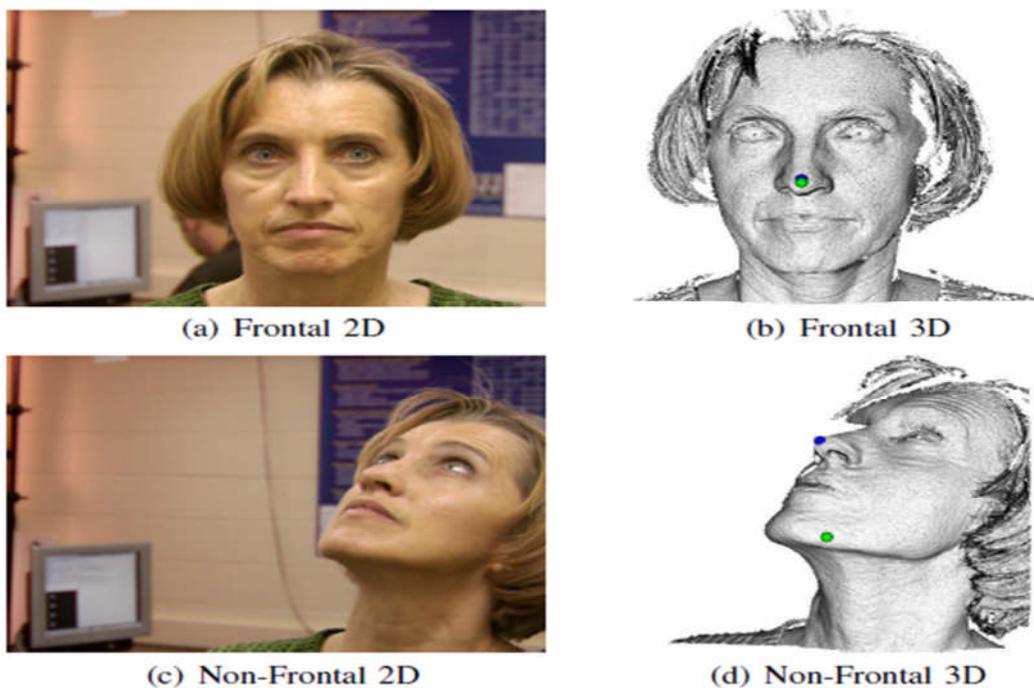


Figure 2.3 - Détection de nez utilisant des images frontales et non frontal (Sous réserve 04385).

Le bout du nez détecté par l'algorithme RPS est représenté par la sphère bleue, et la pointe du nez située par le Z-heuristique est représentée par la sphère verte.

Les images pour les expériences réalisées dans le présent document sont tirées de la plus grande base de données 3D pose de visage de variation connu au moment de la rédaction: NDOff2007. La norme 3D bases de données de visage comme le Face Recognition Grand Challenge (FRGC) v2.0 [71] ne contiennent pas une composante importante de non-pose frontale images. L'ensemble de données contient NDOff2007 un total de 6911 images non-frontales contenant des expressions neutres et une image neutre frontale unique pour chacun des 406 sujets distincts. Les exemples de poses différentes dans l'ensemble de données NDOff2007 peut être vu dans la figure 2.4 et le nombre d'images par pose se trouvent dans le tableau 2.1.

L'espace pose n'est pas uniformément échantillonné, mais, en raison de la symétrie du visage, les résultats de la gauche devraient être similaires à ceux de la droite.

Les images ont été acquises avec un Vivide 910 scanné de gamme Minolta [72]. Le Minolta 910 utilise une bande de laser et la triangulation pour construire un scan d'intervalle. L'image est initialement dans un format d'image de gamme de 640x480 avec un champ indiquant si un point 0 3D n'est pas présent, et 1 si il est présent. Les deux textures de couleur (r, g, b) et l'emplacement 3D (x, y, z) sont produites, mais pas en même temps, comme le ruban laser nécessite quelques secondes pour franchir l'image de la couleur visage et est prise après la forme. Cela peut entraîner des erreurs de repérage de texture ou de déformation du maillage de mouvement du sujet. Le nombre de points 3D sur une image de face frontale est généralement autour de 110.000, et dépend de l'objectif utilisé ainsi que le capteur à distance du sujet.

TABLE 2.1 - DISTRIBUTION DES 7317 SCANS PAR POSE

		Yaw										
		90	60	45	30	15	0	-15	-30	-45	-60	-90
Pitch	60		140									
	45			163	155	165	88			160		
	30			204	193	186	201		203			
	15			205	183	187	189	192				
	0	175	158	155	173	186	406	180	177	160	173	170
	-15			163	171	123	140	152				
	-30			154	165	173	172		175			
-45			183	161	141	160			157			



Figure 2.4 - Exemple varie images correspondant à la figure 2.4 (Objet 05078). Les images montrent l'évolution de lacet (-90° à +90°), et les changements de hauteur (-45° à +45°).

Dans cet article, on démontre une méthode précise et efficace pour la détection de nez 3D et catégorisation de pose. Sur un processeur 2,4 Ghz Xeon après prétraitement de l'image, la détection de nez peut être effectuée en moins d'une seconde par image. Ce temps de traitement peut être réduit encore davantage si une approche multi-résolution est employée. Les expériences dans ce document sont réalisées sur le plus grand ensemble à ce jour, avec des degrés divers de variation de pose et est composé de plus de 7.300 images de visages 3D.

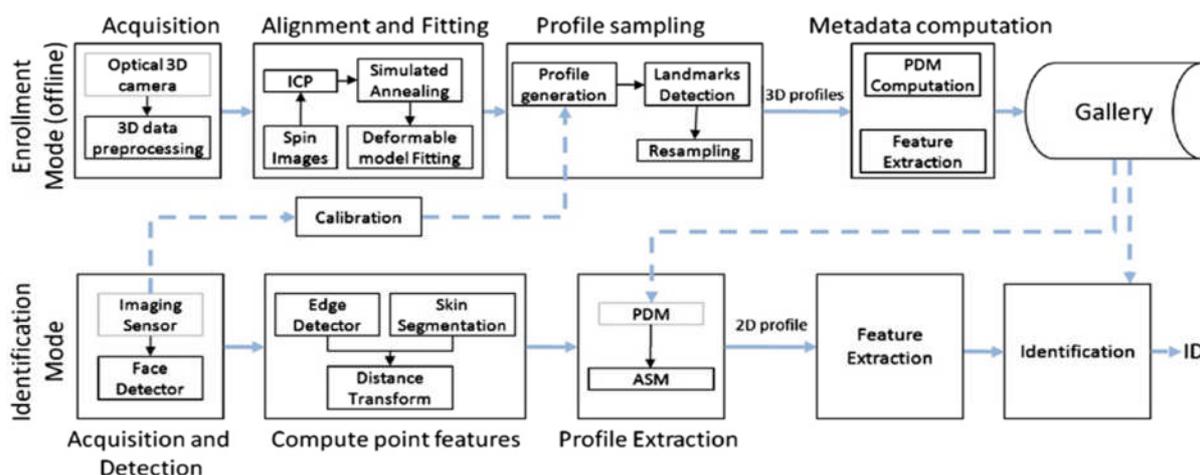


Figure 2.5 - Les composants du système intégré de reconnaissance profil 3D assistée par ordinateur (URxD-PV).

La figure ci dessus présente un système entièrement automatique (URxD-PV) pour la 3D assistée par vision latérale de reconnaissance du visage à base de profils. Il est destiné à être utilisé en l'absence de vue frontale des images de visage. Le système utilise un balayage 3D de la face pour modéliser la géométrie du profil de contour du visage sous diverses poses. Des méthodes sont

fournies pour toutes les étapes du système intégré avec précision de repères, la modélisation statistique des formes, la génération de profils 3D, et l'extraction des profils du côté des images en vue.

De nombreuses expériences sont effectuées pour étudier les facteurs de base (tels que la qualité des données 3D, schémas de fusion pour multi trames de reconnaissance et la configuration du domaine d'échantillonnage) qui peuvent influencer la performance d'un Système de reconnaissance basé sur les profils 3D assistée par ordinateur.

Les auteurs **Quan Ju et al** [73] en **2009** ont utilisé une localisation de visage 3D par une méthode appelée Binary Réseaux de Neurones (BNR). Dans cet article, une méthodologie pour l'identification des caractéristiques du visage et de l'approche de localisation est proposée basé sur des algorithmes de réseaux de neurones binaires. On présente une pose de la tête et l'expression du visage comme descripteur invariant de forme 3D appelé Maillé Multi Cercle Descripteur de Courbure (MMCCD), qui fournit plus d'attributs de courbure 3D que d'autres approches similaires. Pour faire correspondre les modèles dotés de plusieurs attributs, on utilise l'architecture de raisonnement incertain avancée (AURA) avec codage par algorithme KNN (proches voisins). Des expériences effectuées sur l'ensemble des données FRGC (4950 3D) variation de poses et d'expressions montrent que l'approche est capable d'atteindre une identification précise (plus de 99,69% de l'identification de la pointe du nez) et une localisation des traits du visage robuste.



Figure 2.6 - Un exemple de mauvaise correspondance 2D-3D

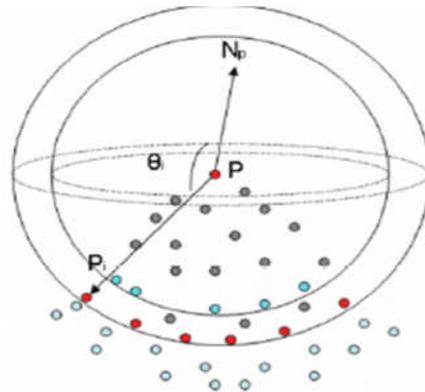


Figure 2.7 - P et son point voisin dans les deux sphères.

Ce travail présente une méthode basée sur MMCCD et l'algorithme k-NN AURA pour identifier et localiser des caractéristiques de visage. Contrairement à d'autres techniques similaires utilisant des visages 2D, cette approche est une analyse de la forme 3D pure qui est naturellement invariable aux variations d'illumination. Une identification avec un taux de localisation de pointe de nez de 99,69% dans une grande base de données avec des variations d'expression ont démontré la robustesse et l'efficacité de notre méthode. Dans les expériences futures, cette méthode sera mise en œuvre sur des ensembles de données plus difficiles à vérifier sa robuste possibilité de poser une plus grande variation.

Les auteurs **Ognjen Rudovic et al** [74] en **2009** ont étudiés les expressions faciales avec normalisation Gaussienne pour l'invariance de poses en utilisant un processus de régression. On présente un système fondé sur la régression pour expressions faciales avec normalisation de poses invariantes de la tête. Le problème par la cartographie des emplacements des points de visage 2D (par exemple coins de la bouche) de pose non frontale à la pose frontale est abordé. Cela se fait en deux étapes : premièrement, on propose un estimateur de pose de la tête (cartes d'entrée des emplacements de points du visage 2D) dans un espace pose-tête LDA multi-classe défini par une variété de dimension basse ou moyenne. Ensuite, pour apprendre les correspondances entre un ensemble discret de pose de tête non-frontale et la pose frontale, on propose d'utiliser une régression du processus gaussien (GPR) pour chaque paire de cible pose (c'est à dire une non-frontale et la pose frontale). Pendant le test, l'estimateur la pose de la tête est utilisée pour activer le modèle GPR le plus pertinent qui est ensuite appliqué pour projeter les lieux des points de repère du visage 2D à partir d'une pose arbitraire (qui ne doivent pas être une pose de la formation) de la pose frontale. Dans ces expériences, on montre que le système

proposé (i) effectue avec précision pour la pose de la tête en continu dans la gamme de 0° à 45° rotation et de 0° à 30° rotation d'inclinaison en dépit du fait que la formation a été réalisée uniquement sur un ensemble discret de poses, (ii) gère avec succès deux visages expressifs et inexpressif (même dans les cas où certaines catégories d'expression manquaient dans certaines poses pendant la formation), et (iii) surpasse à la fois du modèle de distribution de points 3D (3D-PDM) et de régression linéaire (LR) modèle qui sont utilisées comme méthodes de référence pour poser la normalisation.

La méthode proposée est évaluée expérimentalement sur les données du 3DFE base de données de l'expression du visage BU.

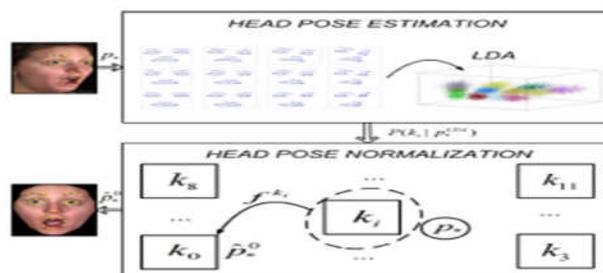


Figure 2.8 - L'approche proposée.

Vingt points marqués avec une couleur jaune sont utilisés comme entrée pour l'estimation de pose. Dans le second bloc, couple à deux GP permet d'obtenir des prévisions plus précises dans la pose frontale.

Les résultats expérimentaux montrent que la méthode proposée surpasse à la fois la méthode 3D face forme de base, la 3D PDM, et la méthode de base 2D, c'est à dire fondée sur la méthode LR. Enfin, les résultats expérimentaux montrent clairement que la méthode proposée peut réussir à traiter le problème de disposer d'un ensemble de données de formation incomplète, par exemple, lorsque des exemples de certaines expressions faciales ne sont pas inclus dans l'ensemble de données de formation pour une pose donnée

Les auteurs **Gabriele Fanelli et son équipe** [75] ont travaillé en **2010** sur l'estimation de la pose avec la régression aléatoire. Des algorithmes rapides et fiables pour estimer la pose du visage sont essentiels pour de nombreuses applications et tâches d'analyse de visage de niveau supérieur. Dans ce travail le problème de la pose de la tête est estimé à partir des données de profondeur, qui peuvent être capturées à l'aide de technologies de détection 3D de plus en plus abordables et disponibles de nos jours. Pour assurer la robustesse, on formule l'estimation de la pose en tant que problème de régression. Bien que la détection des parties du visage spécifiques comme le nez est sensibles aux occlusions, l'apprentissage de la régression sur les correctifs de

surface plutôt génériques exige d'énormes quantité de données de formation afin de parvenir à des estimations précises. On propose d'utiliser les régressions aléatoires pour la tâche à accomplir, compte tenu de leur capacité à gérer de grands ensembles de données. En outre, on synthétise une grande quantité de données d'apprentissage en utilisant un modèle statistique du visage. Ces caractéristiques aléatoires (comme toute variable explicative) doivent être formés sur des données étiquetées et la précision dépend de la quantité de la formation, ceci peut enfin nous permettre de surmonter certains des problèmes des méthodes basées sur des données 2D. Cependant, les méthodes basées sur la profondeur ont besoin d'initialisation manuelle, et ne peuvent pas traiter de grandes variations pose en temps réel. Une exception faite sur les approches comme celle présentée par [76], où les auteurs atteignent des performances en temps réel pour exploiter la puissance de traitement massivement en parallèle sur GPU. Leur approche repose sur un descripteur géométrique qui fournit des hypothèses de localisation de nez qui sont ensuite comparées pour un grand nombre d'enduits d'un modèle de visage générique, effectuée en parallèle sur le GPU. Le temps de calcul rapide rapporté est uniquement réalisable à condition que le matériel graphique spécifique est disponible.

Dans ces expériences, on montre que l'approche peut manipuler les données réelles présentant de grand changements de pose, des occlusions partielles, et les expressions faciales, même s'il est formé uniquement sur des données de visage neutre synthétiques. Le système a été soigneusement évalué sur une base de données accessible au public on atteint les performances semblable à ceux de l'état de l'art sans avoir de recourir aux cartes graphiques.



Figure 2.9 - Tête en temps réel pose exemple d'estimation.

La précision dépend de la quantité des données d'acquisition de la formation, qui est une question clé. On résout ce problème en formation uniquement sur des données synthétiques, la

génération d'un nombre arbitraire d'exemples d'entraînement sans avoir besoin de laborieux et annotations d'erreurs. Ce système fonctionne en temps réel sur un base trame par trame, sans initialisation manuelle ou calculs coûteux. Dans ces expériences, on montre qu'on peut gérer de grandes changements de poses et des variations telles que les poils du visage et des occlusions partielles, par exemple, en raison de lunettes, les mains, ou des pièces manquantes dans la reconstruction 3D. De plus, comme il ne repose pas sur des caractéristiques spécifiques, par exemple, pour la détection de la pointe du nez, ce procédé peut être adapté à la localisation d'autres parties du visage. La performance du système est évaluée sur une base de données accessible au public les résultats sont comparables ou supérieurs à ceux obtenus dans l'état de l'art.

Les auteurs **Boris Efraty et al** [77] en **2011** ont abordé la reconnaissance faciale 3D assistée par ordinateur basée sur les profils. Ce travail présente un cadre automatique pour la reconnaissance faciale basé sur un profil de visage silhouette (URxD-PV). Des recherches antérieures ont démontré le potentiel discriminant élevé de cette biométrie. Par rapport aux approches traditionnelles de reconnaissance basée sur les profils, cette approche n'est pas limitée au seul standard de vue latérale de visages. On propose d'explorer l'espace des caractéristiques des profils sous diverses rotations à l'aide d'un modèle 3D du visage. Dans le mode d'inscription, les données 3D de sujets sont acquises et utilisées pour créer des profils sous différentes rotations. Les caractéristiques extraites à partir de ces profils sont utilisés pour former un classificateur. Dans le mode d'identification, le profil est extrait de l'image en vue latérale, et les métadonnées sont mises en correspondance avec les métadonnées de la galerie. On valide l'exactitude des URxD-PV en utilisant des données provenant de bases de données accessibles au public.

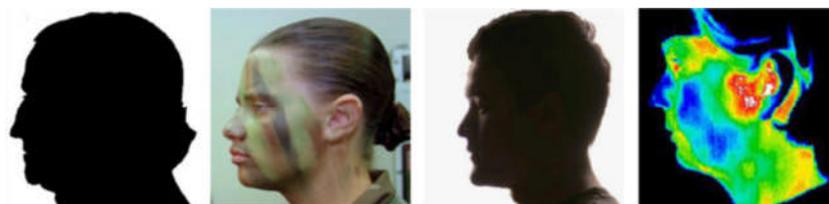


Figure 2.10 - Exemples de vue latérale images où la texture des couleurs fiables ne sont pas disponibles (de gauche à droite): image binaire, camouflage, rétro-éclairage, et près de l'imagerie infrarouge.

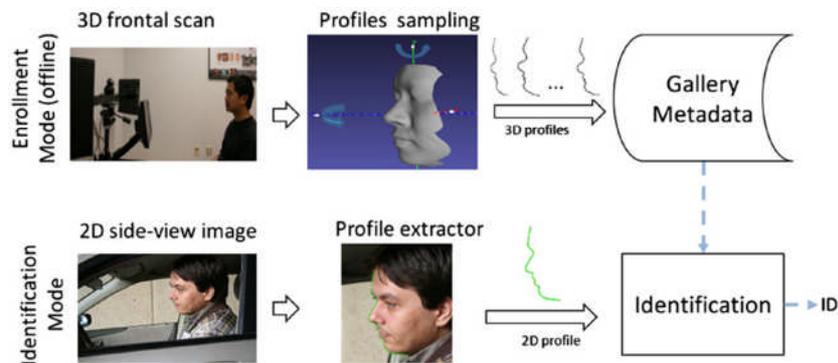


Figure 2.11 - D'inscription et d'identification des modes du système de reconnaissance faciale à base

Mode d'enrôlement

Pendant le mode d'inscription (E) des données brutes acquises à partir de chaque objet sont converties à des métadonnées et stockées dans la base de données. Les étapes de cette conversion sont les suivantes:

E1. Acquérir une forme du visage avec un scanner 3D et le convertir en une représentation de maillage polygonal.

E2. Aligner et ajuster les données 3D à un modèle de référence commun.

E3. Générer plusieurs profils synthétiques en échantillonnant une plage prédéfinie d'angles de rotation et de localiser un ensemble anatomiques sur les profils générés

E4. Dériver un ensemble de caractéristiques en fonction de la géométrie du profil et l'emplacement de point de repère à partir de profilés et de les stocker sous forme de métadonnées devant être utilisé dans la phase d'identification.

Étape E1: Une condition essentielle du système est grande précision dans les régions le long de la ligne centrale de symétrie du visage.

En fait, la précision de la surface dans certaines régions (par exemple, près des oreilles) n'assume aucune importance parce que ces régions n'apparaissent jamais sur les profils silhouette. On applique un certain nombre d'étapes de prétraitement tels que le trou de remplissage, lissage, et l'échantillonnage [9], dans le but d'atténuer les problèmes spécifiques au capteur.

Étape E2: Ce processus vise à s'adapter à un modèle de visage générique au premier balayage 3D de l'objet.

Cette étape est nécessaire pour obtenir des modèles de visage 3D qui sont adaptés pour constituer la base de données de la galerie. On utilise le modèle annoté de visage (AFM) [78] qui définit les points de contrôle d'une surface de subdivision et est annoté dans différents domaines (par exemple, bouche, nez, yeux). Plus précisément, le modèle est d'abord globalement aligné sur le sujet, puis des déformations locales élastiques sont appliquées pour ajuster le modèle aux données. Le résultat de cette méthode lorsqu'il est appliqué sur des données 3D brutes est illustré dans la figure 2.12 (plus de détails peuvent être trouvés dans [78]).

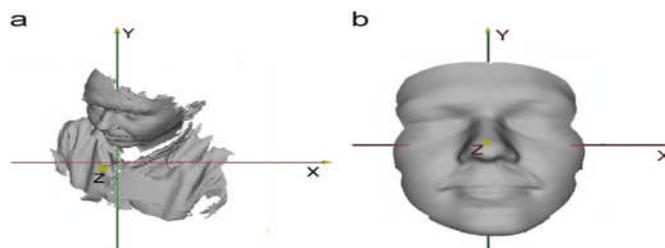


Figure 2.12 - Représentation du procédé de montage d'un modèle de visage générique pour la première 3D scan: (a) les données brutes, et (b) équipée AFM.

Étape E4: Les métadonnées sont stockées dans la base de données de la galerie, qui sera utilisée dans deux stades différents de la phase d'identification: (i) les métadonnées utilisées pour inclure l'appariement des informations disciplinaires et dépendent de la mesure de classificateur ou distance spécifique, (ii) les métadonnées utilisées pour l'extraction du profil reflètent des propriétés communes à tous les profils de la galerie actuelle - un modèle de forme statistique. Ces données comprennent une variante hiérarchique d'un modèle déformable linéaire [79].

Les auteurs **Enrique D et al** [80] ont travaillé en 2011 sur l'étude de reconnaissance des visages de mauvaise qualité images en utilisant Gabor Wavelet et le Support Vecteurs Machines (SVM). La reconnaissance faciale est un problème qui se pose à de nombreuses applications du monde réel, tels que ceux liés à l'intelligence ambiante (AMI). La nature spécifique et objectifs d'applications AMI nécessite de minimiser le caractère invasif des données. Les méthodes de collecte entraînent souvent une réduction drastique de la qualité des données et les effets d'imprévus peuvent mettre des systèmes de reconnaissance faciale standards non fiables. Pour faire face à cela, un système de reconnaissance faciale pour les applications AMI ne doit pas seulement être conçu avec soin, mais également faire l'objet d'un plan de configuration exhaustive et assurer une précision requise, la robustesse et la performance en temps réel. Dans

leurs travaux il s'agit d'une conception et mise au point d'un système de reconnaissance faciale holistique qui doit travailler sous des conditions de capture partiellement incontrôlée: images frontales avec variation de pose jusqu'à 40 degrés, le changement l'éclairage, taille d'image variable et une qualité dégradée par une grande variété d'images artefacts dus à la configuration lumière, le mouvement et la compression.

Le système proposé est basé sur Support Vector Machine (SVM classificateurs) et applique les filtres de Gabor intensive. Une analyse complète de sensibilité montre comment la précision de la reconnaissance peut être augmentée grâce à la configuration et paramétrage adéquat, bien que la plupart de réglage adéquat dépend des exigences pour le système final. Depuis la publication de l'architecture Dynamic Link [84], on a démontré que les filtres de Gabor sont un outil précieux pour l'identification faciale. Un grand nombre de recherches indiquent que cela augmente la robustesse des algorithmes de reconnaissance de forme beaucoup mieux que de nombreuses autres méthodes de transformation d'informations [81, 83, 86]. La décision d'utiliser un SVM comme classificateur est basée sur la performance, il a été démontré avec ce genre de problèmes dans plusieurs recherches [82, 85, 87]. Par ailleurs, il réduit la charge de travail énorme expérimentale. En dépit de sa simplicité, la combinaison Gabor + SVM peut surpasser certaines techniques de prétraitement modernes axées sur l'invariance de l'éclairage. Il est robuste contre nombreuses formes d'artefacts de bruit et problèmes d'images entrelacées présentées dans l'ensemble de données utilisé. Une fois l'architecture est sélectionnés, les paramètres de configuration seront intégralement analysés afin de savoir comment ils affectent le résultat et extrait le potentiel du système. Une décision importante est de savoir comment utiliser les informations disponibles du visage, que ce soit locales ou une approche holistique (l'image dans son ensemble). Plus concrètement, le système s'avère être valable dans une application basée sur la vidéo, le traitement des images de visage autour de 100×100 pixels dans des conditions un peu limitées (pose frontale avec jusqu'à 30-40 ° rotation, ce qui est assez fréquent dans de nombreux scénarios). Étant donné un taux de fausse acceptation de 1%, il peut atteindre une précision de reconnaissance autour de 70% avec des visages en présence d'expression et ayant de petits changements de poses sous un éclairage variable. Le taux d'erreur avec des personnes connues est inférieur à 0,5%. En ce qui concerne la mise au point du système, le sous-échantillonnage caractéristiques Gabor est un moyen efficace pour effectuer une réduction de dimension. En effet, grâce à la bonne apparence des éléments Gabor, elle conduit souvent à des résultats plus robustes dans tous les scénarios. La taille optimale pour les fonctions de Gabor extraits varie entre 16×16 et 32×32 . La banque de filtre de Gabor couvrant 7 fréquences se

comporte au moins aussi bien qu'une version avec 5 fréquences, mais peut surperformer largement lorsque le classificateur SVM s'approche de ses limites. Ce fait contredit la conclusion remarquée dans le travail pionnier de l'application de filtres de Gabor en reconnaissance de visage [84]. Étant donné que presque toutes les œuvres postérieures ont utilisé la même configuration, ce résultat suggère qu'une analyse plus approfondie de cette question peut être d'intérêt pour les chercheurs.

Les auteurs **Meltem Demirkus et al** [88] ont utilisés en **2011** les modèles spatiaux et probabilistes basés sur l'estimation de la pose des environnements sans contraintes. Dans ces environnements, la détection de la pose de la tête peut être très difficile en raison de la présence conjointe et arbitraire des expressions faciales, l'encombrement de fond, des occlusions partielles et les conditions d'éclairage. Malgré le large éventail de littérature de la pose de la tête, la plupart des méthodes actuelles ne peuvent répondre à ce problème que jusqu'à un certain degré, et surtout pour les scénarios restreints. Dans ces travaux, on aborde le problème de la classification de la pose de la tête à partir des images du monde réel avec une grande variation d'apparence. On représente chaque pose avec un modèle probabiliste et spatiale appris par mots de code du visage. L'inférence du meilleur modèle représentant une image de test est obtenue d'une façon probabiliste et spatiale par table de codage. Les résultats expérimentaux sont obtenus à partir de 5500 images vidéo recueillies sous différents éclairages et conditions de fond.

Dans cet article, on considère le problème particulier de déduire automatiquement la pose de la tête à partir d'images de visage à partir des données vidéo de surveillance recueillies dans des environnements non contraints. Développer un système automatique de classification de pose de la tête des données de vidéosurveillance n'est pas une tâche facile, étant donné les défis de conditions sans contraintes présentes dans le monde réel, qui comprennent: échelle arbitraire du visage, conditions d'éclairage non uniforme, occlusions partielles arbitraires et l'encombrement de fond ainsi que d'une grande variabilité dans une qualité d'image de visage (voir Figure 2.13).

Actuellement les méthodes d'estimation de la pose de la tête à partir d'images 2D peuvent être divisées en plusieurs groupes, à savoir les méthodes de modèle d'apparence [89,90], les méthodes d'incorporation du collecteur [91,92], méthodes de suivi [93], et les méthodes géométriques [94]. Cependant, la plupart des approches de la littérature ne sont pas construits pour les environnements non contraints.



Figure 2.13 - Exemples d'images (avec des échelles de visage originales conservées) issus de sa propre base de données sans contrainte du visage de l'environnement.

Un code de couleurs indique la tête assigné pose: - 90° (rouge), - 45°(vert), 0° (bleu), +45° (jaune), +90° (magenta)

Ainsi, la plupart des approches supposent que l'ensemble des traits du visage d'une vue frontale est visible. La plupart des approches sont formées et testées sur des images qui présentent des variations d'apparence comme les expressions faciales et l'éclairage.

Les bases de données testées contiennent principalement des images avec un fond fixe ou constant, l'expression faciale limitée, pas d'éclairage aléatoire, et avec peu ou pas d'occlusion du visage. Les résultats expérimentaux indiquent que l'approche dans [95] nécessite une bonne localisation de visage pour être en mesure d'éviter l'encombrement de fond et de diviser correctement une image de visage en patches non-recouvrement. Les images de visage sont représentées sous forme de dictionnaire, c'est à dire un ensemble de caractéristiques invariantes locales, à savoir les mots de code. La motivation derrière l'utilisation des invariants locaux est due à leur haut degré de robustesse aux différentes transformations, telles que les changements d'échelle, point de vue, rotation, translation, et occlusion. La méthodologie proposée "modèle probabiliste de répertoire spatial" pour chaque pose de tête est présentée sur la figure 2.14. Ces modèles sont inspirés des régions anatomiques du visage (par exemple : nez, bouche, oreille et des yeux) la distribution spatiale des régions anatomiques du visage est unique pour chaque pose de tête (Figure 2.14). Cette nouvelle approche diffère des méthodes actuelles sur plusieurs aspects. Les modèles actuels se composent généralement d'un ensemble d'images d'apprentissage (ou sur mesure (images déterministes)) avec pose d'étiquettes correspondant à laquelle une image d'essai est comparée à des techniques de comparaison basés sur l'image.

Ainsi, ils ne sont pas bien adaptés pour les images provenant d'environnements non contraints. Les modèles probabilistes proposés, d'autre part, sont des cartes de bibliothèque spatiales tirées de données d'entraînement où chaque mot de code contient les fonctions de densité de probabilité de la classe de pose de tête et l'étiquetage anatomique (voir Figure 2.14). À notre connaissance, l'approche proposée fournit la première adaptation de la représentation du dictionnaire, qui prévaut en matière de détection et de classification d'objet, le problème d'inférence de pose de la tête. De plus, contrairement à l'approche commune sac-de-fonctionnalités, cette approche étudiée attribue l'information spatiale et fonctionnelle (étiquetage anatomique) pour chaque mot de code. En outre, le dictionnaire basé sur la formulation Bayésienne proposée permet l'identification en présence d'occlusions partielles arbitraires contrairement à la plupart des approches disponibles dans la littérature actuelle. Sur une vaste base de données de 5500 sans contrainte images et vidéo, l'approche fournit un taux de précision élevé dans la classification de pose de la tête par rapport à l'état de l'art actuel [95].

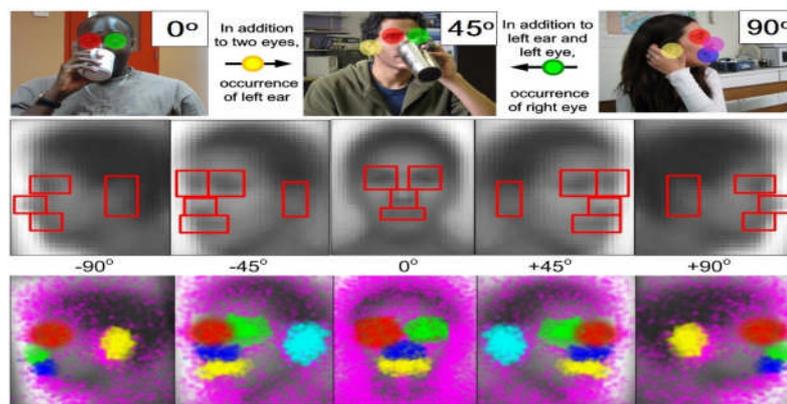


Figure 2.14 - L'utilisation de régions anatomiques pour la tête pose.

La première rangée: l'effet de pose de la tête (région anatomique) en présence d'occlusion.

La deuxième rangée: images de visage moyen de la base de données de formation (FERET) pour les cinq poses des classes et l'étiquetage manuel des régions anatomiques.

Les troisièmes rangées : modèles de pose de la tête où chaque couleur est affectée à une région anatomique différente.

On peut finalement dire que cette approche probabiliste permet de déduire la pose de la tête à partir d'images de visage 2D recueillies dans des environnements non contraints. Les résultats expérimentaux ont montré que ce siège basé sur un dictionnaire probabiliste et spatial de pose est une représentation qui surpasse nettement l'état de l'art actuel. Des études actuellement sont

menées de comment étendre cette formulation de classification. Une base de données de visage sans contrainte sera disponible pour une utilisation académique bientôt.

2.3 Conclusion :

Ce chapitre porte sur la mise au point détaillée d'un système de reconnaissance des visages SRV 2D et 3D et 2.5D (profondeur) et profil 2D avec les différentes techniques appliquées, qui sont décrit par les auteurs basée sur les approches par comparaison avec des prototypes, par classification, régression et alignement.

Nous avons remarqués que les méthodes par régression elles sont rapides et précises par rapport aux autres approches globales.

Pour obtenir de bons résultats il faut faire un bon choix sur les techniques utilisées, dont la pose est déduite à partir des positions relatives de certains éléments du visage tels que le nez la bouche, l'oreille et les yeux.

3.1 Introduction

Après avoir présenté les différentes méthodes de la reconnaissance faciale dans les deux chapitres précédents et en se basant sur les avantages et les limites de chacune d'elles. Nous nous orientons vers les méthodes globales qui préservent implicitement les informations locales sur le visage. Dans ce chapitre, nous allons donner une description globale sur les étapes essentielles de notre démarche, ainsi que l'explication de chaque méthode choisie.

- Nous choisissons d'extraire les informations pertinentes par le filtre de moyennage qui Sélectionne un bloc et choisisse la valeur moyenne des pixels
- La méthode d'analyse en composantes principales (PCA) pour garder les informations les plus importantes.
- La méthode d'analyse discriminante linéaire EFM pour discriminer les informations extraites et accélérer le processus d'apprentissage qui crée un modèle de classification (classificateur).
- La classification se fait à l'aide de la calcul des distances (la mesure de similarité)
- Finalement une Fusion par SVM (Support Vector Machine) supervisée est prévue. Pour cela, nous présentons deux modèles de fusion : Fusion (max, min).

3.2 Présentation de l'approche SRV3D étudiée

Le schéma de principe suivant illustre l'essentiel des étapes de notre système de reconnaissance de visage 3D.

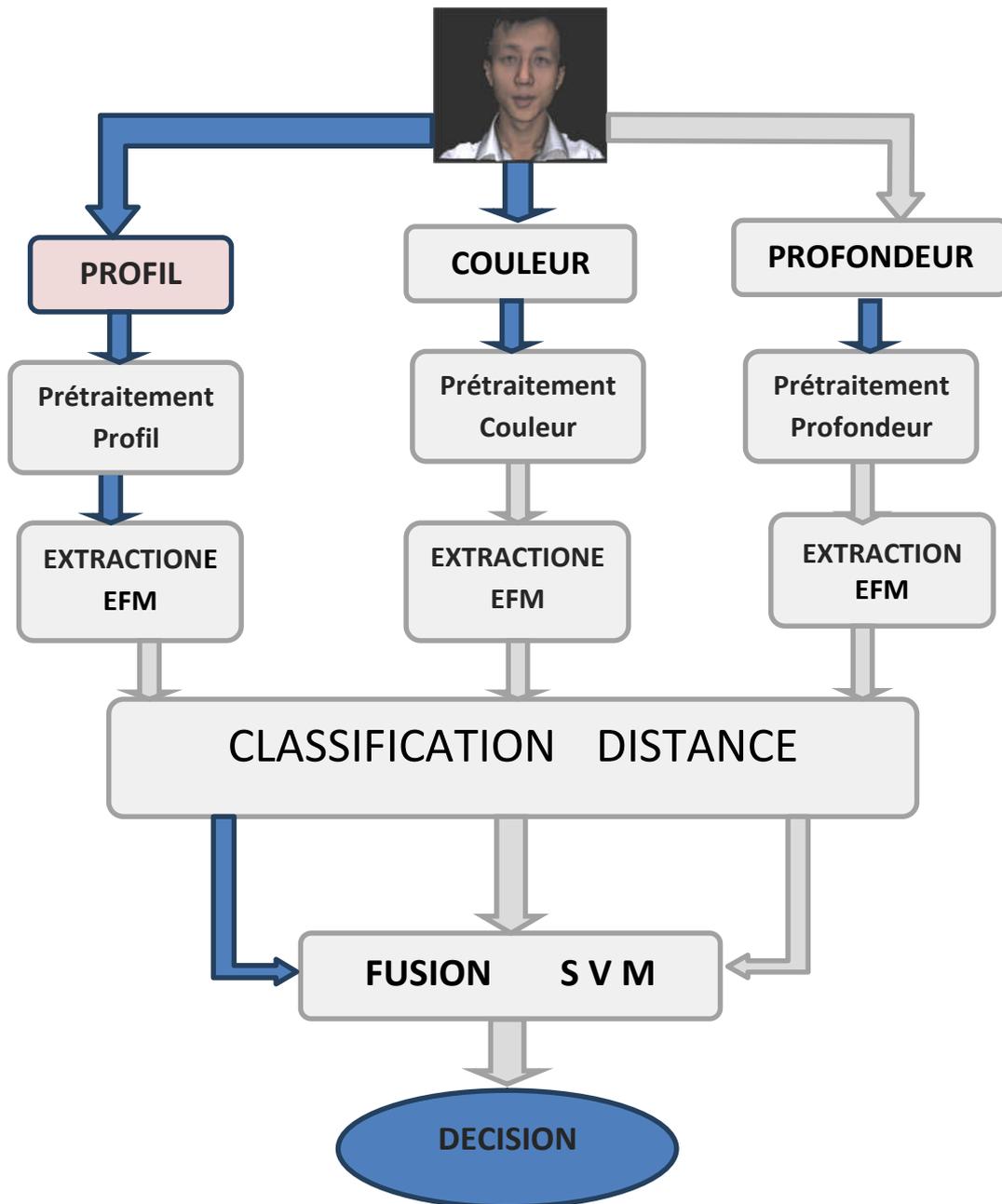


Figure 3.1 Schéma de principe de l'approche SRV3D

Dans tout système SRV3D la phase d'acquisition est primordiale pour cela nous présentons un bref aperçu sur ces capteurs. Malgré que notre travail se limite à l'utilisation de bases de données 3D déjà acquises. Dans ce qui suit des généralités et un exemple de système d'acquisition de nos images de visages étudiées.

3.3 Système d'acquisition 3D

Généralement le modèle du visage est représenté par des images 2.5D et 3D (voir figure 3.2). L'image 2.5D (image de profondeur) correspond à une représentation bidimensionnelle d'un ensemble de points 3D (x,y,z) où chaque pixel dans le plan (X,Y) stocke la valeur de sa profondeur Z . On peut assimiler une image 2.5D à une image en niveau de gris où les pixels noirs correspondent au fond tandis que les pixels blancs représentent les points de surface les plus proches de la caméra. Par ailleurs, la méthode la plus simple pour représenter un visage 3D est le maillage polygonal 3D, ce dernier correspond à une liste de points connectés par des arêtes (polygones). Il existe plusieurs techniques pour construire un maillage 3D, les plus utilisées combinent des images 2.5D ou bien exploitent des systèmes d'acquisition 3D tel que le scanner 3D.

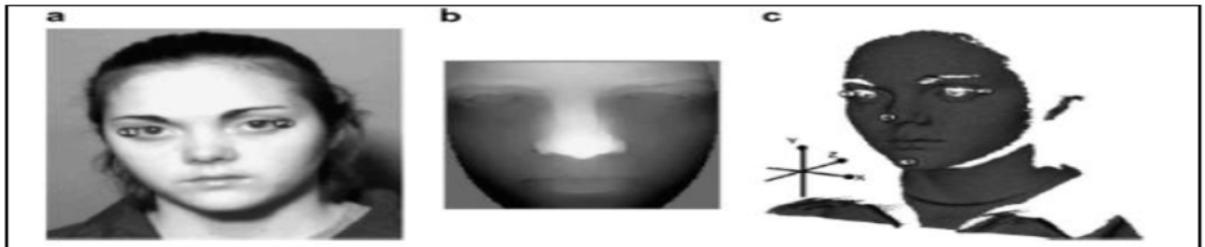


Figure 3.2: (a) Image texture (b) Image 2.5D (c) Image 3D [GUE 08].

De nombreux travaux de recherche visent le développement des méthodes d'acquisition 3D. Elles sont intensivement utilisées dans des domaines tel que : la rétro projection, le contrôle de qualité, la médecine, la réalité virtuelle, la robotique et l'archéologie. Ici, on restreint notre état de l'art aux méthodes dites optiques qui semblent être les plus adaptées à être utilisées dans notre contexte d'application qui est la numérisation 3D de visages. Cette catégorie d'approches comporte elle-même deux sous catégories qui sont : les méthodes **passives** et les méthodes **actives**. (La figure 3.3) présente cette catégorisation des approches d'acquisition de formes 3D [AMO 08].

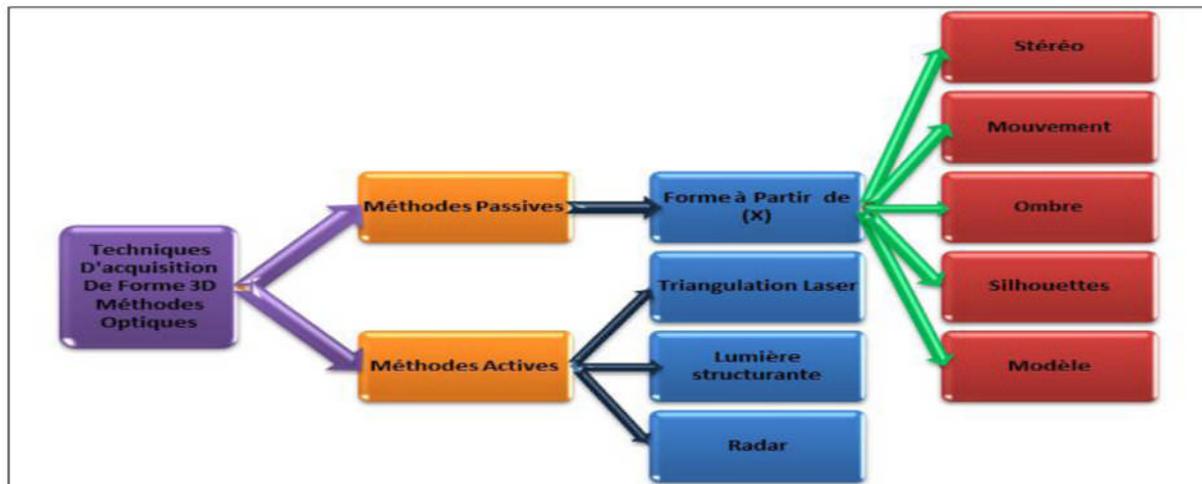


Figure 3.3 : Taxonomie des techniques optiques d'acquisition de formes 3D [AMO 08].

- La **vision active** consiste à combiner un capteur photographique avec une source lumineuse, afin de mesurer les coordonnées tridimensionnelles des points sur la surface d'un objet. Dans cette catégorie d'approches, deux types de sources lumineuses sont généralement employé : le laser et plusieurs types de lumières structurantes. Le laser, une lumière monochromatique, qui en passant à travers une lentille cylindrique, crée un plan lumineux éclairant une tranche de la scène. Ce qui permet de n'observer par une caméra que la partie de la scène ainsi éclairée. Connaissant l'équation du plan lumineux dans l'espace de la scène et ses coordonnées dans l'image, on peut alors calculer la position dans l'espace de la tranche de la scène captée par la caméra associée au dispositif laser. Ce principe est exploité dans plusieurs produits industriels, qui dotés d'un miroir rotatif, effectuent un balayage de l'objet placé devant. En revanche, les méthodes utilisant une lumière structurante, projettent sur la scène entière, un motif structurant qui est souvent une grille ou des bandes de lumière parallèles codées différemment. La géométrie de la scène est obtenue en analysant les déformations du motif lumineux ou en décodant la succession de lumières envoyées sur l'objet à numériser. [SAL 04], [ZHA 02], [GAR 01].

- La **vision passive**, contrairement à la vision active, n'utilise aucune structuration particulière de la scène. Les seules données disponibles sont, une ou plusieurs images. Parmi les différentes méthodes développées dans cette catégorie, la vision multi-images est la plus connue. Les approches multioculaires (plusieurs caméras) ou monoculaire (séquence vidéo) de cette catégorie reposent sur le principe de la triangulation optique pour localiser les objets. Cette géométrie multi-images permet, par l'intermédiaire de l'intersection de plusieurs rayons optiques, de retrouver l'information profondeur, ambiguë sur un seul rayon optique. Le défi

relevé par cette technique réside principalement dans la résolution du problème de la mise en correspondance dense ou la mise en correspondance de primitives dans les images. Dans la littérature de cette catégorie, un spectre très large de travaux a été développé pour traiter ce problème clé de la vision [AMO 08].

D'autres approches ont été proposées pour répondre au problème de la reconstruction tridimensionnelle de visage. En passant par un modèle générique ou par un modèle obtenu par ACP (Analyse en composante principale) comme modèle initial, certains algorithmes raffinent cette solution initiale, par une étape d'optimisation. Il s'agit de déformer le modèle de référence pour le rapprocher au mieux aux mesures réelles extraites par stéréovision ou par flot optique. Le résultat de ces algorithmes reste étroitement lié au modèle de référence pris comme solution initiale. Une autre technique, consiste à donner une approximation du visage, à partir de deux vues orthogonales : une vue frontale et une deuxième de profil. On trouve également, des approches moins récentes basées sur la texture où les variations globales de luminosité qui sont censées décrire les variations des normales à une surface observée. On extrait alors la forme générale des objets à partir des variations photométriques observées. Des techniques basées sur les silhouettes sont aussi présentes, et une combinaison avec des algorithmes multi-images, donne de bons résultats, néanmoins, ces techniques sont gourmandes en temps de calcul [AMO 08].

3.4 Prétraitement de l'image 3D

Cette phase est aussi très importante dans le SRV3D et demande une attention très importante. Nous détaillerons cela dans le chapitre 4 où chaque modalité est prétraitée séparément. Dans notre étude nous nous sommes basé sur trois modalités que nous jugeons comme informations pertinentes dans la reconnaissance de visage : le profil, la profondeur et la couleur. Nous jugeons utile de présenter l'utilité de ces modalités à travers certains travaux récents ce que nous présentons dans la section 3.5.

3.5 Descripteurs du visage 3D

3.5.1 Utilisation du profil dans la reconnaissance du visage 3D

La plupart de ces méthodes et produits sont conçus pour fonctionner avec les images de visages en frontales. Le profil de visage fournit des informations complémentaires de

structure en forme de ce qui n'est pas visible dans la vue frontale, mais peut souvent être obtenu à partir d'autres vues. Selon une étude de **Davidenko**, le profil silhouette des visages joue un rôle important dans la perception humaine de la tâche de reconnaissance de l'identité de genre.

L'utilisation du profil de visage en biométrie est particulièrement attrayante pour les scénarios où seules les images de vision latérale sont disponibles. (Figure 3.4) représente des images exemplaires où le profil silhouette contient des informations plus fiables que la texture, ce qui rend les informations de profil particulièrement précieuses pour la reconnaissance [EFR, 11].



Figure 3.4 Exemples de vue de côté l'image où la texture des couleurs fiables ne sont pas disponibles (de gauche à droite): image binaire, camouflage, rétro-éclairage, et dans l'imagerie infrarouge.

3.5.2 Utilisation de la couleur et profondeur de visages pour la reconnaissance faciale

Une technique de reconnaissance faciale est développée sur la base de la profondeur et l'information de couleur. L'objectif principal est l'évaluation de trois approches différentes (couleur, profondeur, combinaison de la couleur et la profondeur) pour une reconnaissance faciale et quantifier la contribution de la profondeur dans la reconnaissance du visage. La technique de reconnaissance faciale proposée est basée sur le composant principal algorithme d'analyse et l'extraction de la profondeur et visages propres couleur. Les résultats expérimentaux montrent de significatifs gains obtenus avec l'addition de l'information profondeur.

La méthode visages propres pour une utilisation dans la reconnaissance des visages sur la base de la couleur et l'information de profondeur est développée dans la (figure 3.5). La technique de reconnaissance est présentée basée sur la mise en œuvre de l'algorithme d'analyse en composantes principales et l'extraction de la couleur et visages propres profondeur. Le principal objectif était d'évaluer trois différentes approches couleur, la profondeur, la combinaison de la couleur et la profondeur pour la reconnaissance faciale et de quantifier la contribution de la profondeur dans ce domaine. Les résultats expérimentaux

montrent des gains significatifs avec l'utilisation des informations de profondeur. Le procédé d'amélioration choisie varie et est compris entre 2% et 5%.

La technique basée sur la profondeur seule semble être robuste aux variations dues au bruit, au changement de pose et les rotations de la tête humaine. L'algorithme sur la base de l'utilisation de la couleur et de luminance ainsi que la profondeur a été vu pour mener à la reconnaissance un taux très élevé [TSA 02] .

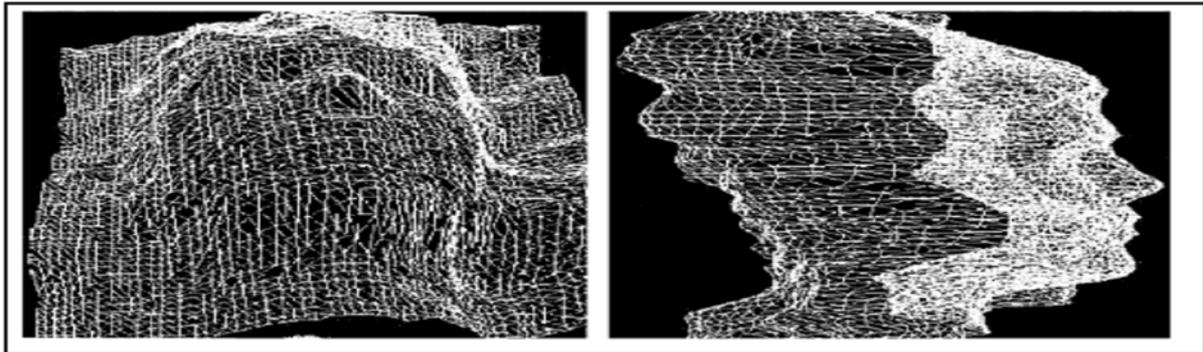


Figure 3.5: Modèle 3D de visage humain constitué de triangles

3.6 Extraction des caractéristiques

Les techniques d'extractions des caractéristiques permettent d'éliminer l'information redondante et de ne garder que l'information pertinente pour but de recueillir des informations biométriques sur les personnes à authentifier. Nous avons combiné trois méthodes : Les ondelettes de Gabor, PCA et LDA afin d'atteindre ce but.

L'algorithme LDA est né des travaux de Belhumeur et al. De la Yale Université (USA), en 1997. Il est aussi connu sous le nom de Fisher faces. Contrairement à l'algorithme PCA, l'algorithme LDA effectue une véritable séparation de classes (figure 3.6).

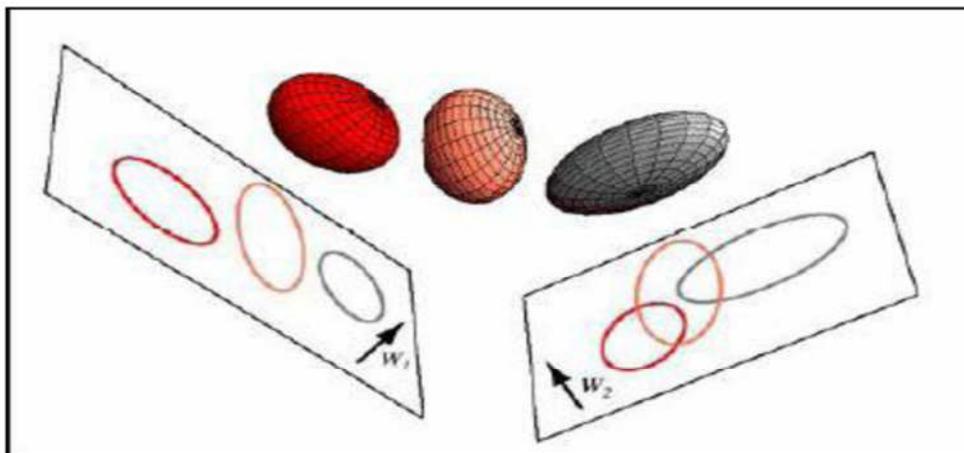


Figure 3.6: Illustration du principe de séparation optimale des classes par le LDA.

Trois distributions 3D sont projetées sur deux sous-espaces 2D décrits par les vecteurs W_1 et W_2 . Puisque le LDA essaye de trouver la plus grande séparation parmi les classes, on voit bien que W_1 est ici le vecteur optimal.

Tout comme dans le PCA, on rassemble les images de la base d'apprentissage dans une grande matrice d'images T où chaque colonne représente une image Γ_i , puis on calcule l'image moyenne.

Ensuite, pour chaque classe C_i , on calcule l'image moyenne Ψ_{C_i} :

$$\Psi_{C_i} = \frac{1}{q_i} \sum_{k=1}^{q_i} \Gamma_k \quad (3.1)$$

Avec q_i , le nombre d'images dans la classe C_i .

Chaque image Γ_i de chaque classe C_i est ensuite recentrée par rapport à la moyenne Ψ_{C_i} . On obtient alors une nouvelle image Φ_i :

$$\Phi_i = \Gamma_i - \Psi_{C_i} \quad (3.2)$$

Vient ensuite le calcul de nos différentes matrices de dispersion. On notera c le nombre total de classes (i.e. le nombre d'individus), q_i le nombre d'images dans la classe C_i et M le nombre total d'images.

La Matrice de Dispersion Intra-Classe (S_w) :

$$S_w = \sum_{i=1}^c \sum_{\Gamma_k \in C_i} (\Gamma_k - \Psi_{C_i})(\Gamma_k - \Psi_{C_i})^T \quad (3.3)$$

La Matrice de Dispersion Inter-classe (S_b) :

$$S_b = \sum_{i=1}^c q_i (\Psi_{C_i} - \Psi)(\Psi_{C_i} - \Psi)^T \quad (3.4)$$

La Matrice de Dispersion Totale (S_T):

$$S_T = \sum_{i=1}^M q_i (\Gamma_i - \Psi)(\Gamma_i - \Psi)^T \quad (3.5)$$

Une fois ces matrices calculées, nous devons trouver une projection optimale W qui maximise la dispersion intra-classe, relative à la matrice S_w , tout en minimisant la dispersion interclasse, relative à la matrice S_b .

- **Le modèle discriminant linéaire amélioré de Fisher (Enhanced Fisher linear discriminant Model (EFM))**

Ce modèle discriminant linéaire de **Fisher** améliore la capacité de généralisation de la **FLD** en décomposant la procédure **FLD** en diagonalisation simultanée des deux matrices de dispersion intra-classe et interclasse. La diagonalisation simultanée est une étape sagement

équivalente à deux opérations comme l'a souligné Fukunaga. Blanchiment de la matrice de dispersion intra-classe et l'application de la **ACP** sur la matrice de dispersion intra-classe en utilisant les données transformées. Durant l'opération du blanchiment de la matrice de dispersion intra-classe apparaisse dans le dénominateur de la séparabilité des petites valeurs propres qui tendent à capturer du bruit. Pour atteindre des performances améliorées l'**EFM** préserve un équilibre approprié entre la sélection des valeurs propres (correspondant à la composante principale de l'espace de l'image originale) qui tiennent compte de la plupart de l'énergie spectrale des données brutes, c'est à dire, représentation adéquate et l'exigence que les valeurs propres de la matrice de dispersion intra-classe (de l'espace ACP réduit) ne sont pas trop petites, c'est à dire, meilleure généralisation. En particulier, ces étapes **FLD** permettent de trouver les valeurs propres et les vecteurs propres de $S_W - 1S_b$ comme résultat de la diagonalisation simultanée de S_W et S_b . Alors l'étape de EFM sont présentées comme suit :

- Blanchissons d'abord la matrice de dispersion intra-classe :

$$S_W V = V \Lambda \text{ et } V^T V = I \quad (3.5)$$

$$\Lambda^{-1/2} V^T S_W V \Lambda^{-1/2} = I \quad (3.6)$$

Où $V, \Lambda \in \mathbf{R}^{m \times m}$ sont la matrice des vecteurs propres et la matrice diagonale des valeurs propres de S_W respectivement.

Les valeurs propres de la matrice de dispersion intra-classe dans l'espace ACP réduit peut être obtenu en (équation 3.5). Donc, EFM diagonalise en premier lieu la matrice de dispersion intra-classe S_W (3.6) et (3.6). Notez que V et Λ sont les matrices des vecteurs propres et des valeurs propres correspondants aux vecteurs caractéristiques.

- En second lieu **EFM** procède à calculer la nouvelle matrice de dispersion inter-classe comme suit:

$$\Lambda^{-1/2} V^T S_b V \Lambda^{-1/2} = K_b \quad (3.7)$$

- Diagonalisons maintenant la nouvelle matrice de dispersion inter-classe K_b :

$$K_b V_b = V_b \Lambda_b \text{ et } V_b^T V_b = I \quad (3.8)$$

Où $V_b, \Lambda_b \in \mathbf{R}^{m \times m}$ sont la matrice des vecteurs propres et la matrice diagonale des valeurs propres de K_b respectivement.

- La matrice de transformation globale de l'**EFM** est définie comme suit:

$$U = V \Lambda^{-1/2} V_b \quad (3.9)$$

3.7 La fusion des scores

Une caractéristique unique fournit de l'information discriminante limitée et une applicabilité locale. La reconnaissance du visage basée sur une seule caractéristique visage ne peut pas atteindre des performances satisfaisantes. Donc, il est une approche raisonnable à fusionner plusieurs types de traits du visage ou bien plusieurs modalités ou algorithmes. La stratégie de fusions multi-caractéristiques est capable de synthétiser toutes les caractéristiques informations, données sur les réserves visage plus originaux, et d'obtenir des résultats satisfaisants. Les stratégies de fusion populaires sont la série conjointe de caractéristiques et le joint pondéré et ainsi de suite. La stratégie de fusion est vouée à l'ancienne face à haute vecteurs de caractéristiques dimensionnelles. Si la réduction de dimension est mise en oeuvre, les vraies significations du nouveau vecteur sont ambigus. La stratégie de fusion est linéaire pondérée stratégie basée sur les résultats de la reconnaissance de visage de différentes caractéristiques.

L'idée principale est basée sur la théorie suivante. La similitude entre les différents échantillons de la même personne (ou les échantillons de mêmes classe) est supérieure à la similarité entre des échantillons différents de différents individus (ou des échantillons de classes différentes) [SUN 08]. On peut réaliser une fusion par combinaison en utilisant des opérateurs simples ou la logique floue, comme on peut le faire par une fusion par classification par des classifieurs telles Fisher, réseaux de neurones ou SVM. En se basant sur les travaux de M.Belahcène et A.Ouamane [Euvip 11] sur la même base de données Casia3D qui affirment que le SVM donne les meilleurs résultats de performance systèmes. Nous retenons donc cette méthode de fusion que nous présentons dans ce qui suit.

- **Machines à Vecteurs de Support (SVM)**

Le SVM est une approche non supervisée basée sur la théorie de l'apprentissage statistique. Il estime que le optimale frontière dans l'espace des caractéristiques en combinant une stratégie de marge maximale avec une méthode du noyau, ce processus est appelé une machine noyau.

La machine est formée en fonction de la minimisation du risque structurel (SRM), le critère [BUR 98], [VAP 98] Les limites de décision sont directement dérivées à partir des données de formation fixées par l'apprentissage. Les cartes SVM les entrées dans un espace de caractéristiques de grande dimension par le biais d'une fonction noyau sélectionné. Puis, il construit une séparation optimale hyper-plan dans l'espace caractéristique. La dimensionnalité

de la fonction espace est déterminée par le nombre de vecteurs de support extraites des données de formation voir la (figure 3.7).

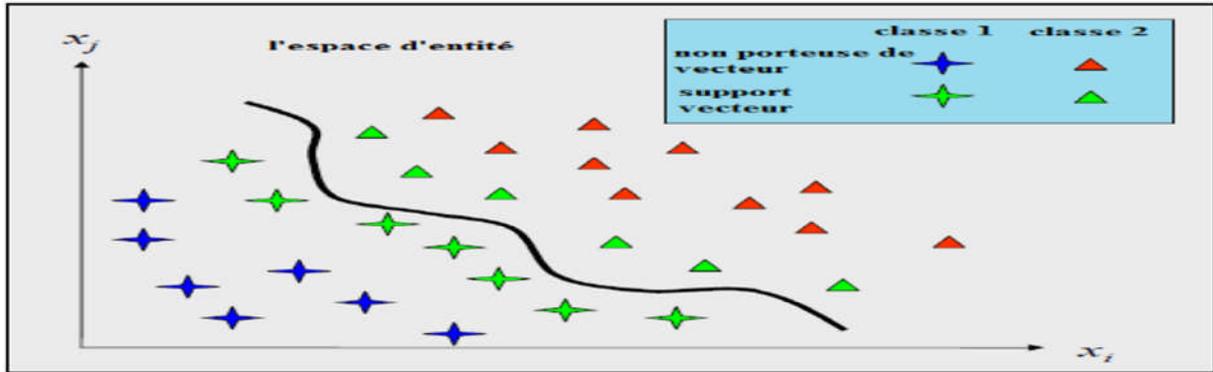


Figure 3.7 Limite optimale recherchée par la SVM.

Le principe des SVM repose sur la recherche d'un séparateur linéaire dans un espace de grande dimension qu'on construit parce que les données d'entrée ne sont pas linéairement séparables dans l'espace d'origine. Nous maximisons alors la marge, grandeur mesurant l'écart du modèle aux données, ce qui assure alors en principe de bonnes performances en généralisation.

Soit $X_l = (x_i)_{i=1}^l \subset I$ l'ensemble des données étiquetées suivant $Y = (y_i)_{i=1}^l \in \{-1, +1\}^l$

Qui représente la classe de chaque individu, et $\phi: I \rightarrow F$ la fonction qui envoie les données d'entrée dans l'espace F des caractéristiques. Comme défini sur la (figure 3.8), la distance entre l'hyperplan $H_{w,b} = \{X \in F \mid \langle w, X \rangle + b = 0\}$ et X est appelée la marge Δ . Avec le principe de Minimisation du Risque Structurel (SRM), Vapnik [VAP 99], a montré que maximiser la marge (ou minimiser $\|w\|$) permet de minimiser la VC-dimension du modèle de séparation, qui est un critère de généralisation efficace.

On définit alors le noyau K comme $K(X, Y) = \langle \phi(X), \phi(Y) \rangle_F$. Cette fonction est en fait un raccourci mathématique qui nous évite de manipuler directement.

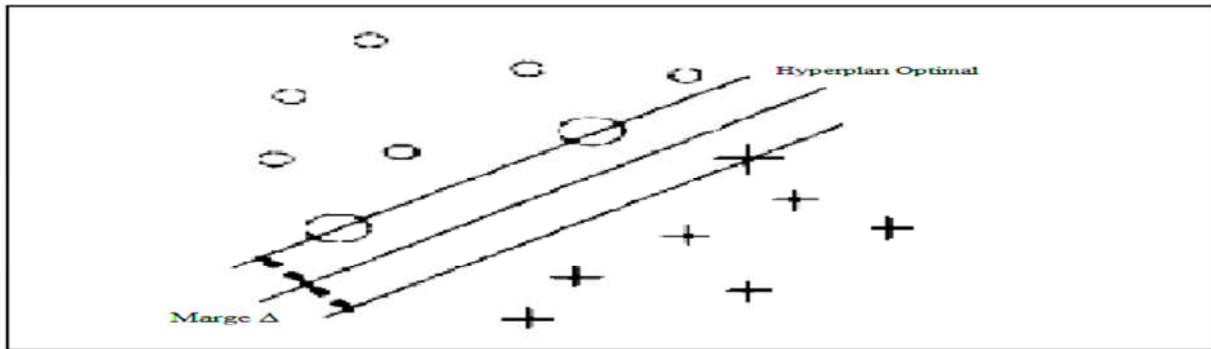


Figure 3.8 : L'hyperplan optimal H et la marge

- **SVM utilisé dans une application SRV**

L'image support de l'apprentissage est extraite du flot vidéo par la méthode, à base de filtres de Gabor. Nous sélectionnons les images de face en écartant les images de profil, par un procédé de reconnaissance de formes de type SVM que nous détaillerons dans ce chapitre. La (figure 3.9) montre un exemple d'images extraites :



Figure 3.9: Images caractéristiques de visages

3.8 Conclusion

Dans ce chapitre, nous avons décrit les démarches pour réaliser notre application. Nous avons commencé par sa description globale. Ensuite, nous avons mis en relief la description des méthodes utilisées à la reconnaissance des visages 3D, l'extraction des caractéristiques et la classification. Le chapitre suivant sera consacré essentiellement à l'implémentation de ces méthodes et le test de la performance de notre application.

4.1 Introduction

Jusqu'ici nous avons approché la reconnaissance automatique des visages d'un point de vue théorique. Comme l'authentification est un problème pratique, il est maintenant d'examiner les différentes alternatives de solutions et d'essayer de tirer quelques conclusions. Les résultats expérimentaux que nous allons présenter sont divisés en deux parties. Nous donnerons tout d'abord les résultats des comparaisons des systèmes monomodaux d'authentification de visage 3D, visage 2D et profile 2D. Nous présentons ensuite la fusion de scores par SVM des quatre combinaisons des systèmes (fusion 3D visage et 2D visage, fusion 3D visage et 2D profile, fusion 2D visage et 2D profile, fusion 3d et 2D visage et 2D profile).

4.2 La base de données CASIA 3D

Nous utilisons la base de données des visages CASIA 3D pour tester notre système d'authentification proposé [96].

La base est construite par un numériseur 3D sans contact Minolta VIVID 910, travaille sur le mode rapide. Cette base de données contient 123 sujets, chaque sujet ayant 37 ou 38 images avec les variations individuelles de poses, les expressions et les illuminations, les variations combinées des expressions sous illumination et pose sous des expressions.

Cette base de données contient des variations complexes qui sont difficiles à n'importe quel algorithme.

Dans notre travaille nous avons étudiées les variations d'illuminations (Figure.4.1), expressions (Figure.4.2) et les variations combinées des expressions sous illumination, donc nous avons utilisé 10 image par chaque sujet. La base de données de 1230 images est divisée en deux sous-ensembles, l'ensemble d'apprentissage et l'ensemble de test. Nous avons utilisé un protocole d'évaluation qui consiste à séparer les individus en deux classes, client et imposteur. Le groupe client contient 100 sujets, alors que le groupe imposteur contient 23 sujets. La répartition des photos dans les différents ensembles sont présentée dans le tableau 4.1 pour les images 3D et 2D et le tableau 4.2 pour les images de profile 2D.



Figure 4.1- Illumination variations of the CASIA 3D Face Database

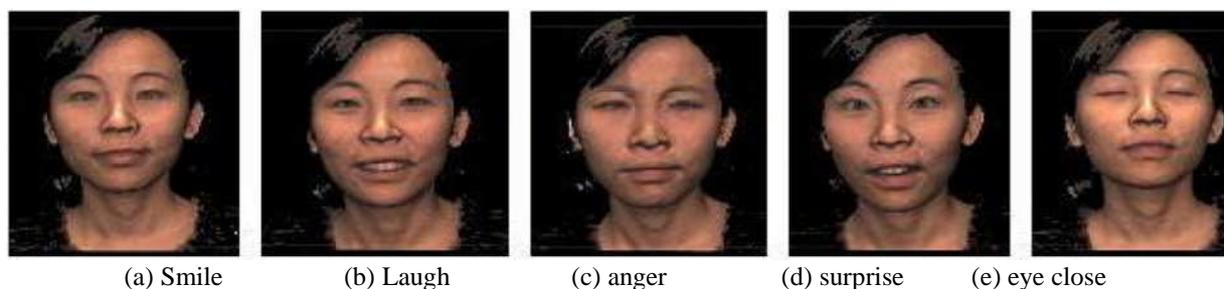


Figure 4.2 - Expression variations of the CASIA 3D Face Database

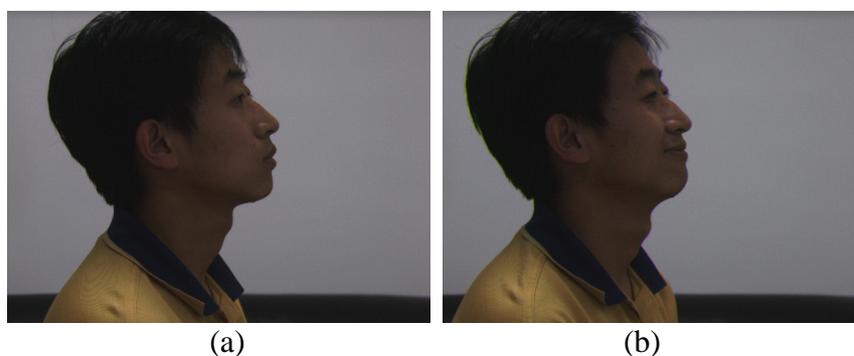


Figure 4.3 - Exemple d'image de profil 2D, (a) image 19, (b) image 30.

Ensemble	Clients	Imposteurs
Apprentissage	500 images (1, 4, 8, 9, 10)	0 images
Test	500 images (2, 6, 7, 14, 15)	230 images (1, 2, 4, 6, 7, 8, 9, 10, 14, 15)

Tableau 4.1- Répartition des photos dans les différents ensembles pour les images 3D et 2D de visage

Ensemble	Clients	Imposteurs
Apprentissage	100 images (19)	0 images
Test	100 images (30)	26 images (19, 30)

Tableau 4.2 - Répartition des photos dans les différents ensembles pour les images 2D de profile

4.3 Mesure de la performance des systèmes d'authentification biométriques

Tout d'abord, afin de comprendre comment déterminer la performance d'un système biométrique, il nous faut définir clairement **trois critères principaux** :

- Le premier critère s'appelle le **taux de faux rejet** ou **TFR** ("False Reject Rate" ou **FRR**). Ce taux représente le pourcentage de personnes censées être reconnues mais qui sont rejetées par le système,
- Le deuxième critère est le **taux de fausse acceptation** ou **TFA** ("False Accept Rate" ou **FAR**). Ce taux représente le pourcentage de personnes censées ne pas être reconnues mais qui sont tout de même acceptées par le système,
- Le troisième critère est connu sous le nom de **taux d'égale erreur** ou **TEE** ("Equal Error Rate" ou **EER**). Ce taux est calculé à partir des deux premiers critères et constitue un point de mesure de performance courant. Ce point correspond à l'endroit où **TFR = TFA**, c'est-à-dire le meilleur compromis entre les faux rejets et les fausses acceptations.

La **figure 4.4** illustre le **TFR** et le **TFA** à partir de distributions des scores clients et imposteurs tandis que le **TEE** est représenté sur la **figure 4.5**

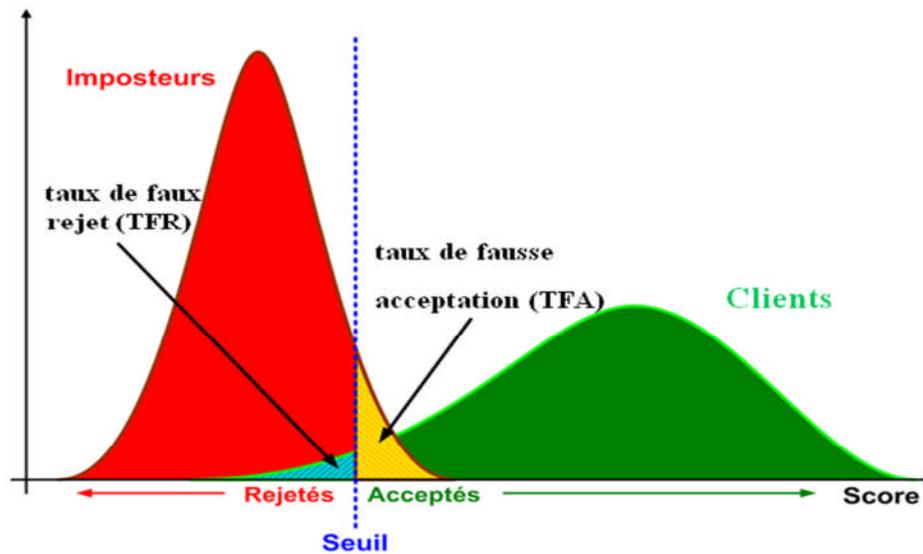


Figure 4.4 : Illustration du TFR et du TFA.

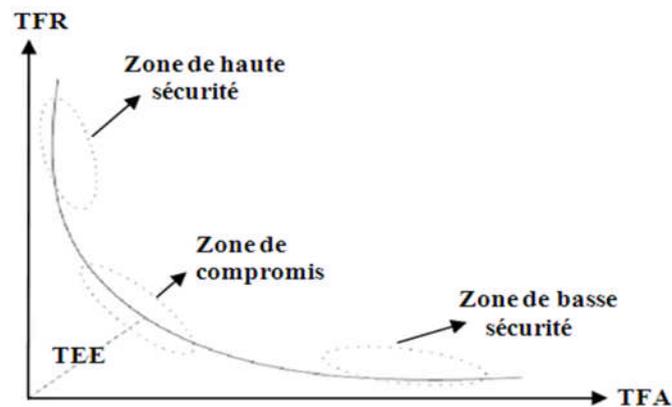


Figure 4.5 : Courbe caractéristique de fonctionnement (ROC) d'un système de Vérification Biométrique.

La performance globale d'un système de vérification d'identité est mieux caractérisée par sa courbe caractéristique de fonctionnement (**Receiver Operating Characteristic** ou **ROC**), La courbe ROC (**figure 4.5**) trace le taux de faux rejet en fonction du taux de fausse acceptation [97]. Plus cette courbe tend à épouser la forme du repère, plus le système est performant, c'est-à-dire possédant un taux de reconnaissance global élevé.

Le seuil de décision doit donc être ajusté en fonction de l'application ciblée : haute sécurité, basse sécurité ou compromis entre les deux.

4.4 Architecture structurelle du système de conception

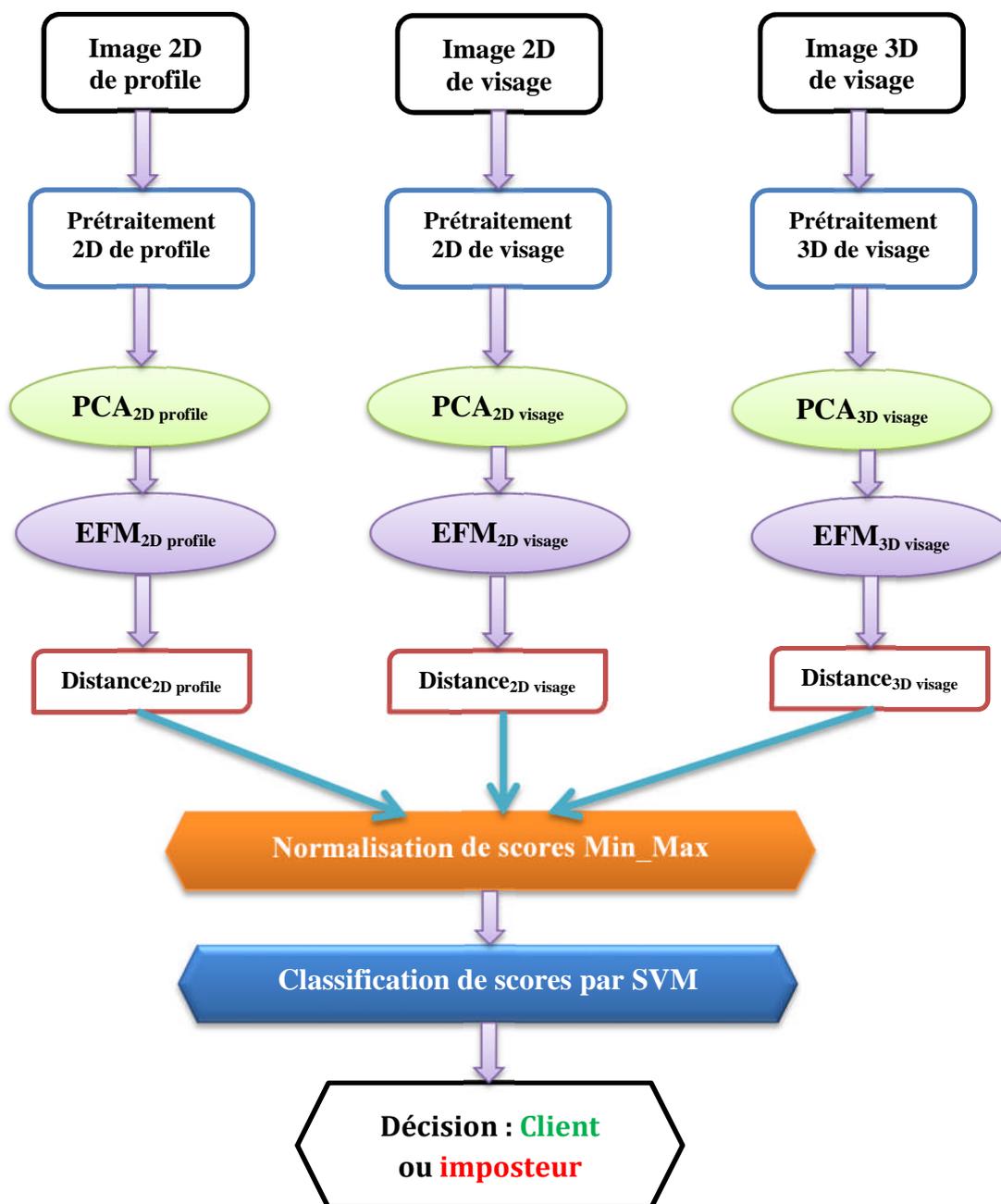


Figure 4.6 : Schéma de l'architecture structurelle du système de conception.

4.4.1 Prétraitement d'image

Le prétraitement est une phase importante dans le processus globale d'identification, C'est une méthode simple qui augmente en général les performances du système.

Prétraitement d'image de visage 3D et 2D [3]

Il est supposé dans ce travail que l'une des faces est décrite par un nuage de points 3D capturés par un scanner laser 3D. Chaque nuage de points constitué de milliers de points dans l'espace 3D. Ces points discrets décrivent la surface du visage. Dans notre base de données de visage 3D CASIA chaque point est décrit avec des coordonnées spatiales 3D et les coordonnées couleur RGB correspondants. Dans cette partie, nous décrivons comment les données originales en 3D sont prétraitées.

Les données 3D est convertie en une image en profondeur (voir Figure. 4.7 (a)) et une image de couleur (voir Figure.4.8(a)). Dans la plupart des images, le nez est la plus proche partie du visage au scanner 3D, c'est à dire, il a la valeur la plus élevée en profondeur entre tous les points du visage. En utilisant une fenêtre de taille 3x3 qui calcule la somme des valeurs de profondeur de ses pixels correspondants, le nez est détecté comme le point de coordonnées du pixel central de la fenêtre qui renvoie la valeur maximale (voir Figure.4.7 (b)). Après avoir détecté le nez, toutes les images dans la base de données sont découpé par une fenêtre rectangulaire de taille 57x47 centrée autour le centre de nez (voir Figure.4.7 (c), Figure.4.8 (b)). Pour l'image de couleur RGB nous avons utilisée l'espace de couleur HSV (Hue, Saturation, Value) (voir Figure.4.8 (c)) et nous considérons la composante V " Value " parce que est moins sensible à la variation d'éclairage (voir Figure.4.8 (d)).

Toutefois, en raison de la qualité de données originales en 3D, les images de profondeur et couleur que nous utilisons contiennent généralement beaucoup de bruit, tels que les trous et les valeurs aberrantes. Nous pouvons obtenir des images claires par les procédés suivants. Le prétraitement des images de profondeur et l'image de la composante V de l'espace de couleur HSV comprend la suppression du bruit et le trou de remplissage. Pour chaque pixel, la moyenne est calculée pour sous-fenêtre de taille 5×5 autour de lui. Le résultat est montré dans les figures.4.7 (d), 4.8 (e).

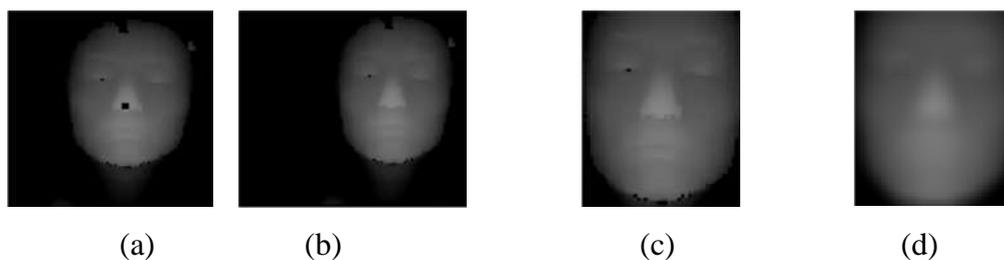


Figure 4.7- Prétraitement d'image de profondeur: (a) l'image de profondeur; (b) détection de bout de nez ; (c) Image découpée ; (d) après avoir enlevé le bruit et le remplissage de trous.

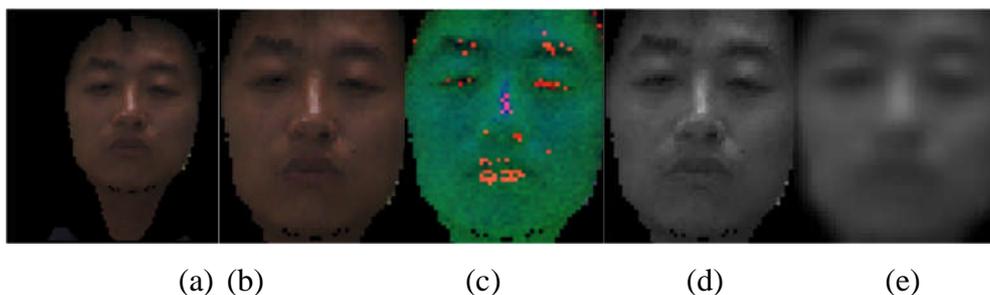


Figure 4.8 - Prétraitement d'image de couleur: (a) l'image de couleur; (b) Image découpée; (c) Image en système HSV; (d) La composante V de HSV ; (e) après avoir enlevé le bruit et le remplissage de trous.

Prétraitement d'image de profile 2D

Nous avons utilisé deux étapes la convention de l'image couleur en niveaux de gris (figure 4.9 (b)) et la décimation qui consiste à ne prendre qu'un pixel sur quatre. Cela réduit bien entendu la résolution des images (figure 4.9 (c)).

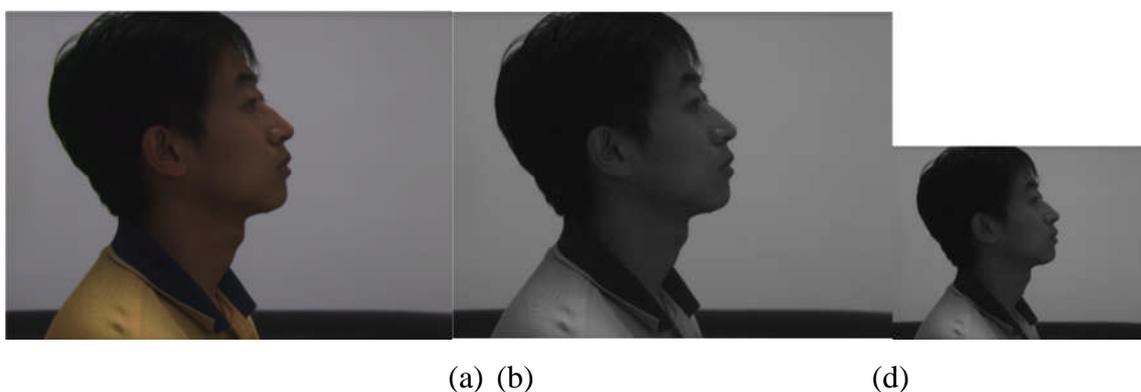


Figure 4.9 - Prétraitement d'image de profile 2D : (a) image de profile, (b) image en niveaux de gris, (c) image de profile après décimation.

4.4.2 L'analyse en composantes principales(PCA)

Après l'étape de prétraitement des visages 2D et 3D et profile 2D nous tentons de réduire par un petit programme en Matlab calculant la **PCA (Analyse en Composante Principale)**. Les étapes de cet algorithme sont les suivants :

- 1) *Prétraitement.*
- 2) *Création de la matrice des données X (L'ordonnement des vecteurs images des clients coté à coté).*
- 3) *Centrer la matrice des données \bar{X}*
- 4) *Calcul de la matrice de covariance des données Σ .*
- 5) *Calcul des valeurs et vecteurs propres de la matrice de covariance.*
- 6) *La matrice de transformation de la PCA est les m premiers vecteurs propres (U_{PCA}).*

4.4.3 Le Modèle Discriminant Linéaire Amélioré de Fisher (EFM)

- 1) *Prétraitement.*
- 2) *Calcul la matrice de dispersion intra-classes.*
- 3) *Calcul de la matrice de dispersion inter-classes S_B .*
- 4) *Calcul des valeurs (Λ) et vecteurs (V) propres de la matrice S_w .*
- 5) *Calculer la nouvelle matrice de dispersion inter-classe $K_b = \Lambda^{-1/2} V^T S_b V \Lambda^{-1/2}$*
- 6) *Calcul des valeurs (Λ_b) et vecteurs (V_b) propres de la matrice K_b .*
- 7) *Calcul de la matrice $U = V \Lambda^{-1/2} V_b$.*
- 8) *La matrice de transformation de l'EFM est les m premiers vecteurs de U (U_{EFM}).*

Enfin, nous réalisons notre base de données indexée dans la phase enrôlement selon la figure 4.10 .

4.4.4 Distance

Nous utilisons pour la comparaison la distance de **corrélation normalisée** (normalized correlation) [99] donnée par l'équation 4.1 pour comparer deux vecteurs caractéristiques réduits.

$$S(A, B) = \frac{A^T B}{\|A\| \|B\|} \quad (4.1)$$

Cette fonction calcule simplement le cosinus de l'angle entre les deux vecteurs caractéristiques A et B. Une valeur élevée de corrélation normalisée correspond à une bonne similarité entre les deux vecteurs.

4.4.5 Normalisation par la méthode Min-Max

Soit s_{ij} le $j^{\text{ème}}$ score de sortie du $i^{\text{ème}}$ système, où $i = 1, 2, \dots, N$ et $j = 1, 2, \dots, M$ (N est le nombre de systèmes et M le nombre de scores disponibles dans l'ensemble de données d'entraînement). Le score normalisé Min-Max pour le score de test s_{ik} est donné par:

$$= \frac{s_{ik} - \min(\{s_{ij}\})}{\max(\{s_{ij}\}) - \min(\{s_{ij}\})} \tag{4.2}$$

Où $\{s_i\} = \{s_{i1}, s_{i2}, \dots, s_{iM}\}$. Quand les valeurs minimales et maximales sont estimées à partir du jeu d'entraînement de scores donné.

4.4.6 Classification des scores par SVM

Les méthodes de fusion de scores combinent les informations au niveau des scores issus des modules de comparaison comme indiqué sur la **figure 4.10**. Un système de fusion est constitué de deux modules, un module de **fusion** et un module de **décision**. Le problème devient donc un problème de classification à 2 classes (**OUI** ou **NON**, Client ou Imposteur) à partir d'un vecteur de nombre réels dont la dimension est égale au nombre des sous-systèmes locale et globale de classification des image de profondeur et d'intenté du visage.

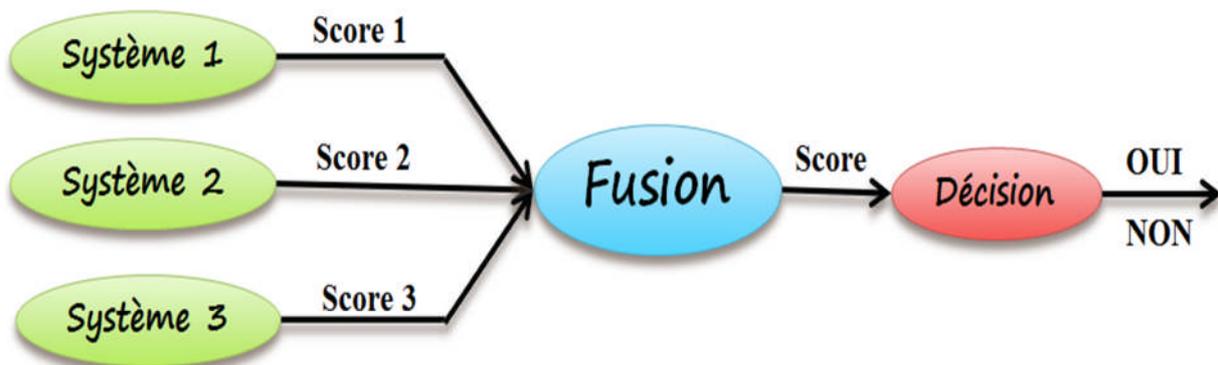


Figure 4.10 : Schéma de la fusion de scores.

Il existe deux approches pour combiner les scores obtenus par différents systèmes :

- i) La première approche est de traiter le sujet comme un problème de **combinaison**, les scores individuels sont combinés de manière à former un unique score.
- ii) tandis que l'autre approche est de voir cela comme un problème de **classification**, un vecteur de caractéristiques est construit en utilisant les scores donnés en sortie par les systèmes individuels ; ce vecteur est ensuite attribué à une des deux classes.

Dans [98], [100], [101], [102] les auteurs ont étudié les deux approches de fusion de score et plusieurs méthodes de normalisation de scores, mais dans ce travail nous avons utilisé la normalisation de score la Min-Max pour sa simplification et pour la fusion une méthode de classification appelée les machines à vecteurs de support SVM c'est une méthode très connue dans le domaine de classification à deux classes.

4.5 Résultats obtenus

4.5.1 Résultats obtenus pour l'authentification monomodaux

La figure.4.11 et le tableau. 4.3 présente la comparaison de TEE des trois systèmes d'authentification de visage 3D, visage 2D et profil 2D.

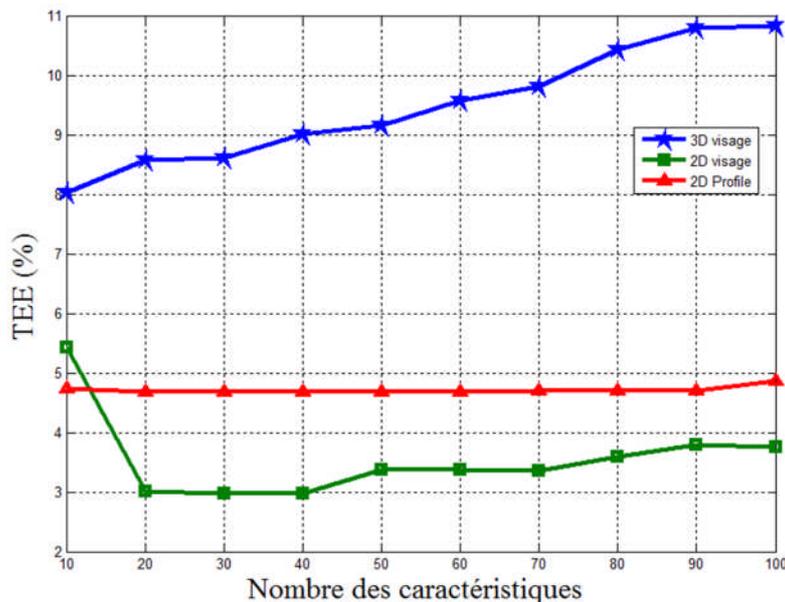


Figure. 4.11 Comparaison de TEE en fonction de nombre des caractéristiques pour les trois systèmes d'authentification de visage 3D, visage 2D et profil 2D.

Nombre des caractéristiques	TEE (%) 3D visage	TEE (%) 2D visage	TEE (%) 2D profile
10	8.03	5.42	4.73
20	8.56	3.01	4.67
30	8.6	2.97	4.67
40	9.01	2.97	4.68
50	9.15	3.36	6.67
60	9.57	3.37	4.68
70	9.81	3.36	4.69
80	10.43	3.58	6.69
90	10.79	3.78	4.69
100	10.83	3.75	4.85

Tableau 4.3 - Comparaison de TEE en fonction de nombre des caractéristiques pour les trois systèmes d'authentification de visage 3D, visage 2D et profile 2D.

La figure 4.11 et le tableau montre que la meilleure performance est obtenue par le système d'authentification de visage 2D avec un TEE=2.97%, suivi par le profile avec un TEE=4.67%. Enfin le dernier système et le système d'authentification de visage 3D, car notre système utilise la méthode globale PCA+EFM qui donne de bons résultats pour les images 2D, Pour augmenter la performance des systèmes de reconnaissance 3D il faut utiliser des méthodes locales d'extraction des caractéristiques.

La figure 4.12 présente les courbes ROC pour les trois systèmes d'authentification avec :

Nombre des caractéristiques 3D visage=10

Nombre des caractéristiques 2D visage =30

Nombre des caractéristiques 2D Profile =20

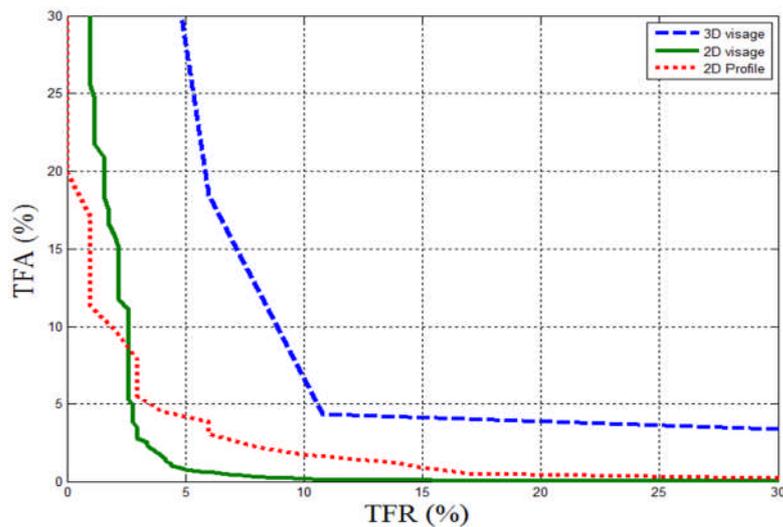


Figure 4.12 - Courbes ROC pour les trois systèmes d'authentification de visage 3D, visage 2D et profil 2D.

4.5.2 Résultats obtenus pour la fusion

Dans cette partie nous présentons la fusion des scores par un système de fusion composé par la normalisation des scores Min_ Max et un classifieur à deux classes. L'objectif est de classifier les scores client et imposteur.

La figure 4.13 et le tableau 4.4 présente la comparaison de TEE pour la fusion des scores :

- 3D visage et 2D visage,
- 3D visage et 2D profil,
- 2D visage et 2D profil,
- 3D visage et 2D visage et 2D profil,

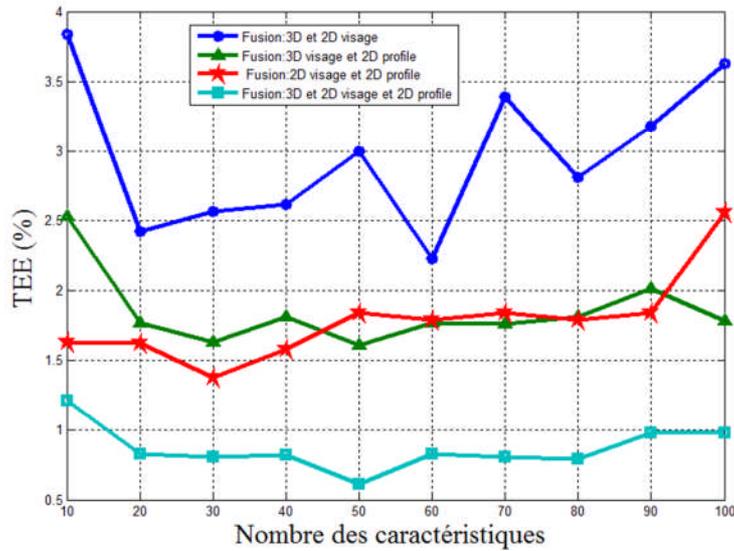


Figure 4.13 - Comparaison de TEE pour la fusion des scores

Nombre des caractéristiques	TEE Fusion 3D et 2D visage	TEE Fusion 3D visage et 2D profile	TEE Fusion 2D visage et 2D profile	TEE Fusion 3D et 2D visage et 2D profile
10	3.84	2.53	1.63	1.21
20	2.42	1.77	1.62	0.83
30	2.57	1.63	1.38	0.81
40	2.62	1.81	1.58	0.82
50	3	1.61	1.84	0.61
60	2.23	1.77	1.79	0.83
70	3.39	1.76	1.84	0.81
80	2.81	1.81	1.79	0.79
90	3.18	2.01	1.84	0.98
100	3.63	1.78	2.56	0.98

Tableau 4.4 - Comparaison de TEE pour la fusion des scores

La figure 4.13 et le tableau 4.4 montre que la fusion des trois système d’authentification 3D et 2D de visage et 2D de profile donne la meilleur performance avec un TEE=0.61%.

La figure 4.14 présente les courbes ROC pour les quatre combinaisons de fusion de scores avec :

Nombre des caractéristiques 3D visage et 2D visage=20

Nombre des caractéristiques 3D visage et 2D profile =50

Nombre des caractéristiques 2D visage et 2D profile =30

Nombre des caractéristiques 3D visage et 2D visage et 2D profile =50

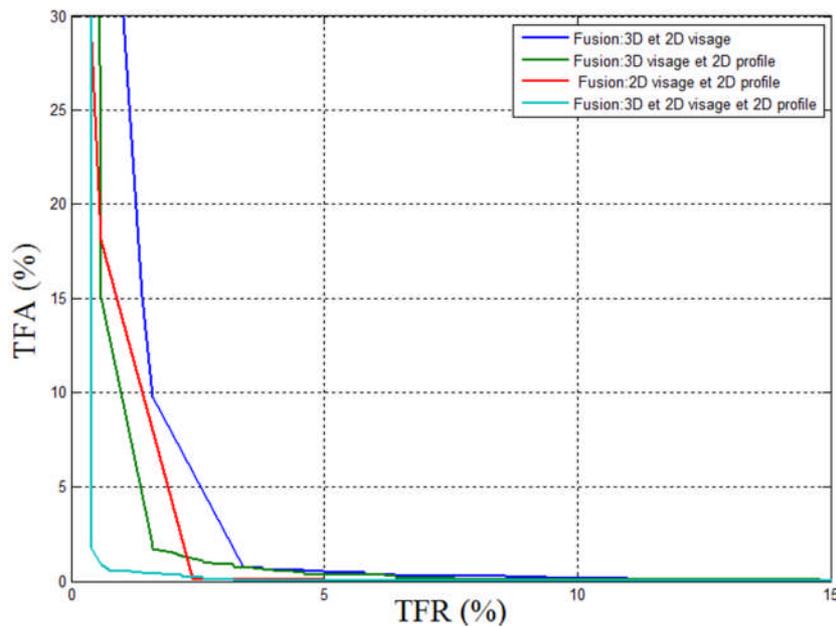


Figure 4.14 - Courbes ROC pour les quatre combinaisons de fusion de scores.

La figure 4.15 et le tableau 4.5 présente la comparaison de TEE en fonction de nombre des caractéristiques pour les trois systèmes d’authentification monomodaux et les quatre combinaisons de fusion de scores.

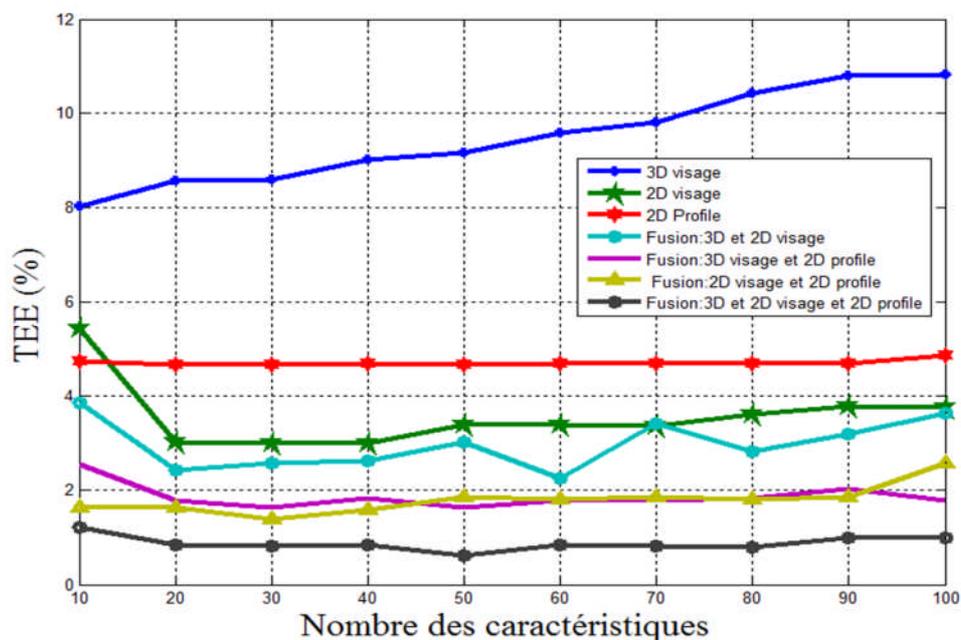


Figure 4.15 - Comparaison de TEE pour les trois systèmes d’authentification monomodaux et les quatre combinaisons de fusion de scores.

Nombre des caractéristiques	TEE (%) 3D visage	TEE (%) 2D visage	TEE (%) 2D profile	TEE Fusion 3D et 2D visage	TEE Fusion 3D visage et 2D profile	TEE Fusion 2D visage et 2D profile	TEE Fusion 3D et 2D visage et 2D profile
10	8.03	5.42	4.73	3.84	2.53	1.63	1.21
20	8.56	3.01	4.67	2.42	1.77	1.62	0.83
30	8.6	2.97	4.67	2.57	1.63	1.38	0.81
40	9.01	2.97	4.68	2.62	1.81	1.58	0.82
50	9.15	3.36	6.67	3	1.61	1.84	0.61
60	9.57	3.37	4.68	2.23	1.77	1.79	0.83
70	9.81	3.36	4.69	3.39	1.76	1.84	0.81
80	10.43	3.58	6.69	2.81	1.81	1.79	0.79
90	10.79	3.78	4.69	3.18	2.01	1.84	0.98
100	10.83	3.75	4.85	3.63	1.78	2.56	0.98

Tableau 4.5 - Comparaison de TEE pour les trois systèmes d'authentification monomodaux et les quatre combinaisons de fusion de scores.

La figure 4.15 et le tableau 4.5 montre que la fusion donne meilleur résultat que tous les systèmes monomodaux.

La figure 4.16 présente les courbes ROC pour les trois systèmes d'authentification monomodaux et les quatre combinaisons de fusion de scores.

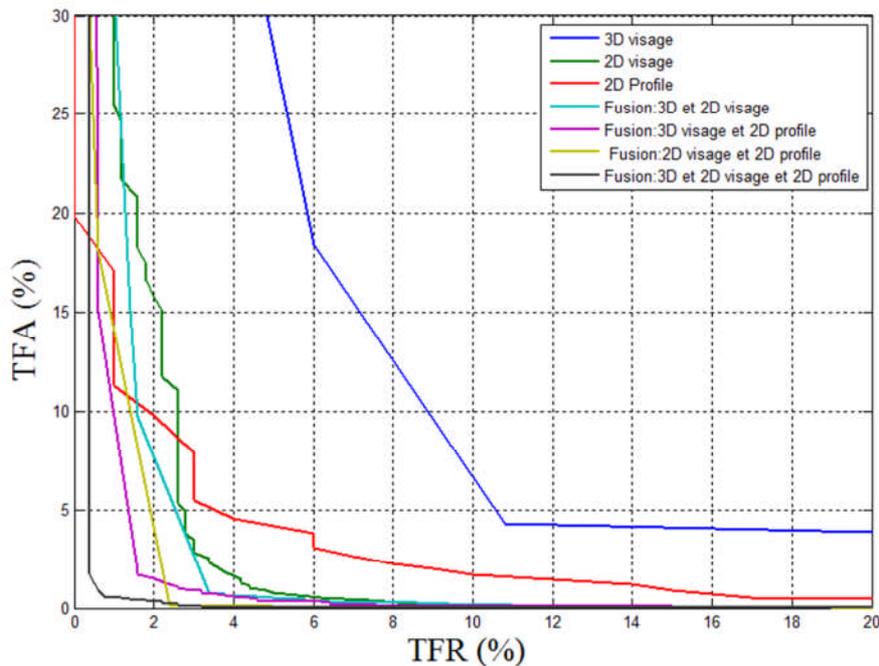


Figure 4.16 - Courbes ROC pour les trois systèmes d'authentification monomodaux et les quatre combinaisons de fusion de scores.

4.6 Conclusion

Dans ce chapitre nous avons présenté les résultats obtenus en effectuant une étude comparative tout d'abord sur les trois systèmes d'authentification de visage 3D, 2D et profile 2D. Puis une comparaison entre les quatre combinaisons de scores qui sont :

- 3D visage et 2D visage,
- 3D visage et 2D profile,
- 2D visage et 2D profile,
- 3D visage et 2D visage et 2D profile,

Enfin nous avons présenté une comparaison entre les trois systèmes d'authentification monomodaux et les quatre systèmes multimodaux.

Les résultats obtenus dans notre travail sont :

- La meilleur performance est obtenue par le système d'authentification de visage 2D avec un TEE=2.97%, suivi par le profile avec un TEE= 4.67%.
- Le dernier système et le système d'authentification de visage 3D, car notre système utilise la méthode globale PCA+EFM qui donne des bons résultats pour les images 2D, Pour augmenter la performance des systèmes de reconnaissance 3D, il faut utiliser des méthodes locales d'extraction des caractéristiques.
- La fusion des trois systèmes d'authentification 3D et 2D de visage et 2D de profile donne la meilleur performance avec un TEE=0.61%.
- La fusion donne meilleur résultat que tous les systèmes monomodaux.

- [1] Chenghua Xu, Yunhong Wang, Tieniu Tan and Long Quan, “ 3D Face Recognition Based on G-H Shape Variation,” *Springer*, LNCS 3338, pp.233-243, 2004.
- [2] F. Perronnin and J.-L. Dugelay. Introduction à la biométrie Authentification des individus par traitement audio-vidéo. *Traitement du signal*, 19(4), 2002.
- [3] Abdelmalik OUAMANE, Mébarka BELAHCENE, Abdelhamid BENAKCHA, Mohamed BOUMEHREZ, Abdelmalik TALEB AHMED, “Identification of Faces by Multimodal Information Fusion of Depth and Color,” *Sensors & Transducers Journal* (ISSN 1726- 5479), Vol. 140, No. 5, pp. 74-87, May 2012
- [4] Wonjun Hwang, Haitao Wang, Hyunwoo Kim, Member, IEEE, Seok-Cheol Kee, and Junmo Kim, Member, IEEE. Face Recognition System Using Multiple Face Model of Hybrid Fourier Feature under Uncontrolled Illumination Variation. *IEEE TRANSACTIONS ON IMAGE PROCESSING* 2011.
- [5] A.Ouamane, M.Belahcene, A.Benakcha, M.Boumehrez, A.Taleb-Ahmed, “The Classification of Scores from Multi-classifiers for Face Verification,” *Soft Sensors and Artificial Neural Networks*, *Sensors & Transducers Journal*, Vol. 145, No. 10, pp. 116-118, October 2012.
- [6] M. Belahcène, A. Ouamane, A.Taleb-Ahmed, “fusion by combination of scores multi-biometric systems,” *3rd European Workshop on Visual Information Processing (EUVIP)*, IEEE, pp. 252 – 257, 2011.
- [7] A. Ouamane, M. Belahcene, A. Benakcha, A.Taleb-Ahmed, “Fusion By Classification Of Scores Multi-Biometric Systems ,” *The Third International Conference on Image and Signal Processing and their Applications*, from Mostaganem 2,4 December 2012.

Conclusion général

Dans ce travail, nous avons présenté les résultats obtenus en effectuant une étude comparative tout d'abord sur les trois systèmes d'authentification de visage 3D, 2D et profile 2D. Puis une comparaison entre les quatre combinaisons de scores qui sont : (3D et 2D) - (3D et 2D profile) (2D et 2D profile) - (3D et 2D et 2D profile). Les résultats de ces tests ont montré l'apport de la fusion aux performances, néanmoins on a constaté que la supériorité d'une telle technique sur une autre dépend étroitement des caractéristiques de la base de données étudiée.

Dans ce mémoire nous avons étudié la mise en place et l'évaluation de systèmes SRV 3D. Après une présentation générale du contexte de ce mémoire, le profile, ses caractéristiques, ses avantages et ses inconvénients, les détails de la reconnaissance des visages 2D et 3D et 2.5D et profil 2D effectuée au chapitre 1 et chapitre 2 nous avons présenté les outils généraux d'évaluation des performances des systèmes de fusion dans le chapitre 3. Ce chapitre définit le contexte de notre travail, Ensuite, nous avons présenté les étapes de la réalisation de notre système et les méthodes utilisées et les différentes techniques d'extraction des paramètres et les types de calculs des distances, puis les types de fusion des scores dans le chapitre 3.

Le quatrième chapitre contient tous les résultats obtenus lors de notre expérience, commençant par les résultats de prétraitement des images de visages et de profil et de profondeur puis la fusion, puis l'extraction des paramètres par l'ACP & EFM et le calculs des distances, puis la fusion des distances et le calculs des erreurs et des taux de précision.

Dans ce travail nous avons obtenu les résultats suivants :

- La meilleure performance est obtenue par le système d'authentification de visage **2D couleur** avec un **TEE=2.97%**, suivi par **2D profil** avec un **TEE= 4.67%**.
- Le plus mauvais est le système d'authentification de visage **3D** avec un **TEE=8.03%**, car notre système utilise la méthode globale PCA+EFM qui donne des bons résultats pour les images 2D, pour augmenter la performance des systèmes de reconnaissance 3D, il faut utiliser des méthodes locales d'extraction des caractéristiques.
- La **fusion** des trois systèmes d'authentification **3D et 2D** de visage et 2D de profile donne la meilleure performance avec un **TEE=0.61%**.

Pour améliorer de plus les performances du système d'authentification nous proposons

Plusieurs types d'amélioration :

- Améliorer les méthodes d'acquisition des images lors de l'identification et l'authentification.
- Améliorer la détection.
- Améliorer le prétraitement des images d'identification et d'authentification
- Amélioré la classification.
- Appliquer plusieurs modèles de fusion sur des différents niveaux pour chaque Système biométrique.

Bibliographie

- [1] G. Daugman : Complete discrete 2-d gabor transform by neural networks for image analysis and compression. *IEEE Transaction on Acoustics, Speech and Signal Processing*, 36:1169–1179, 1988.
- [2] Wei, L. Fradet et T. Tan : Head pose estimation using gabor eigenspace modeling. *In International Conference on Image Processing (ICIP'02)*, 2002.
- [3] Wu et M.M. Trivedi : A two-stage head pose estimation framework and evaluation. *Pattern Recognition*, 41(3):1138–1158, 2008.
- [4] Sherrah, S. Gong et E.-J. Ong : Face distributions in similarity space under varying head pose. *Image and Vision Computing*, 19(12):807–819, 2001.
- [5] Ojala, M. Pietikainen et D. Harwood : A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29(1):51–59, 1996.
- [6] Zhang, R. F. Chu, S. M. Xiang, S. C. Liao et S. Z. Li : Face detection based on multi-block lbp representation. *In International Conference on Biometrics (ICB'07)*, pages 11–18, 2007.
- [7] Tan et B. Triggs : Enhanced local texture feature sets for face recognition under difficult lighting conditions. *In International Workshop on Analysis and Modeling of Faces and Gestures (AMFG'07)*, pages 168–182, 2007
- [8] Ma, W. Zhang, S. Shan, X. Chen et W. Gao : Robust head pose estimation using lgbp. *In International Conference on Pattern Recognition (ICPR '06)*, pages 512–515, 2006a.
- [9] Mokhber, C. Achard et M. Milgram : Recognition of human behavior by space-time silhouette characterization. *Pattern Recognition Letters*, 29:81–89, 2008.
- [10] Zhang : *Image Retrieval Based on Shape*. Thèse de doctorat, Monash University, 2002.
- [11] Gond, P. Sayd, T. Chateau et M. Dhome : A 3d shape descriptor for human pose recovery. *In International Conference on Articulated Motion and Deformable Objects (AMDO'08)*, 2008.
- [12] J. Swain et D. H. Ballard : Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

- [13] Schiele et J. L. Crowley : Recognition without correspondence using multidimensional receptive field histograms. *International Journal of Computer Vision*, 36(1):31–50, 2000. ISSN 0920-5691.
- [14] Dalal et B. Triggs : Histograms of oriented gradients for human detection. In *Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 886–893, June 2005.
- [15] Lowe : Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.
- [16] Murphy-Chutorian, A. Doshi et M.M. Trivedi : Head pose estimation for driver assistance systems : A robust algorithm and experimental evaluation. In *Intelligent Transportation Systems Conference (ITSC'07)*, pages 709–714, 2007.
- [17] Tuytelaars et K. Mikolajczyk : Local invariant feature detectors : A survey. *Foundations and Trends in Computer Graphics and Vision*, 3:177–280, 2008.
- [18] Mikolajczyk et C. Schmid : A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.
- [19] Belongie, J. Malik et J. Puzicha : Matching shapes. In *International Conference on Computer Vision (ICCV'01)*, 2001.
- [20] Freeman et E.H. Adelson : The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):891–906, 1991.
- [21] Viola et M. Jones : Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.
- [22] Papageorgiou et T. Poggio : A trainable system for object detection. *International Journal of Computer Vision*, 38(1):15–33, 2000.
- [23] Freund et R. Schapire : A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55:119–139, 1997.
- [24] Harris et M. J. Stepheds : A combined corner and edge detector. In *Alvey Vision Conference*, 1988.

- [25] Achard, E. Bigorgne et J. Devars : A sub-pixel and multispectral corner detector. *In International Conference on Pattern Recognition (ICPR'00)*, volume 3, page 3971, 2000. ISBN 0-7695-0750-6.
- [26] Bay, A. Ess, T. Tuytelaars et L. Van Gool : Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346–359, 2008.
- [27] Burges : *Data Mining and Knowledge Discovery Handbook : A Complete Guide for Researchers and Practitioners*, chapitre Geometric Methods for Feature Extraction and Dimensional Reduction : A Guided Tour, pages 1–34. Kluwer Academic Publishers, 2005
- [28] Brun : *Manifolds in Image Science and Visualization*. Thèse de doctorat, Linköping University, 2007
- [29] van der Maaten, E. O. Postma et H. J. van den Herik : Dimensionality reduction : A comparative review. Submitted to *Journal of Machine Learning Research*, 2009.
- [30] Pearson : On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(6):559–572, 1901.
- [31] Turk et A. Pentland : Face recognition using eigenfaces. *In Conference on Computer Vision and Pattern Recognition*, 1991.
- [32] Gong, S. McKenna et J. J. Collins : An investigation into face pose distributions. *In International Conference on Automatic Face and Gesture Recognition (FG '96)*, page 265, 1996.
- [33] Srinivasan et K. L. Boyer : Head pose estimation using view based eigenspaces. *In International Conference on Pattern Recognition (ICPR '02)*, volume 4, 2002.
- [34] Torgerson : Multidimensional scaling. *Psychometrika*, 17:401–419, 1952.
- [35] Cox et M. Cox : *Multidimensional scaling*. Chapman and Hall, 2000.
- [36] Williams : On a connection between kernel pca and metric multidimensional scaling. *Machine Learning*, 46(1-3):11–19, 2002. ISSN 0885-6125.
- [37] Schölkopf, A. Smola et K.-R. Müller : Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5):1299–1319, 1998. ISSN 0899-7667.

- [38] Weinberger, F. Sha et L. K. Saul : Learning a kernel matrix for nonlinear dimensionality reduction. *In International Conference on Machine Learning (ICML'04)*, page 106, 2004.
- [39] Tenenbaum, V. de Silva et J. C. Langford : A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.
- [40] Dijkstra : A note on two problems in connection with graphs. *Numerische Mathematik*, 1:269–271, 1959.
- [41] Floyd : Algorithm 97 : Shortest path. *Communications of the ACM*, 5(6):345, 1962.
- [42] Hu, W. Huang et S. Ranganath : Head pose estimation by non-linear embedding and mapping. *In International Conference on Image Processing (ICIP '05)*, 2005.
- [43] Balasubramanian et E. L. Schwartz : The isomap algorithm and topological stability. *Science*, 295(5552):7, 2002.
- [44] Lee et M. Verleysen : Nonlinear dimensionality reduction of data manifolds with essential loops. *Neurocomputing*, 67:29–53, 2005.
- [45] Roweis et L. K. Saul : Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.
- [46] Belkin et P. Niyogi : Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15:1373–1396, 2003.
- [47] Zhang et H. Zha : Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM Journal on Scientific Computing*, 26(1):313–338, 2005.
- [48] Coifman et S. Lafon : Diffusion map. *Applied and Computational Harmonic Analysis*, 21:5–30, 2006.
- [49] Raytchev, I. Yoda et K. Sakaue : Head pose estimation by nonlinear manifold learning. *In International Conference on Pattern Recognition (ICPR'04)*, pages 462–466, 2004.
- [50] Bengio, J. Paiement, P. Vincent, O. Delalleau, N.L. Roux et M. Ouimet : Outof-

- sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering. *In Neural Information Processing Systems (NIPS'03)*, 2003.
- [51] Yan, D. Xu, B. Zhang, H.-J. Zhang et S. Lin : Graph embedding and extensions : A general framework for dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1):40–51, 2007.
- [52] He, S. Yan, Y. Hu et H. J. Zhang : Learning a locality preserving subspace for visual recognition. *In International Conference on Computer Vision (ICCV'03)*, page 385, 2003.
- [53] Kwak, S.-I. Choi et C.-H. Choi : Feature extraction for regression problems and an example application for pose estimation of a face. *In International Conference on Image Analysis and Recognition (ICIAR'08)*, pages 435–444, 2008.
- [54] Li : Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86:316–342, 1991.
- [55] Li et J. Guo : Supervised isomap with explicit mapping. *In International Conference on Innovative Computing, Information and Control (ICICIC'06)*, pages 345–348, 2006.
- [56] Geng, D.-C. Zhan et Zhou Z.-H. : Supervised nonlinear dimensionality reduction for visualization and classification. *IEEE Transactions on Systems, Man, and Cybernetics - Part B : Cybernetics*, 35(6):1098–1107, 2005.
- [57] Li, Y. Fu, J. Yuan, T. S. Huang et Y. Wu : Query driven localized linear discriminant models for head pose estimation. *In International Conference on Multimedia and Expo (ICME'07)*, pages 1810–1813, 2007b.
- [58] Gourier, J. Maisonnasse, D. Hall et J. Crowley : Head pose estimation on low resolution images. *In Multimodal Technologies for Perception of Humans*, LNCS, pages 270–280, 2007.
- [59] Tu, Y. Fu, Y. Hu et T. S. Huang : Evaluation of head pose estimation for studio data. *In Multimodal Technologies for Perception of Humans*, LNCS, pages 281–290, 2007.
- [60] Voit, K. Nickel et R. Stiefelhagen : Neural network-based head pose estimation and multi-view fusion. *In Multimodal Technologies for Perception of Humans*, LNCS, pages 291–298, 2007.

- [61] Baker, S., Kanade, T.: Hallucinating Faces. Robotics Institute, Carnegie Mellon University. 1999. CMU-RI-TR-99-32 (1999)
- [62] Kemal Ekenel, H., Fischer, M., Stiefelhagen, R.: Face recognition in smart rooms. Lecture Notes in Computer Science. s.l. : Springer Berlin/Heidelberg. **4892/2008**, 120–131 (2008).
- [63] Levine, M.D., Yu, Y.: State-of-the-art of 3D facial reconstruction methods for face recognition based on a single 2D training image per person. Pattern Recogn. Lett. **30**(10), 908–913 (2009).
- [64] Li, B., Chang, H., Shan, S., Chen, X., Gao, W.: Hallucinating facial images and features. In: 2008 19th International Conference on Pattern Recognition, pp. 1–4. IEEE (2008)
- [65] O’toole, A.J., et al.: Face Recognition Algorithms surpass Humans Matching Faces over Changes in Illumination. IEEE Trans. Pattern Anal. Mach. Intell. **29**, 1642–1646 (2005)
- [66] Park, U., Jain, J.A.: 3D model-based face recognition in video. 2nd International Conference on Biometrics. Seoul : s.n. (2007)
- [67] Pentland, A., Choudhury, T.: Face recognition for smart environments. Computer. **33**(2), 50–55 (2000).
- [68] Wang, X., Xiaoou, T.: Face Hallucination and Recognition. (ed.) Springer Berlin/Heidelberg. Lecture Notes in Computer Science. ISBN 978-3-540-40302-9 (2003)
- [69] Les auteurs A-Nasser Ansari et tout Ont Travaillé en 2008 sur" Une approche multimodale pour la modélisation 3D du visage et la reconnaissance en utilisant le masque facial 3D déformable. Springer-Verlag 2008
- [70] Les auteurs Timothy C et tout Ont Travaillé en 2008 sur Profil tourné Signatures pour La détection robuste d'entité 3D
- [71] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," IEEE Con
- [72] Konica-Minolta USA. Minolta Vivid 910 non-contact 3D laser scanner, "http://www.minoltausa.com/vivid/. accessed 2004."
- [73] Quan Ju et tout [1] Ont Travaillé en 2009 sur Binary neurones en fonction du visage localisation d'entité 3D Network
- [74] Ognjen Rudovic et tout Ont Travaillé en 2009 sur Facial expression Invariant normalisation gaussienne utilisant la régression du processus

- [75] Gabriele Fanelli et son équipe ont travaillé en 2010 sur l'estimation de la pose avec la régression aléatoire.
- [76] M. D. Breitenstein, D. Kuettel, T. Weise, L. Van Gool, and H. Pfister. Real-time face pose estimation from single range images. In CVPR, 2008.
- [77] Boris Efraty et tout Ont Travaillé en 2011 sur Reconnaissance faciale 3D assistée par ordinateur basée sur les profils
- [78] I.A. Kakadiaris, G. Passalis, G. Toderici, M. Murtuza, Y. Lu, N. Karampatziakis, T. Theoharis, Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29 (4) (2007) 640–649.
- [79] B. Efraty, E. Ismailov, S. Shah, I. A. Kakadiaris, Profile-based 3D aided face recognition, Technical Report UH-CS-10-08, University of Houston, 2010.
- [80] Enrique D et tout [1] Ont Travaillé en 2011 sur " Une étude de cas pratique: Reconnaissance des visages sur la qualité des images à faible Utilisation Gabor Wavelet et soutien Vector Machines" Springer Science+Business Media B.V. 2011
- [81] Hamamoto, Y., et al.: A Gabor filter-based for recognizing handwritten numerals. *Pattern Recogn.* **31**(4), 395–400 (1998)
- [82] Heisele, B., Ho, P., Poggio, T.: Face recognition with Support Vector Machines: global vs. component-based approach. *International Conference on Computer Vision*. pp. 688–694 (2001)
- [83] Jain, A.K., Farrokhnia, F.: Unsupervised texture segmentation using Gabor filters. *Pattern Recogn.* **24**(12), 1167–1186 (1991)
- [84] Lades, M., et al.: Distortion invariant object recognition in the Dynamic Link Architecture. *IEEE Trans. Comput.* **42**(3), 300–311 (1993)
- [85] Phillips, P.J.: Support vector machines applied to face recognition. *Adv. Neural Inf. Process. Syst.* **11**, 803–809 (1999)
- [86] Wang, X., Xiaoou, T.: *Face Hallucination and Recognition*. (ed.) Springer Berlin/Heidelberg. *Lecture Notes in Computer Science*. ISBN 978-3-540-40302-9 (2003)
- [87] Weldon, T.P., Higgins, W.E., Dunn, D.F.: Efficient Gabor filter design for texture

- segmentation. *Pattern Recogn.* **29**(12), 2005–2015 (1996).
- [88] Meltem Demirkus et tout Ont Travaillé en 2011 sur" SPATIAL et probabilistes CODES MODELE TETE BASE estimation de la pose des environnements sans contrainte.
- [89] J. Orozco, S.G. Gong, and T. Xiang, "Head pose classification in crowded scenes," in *BMVC*, 2009.
- [90] J. Sherrah and S. Gong, "Fusion of perceptual cues for robust tracking of head pose and position," *Pattern Recognition*, vol. 34, no. 8, pp. 1565–1572, 2001.
- [91] C. BenAbdelkader, "Robust head pose estimation using regression-based supervised manifold learning," in *ECCV*, 2010.
- [92] D. Tosato, M. Farenzena, M. Spera, V. Murino, and M. Cristani, "Multi-class classification on riemannian manifolds for video surveillance," in *ECCV*, 2010, pp. 378–391.
- [93] Gangqiang Zhao, Ling Chen, Jie Song, and Gencai Chen, "Large head movement tracking using sift-based registration," in *IEEE ICME*, 2007, pp. 807–810.
- [94] J.-G. Wang and E. Sung, "Em enhancement of 3d head pose estimated by point at infinity," *Image and Vision Computing*, vol. 25, no. 12, pp. 1864–1874, 2007.
- [95] J. Aghajanian and S.J.D. Prince, "Face pose estimation in uncontrolled environments," in *BMVC*, 2009, pp. 1–11.
- [96] Chenghua Xu, Yunhong Wang, Tieniu Tan and Long Quan, "3D Face Recognition Based on G-H Shape Variation," *Springer*, LNCS 3338, pp.233-243, 2004.
- [97] F. Perronnin and J.-L. Dugelay. Introduction à la biométrie Authentification des individus par traitement audio-vidéo. *Traitement du signal*, 19(4), 2002.
- [98] Abdelmalik OUAMANE, Mébarka BELAHCENE, Abdelhamid BENAKCHA, Mohamed BOUMEHREZ, Abdelmalik TALEB AHMED, "Identification of Faces by Multimodal Information Fusion of Depth and Color," *Sensors & Transducers Journal (ISSN 1726-5479)*, Vol. 140, No. 5, pp. 74-87, May 2012
- [99] Wonjun Hwang, Haitao Wang, Hyunwoo Kim, Member, IEEE, Seok-Cheol Kee, and Junmo Kim, Member, IEEE. Face Recognition System Using Multiple Face Model of Hybrid Fourier Feature under Uncontrolled Illumination Variation. *IEEE TRANSACTIONS ON IMAGE PROCESSING* 2011.

-
- [100] A.Ouamane, M.Belahcene, A.Benakcha, M.Boumehez, A.Taleb-Ahmed, "The Classification of Scores from Multi-classifiers for Face Verification," *Soft Sensors and Artificial Neural Networks*, Sensors & Transducers Journal, Vol. 145, No. 10, pp. 116-118, October 2012.
- [101] M. Belahcène, A. Ouamane, A.Taleb-Ahmed, "fusion by combination of scores multi-biometric systems," *3rd European Workshop on Visual Information Processing (EUVIP)*, IEEE, pp. 252 – 257, 2011.
- [102] A. Ouamane, M. Belahcene, A. Benakcha, A.Taleb-Ahmed, "Fusion By Classification Of Scores Multi-Biometric Systems ," *The Third International Conference on Image and Signal Processing and their Applications*, from Mostaganem 2,4 December 2012.