

Université MohamedKhider – Biskra

Faculté des Sciences et de la technologie

Département :Génie Electrique

Ref :.....



جامعة محمد خيضر بسكرة

كلية العلوم و التكنولوجيا

قسم:.....

المرجع:.....

Thèse présentée en vue de l'obtention
du diplôme de
Doctorat en sciences en Automatique

**Authentification et
Identification en Biométrie**

Présentée par :

Mébarka BELAHCENE

Soutenue publiquement le 14 Janvier 2013

Devant le jury composé de :

Mr DEBILOU AbdErrazak	Maitre de Conférences 'A'	Président	Université de BISKRA
Mr BOUMEHRAZ Mohamed	Maitre de Conférences 'A'	Rapporteur	Université de BISKRA
Mlle HACHOUF Fella	Professeur	Examinatrice	Université de CONSTANTINE
Mme EL KOURD Kawther	Maitre de Conférences 'A'	Examinatrice	Université de BISKRA
Mr BOUGUECHAL NourEddine	Professeur	Examineur	Université de BATNA
Mr BENYOUCEF Moussa	Professeur	Examineur	Université de BATNA

Introduction

- 1.1 Contexte et motivations
- 1.2 Problématique de l'identification dans les bases de données biométriques
- 1.3 Objectifs principaux des travaux
- 1.4 Organisation du manuscrit

Introduction

Dans cette introduction, nous donnons une vision globale de l'ensemble des travaux développés durant cette thèse, en présentant le contexte dans lequel se sont déroulées ces recherches, les motivations ainsi que leur intérêt pour la sécurité et l'identification des personnes par la biométrie. Nous présentons ensuite les problèmes majeurs à résoudre et les contributions apportées par nos travaux.

1.1 Contexte et motivations

Les caractéristiques biométriques ne peuvent pas être facilement volées, falsifiées, ou partagées. Ainsi, elles sont plus fiables et sécurisées pour la reconnaissance de personne que les méthodes traditionnelles basées sur la connaissance ou la possession. Cependant ces caractéristiques physiques et comportementales doivent satisfaire plusieurs contraintes pour une grande fiabilité des systèmes biométriques. En effet, les objectifs de la reconnaissance biométriques ont la facilité d'utilisation par une reconnaissance sans carte ou code PIN, la sécurité accrue qui est traduite par la difficulté de contourner le contrôle d'accès ainsi que la plus grande performance quant à la précision et la rapidité des traitements. Ainsi, chaque caractéristique physiologique et/ou comportementale peut être utilisée comme une biométrie pour reconnaître une personne dès lorsqu'elle répond à ces exigences [1,2].

La performance d'un système dépend des ressources nécessaires pour atteindre la vitesse et la précision escomptées, ainsi que les facteurs opérationnels et environnementaux qui influent sur la vitesse et la précision de la reconnaissance [3].

Cependant, dans un système biométrique pratique, il existe un certain nombre d'autres paramètres qui doivent être examinés, notamment:

L'acceptation par l'utilisateur qui reflète l'ampleur des inquiétudes et des objections que l'utilisation d'une technologie biométrique donnée tend à susciter. La mesure de l'acceptation reste très subjective et varie d'une personne à l'autre et d'un pays à l'autre, suivant le régime de protection des données en vigueur, le contexte culturel et les attentes personnelles des utilisateurs.

Le contournement, qui reflète combien il est facile de tromper le système par des méthodes

frauduleuses.

La permanence ou la stabilité qui est définie par la constance d'une caractéristique biométrique au cours du développement normal et du vieillissement d'une personne. En principe, plus une caractéristique est stable, moins il est nécessaire de mettre à jour les caractéristiques personnelles ou de réenregistrer la personne.

Un système biométrique pratique doit avoir une précision acceptable et une vitesse de reconnaissance raisonnable vis-à-vis des ressources requises, inoffensif pour les utilisateurs, accepté par la population, et suffisamment robuste à l'encontre des méthodes frauduleuses.

Pour y parvenir, les spécialistes ont souvent recours à la reconnaissance des visages avec diverses techniques traditionnelles qui étaient au départ essentiellement manuelles, subjectives, assez fastidieuses, longues à réaliser, et ne pouvaient aboutir qu'à une seule représentation à la fois (mesures des régions d'intérêts). De plus, le taux de succès d'identification des visages reconstruits par les méthodes manuelles est de l'ordre de 50% seulement [3-1]. Ce faible taux est dû d'une part à la difficulté de déterminer l'aspect des traits faciaux tels que le nez, les yeux et les lèvres, et d'autre part à la variabilité des interprétations des spécialistes du domaine de la reconnaissance faciale. Celle-ci est d'autant plus difficile qu'il existe de nombreuses variations faciales, particulièrement liées à l'état nutritionnel de l'individu et aux différents stades de vieillissement. De plus, les détails du nez, des yeux, des oreilles, des lèvres, du menton sont difficiles à caractériser pour les méthodes locales. L'évolution des techniques logicielles et matérielles informatiques ainsi que l'apparition des nouvelles sources d'images numériques de haute définition, relancent depuis quelques années ce domaine à des fins judiciaires pour aider à l'identification et la recherche de personnes.

1.2 Problématique de l'identification dans les bases de données biométriques

Dans son environnement quotidien, un individu a besoin de s'identifier dans une multitude de contextes : accéder à un aéroport, entrer dans des endroits militaires, entrer à son lieu de travail, pour retirer de l'argent à un distributeur ou payer en magasin, pour demander un service social... Autant de codes et de mots de passe à mémoriser et à protéger. Afin de développer les moyens de reconnaissance, la recherche connaît de puis quelques années un renouveau spectaculaire et manifeste un intérêt majeur aux données "biométriques", c'est-à-dire aux caractéristiques propres à chaque personne: sa voix, ses

Introduction

empreintes digitales, les traits de son visage, la forme de sa main, sa signature et même son ADN. Ainsi, les techniques basées sur la biométrie jouissent d'un engouement général favorisé par un phénomène de mode, principalement véhiculé par les films au cinéma et à la télévision. Cependant, plus récemment, l'augmentation de la fraude à l'identité a créé un besoin croissant de la technologie biométrique dans un certain nombre d'applications nécessitant un haut degré de sécurité: accès à des sites sensibles, surveillance d'aéroport...

Il existe traditionnellement deux manières d'identifier une personne:

1. Méthodes basées sur une connaissance (knowledge-based). Cette connaissance correspond par exemple à un mot de passe pour ouvrir une session ou un code SIM pour un téléphone portable.
2. Méthodes basées sur une possession (token-based). Il peut s'agir d'une pièce d'identité, d'un badge, d'une clé...

Ces deux modes d'identification peuvent être utilisés d'une manière complémentaire afin d'obtenir une sécurité accrue comme pour la carte bleue. Cependant elles ont leurs faiblesses respectives. Le mot de passe peut être oublié ou deviné par une autre personne. On estime ainsi qu'une personne sur quatre écrit son mot de passe sur sa carte bleue pour ne pas l'oublier. D'autre part, la pièce d'identité peut être perdue ou volée. La biométrie exploite à l'encontre de ces deux méthodes les caractéristiques d'une personne qu'elles soient innées comme les empreintes digitales ou acquises comme la signature. Ces caractéristiques sont attachées à chaque individu et ne souffrent donc pas des faiblesses des méthodes basées sur une connaissance ou une possession. En effet les caractéristiques biométriques ne peuvent être oubliées ou perdues. De plus, elles sont très difficiles à deviner, à voler, ou à dupliquer.

S'investir dans le domaine de la reconnaissance de visages est sans doute motivé par la multiplicité et la variété des champs d'application (haute sécurité, télésurveillance et contrôle d'accès...). Les travaux portant sur ce domaine, sous différentes conditions d'éclairage, d'expressions faciales et d'orientations, peuvent être répertoriés en deux catégories distinctes suivant qu'elles se rapportent à une approche géométrique ou une approche globale. Un système automatique de reconnaissance de visages se décompose en trois sous-systèmes : détection de visage, extraction des caractéristiques et reconnaissance de visages. La mise en œuvre d'un système automatique et fiable de reconnaissance faciale est un verrou technologique qui n'est toujours pas résolu.

La reconnaissance du visage automatique à l'aide d'images 2D a été soigneusement étudiée au cours des dernières décennies. Toutefois, en raison de l'insuffisance intrinsèque de

Introduction

l'information faciale dans les images 2D, des algorithmes de reconnaissance de visage ont beaucoup de difficultés à la manipulation des variations d'éclairage, changements de pose, expressions faciales, temps de retard et des occlusions. La surface 3D du visage offre la possibilité de surmonter les difficultés causées par la variation de la tête, de pose et d'éclairage. Avec les progrès de l'imagerie 3D techniques, les données 3D du visage sont devenues de plus en plus disponibles dans ces dernières années. Il y a maintenant plusieurs bases de données publiques de référence 3D du visage. Nous présenterons donc un état de l'art sur les techniques de détection de visage 3D, nous donnerons une taxonomie des approches, des techniques et des algorithmes existants. Nous exposerons aussi les méthodes de reconnaissance de visage 3D. La reconnaissance de personnes basée sur le visage 3D présente des limites. Elle ne peut faire une identification optimale en présence d'expressions faciales. Pour cette raison nous tentons d'aborder ce domaine. La reconnaissance des expressions faciales a débutée depuis une vingtaine d'années car les expressions faciales jouent un rôle important dans la conversation humaine. Donc, l'interface homme machine du futur utilisera non seulement la souris, le clavier, mais aussi de nouvelles sources d'informations telles que les expressions faciales de six cas (colère, dégoût, joie, peur, surprise, tristesse), la posture humaine ou parole. Pour toutes ces raisons nous exposerons différentes techniques de reconnaissance de visage 3D expression t nous détaillerons en particulier les méthodes basées sur la réduction d'espace.

1.3 Objectifs principaux des travaux

La reconnaissance faciale fait partie des techniques biométriques. On remarque que dans la vie quotidienne chacun de nous identifie tout au long de la journée différents visages. Ainsi lorsque nous rencontrons une personne, notre cerveau va chercher dans notre mémoire et vérifier si cette personne est répertoriée ou non. La difficulté de la reconnaissance de visage par ordinateur varie énormément suivant les conditions d'acquisition. Dans un environnement contrôlé, des paramètres tels que l'arrière plan, la direction et l'intensité des sources lumineuses, l'angle de la prise de vue, la distance de la caméra au sujet sont des paramètres maîtrisés par le système. Dans un environnement non contrôlé, une série de prétraitements sont souvent nécessaires avant de faire la reconnaissance à proprement parler. Il faut tout d'abord *détecter* la présence ou l'absence de visage dans l'image. Le visage doit ensuite être localisé. Plusieurs méthodes ont été développées pour la reconnaissance de visage 2D. Cependant, elles présentent un certain nombre de limitations. Récemment les techniques de reconnaissance de

Introduction

visage 3D sont comme une solution alternative pour résoudre les problèmes. En effet, la richesse de l'information fournie par les mesures 3D permet de reconstruire la forme tridimensionnelle du visage. Ce type de représentation du visage est invariant aux changements d'illumination et de pose.

Dans cette thèse nous nous intéressons particulièrement au visage et nous proposons d'améliorer le prétraitement phase primordiale et aussi à la phase d'extraction de paramètres réduits du **SRV** (**S**ystème de **R**econnaissance de **V**isages).

Notre but étant donc de développer un système d'authentification de visage simple et efficace dans le cadre du projet REC_BIOMET (MERS ALGERIE). Il est nécessaire de travailler à tous les niveaux du système (détection, extraction des caractéristiques et reconnaissance) en apportant des contributions pertinentes à différents points de la chaîne de traitement. Ainsi, nous commençons par développer une technique « bas niveau » pour la localisation de visage unique sur fond uniforme par un procédé simple, et le plus robuste possible aux variations d'éclairage et de distances, d'où l'exploitation de techniques basées sur l'apparence. Particulièrement, celles utilisant l'analyse statistique et l'apprentissage automatique pour construire des machines capables de séparer les visages des non-visages. Les réseaux de neurones, les machines à vecteurs de support (**SVM**), les classificateurs bayésiens, les modèles de Markov cachés (**HMM**) sont parmi les techniques d'apprentissage automatique les plus souvent utilisées. Nous présentons un système à base de **SVM**, les vecteurs de support sont appris sur des images extraites de la base de données de visages.

Ce choix se justifie par le fait que la méthode **SVM** est très efficace pour localiser/classifier les visages d'une façon automatique et nous permettre de ne pas s'encombrer des non visages dans notre système de reconnaissance. Nous nous intéressons aux descripteurs issus de filtres de Gabor en se basant sur des travaux et résultats récents dans le domaine.

1.4 Organisation du manuscrit

Nous avons choisi d'articuler notre étude autour de **six chapitres** principaux.

Le **premier chapitre** est consacré à la présentation générale de la biométrie. Il décrit tout d'abord le principe de fonctionnement des systèmes biométriques puis définit les outils utilisés pour évaluer leurs performances. Ensuite, la place de la reconnaissance faciale parmi les autres techniques biométriques est analysée. A travers ce chapitre, nous voulons positionner le problème de la reconnaissance faciale et présenter ses enjeux et intérêts par rapport aux autres techniques. Nous étudions également les principales composantes d'un système de

Introduction

reconnaissance faciale, notamment la détection, l'extraction de caractéristiques et la reconnaissance. Enfin, nous mettons en lumière les difficultés rencontrées par les systèmes de reconnaissance de visage que nous proposons d'étudier dans le cadre de cette thèse.

Dans le **second chapitre** une analyse détaillée des différentes techniques développées au cours de ces dernières années dans les domaines de la reconnaissance 2D, 3D et 3D expressions et de la fusion est présentée.

Notre objectif est de classifier ces méthodes de manière efficace afin de mettre en évidence les particularités ainsi que les avantages et les inconvénients de chacune d'entre elles. Ceci nous permettra, par la suite, de mettre en avant les critères de performances qui vont guider le choix des solutions retenues dans le cadre de notre problématique.

Dans le **troisième chapitre** nous étudions les différentes méthodes de réduction et de classification utiles pour notre système de reconnaissance de visage.

Le **quatrième chapitre** est consacré à l'étude de la fusion et à la normalisation.

Le **cinquième chapitre** est dédié à notre approche monomodale d'extraction de visage et de ses paramètres caractéristiques. Nous présentons, dans un premier temps, le problème de la détection de visage dans une image couleur. Ensuite, nous détaillons la méthode de détection de contours que nous avons élaborée, basée sur le filtre de **Gabor**. Puis la méthode de réduction d'espace basée sur l'**EFM**. Ainsi que les méthodes de classification retenues : mesure de similarité et **SVM**. Nous présentons les bases utilisées dans le cadre de cette thèse : **XM2VTS**, **CASIA 3D** et **3D expressions**.

Par ailleurs, nous décrivons les différents outils que nous avons mis en place afin d'implémenter notre système de reconnaissance sur l'espace de couleur **HVS** et particulièrement la composante **S**. Puis, nous détaillons les quatre étapes de notre algorithme d'extraction, à savoir : extractions des régions caractéristiques, réduction d'espace, classification et décision. Notre approche est appliquée tout d'abord à la base **XM2VTS** sur visages **2D** neutre. Un choix des meilleurs filtres de **Gabor** et des meilleures méthodes de réduction d'espace linéaires est réalisé. Notre étude s'est étendue aux méthodes non linéaires et nous nous sommes limités à la **KPCA** qui est une extension de la **PCA** (méthode très répandue en reconnaissance de visages). Puis nous nous sommes intéressés au domaine **3D**, nous étudions les deux aspects : **3D** (profondeur) et **3D expressions**. Notre approche est appliquée à la **BDD CASIA 3D**. Les résultats obtenus à chaque étape sont examinés et des conclusions en sont tirées.

Dans le **sixième chapitre**, nous nous intéressons à l'exploitation des informations **2D**, **3D** et **3D expressions** du visage comme moyen biométrique pour l'authentification de

Introduction

personnes. Nous détaillons les différentes approches de reconnaissance faciale que nous avons élaborées, à savoir : l'approche **2D** basée **ACP+EFM**. Ce système de reconnaissance est appliqué en multi biométries. Nous nous intéressons tout d'abord à la fusion multi systèmes; puis au système multi algorithmes ensuite nous appliquons notre approche à une bimodalité visage-parole et finalement à la fusion des informations « couleur » et « profondeur » des images de visages **3D**. Dans ce chapitre nous étudions d'une part la fusion des caractéristiques pour son apport en information et d'autre part celle des scores pour sa rapidité. Les deux approches de fusion sont mises en œuvre avec une étape de normalisation au préalable est réalisée. Plusieurs expériences sont menées et la méthode de fusion associée à une normalisation la plus performante est retenue.

Pour éviter les problèmes d'illumination et rotation de visage nous étions tenté d'approfondir nos recherches et les orienter vers le visage **3D** et **3D** Expressions. Pour cela, les approches **3D** expressions, **3D** et l'approche bimodale **2D-3D** sont aussi étudiées et ceci toujours dans le souci d'améliorer les performances de notre système de reconnaissance. Dans toutes ces investigations la couleur est prise en considération. Nous présentons les résultats expérimentaux obtenus par chaque méthode et analysons les performances.

Enfin, la **conclusion générale** résumera nos contributions et donnera quelques perspectives sur les travaux futurs.

Chapitre 1

Système de Reconnaissance de Visage.

1.1 Introduction

Dans tous les domaines de sécurité et de contrôle d'accès, on utilise des mots de passe ou des clés qui se compose de chiffres ou de lettres. Mais, dans ces derniers temps avec l'avancement de la technologie ces mots de passe sont devenus facilement falsifiables et franchissables. C'est pour ça, que les chercheurs de différents domaines ont orientés leurs travaux sur des clés et mots de passe impossible à falsifier, sûr et surtout efficace. La biométrie est devenue à la mode dans les domaines qui requiert un niveau élevé de sécurité et de contrôle [1][2]. Et parmi toutes les technologies biométriques qui existent, la reconnaissance des visages est l'une des technologies les plus utilisées et les plus adaptées. Dans ce chapitre, nous allons mettre en relief quelques notions de base liées à la biométrie. Nous donnerons le principe de fonctionnement des systèmes biométriques, les diverses technologies et les outils utilisés pour mesurer leurs performances ainsi que les domaines d'applications. Plusieurs méthodes de reconnaissance de visages ont été proposées durant ces trente dernières années, suivant deux grands axes : la reconnaissance à partir d'images fixes et la reconnaissance à partir de séquence d'images (vidéo). La reconnaissance de visages basée sur la vidéo est préférable à celle basée sur des images fixes, puisque l'utilisation simultanée des informations temporelles et spatiales aide dans la reconnaissance. *Dans notre étude nous nous intéressons à la reconnaissance basée sur les images fixes, puisqu'elle représente la base de tous les travaux.* La reconnaissance de visages est un gros challenge tellement intéressant, qu'il a attiré les chercheurs des différents domaines : psychologie, identification de modèles, réseaux de neurones, vision d'ordinateur, infographie ... C'est pour cela, que la littérature est si vaste et si diverse. On peut répartir l'ensemble des techniques de reconnaissance de visages basées sur les images fixes, en trois grandes catégories : les *méthodes globales*, les *méthodes locales* et les *méthodes hybrides*. [3,4]

Relation entre l'identification et l'authentification Ces deux approches sont complémentaires. D'une manière générale, les systèmes de sécurité reposent sur l'usage judicieux de ces deux fonctions combinées. L'identification servira à vérifier que le demandeur n'a pas sa photo dans la base de données des interdits. L'authentification servira à vérifier que le demandeur est autorisé pour l'utilisation d'un droit donné. Grâce à cette

approche globale, la délivrance de vrais droits à partir de demandes reposant sur de fausses identités est exclue ou extrêmement difficile.

1.2 Système biométrique

1.2.1 Les phases d'un système biométrique

Chaque système biométrique comporte deux phases d'exécution :

- La phase d'apprentissage
- La phase de test.

Chacune contient trois modules de base plus un module propre à la phase de test.

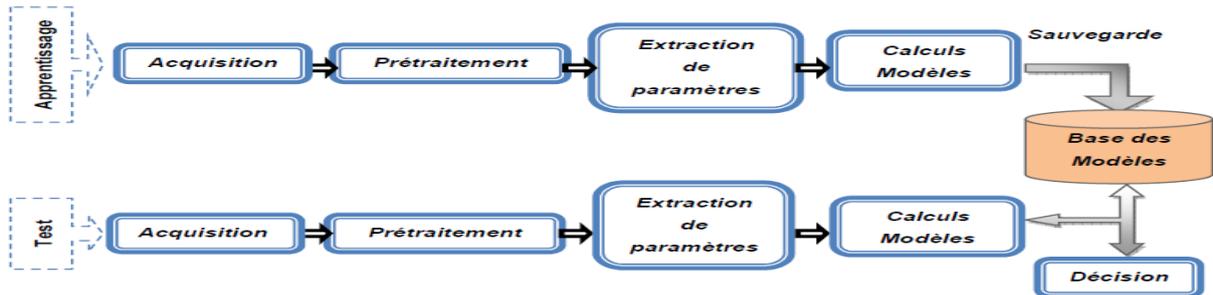


Figure 1.1 Architecture d'un système biométrique.

1.2.2 Modules de systèmes biométriques

Un système biométrique typique peut être représenté par quatre modules principaux :

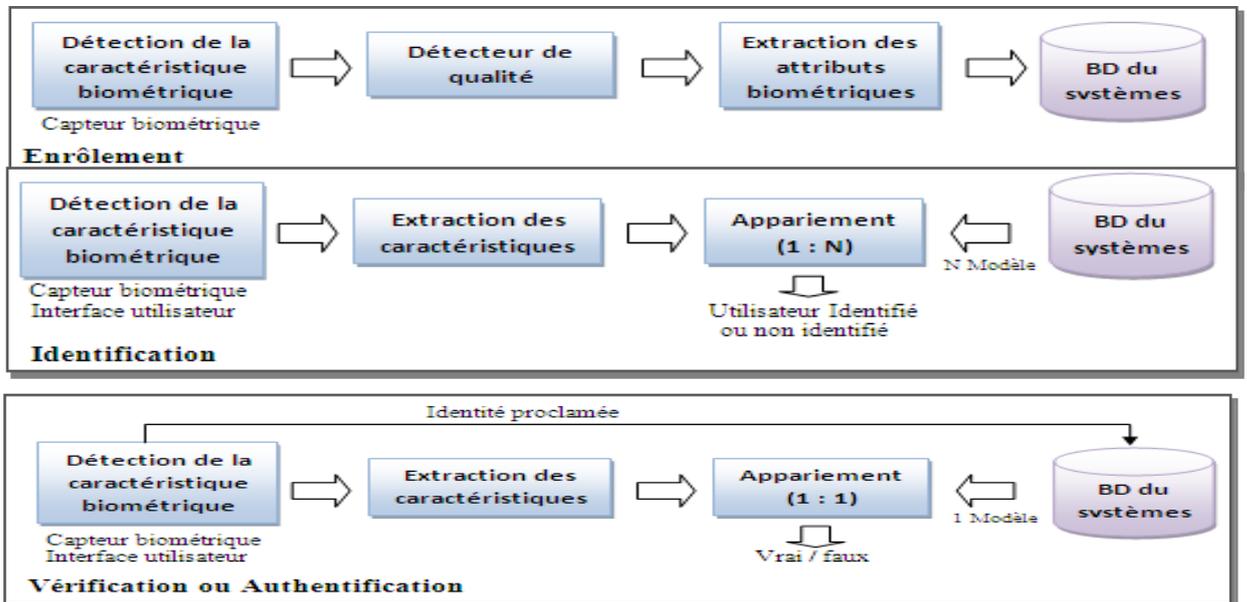


Figure 1.2 Les différents modules d'un système biométrique.

1. **Module de capture** : responsable de l'acquisition des données biométriques d'un individu (cela peut être un appareil photo, un lecteur d'empreintes digitales, une caméra de sécurité, etc).

2. **Module d'extraction de caractéristiques** : Qui prend en entrée les données biométriques acquises par le module de capture et extrait seulement l'information

pertinente afin de former une nouvelle représentation des données. Idéalement, cette nouvelle représentation est censée être unique pour chaque personne et relativement invariante aux variations intra-classe.

3. Module de correspondance : Il compare l'ensemble des caractéristiques extraites avec le modèle enregistré dans la base de données du système et détermine le degré de similitude (ou de divergence) entre les deux.

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

1.3 Pourquoi choisir le visage ?

La reconnaissance de visages est la technique la plus commune et populaire. Elle reste la plus acceptable puisqu'elle correspond à ce que les humains utilisent dans l'interaction visuelle; et par rapport aux autres méthodes, la reconnaissance du visage s'avère plus avantageuse, d'une part c'est une méthode non intrusive, c'est-à-dire elle n'exige pas la coopération du sujet (en observant les individus à distance), et d'une autre part les capteurs utilisés sont peu coûteux (une simple caméra) contrairement à l'empreinte digitale et l'iris où le sujet devra être très proche du capteur et devra coopérer pour l'acquisition de l'image sans oublier le coût de l'équipement nécessaire pour l'acquisition (équipement spécial coûteux). Malgré que certains disent que la reconnaissance de visages est une biométrie relativement peu sûre, sur le fait que le signal acquis est sujet à des variations beaucoup plus élevées que d'autres caractéristiques, comme la variation de l'éclairage, le changement de la position du visage, la présence ou l'absence de lunettes et autres; mais, au cours de ces dernières années plusieurs techniques de traitements d'images sont apparues, telle que la détection du visage, la normalisation de l'éclairage, etc. Sans oublier le développement considérable des technologies des caméras numériques, ce qui néglige l'effet de ces problèmes.

Parmi les six attributs biométriques considérés, les caractéristiques faciales marquent un score de compatibilité le plus élevé dans un *système MRTD* ("Machine Readable Travel Documents"), ce score étant basé sur **plusieurs facteurs d'évaluation** tels que l'enrôlement, le renouvellement des données, les requis matériels et la perception des utilisateurs [3,5]

Tableau 1.1 Avantages et inconvénients de la Reconnaissance de Visage.

Avantages	Inconvénients
<p>Bien accepté par le public. Aucune action de l'utilisateur (peu intrusive). Pas de contact physique. Technique peu coûteuse.</p>	<p>Technologie sensible à l'environnement (éclairage, position, expression du visage...) Difficultés de différencier de vrais jumeaux. Sensible aux changements. (barbe, moustache, lunettes, piercing, chirurgie...)</p>

1.4 Système de reconnaissance de visages

Comme nous l'avons déjà cité un système automatique de reconnaissance de visages doit intégrer une étape d'apprentissage durant laquelle il associe l'allure du visage à l'identité d'une personne. Cette étape est réalisée chez les êtres humains d'une façon spontanée et évolutive [6]. Dans un système artificiel, cette étape permet de construire une base de données des personnes connues, stockant des images étiquetées des identités. Pour ce faire, un système automatique comporte deux modes de fonctionnement : un mode **enrôlement** et un mode **identification**. Le premier mode sert à extraire pour chaque personne les éléments caractéristiques et les met sous la forme d'un vecteur caractéristique, appelé par la suite signature. Cette dernière, associée à une étiquette d'identité, sera stockée dans une base de données dédiée. Le mode d'identification permet de reconnaître une personne à partir de son image faciale, c'est à dire de retrouver l'identité associée à l'image. [7]

1.4.1 Architecture générale Idéalement, un système de reconnaissance faciale doit pouvoir identifier des visages présents dans une image ou une vidéo de manière automatique. Le système peut opérer dans les deux modes suivants : authentification ou identification. On peut également noter qu'il existe un autre type de scénario de reconnaissance faciale mettant en jeu une vérification sur une liste de surveillance, où un individu est comparé à une liste restreinte de suspects. Le principe de fonctionnement de base d'un système de reconnaissance faciale (**figure 1.3**) peut être résumé **en trois étapes** après l'acquisition.

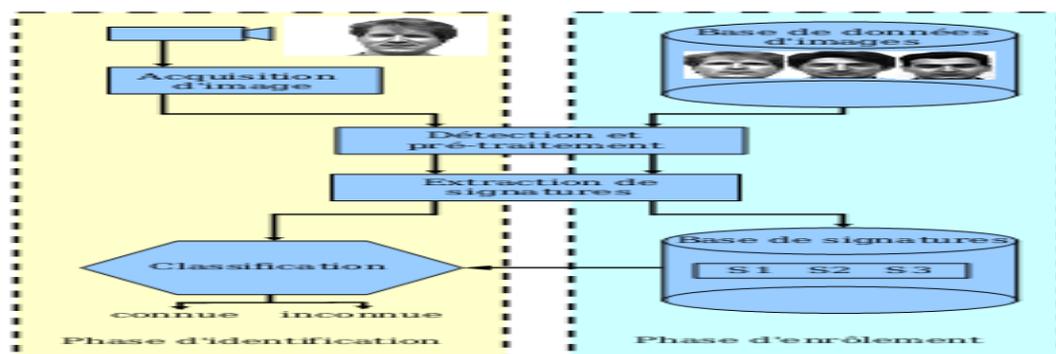


Figure 1.3 Fonctionnement d'un système de reconnaissance automatique de visages.

Les méthodes appliquées dans les modules formant le système dépendent essentiellement de la technique d'acquisition d'images de visage. En effet, s'il s'agit d'un scanner tridimensionnel ou d'un système d'acquisition stéréoscopique, la détection de visages et l'extraction de signatures reposent sur des techniques de traitement 3D [8]. Cette thématique est actuellement en pleine expansion. Elle met en évidence l'information de profondeur qui enrichit les données utilisées aussi bien dans la phase d'enrôlement que dans la phase d'identification. Toutefois, elle est limitée par le prix élevé de l'appareil d'acquisition (s'il s'agit d'un scanner 3D) et la difficulté d'installation (s'il s'agit d'un système stéréoscopique). Cette limite laisse le traitement 2D des images de visages plus accessible et encore très exploré, tant dans le domaine académique que dans le milieu industriel. Les travaux de recherche menés dans cette thèse portent sur l'utilisation des images 2D et 3D de visages. Le système comporte deux modules qui appliquent les mêmes traitements dans les deux modes de fonctionnement.

Le premier module permet de détecter le visage dans l'image brute captée par le système d'acquisition, il fournit ainsi au second module une image contenant seulement le visage. À partir de celle-ci, le second module extrait la signature discriminante. Elle sera soit stockée soit utilisée pour la classification. En **mode hors-ligne (enrôlement)**, elle sera étiquetée et stockée dans une base de données dédiée. En **mode en ligne (identification)**, elle servira comme entrée à un troisième module qui s'occupe de la classification de cette signature et de fournir la décision finale : personne inconnue ou personne connue.

1.4.1.1 Acquisition de l'image

La *capture* est la première étape dans le processus. Il faut réussir à capter l'information pertinente sans bruit. Dans la reconnaissance de visage on peut utiliser *les capteurs 3D* par exemple pour s'affranchir des problèmes de pose. Mais leur prix excessif ne permet pas une utilisation à grande échelle. Les capteurs en proche infrarouge sont utilisés pour éliminer les problèmes de l'illumination. Le *codage* consiste en l'acquisition d'image et sa digitalisation, ce qui donne lieu à une représentation bidimensionnelle au visage. L'image dans cette étape est dans un état brut ce qui engendre un risque de bruit qui peut dégrader les performances du système et donne lieu à une représentation **2D** (la matrice des niveaux de gris) pour un objet **3D** (le visage). [9]

1.4.1.2 Détection de visage et prétraitement

1) Détection : La détection de visage peut se faire par détection de la couleur de la peau, la forme de la tête ou par des méthodes détectant les différentes caractéristiques du

visage. Cette étape est autant plus délicate quand l'image acquise contient plusieurs objets de visage ou un fond non uniforme qui crée une texture perturbant la bonne segmentation du visage. Cette étape est dépendante de la qualité des images acquises. Dans la littérature scientifique, le problème de localisation de visages est aussi désigné par la terminologie "détection de visages". Les performances globales de tout système automatique de reconnaissance dépendent amplement des performances de la détection de visages. Dans l'étape de détection, on identifie et on localise le visage dans l'image acquise au départ, indépendamment de la position, de l'échelle, de l'orientation et de l'éclairage. C'est un problème de classification où on assigne l'image à la classe visage ou à la classe non visage. On peut diviser les approches de détection en quatre catégories : les méthodes basées sur la connaissance où on code la connaissance humaine du visage, les méthodes de correspondance de masques, les méthodes à caractéristiques invariables où on utilise la couleur, les textures et les contours et finalement les méthodes les plus répandues et qui sont ceux basées sur l'apprentissage ou les statistiques comme **PCA**, **SVM**, **Graph matching...**[10,11,12,13,14,15] Plusieurs travaux de recherches ont été effectués dans ce domaine. Ils ont donné lieu au développement d'une multitude de techniques allant de la simple détection du visage, à la localisation précise des régions caractéristiques du visage, tels que les yeux, le nez, les narines, les sourcils, la bouche, les lèvres, les oreilles, etc. Cependant, les solutions proposées jusqu'à maintenant sont loin d'être satisfaisantes car elles fonctionnent uniquement dans des environnements contrôlés, et par conséquent elles ne gèrent pas la variabilité des conditions d'acquisition de la vie quotidienne, notamment : *La pose, la présence ou absence des composantes structurales* (la barbe, la moustache et les lunettes), *les occultations et les conditions d'illumination*.

2) Prétraitement : Dans le **monde physique**, il y a trois paramètres à considérer : l'éclairage, la variation de posture et l'échelle. La variation de l'un de ces trois paramètres peut conduire à une distance entre deux images du même individu, supérieure à celle séparant deux images de deux individus différents. Le rôle de cette étape est d'éliminer les parasites causés par la qualité des dispositifs optiques ou électroniques lors de l'acquisition de l'image en entrée, dans le but de ne conserver que les informations essentielles et donc préparer l'image à l'étape suivante. Elle est indispensable car on ne peut jamais avoir une image sans bruit à cause du fond (background) et de la lumière qui est généralement inconnue. Il existe plusieurs types de traitement et d'amélioration de la qualité de l'image, telle que : la **normalisation**,

l'égalisation d'histogramme, le filtrage, la correction gamma ou par des méthodes plus complexes tel que le lissage anisotropique ou la méthode retinex.

1.4.3 L'extraction des caractéristiques

Cette étape représente le cœur du système de reconnaissance, on extrait de l'image les informations qui seront sauvegardées en mémoire pour être utilisées plus tard dans la phase de décision. Le choix de ces informations utiles revient à établir un modèle pour le visage, elles doivent être discriminantes et non redondantes. L'analyse est appelée indexation, représentation, modélisation ou extraction de caractéristiques. L'efficacité de cette étape a une influence directe sur la performance du système de reconnaissance de visage.

1.4.4 La comparaison des caractéristiques (classification) et décision

Elle consiste à modéliser les paramètres extraits d'un visage ou d'un ensemble de visages d'un individu en se basant sur leurs caractéristiques communes. Un modèle est un ensemble d'informations utiles, discriminantes et non redondantes qui caractérise un ou plusieurs individus ayant des similarités, ces derniers seront regroupés dans la même classe, et ces classes varient selon le type de décision. Selon les caractéristiques extraites précédemment, les algorithmes de comparaison diffèrent. On trouve dans la littérature plusieurs approches dont la plus simple est le calcul de distance (recherche de similarité). D'autres méthodes se basent sur la classification des caractéristiques par un seul classifieur (SVM[16], classifieur bayésien [17], réseau de neurones RN [18] etc) ou par plusieurs (Adaboost) [19]. L'apprentissage consiste donc à mémoriser les représentations calculées dans la phase analyse pour les individus connus. Généralement les deux étapes d'analyse et d'apprentissage sont confondues et regroupées en une seule étape.

La décision : C'est l'étape qui fait la différence entre un système d'identification d'individus et un système de vérification. Dans cette étape, un système d'identification consiste à trouver le modèle qui correspond le mieux au visage pris en entrée à partir de ceux stockés dans la base de données, il est caractérisé par son taux de reconnaissance. Par contre, dans un système de vérification il s'agit de décider si le visage en entrée est bien celui de l'individu (modèle) proclamé ou il s'agit d'un imposteur. Pour estimer la différence entre deux images, il faut introduire une mesure de similarité.

1.5 Techniques de reconnaissance de visage 2D et 3D

Les différentes techniques utilisées dans la littérature sont recensées et se présentent selon la taxonomie suivante :

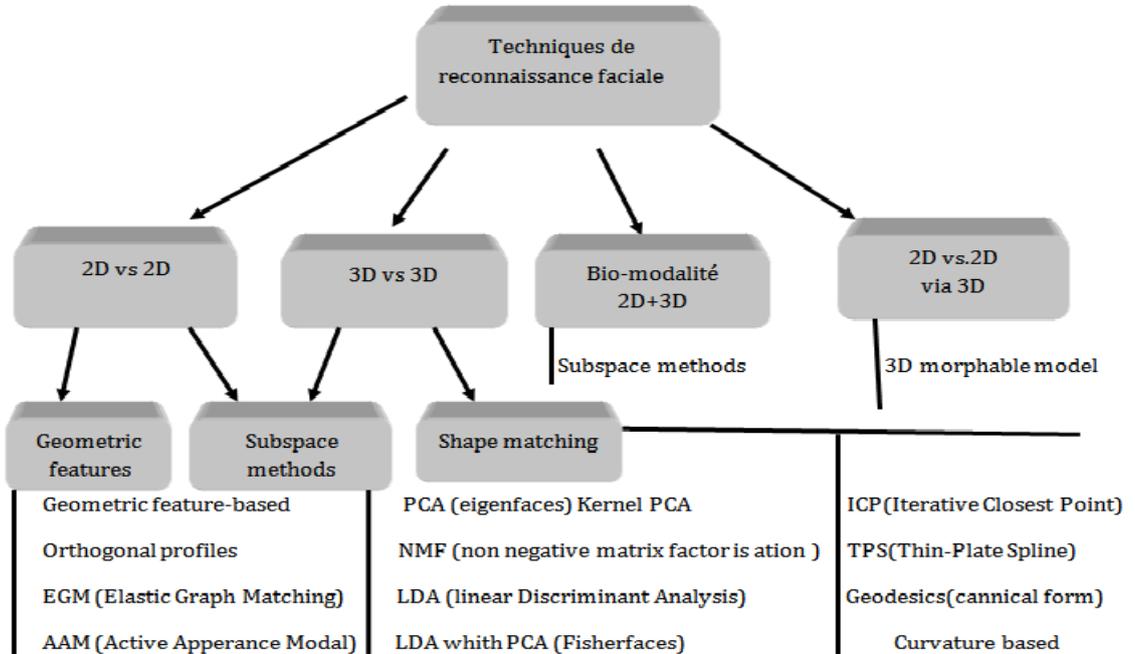


Figure 1.4 Taxonomie des approches de reconnaissance de visages 2D,2D/3D, 3D et 2D+3D. [20]

Nous classons toutes ces approches en deux sous catégories à savoir : (a) les méthodes de sous espace (Subspace Methods) et (b) les approches à base de caractéristiques géométriques (Geometric Feature Based Methods). Dans la suite, nous détaillons ces deux sous catégories et les approches qui en découlent. Il est à noter que quelques unes de ces approches ont été appliquées sur des images de profondeurs, profitant ainsi du développement mathématiques considérable que les approches 2D ont gagnées, ces quelques dernières années.

1.5.1 Techniques 2D de reconnaissance de visage

Plusieurs méthodes d'identification de visages ont été proposées durant les vingt dernières années. Avant de détailler les différentes techniques liées à la reconnaissance de visage 2D, nous allons d'abord présenter un aperçu des études faites par les chercheurs en reconnaissance faciale. En effet, la connaissance des résultats de ces études est importante car elle permet le développement de nouvelles approches. Le but ultime de la reconnaissance faciale est de rivaliser, voir même dépasser, les capacités humaines de reconnaissance. Les résultats fondamentaux de ces études sont comme suit :

- Les humains peuvent reconnaître des visages familiers dans des images de faible résolution.

- La capacité de tolérer les dégradations des images augmente avec la familiarité.
- Les informations hautes fréquences seules, soit les contours, sont insuffisantes pour obtenir une reconnaissance faciale performante.
- Les caractéristiques faciales sont traitées de manière holistique.
- Parmi les différentes caractéristiques faciales, les sourcils sont les moins importants pour la reconnaissance.
- La forme du visage est généralement codée de manière caricaturale.
- La pigmentation du visage est aussi importante que sa forme.
- La couleur joue un rôle important spécialement lorsque la forme est dégradée.
- Les changements d'illumination influencent la capacité de généralisation.
- Le mouvement des visages semble faciliter la reconnaissance de manière conséquente.
- Le système visuel progresse d'une stratégie locale vers une stratégie holistique au cours des premières années de la vie.
- L'identité faciale et les expressions sont traitées par des systèmes séparés. [21]

On distingue trois catégories de méthodes : les *méthodes globales*, les *méthodes locales* et les *méthodes hybrides*. [3,22].

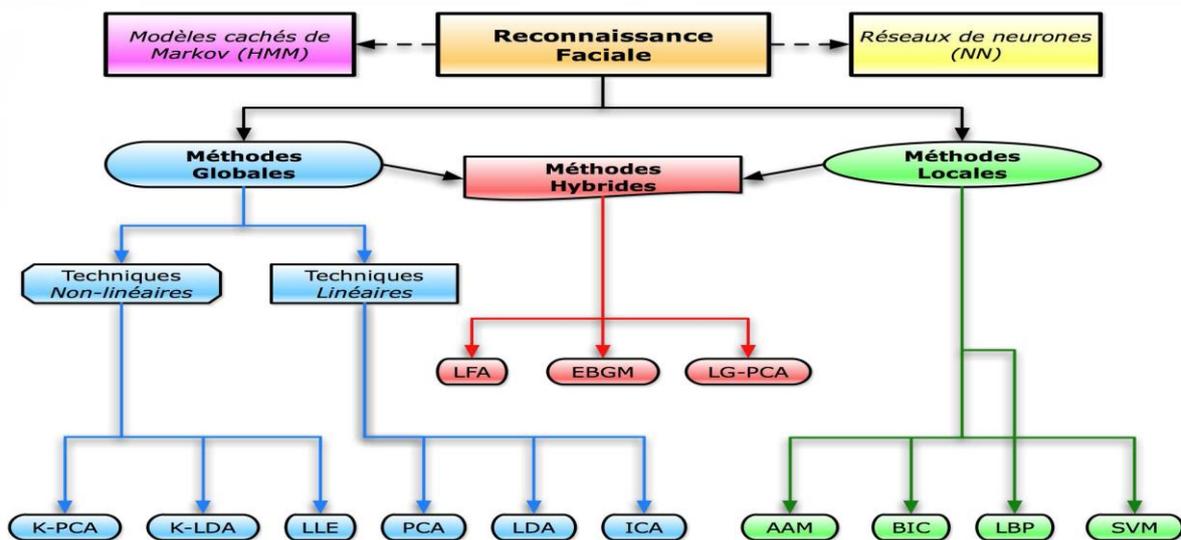


Figure 1.5 Classification des algorithmes principaux utilisés en reconnaissance faciale.

1.5.1.1 Les méthodes globales : Le principe de ces approches est d'utiliser toute la surface du visage comme source d'information sans tenir compte des caractéristiques locales comme les yeux, la bouche, etc...L'une des méthodes la plus largement utilisée pour la représentation du visage dans son ensemble est l'ACP. Les algorithmes globaux s'appuient sur des

propriétés statistiques bien connues et utilisent l'algèbre linéaire. Ils sont relativement rapides à mettre en œuvre, mais sont sensibles aux variations d'illumination, de pose et d'expression faciale. Parmi les approches les plus importantes réunies au sein de cette classe on trouve:

- L'Analyse en Composantes Principales (**PCA** ou **Eigen Faces**),
- L'Analyse Discriminante Linéaire (**LDA**),
- Machine à Vecteurs de Support (**SVM**),
- Les Réseaux de Neurones (**RNA**),
- Mélange de Gaussiennes (**GMM**),
- Modèle Surfamique du Visage (**3D**),
- L'approche **statistique et probabiliste**.

1.5.1.2 Les méthodes locales : On les appelle aussi les méthodes à traits, géométriques, à caractéristiques locales, ou analytiques. Ce type consiste à appliquer des transformations en des endroits spécifiques de l'image, le plus souvent autour des points caractéristiques (coins des yeux, de la bouche, le nez, ...), l'énergie sera accordée aux petits détails locaux évitant le bruit engendré par les cheveux, les lunettes, les chapeaux, la barbe, etc. Mais leur difficulté se présente lorsqu'il s'agit de prendre en considération plusieurs vues du visage ainsi que le manque de précision dans la phase "extraction" des points constituent leur inconvénient majeur. Précisément, ces méthodes extraient les caractéristiques locales de visage comme les yeux, le nez et la bouche, puis utilisent leur géométrie et/ou l'apparence comme donnée d'entrée du classificateur. On peut distinguer **deux pratiques différentes** :

- La première repose sur l'**extraction de régions** entières du visage, elle est souvent implémentée avec une approche globale de reconnaissance de visage.
- La deuxième pratique **extraie des points** particuliers des différentes régions caractéristiques du visage, tels que les coins des yeux, de la bouche et du nez.

Parmi ces approches on peut citer :

- Modèles de Markov Cachés (**Hidden Markov Models (HMM)**),
- L'Algorithme Elastic Bunch Graph Matching (**EBGM**),
- Eigen Object (**EO**),
- L'appariement de gabarits.

1.5.1.3 Les méthodes hybrides : Comme on a vu précédemment plusieurs approches ont été proposées pour la reconnaissance de visages, sauf qu'aucune d'elle n'est capable de s'adapter aux changements d'environnements tels que la pose, expression du visage, éclairage, etc. La

robustesse d'un système de reconnaissance peut être augmentée par la fusion de plusieurs méthodes. Il est par ailleurs possible d'utiliser une combinaison de classificateurs basés sur des techniques variées dans le but d'unir les forces de chacun et ainsi pallier à leurs faiblesses. Les techniques hybrides combinent les deux méthodes précédentes pour une meilleure caractérisation des images de visages. [23,24, 25,26]

1.5.2 Techniques 3D de reconnaissance de visage

Nous avons passé en revue les méthodes de reconnaissance **2D** de visage. Malgré les avancées réalisées ces dernières années, les techniques de reconnaissance **2D** de visages robustes aux différents facteurs de variabilité (éclairage, pose, occultation) sont loin d'être développées. La reconnaissance **3D** de visages constitue une alternative prometteuse pour surmonter ces problèmes, surtout depuis l'apparition de dispositifs d'acquisition **3D** performant. L'avantage principal des approches basées modèle **3D** réside dans le fait que le modèle **3D** conserve toutes les informations sur la géométrie de visage, ce qui permet d'avoir une représentation réelle de ce dernier. Dans cette section, après avoir rapidement évoqué les systèmes d'acquisition **3D**, nous présenterons les travaux récents sur la reconnaissance **3D** de visages (dans le **chapitre 2**). Généralement le modèle du visage est représenté par des images **2.5D** et **3D**. 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.



Figure 1.6 : (a) Visage normalisé; (b) **2D** Image; (c) **2.5D** Image.

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**. Différents systèmes d'acquisition sont utilisés pour la reconstruction

3D de visages. Ces derniers peuvent être décomposés, en fonction des capteurs de vision qu'ils utilisent, en deux grandes classes: la *vision dite active* et la *vision dite passive*.

Nous appelons *vision active* tout système de vision utilisant un capteur actif qui permet d'illuminer les objets de la scène d'une manière particulière en projetant sur leur surface des motifs structurés (un spot, un plan lumineux, une grille, etc.), ceci permet d'en extraire des informations **3D** indépendamment des conditions propres d'éclairage de la scène. [27] Ainsi, l'objectif de la *vision active* est de **réduire la complexité des traitements** à effectuer pour reconstruire la scène. En effet, les motifs projetés sur la scène sont facilement identifiables [28] à l'inverse des primitives « naturelles » tel que les contours. Le **scanner 3D**, par exemple, est un système de vision active en lumière structurée qui est capable de scanner les formes 3D via un laser et les textures (couleur de la peau) via une caméra simultanément.

En *vision passive*, la vision binoculaire ou vision stéréoscopique (utilisation de deux caméras fixes).

Un modèle de visage doit tenir compte de chacun de ces variations de façon indépendante par des paramètres explicites. Le principal défi pour la conception de tels systèmes est de trouver ou de choisir une description de ces paramètres qui permet à la fois, la modélisation appropriée d'images sur un côté et la description avec précision d'une image de l'autre. Quelques unes des sources de variation, comme l'éclairage et la pose, obéissent à l'intégrité physique des lois de la nature. Ces lois reflètent les contraintes dérivées de la géométrie tridimensionnelle des visages et l'interaction de leurs **surfaces** avec la **lumière**. Elles sont de façon optimale imposées par la représentation en **3D** qui a donc été choisie pour le modèle **morphable** [29]. D'un autre côté il ya des régularités supplémentaires entre les faces qui ne sont pas formulées que par les lois physiques, mais peuvent être obtenues en exploitant les **statistiques** générales de visages. Ces méthodes sont également représentées comme l'apprentissage à partir d'exemples. On s'attend à ce que l'apprentissage qui se conforme aux contraintes physiques est plus efficace dans les tâches comme la généralisation à partir d'une seule image d'un visage à différents éclairage et conditions. En conséquence, le modèle **3D morphable** utilise des lois physiques de modèle pose et l'éclairage et des méthodes statistiques du modèle d'identité et d'expression. Ces méthodes statistiques exigent que les faces à mettre en correspondance avec les données de chaque **visage** en **3D** soient composées de plus de **200 000 sommets** et **400 000 correctifs triangulaires** avec une collection de bords

connectés et les sommets comme dans la **figure 1.7**.

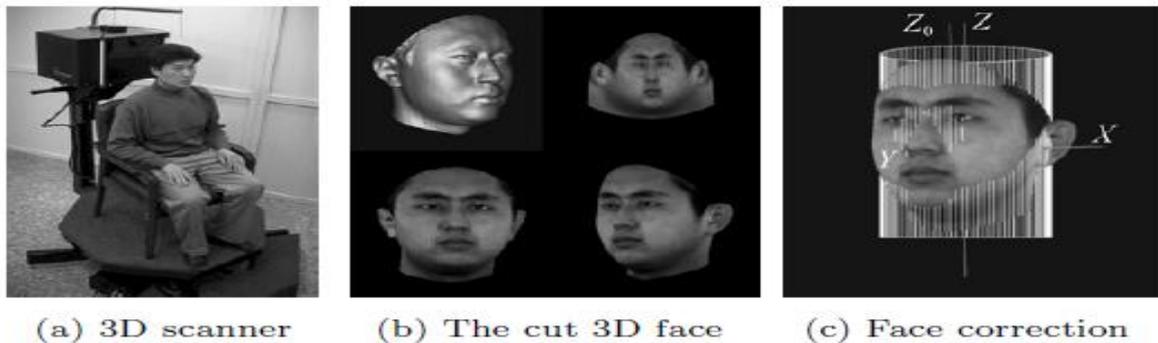


Figure 1.7 Scanner 3D et traitement de données de face originale.

C'est parce que les échantillons acquis visage **3D** sont bruyants et épineux, un préprocesseur est nécessaire pour enlever ces pointes, à remplir les trous qui résultent de la suppression des pointes par interpolation et pour lisser la surface de tout le visage. Puis, le visage est coupé à partir de l'ensemble des données numérisées par une coupe limite supprimant les données **3D** se trouvant sur les cheveux et l'épaule voir **figure 1.7** (b). Les données visage ainsi obtenus sont corrigées par un système de coordonnées homogènes. Les sommets discrets visage **3D** correspondent à un cylindre. L'axe central du cylindre est défini comme l'axe Z avec la direction positive vers le haut, et la direction à travers la pointe du nez de l'avant et perpendiculaire à l'axe Z est l'axe Y, et l'axe X est obtenu par produit vectoriel de l'axe Z et l'axe Y voir **figure 1.7** (c). Comme chaque échantillon visage **3D** ne peut avoir les mêmes quantités de données : nombre de sommets en reconnaissance de visage, on prête attention à la fonctionnalité de visage des domaines tels que les yeux, front, nez, bouche, et ainsi de suite. Dans l'ordre de réduire les informations inutiles, les échantillons face sont traités comme suit. Tout d'abord, deux points caractéristiques clés (les bout du nez et un autre hors-bord sommet sur le front à gauche) sont situés voir **figure 1.7** (b). Parce que l'axe des Y représente les informations sur la profondeur du visage, le bout du nez sommet peut être recherché facilement que le sommet avec le maximum de la valeur y de **3D** des données de face. La texture extérieure sommet sur le sourcil gauche peut être déterminée par l'algorithme de détection de visage. La correspondance de coordonnées peut être acquise par le rapport entre la texture et l'information de forme. Le procédé de chercher les deux points clés (caractéristiques) doit être mené une seule fois, et les numéros de séquence des deux points sur la surface de face peuvent être enregistrés. Étant donné un nouveau échantillon visage **3D**, les deux points clés peuvent être repérés par leurs numéros de séquence directement et par patches des triangulaire.

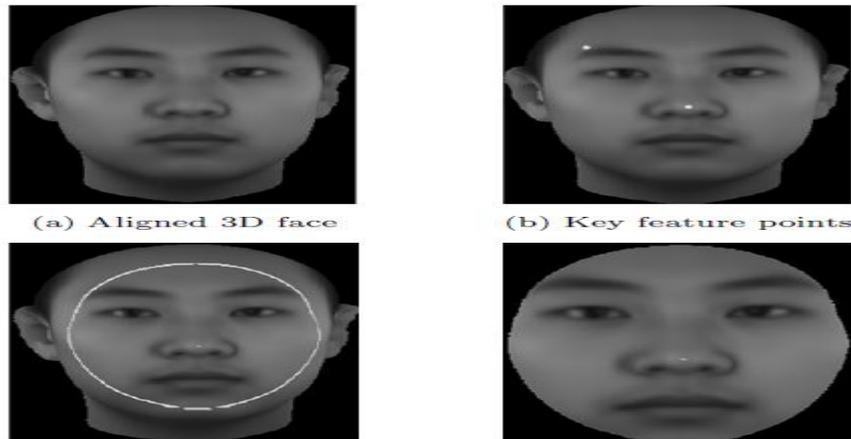


Figure 1.8 Détection de visage 3D (vue frontale)

Les techniques de reconnaissance **3D** de visage peuvent être regroupées en trois catégories principales : **approches** basées **modèle** [30,31,32], **approches 3D** [33,34,13] et **approches multimodales 2D + 3D** [35,36,37]. Ces approches sont détaillées dans le **chapitre 2**. Nous nous limitons à les exposer brièvement dans ce présent chapitre.

1.5.2.1 Approches modèles

Ces approches construisent, à partir des points **3D**, des modèles de visages qu'elles utilisent par la suite pour la reconnaissance. L'ensemble des visages est représenté par un espace vectoriel. Les points **3D** des modèles de visages générés sont représentés par leurs coordonnées cylindriques définies par rapport à un axe vertical.

1.5.2.2 Approches 3D

Elles sont subdivisées en deux catégories : les approches basées surface qui utilisent la **géométrie** de la **surface du visage** [38,39,40,41] et les **approches holistiques 3D** [42] .

1.5.2.2.1 Approches surface

Dans ce cas, le problème de la reconnaissance **3D** de visage est celui de l'alignement de deux surfaces **3D** qui modélisent les deux visages à appairer. L'algorithme généralement utilisé est l'**algorithme du plus proche voisin itéré**, ou **ICP (Iterative Closest Point)**, qui a été introduit par [43]. Il consiste en une optimisation alternée d'appariements et de transformations. Ainsi, à partir d'une transformation initiale, les deux étapes suivantes sont réitérées [44]:

- *mise en correspondance* (plus proche voisin) : on apparie chaque primitive du modèle transformé avec la primitive la plus proche dans la scène.

- *recalage* : la transformation (translation + rotation) est généralement calculée aux sens des moindres carrés, surtout si l'on travaille avec des points. Si l'on possède une information d'incertitude, on peut l'utiliser dans les étapes terminales pour affiner la solution.

1.5.2.2.2 Approches holistiques 3D

Les techniques holistiques comme l'ACP ont été largement utilisées dans la reconnaissance faciale 2D. Plus récemment, ces techniques ont été aussi étendues aux données 3D de visage.

1.5.2.2.3 Approche géométrique ou locale 3D

Par rapport aux approches « holistiques », les techniques d'identification 3D du visage basées sur les caractéristiques faciales locales de type géométriques restent relativement peu développées, alors que dans le même temps des approches locales d'identification 2D, telles que l'« Elastic Bunch Graph Matching (EBGM)» [26] et l'analyse locale de caractéristiques [39,45], connaissent un grand succès auprès des fournisseurs de données (Recognition Vendors Test 2002 [46]). Par ailleurs, les combinaisons des techniques de reconnaissance 2D et 3D sont de plus en plus étudiées [47]. Ainsi la combinaison des approches EBGM 2D et EBGM 3D (qui correspond à une **extension** de la méthode EBGM 2D aux images de **profondeur** de visage) a donné lieu à une technique d'identification du visage 2D+3D très efficace évaluée sur la base FRGC. Cependant, l'évaluation de cette approche hybride a démontré que les **performances** du EBGM 2D **surpassent** de manière significative celles du EBGM 3D. Elle a aussi mis en avant le besoin de comprendre et d'utiliser les propriétés discriminantes des caractéristiques locales du visage afin de concevoir des techniques efficaces de reconnaissance 3D de visage.

1.5.2.3 Multimodal 2D +3 D reconnaissance de visage

Notre reconnaissance de visage optimise l'avantage d'utiliser à la fois les données en 2D et 3D du visage grâce à l'utilisation de l'information aux niveaux capteurs ou aux niveaux scores. Les acquis en 2D et 3D des données faciales subissent une série de prétraitement afin d'obtenir des images 2D et 3D normalisées tel que décrit dans la section précédente. Au niveau de l'image, les images 2D et 3D sont recombinaés dans un ensemble de données plus indépendant et discriminant par la recombinaison d'image basée sur l'APC. Les données recombinaées peuvent être utilisés individuellement ou en parallèle en tant qu'entrée au classificateur visage. Dans notre système de reconnaissance de visage, deux procédés

correspondants sont effectués indépendamment sur les données recombinaison. Les données sont recombinaison transformées dans leurs espaces respectifs utilisant **Fisher faces**. Les deux fonctions sont utilisées pour déterminer les scores de dissemblance entre visages **2D** et **3D**. Les processus d'appariement donnent lieu à deux notes distinctes. Enfin, au niveau score, le score final est obtenu en tant que somme pondérée des deux partitions. Les pondérations sont déterminées empiriquement par l'analyse discriminante linéaire de **Fisher**. Dans ce qui suit, nous décrivons la recombinaison des images de visage **2D** et **3D**, la réduction d'espace par Fisher faces amélioré et la fusion au niveau caractéristiques et scores du **SRV** multimodal **2D+3D**.

1.6 Principales difficultés de la reconnaissance de visage

Pour le cerveau humain, le processus de la reconnaissance de visages est une tâche visuelle de haut niveau [6]. Bien que les êtres humains puissent détecter et identifier des visages dans une scène sans beaucoup de peine, construire un système automatique qui accomplit de telles tâches représente un sérieux défi. Ce défi est d'autant plus grand lorsque les conditions d'acquisition des images sont très variables. Il existe deux types de variations associées aux images de visages : inter et intra sujet. La variation inter sujet est limitée à cause de la ressemblance physique entre les individus. Par contre la variation intra sujet est plus vaste. Elle peut être attribuée à plusieurs facteurs. Chaque visage individuel peut générer une grande variété d'images différentes. Cette grande diversité d'images de visages rend l'analyse difficile. Outre les différences générales entre les faces des variations dans l'apparence d'images de visage posent de grands problèmes à l'identification. Ces variations sont recensées comme suit:

- Changements d'éclairage influencent l'apparition d'un visage, même si la pose de la face est fixée.
- Variations de pose peuvent entraîner des changements dramatiques dans les images.
- Les expressions faciales un outil important dans la communication humaine sont une autre source de variations dans les images. Seuls quelques points de repère du visage qui sont directement couplés avec la structure osseuse du crâne, comme la distance interoculaire ou la position générale de l'oreille sont constants dans un visage. La plupart des autres caractéristiques peuvent changer leur configuration spatiale ou position en raison de l'articulation de la mâchoire ou à l'action des muscles, comme les sourcils mobiles, les lèvres ou les joues.

- À long terme un visage change en raison du vieillissement, à une coiffure de changer ou selon maquillage ou accessoires. L'isolement et la description explicite de toutes ces différentes sources de variations doivent être le but ultime d'un système d'analyse du visage.

1.6.1 Changement d'illumination

L'apparence d'un visage dans une image varie énormément en fonction de l'illumination de la scène lors de la prise de vue (voir figures 1.9). Les variations d'éclairage rendent la tâche de reconnaissance de visage très difficile. En effet, le changement d'apparence d'un visage dû à l'illumination, se révèle parfois plus critique que la différence physique entre les individus, et peut entraîner une mauvaise classification des images d'entrée. Ceci a été expérimentalement observé dans Adini et al [48] où les auteurs ont utilisé une base de données de 25 individus. L'identification de visage dans un environnement non contrôlé reste donc un domaine de recherche ouvert. Les évaluations FRVT ont révélé que le problème de variation d'illumination constitue un défi majeur pour la reconnaissance faciale. [49]



Figure 1.9 Exemple de variation d'éclairage.

1.6.2 Variation de la pose

Un changement de l'angle d'inclinaison du visage engendre, pour une position fixe du capteur, de nombreux changements d'apparence dans l'image collectée. Nous nous intéressons ici aux rotations du visage en profondeur tels que les mouvements de type hochement de tête ou négation. En effet, la phase préliminaire de normalisation du visage présentée en section 1.4.1.2 permet de corriger d'éventuelles rotations dans le plan de l'image. Le taux de reconnaissance de visage baisse considérablement quand des variations de pose sont présentes dans les images. Cette difficulté a été démontrée par des tests d'évaluation élaborés sur les bases FERET et FRVT [46]. Les rotations en profondeur engendrent l'occultation de certaines parties du visage comme pour les vues de trois quarts. D'autre part, elles amènent des différences de profondeur qui, projetées sur le plan 2D de l'image, résultent en des déformations qui font varier la forme globale du visage. Ces déformations qui consistent à l'étirement de certaines parties du visage et le compactage d'autres régions font

varier aussi les distances entre les caractéristiques faciales. Si la pose du visage-requête diffère significativement de celle des visages enregistrés, les performances des systèmes de reconnaissance baissent sensiblement. En effet, selon le **FRVT 2000**, la **rotation** de la **tête** n'entraîne pas de **baisse** significatives des **taux de reconnaissance** jusqu'à $\pm 25^\circ$. Ce seuil, dépassé, on constate une chute de performances. On montre que, si le seul facteur de variation entre l'image enregistrée et l'image-requête est une rotation en profondeur de la tête inférieure à 30° , les taux de reconnaissance des systèmes actuels sont de l'ordre de **90%**. Des rotations plus importantes engendrent une forte baisse des performances [50].

Quand le visage est de profil dans le plan image (orientation $< 30^\circ$), il peut être normalisé en détectant au moins deux traits faciaux (passant par les yeux). Cependant, lorsque la rotation est supérieure à 30° , la normalisation géométrique n'est plus possible.



Figure 1.10 Exemples de variation de poses.

1.6.3 Expressions faciales

Les visages sont des objets non rigides. L'expression faciale de l'émotion, combinée avec la parole, peut produire des changements d'apparence importants des visages. Le nombre de configurations possibles est incalculable. L'influence de l'expression faciale sur la reconnaissance est donc difficile à évaluer. Puisque l'expression faciale affecte la forme géométrique et les positions des caractéristiques faciales, il semble logique que les techniques globales ou hybrides y soient plus robustes que la plupart des techniques géométriques. On soutient que les expressions faciales n'ont pas une grande influence sur les algorithmes de reconnaissance, pour autant qu'elles restent raisonnables. Alors que les cas extrêmes qui engendrent d'importantes déformations de la bouche (tel que le cri) et le rétrécissement ou la fermeture complète des yeux dégradent considérablement les performances de la reconnaissance automatique. Il peut être utile de repérer en amont de la reconnaissance ces expressions problématiques. Si l'on est capable de catégoriser l'expression faciale du visage-requête, deux approches sont possibles. Soit plusieurs modèles de visages ont été appris, un par catégorie d'expression faciale. Dans ce cas, il serait possible de comparer directement le visage-test à la base des visages arborant la même expression. Par contre, à l'aide d'une technique générative utilisant un modèle de visage suffisamment précis, il serait possible de

transformer le visage-test de manière à ce qu'il se présente dans des conditions moins difficiles et plus favorables [50].

La déformation du visage qui est due aux expressions faciales est localisée principalement sur la partie inférieure du visage. L'information faciale se situant dans la partie supérieure du visage reste quasi invariable. Elle est généralement suffisante pour effectuer une identification. Toutefois, étant donné que l'expression faciale modifie l'aspect du visage, elle entraîne forcément une diminution du taux de reconnaissance.



Figure 1.11 Exemples de variation d'expressions [19].

L'identification de visage avec expression faciale est un problème difficile qui est toujours d'actualité et qui reste non résolu [51]. L'information temporelle fournit une connaissance supplémentaire significative qui peut être utilisée pour résoudre ce problème [52].



Figure 1.12 Exemples de variation d'expressions.

1.6.4 Présence ou absence des composants structurels

La présence des composants structurels telle que la barbe, la moustache, ou bien les lunettes peut modifier énormément les caractéristiques faciales telles que la forme, la couleur, ou la taille du visage. De plus, ces composants peuvent cacher les caractéristiques faciales de base causant ainsi une défaillance du système de reconnaissance. Par exemple, des lunettes opaques ne permettent pas de bien distinguer la forme et la couleur des yeux, et une moustache ou une barbe modifie la forme du visage.

1.6.5 Occultations partielles

Le visage peut être partiellement masqué par des objets dans la scène, ou par le port d'accessoire tels que lunettes, écharpe... Dans le contexte de la biométrie, les systèmes proposés doivent être non intrusifs c'est-à-dire qu'on ne doit pas compter sur une coopération active du sujet. Par conséquent, il est important de savoir reconnaître des visages

partiellement occultés. **Gross et al [53]** ont étudié l'impact du port de lunettes de soleil, et du cache-nez occultant la partie inférieure du visage sur la reconnaissance faciale. Une application est réalisée sur la base de données **AR [54]**. Les résultats expérimentaux semblent indiquer que, dans ces conditions, les performances des algorithmes de reconnaissance restent faibles [19].

1.6.6 Vieillesse et le changement d'aspect

Les visages changent d'apparence au fil du temps. Les modifications concernent la tension des muscles, l'apparence de la peau (apparition de rides), le port de lunettes, éventuellement le maquillage ou la présence d'une frange occultant une partie du front.

Sur la base **AR**, où le délai entre les deux sessions de prises de vue est seulement de deux semaines, la **baisse des taux de reconnaissance** est estimée à **20%**. Dans le **rapport FRVT 2000**, les effets du temps ont été mesurés à l'aide des vues duplicitées de la base **FERET** : celles-ci sont comparées aux vues "fa" de la galerie (BDD d'apprentissage), ce qui permet d'établir un taux de reconnaissance. Les taux fournis par les meilleurs algorithmes sont de **63%** sur les vues "**dup1**", et **64%** pour les vues "**dup2**", contre **58%** et **52%** pour l'évaluation de 1996. Notons que l'ensemble de test **dup1** est formé par toutes les vues prise à des sessions ultérieures que les vues "fa" de la galerie. L'ensemble "**dup2**" est un sous ensemble de "**dup1**" formé par toutes les vues prises au moins **540 jours** après celles qui forment la galerie.

Plus récemment, les systèmes de reconnaissance ont réalisé plus de progrès pour gérer au mieux le délai de temps entre deux prises de vue. Néanmoins, bien que l'intervalle de temps entre les vues "fa" et les vues "dup1" ne soit pas important, les systèmes ont du mal à reconnaître ces dernières. Cela provient certainement des changements dans les conditions de prises de vue, et non d'un vieillissement des visages. Dans le cadre de l'évaluation **FRVT 2002**, la **baisse des taux** de reconnaissance des meilleurs algorithmes testés a été estimée à **5% par année** d'écart entre l'image de référence et l'image à reconnaître [44].

1.7 Mesure de la performance d'un système biométrique

L'exactitude de tout système biométrique est représentée par deux taux d'erreurs qui se produisent. Le système biométrique prend une des quatre décisions possibles pendant la phase **décision [3], [55]**:

- Le véritable sujet est accepté.
- Le véritable sujet est rejeté (**FRE** ou **FR**).

- L'imposteur est accepté (**FAE** ou **FA**).
- L'imposteur est rejeté.

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

1. Le premier critère s'appelle le **taux de faux rejet** ("False Reject Rate" ou **FRR(TFR)**). Ce taux représente le pourcentage de personnes censées être reconnues mais qui sont rejetées par le système,

2. Le deuxième critère est le **taux de fausse acceptation** ("False Accept Rate" ou **FAR (TFA)**). Ce taux représente le pourcentage de personnes censées ne pas être reconnues mais qui sont tout de même acceptées par le système,

3. Le troisième critère est connu sous le nom de **taux d'égale erreur** ("Equal Error Rate" ou **EER (TEE)**). 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 $FRR = FAR$, c'est-à-dire le meilleur compromis entre les faux rejets et les fausses acceptations.

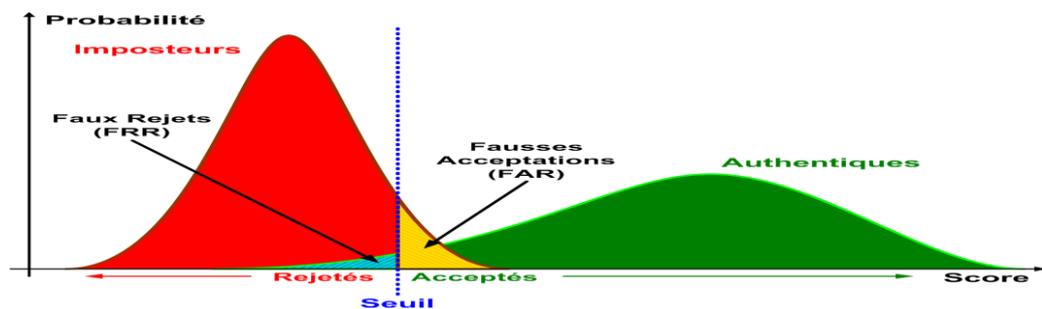


Figure 1.13 Illustration du FRR et du FAR.

La figure 1.13 illustre le **FRR** et le **FAR** à partir de distributions des scores authentiques et imposteurs tandis que l'**EER** est représenté sur la figure 1.14.

Selon la *nature* (*authentification* ou *identification*) du système biométrique, il existe deux façons de **mesurer la performance** en termes de Taux d'acceptation Faux **TFA** le taux de faux rejet **TFR**, défini comme suit:

$$TFR = \frac{\text{Nombre des clients rejetés (FR)}}{\text{Nombre total d'accès clients}} \quad (1.1)$$

$$TFA = \frac{\text{Nombre des imposteurs acceptés (FA)}}{\text{Nombre total d'accès imposteurs}} \quad (1.2)$$

Une vérification parfaite d'identité ($FA=0$ et $FR=0$) est non réalisable dans la pratique. Cependant, comme montré par l'étude l'hypothèse de test binaire, n'importe lequel de ces deux

taux (**TFA**, **TFR**) peut être réduit à une petite valeur arbitraire en changeant le seuil de décision, avec l'inconvénient d'augmenter l'autre. Une seule mesure peut être obtenue en combinant ces deux taux d'erreurs dans le taux erreur totale (**TET**) ou son complément, le taux de réussite total (**TR**) :

$$\mathbf{TET} = \frac{\text{Nombre de fausses acceptations}(\mathbf{FA}) + \text{Nombre de faux rejets}(\mathbf{FR})}{\text{Nombre totale d' accès}} \quad (1.3)$$

$$\mathbf{TR} = 1 - \mathbf{TET} \quad (1.4)$$

Cependant, un soin devrait être pris en utilisant une seule mesure. En effet, cette seule mesure (**TET** ou **TR**) pourrait être fortement pondérée par l'un ou l'autre type d'erreurs (**TFA** ou **TFR**), dépendant seulement du nombre d'accès qui ont été utilisés en obtenant ce type d'erreur. Le **TET** sera toujours plus près de l'erreur (**TFA** ou **TFR**) qui a été obtenue en utilisant le plus grand nombre d'accès dans son type.

Lorsque le système opère en *mode authentification*, on utilise ce que l'on appelle une **courbe ROC** (pour "Receiver Operating Characteristic" en anglais). La courbe **ROC** (**Figure 1.14**) trace le taux de faux rejet en fonction du taux de fausse acceptation [56]. 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. La performance globale d'un système de vérification d'identité est mieux caractérisée par sa courbe caractéristique de fonctionnement (**Receiver Operating Characteristic** ou **ROC**), qui représente le **TFA** en fonction du **TFR**.



Figure 1.14 Courbe ROC.

Le taux d'égale erreur **TEE** correspond au point **TFA = TFR**, c'est-à-dire graphiquement à l'intersection de la courbe ROC avec la première bissectrice. Il est fréquemment utilisé pour donner un aperçu de la performance d'un système. Seuls des systèmes qui produisent des taux **EER** (**Equal Error Rate**) appelé aussi **TEE** Taux d'erreur égale) faibles sont capables d'être déployés en mode identification. Ainsi, les protocoles d'évaluation diffèrent dans le mode identification et le mode vérification. Il faut savoir qu'en mode identification, deux scénarios sont possibles qui induisent deux protocoles différents.

Si l'**identification** se fait dans ce qui est appelé un **mode fermé**, c'est-à-dire que *chaque personne est représentée dans la base de données par au moins une référence*, alors le système n'est pas obligé de faire du rejet. Il peut assigner l'identité de la personne qui a le score le plus élevé (dans le cas de mesure de score de similitude). L'évaluation doit se faire sur toutes les comparaisons possibles entre les éléments de la base des références et les autres éléments de la base afin de simuler au mieux le cas d'un processus d'identification.

Un **deuxième scénario** est possible en mode **identification**, c'est le **mode ouvert**. Dans ce cas, *toutes les personnes de la base de test ne sont pas représentées dans la base de référence*. En mode évaluation, il est primordial de regrouper les personnes de la base de données considérée en deux groupes, celles disposant de références et celles qui n'en ont pas. Ensuite, le système doit pouvoir prendre deux décisions distinctes : soit rejeter un échantillon test ou lui assigner une identité représentée dans la base des référents. Pour mettre en place une stratégie de rejet, il est impératif de fixer un seuil de décision analogue à celui fixé en mode vérification [57,58]. La performance des systèmes biométriques est mesurée par leur exactitude dans l'identification, qui est calculée en utilisant les **TFR** et **TFA**. Le taux où on a la valeur du **TFR** est égal à la valeur du **TFA** s'appelle le taux d'erreur égal "**TEE**" et il est utilisé pour comparer les systèmes. Souvent, le taux véritable d'acceptation **TFA** est également employé comme seule mesure de performance.

Dans tout système, on ne peut pas simultanément diminuer à la fois **TFA** et **TFR**, c'est pour cela que lors de la conception du système, on essaie de dégager un seuil de décision. Et puisque la sécurité est habituellement l'objectif principal, un bas **TFA** est habituellement choisi aux dépens d'un haut **TFR**.

1.7.1 Intervalle de confiance

Le choix des critères d'évaluation des performances des systèmes biométriques, nous devons faire face à une autre difficulté de l'évaluation d'un système biométrique. En général, on ne dispose que de **peu de données** de chaque personne et **de peu de personnes** dans les bases. Certaines bases sont composées de quelques dizaines de personnes seulement et on a peu de données par personne, même si récemment des bases de l'ordre de centaines de personnes ont fait leur apparition. Il est alors impératif de mesurer en plus des taux d'erreurs calculés, une mesure de confiance associée.

Deux types de mesure des intervalles de confiance sont introduits par Mellakh [57] et Allano [58]: l'une basée sur une méthode non-paramétrique appelée "**bootstrap**" et l'autre basée sur une **méthode paramétrique**.

1.7.1.1 Méthode non-paramétrique

Pour utiliser la méthode d'estimation d'intervalles de confiance par le "**bootstrap**", il faut réaliser un ré-échantillonnage de l'ensemble de test. Ce ré-échantillonnage consiste en un tirage aléatoire avec remplacement de M exemples. Dans notre cas, on prendra M égal au nombre d'exemples. Pour l'estimation du **TFR**, on prendra donc $M = N_c$ avec N_c le nombre de scores "client" et pour le **TFA**, on prendra $M = N_i$, N_i étant le nombre de scores "imposteur". Pour chaque tirage, on va donc choisir M exemples avec remplacement, c'est à dire que les exemples peuvent être utilisés plusieurs fois ou jamais. Pour chaque tirage, on estime ensuite les valeurs du **TFR** et du **TFA** pour une valeur du **seuil** θ . On répète ensuite cette procédure T fois avec T tirages (en général, 1000 tirages sont suffisants). A partir de ces T valeurs de **TFA**(θ) et de ces T valeurs de **TFR**(θ), on peut estimer alors l'intervalle de confiance à **90%** ou à **95%** autour de la médiane. Le problème de cette méthode est le changement dans les protocoles d'évaluation biométrique, les bases d'apprentissage et les listes des tests.

1.7.1.2 Méthode paramétrique

L'hypothèse principale qui est faite lorsqu'on utilise cette méthode d'estimation de l'intervalle de confiance est que les scores sont issus de variables indépendantes et identiquement distribuées même pour les scores d'une personne, c'est-à-dire que tous les scores sont issus de la **même loi** et sont mutuellement indépendants. Cette hypothèse semble très stricte mais est nécessaire pour l'application des théorèmes. Ceci dit, elle n'est généralement pas vérifiée en pratique. Il s'agit ensuite d'estimer les taux d'erreur **TFA**(θ) et **TFR**(θ) pour une ou plusieurs valeurs du **seuil** θ . Pour estimer ces erreurs, l'estimation des fonctions de répartition, appelées "fonctions de répartition empiriques", pour les deux classes "client" et "imposteur" doit être réalisée par une **loi normale**. L'estimation des intervalles de confiance des taux d'erreur **TFA** et **TFR** à un **seuil** θ est obtenue par :

$$\text{TFA}(\theta) \in \left[\widehat{\text{TFA}}(\theta) - \lambda \times \sqrt{\frac{\text{TFA}(\theta)(1-\text{TFA}(\theta))}{N_i}}, \widehat{\text{TFA}}(\theta) + \lambda \times \sqrt{\frac{\text{TFA}(\theta)(1-\text{TFA}(\theta))}{N_i}} \right],$$

(1.5)

$$\text{TFR}(\theta) \in \left[\widehat{\text{TFR}}(\theta) - \lambda \times \sqrt{\frac{\text{TFR}(\theta)(1-\text{TFR}(\theta))}{N_c}}, \widehat{\text{TFR}}(\theta) + \lambda \times \sqrt{\frac{\text{TFR}(\theta)(1-\text{TFR}(\theta))}{N_c}} \right],$$

(1.6)

Avec $\widehat{TFA}(\theta)$, $\widehat{TFR}(\theta)$ les valeurs calculées pour un test donné et N_c les scores "client" et N_i les scores "imposteur" du test. Pour des intervalles de confiance à **90%**, on a $\lambda = 1.645$ et pour des intervalles de confiance de **95%**, on a $\lambda = 1.96$.

Cette méthode non paramétrique sera utilisée dans la suite afin de calculer les intervalles de confiance. Ce choix est justifié pour éviter le changement dans les protocoles d'évaluation de notre base de données et cette méthode est la plus utilisée [57,58,59].

1.8 Nouvelles technologies améliorant la reconnaissance faciale

1.8.1 L'utilisation de la technologie infrarouge

Les systèmes de reconnaissance de visage infrarouges utilisent des ondes infrarouges pour mesurer les radiations thermiques émises dans la gamme du spectre infrarouge. Ces systèmes ne sont pas sensibles à la couleur de la peau, bronzage, usage de produits de beauté, les lentilles de l'œil coloré, la chirurgie plastique et les photos peuvent être prises dans l'obscurité. Néanmoins, le coût extrêmement élevé de cette technologie rend son utilisation prohibitive pour les applications [60,61].

1.8.2 L'utilisation de la technologie 3D

La quasi-totalité des systèmes de reconnaissance de visages travaillent à partir d'images fixes. Pour améliorer cette technique, les chercheurs s'orientent vers l'utilisation du modèle 3D du visage, à l'aide de plusieurs caméras, scanners ou caméras spécialisées. L'objectif est de conserver les avantages de la reconnaissance faciale avec des scores de fiabilité proches de ceux obtenus en empreintes digitales par exemple. [1]

1.8.3 Avantages de la technologie 3D

a) Insensibilité à la variation lumineuse

Lorsque la personne se présente de biais, un logiciel ferait « tourner » numériquement le modèle **3D** pour le faire correspondre à l'image **2D** prise en conditions réelles. De même, ce logiciel modifierait les conditions lumineuses de l'image **3D** pour les rendre plus proches de l'éclairage existant.

b) Invariance de l'angle

La diversité des paramètres 3D fait que le dispositif d'enregistrement en temps réel est en mesure de reconnaître des personnes se trouvant dans un champ correspondant à une rotation de 30° de part et d'autre de son axe.

c) Unicité des images 3D

La quantité de points de mesure et de données biométriques saisies est telle qu'elle permet de distinguer des jumeaux monozygotes.

1.9 Conclusion

Dans ce chapitre, nous avons présenté les technologies utilisées dans les systèmes biométriques pour l'identification de personnes. Cette étude nous a permis de constater que la reconnaissance de visage suscite de plus en plus l'intérêt de la communauté scientifique, car elle présente plusieurs challenges et verrous technologiques. Nous avons mis en évidence les différentes difficultés inhérentes à la reconnaissance automatique de visages, notamment l'invariance à l'illumination, pose et expressions faciales. Enfin, nous avons aussi donné un aperçu sur les techniques et la mesure de leurs performances. Les techniques utilisées aux différentes étapes de la reconnaissance de visages sont détaillées dans les **chapitres 3 et 4**.

Chapitre 2

Etat de l'art de la Reconnaissance de Visage.

2.1 Introduction

De nos jours, il existe des ordinateurs capables d'effectuer des tâches faites par l'homme, qui demandent de l'intelligence. Malgré tout ce progrès, la machine est toujours située à des années lumières du cerveau humain si comparaison a lieu entre les deux, ceci est dû principalement à la complexité des tâches que nous cherchons à automatiser et à notre méconnaissance de cet organe mystérieux qu'est le cerveau et des mécanismes qui régissent son activité (l'activité cérébrale). S'investir dans le domaine de la reconnaissance de visages est sans doute motivé par la multiplicité et la variété des champs d'application (haute sécurité, télésurveillance et contrôle d'accès...). La mise en œuvre d'un système automatique et fiable de reconnaissance faciale est un verrou technologique qui n'est toujours pas résolu.

La reconnaissance du visage automatique à l'aide d'images **2D** a été soigneusement étudiée au cours des dernières décennies. Toutefois, en raison de l'insuffisance intrinsèque de l'information faciale dans les images **2D**, des algorithmes de reconnaissance de visage ont beaucoup de difficultés à la manipulation des variations d'éclairage, changements de pose, expressions faciales, temps de retard et des occlusions. La surface **3D** du visage offre la possibilité de surmonter les difficultés causées par la variation de la tête de pose et d'éclairage. Avec les progrès de l'imagerie **3D** techniques, les données **3D** du visage sont devenues de plus en plus disponibles dans ces dernières années. Il y a maintenant plusieurs bases de données publiques de référence **3D** du visage. Dans ce chapitre, nous présenterons un état de l'art sur les techniques de détection de visage **3D**, nous donnons une taxonomie des approches, des techniques, des algorithmes et des systèmes industriels existants. Nous exposerons aussi les méthodes de reconnaissance de visage **3D**. La reconnaissance de personnes basée sur le visage **3D** présente des limites. Elle ne peut faire une identification optimale en présence d'expressions faciales. Pour cette raison nous tentons d'aborder ce domaine. La reconnaissance des expressions faciales a débuté depuis une vingtaine d'années car les expressions faciales jouent un rôle important dans la conversation humaine. Donc, l'interface homme machine du futur utilisera non seulement la souris, le clavier, mais aussi de nouvelles

sources d'informations telles que les expressions faciales de six cas (colère, dégoût, joie, peur, surprise, tristesse), la posture humaine ou parole. Dans ce chapitre, nous détaillerons en particulier les méthodes basées sur la réduction d'espace. Nous exposerons aussi les méthodes de reconnaissance de visage **3D** expression. Enfin, nous terminerons le chapitre par une synthèse des méthodes et techniques étudiées. Le **but de ce chapitre** est de donner un panorama des méthodes les plus significatives en reconnaissance de visages **2D**, **3D** et **3D expressions**. Tout d'abord, une brève présentation des méthodes les plus populaires utilisées en reconnaissance faciale est proposée puis deux des méthodes les plus connues pour l'extraction des caractéristiques faciales, étape indispensable dans les systèmes de reconnaissance de visages, sont décrites enfin un état de l'art de la fusion phase essentielle aux systèmes multi algorithmes et multi biométriques est réalisé.

PARTIE A : Reconnaissance de visage 2D

2.2 Reconnaissance 2D de visages : état de l'art

Plusieurs méthodes de reconnaissance de visages ont été proposées ces dernières années, suivant deux grands axes: la reconnaissance à partir d'images fixes et la reconnaissance à partir de séquence d'images (vidéo)[62,63,64]. Les travaux portant sur ce domaine, sous différentes conditions d'éclairage, d'expressions faciales et d'orientations, peuvent être répertoriés en deux catégories distinctes suivant qu'elles se rapportent à une approche géométrique ou une approche globale [3].

2.2.1 Les principales techniques de reconnaissance faciale [65]

Dans ce chapitre, nous décrivons brièvement quelques techniques parmi les plus importantes ou les plus populaires utilisées en reconnaissance de visages (voir les résumés de l'état de l'art [53] et [66], ou le livre [1] pour plus de détails). Les approches existantes peuvent être grossièrement divisées en trois groupes : les **approches locales**, les **approches globales** ainsi que les **approches hybrides**. Les principales méthodes de ces trois approches sont décrites dans la suite.

Nous **insistons** notamment sur les **méthodes de réduction de dimension** faisant partie des **approches globales**. Les méthodes de réduction de dimension entrent en effet dans le cadre plus vaste des traitements de données en général (et pas seulement des visages) et de la reconnaissance d'objets. Notons que la plupart des méthodes nécessitent une localisation

précise du visage, un recadrage géométrique (pour que celui-ci apparaisse toujours dans la même position), ainsi qu'une mise à l'échelle pour que le visage ait une taille adéquate.

2.2.1.1 Approches locales

Les **approches locales** de la reconnaissance de visages sont basées sur des modèles et reposent sur un traitement séparé des différentes régions de l'image du visage. Les modèles utilisés reposent sur les connaissances que l'on possède à priori de la morphologie des visages. La plupart du temps, cela implique la détection/extraction de caractéristiques faciales locales.

Brunelli et Poggio [67] proposent une technique qui extrait automatiquement un ensemble de **35 caractéristiques géométriques** d'une image de visage (voir la **figure 2.1(a)**). Ces ensembles de caractéristiques sont ensuite comparés deux à deux via la distance de **Mahalanobis** pour réaliser la reconnaissance.

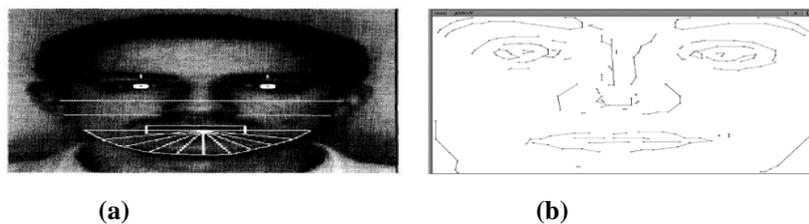


Figure 2.1 (a) Localisation des caractéristiques géométriques utilisées dans [47].
(b) Cartes de contours utilisées dans [68].

Une autre approche géométrique a été proposée par **Takacs** [69]. Des cartes binaires de contour sont extraites des images de visage via un **filtre de Sobel**. La similarité entre deux contours est ensuite calculée en utilisant une variante de la **distance de Hausdorff**. Cette **approche** a été **étendue** par **Gao et al.** [68] qui ont transformé les **cartes de contours** en cartes de lignes de contours (ou **LEM** pour **Line Edge Maps**) contenant des listes de segments (voir la **figure 2.1(b)**). La distance utilisée pour mesurer la similarité est la même que celle de **Takacs**.

L'approche de **Heisele et al.** [70] commence par **détecter la région** contenant le visage dont **dix points caractéristiques** sont extraits. Les zones autour de ces points sont ensuite extraites (voir la **figure 2.2**), et **concaténées** pour former le vecteur caractéristique du visage. La **classification** est finalement réalisée grâce à l'utilisation d'une machine à vecteurs de support (ou **SVM** pour « Support Vector Machine »).

Price et Gee proposent également une méthode [71] se basant sur des **zones extraites** du visage. Ici, **trois régions** sont considérées : une **bande rectangulaire** comprenant les **yeux** et

le nez, une deuxième bande rectangulaire ne contenant que les yeux, et enfin une région contenant le visage entier.

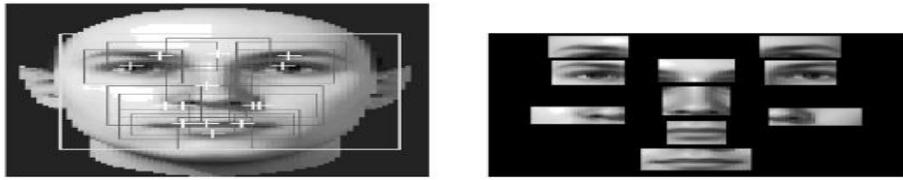


Figure 2.2 Approche proposée dans [70].

Une analyse linéaire discriminante est ensuite appliquée à chacune de ces régions (voir la figure 2.3).

Samaria et al. [72] présentent une approche basée sur les chaînes de Markov cachées (HMM pour Hidden Markov Models). Le visage est segmenté en sous-bandes partiellement recouvrantes, ces sous-bandes étant ensuite concaténées en un vecteur de grande taille ou compressées par DCT. Puis, pour chaque classe (individu), un HMM est créé modélisant la distribution probabiliste des sous-bandes. Les images de visages sont finalement classées en appliquant l'algorithme de Viterbi pour comparer la séquence des sous-bandes de l'image avec les modèles appris. Cette approche a été étendue aux 2D HMM [73].

Perronnin et al. proposent dans [74] une approche basée sur le modèle des 2D-HMM où les expressions faciales et l'illumination sont modélisées indépendamment.

Les approches bayésiennes ont également été explorées via les travaux de Liu et Wechsler [75], où est proposé un cadre bayésien unifiant les méthodes les plus populaires de reconnaissance de visages.

Les machines à vecteur de support (SVM) sont également utilisées pour la reconnaissance faciale par Guo et al. dans [76]. Deux bases y sont utilisés, la base AT&T ainsi qu'une base « maison ». L'approche est comparée à l'ACP.

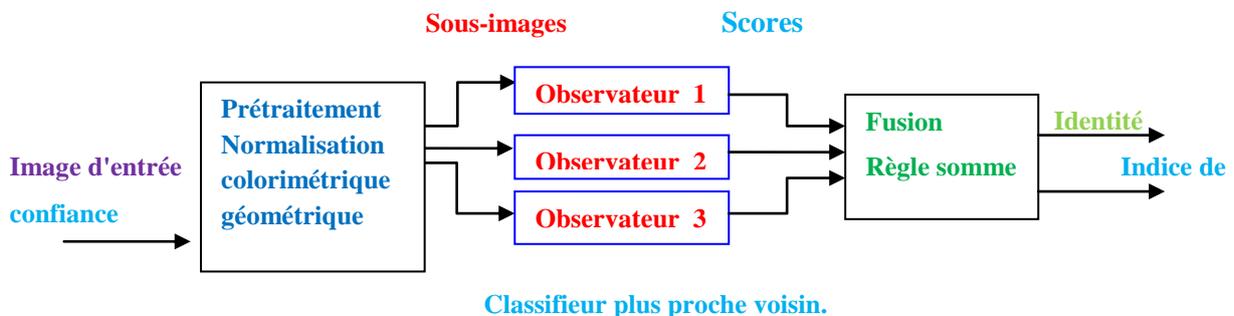


Figure 2.3 Approche de Price et Gee [77].

Les Modèles Actifs d'Apparence (ou AAM pour « Active Appearance Models ») sont présentés par Cootes et al. dans [78]. Ils consistent en la création d'un modèle statistique

d'un visage, voir la **figure 2.4**. Le **modèle** est ensuite déformé pour « **coller** » au **plus près** des **traits du visage**. La **reconnaissance** est effectuée sur le **résidu** calculé correspondant à l'**erreur de prédiction du modèle**.



Figure 2.4 Modèle Actif d'Apparence.

Les **LBP** (pour « Local Binary Patterns ») ont également été utilisés pour la reconnaissance faciale, notamment dans [79]. Le visage est subdivisé en **sous-régions carrées** de taille égale sur lesquelles sont calculées les **caractéristiques LBP**. Les **vecteurs** obtenus sont ensuite **concaténés** pour obtenir le vecteur de caractéristiques final. Des extensions des LBP comme les **MB-LBP** (pour « Multi-Scale Block Binary Pattern ») ont été proposées et appliquées aux visages par Liao et al. [80] (voir la **figure 2.5**).

Le **gros avantage** des **méthodes locales** de reconnaissance de visages est qu'elles peuvent **modéliser facilement les variations** de **pose**, d'**illumination** ou encore d'**expressions** que peut subir un visage. Cependant, elles **nécessitent** souvent le placement **manuel de nombreux points d'intérêts** pour une bonne précision, et sont donc **lourdes à mettre en oeuvre**.



Figure 2.5 Caractéristiques MB-LBP pour un visage, (pour un **masque** de taille $(3, 3)$; $(9,9)$ et $(15,15)$).

2.2.1.2 Approches globales

Les **approches globales** prennent l'image du visage comme un tout et **utilisent** des **techniques d'analyse statistique** bien connues. L'**idée** est généralement de **projeter** l'**image** d'entrée du visage, **préalablement vectorisée**, dans un **espace de plus faible dimension**, où **la reconnaissance est supposée être plus aisée**. La **projection** est souvent **conçue** pour ne sélectionner que **les caractéristiques importantes** et **suffisamment discriminantes** pour différencier les personnes entre elles.

Un des **avantages** des méthodes globales est qu'elles sont **rapides à mettre en oeuvre**, les **calculs** reposant sur des **opérations matricielles** relativement **simples**.

Cependant, étant donné qu'elles considèrent le visage comme un tout, elles sont **sensibles** aux conditions de **luminosité**, de **pose** ou encore d'**expression faciale**.

La plupart de ces méthodes réalisent une analyse de sous-espaces de visage (ou de la variété définie par les visages). Cette analyse découle de la constatation d'un fait relativement simple : la classe des visages réside dans un sous-espace de l'espace de l'image d'entrée. Prenons par exemple une image de taille [100,100] en niveaux de gris. Le nombre de configurations possibles est égal à 25610000. Cependant, parmi toutes ces configurations possibles, seule une petite partie correspond aux visages. L'information contenue dans les images de visages est donc très redondante, la dimension de ces images peut donc être réduite en se concentrant uniquement sur ce qui nous intéresse (les visages). Le sous-espace est souvent appelé espace de visages (ou « facespace »).

Les **méthodes globales** peuvent se décomposer en **deux types de techniques** : les **techniques linéaires** et les **techniques non linéaires**.

1 Techniques linéaires

Les **techniques linéaires réalisent une projection linéaire des visages** (espace dont la dimension est égale à la dimension des images, donc grande) sur un espace de plus faible dimension. Cependant, ces **techniques linéaires** sont **sensibles** aux conditions de **luminosité** notamment, et plus généralement aux **variations non convexes**. **Ainsi, l'utilisation de distances classiques** dans l'espace projeté **ne permet pas** toujours de réaliser une **bonne classification** entre les classes « visages » et « non visages ». **La plus connue** de ces approches est la technique dite des **Eigenfaces** présentée par **Turk et Pentland** dans [4]. Une **ACP** est réalisée sur un ensemble d'apprentissage d'images de visages. Les principaux vecteurs propres résultant de l'**ACP** définissent le nouvel espace. Les images de visages sont ensuite projetées sur cet espace, et les vecteurs obtenus sont utilisés pour la classification.

De nombreux travaux ont été réalisés **sur le choix des vecteurs propres** à retenir pour définir le nouvel espace. Ainsi, **Kirby et al.** [81] proposent un **critère** basé sur l'**énergie des valeurs propres** associées aux vecteurs propres. Les vecteurs propres correspondant aux plus grandes valeurs propres sont retenus jusqu'à ce que la **somme** des valeurs propres **dépasse** un certain **seuil** de l'énergie totale (**90%** dans [81]). **Martinez et al.** montrent dans [82] que les taux de reconnaissance peuvent être améliorés en ignorant les premiers vecteurs propres (ceux dont les valeurs propres associées sont les plus grandes), **ceux-ci encodant** souvent les **variations d'illumination**. Une autre approche bien connue présentée par **Belhumeur et al.** [83] réalise une **Analyse Discriminante Linéaire (LDA)**, elle est ainsi souvent nommée **Fisherfaces**. En

effet, cette technique **consiste** à **maximiser** sur un ensemble d'apprentissage le **critère de Fisher**, à savoir le **quotient** de la **variance inter-classe** par la **variance intra-classe**. Ainsi, **contrairement** à la technique des **Eigenfaces** où la *meilleure représentation (celle maximisant la variance)* est recherchée, le **but** est ici une **meilleure séparation des classes**. Cependant, étant donné que le nombre d'images est souvent inférieur à leur dimension, la matrice de **variance intra-classe** peut être **singulière**, et **son inversion** pose donc **problème**. Ce problème est connu sous le nom de **Small Sample Size Problem**. Des méthodes ont été proposées pour **contourner ce problème**, la plus utilisée étant de *réaliser une ACP au préalable pour diminuer la dimension des échantillons*. De **nombreuses variantes** à ces méthodes linéaires ont été proposées dans la littérature des principales techniques de reconnaissance faciale [84,85,86,87,88].

D'autres techniques linéaires ont également été utilisées pour le calcul de vecteurs caractéristiques :

- l'analyse en composantes indépendantes (**ICA**) dans [89],
- la factorisation de matrices non négatives (**NMF**) dans [90] ou [91],
- l'analyse discriminante bilinéaire (**BDA**) dans [92],
- la technique dite de « Vecteurs communs discriminants » (**DCV**) dans [93].

Certaines méthodes proposées **ne reposent pas sur un seul sous-espace**, mais sur plusieurs, chacun étant caractéristique à une variation [94,95,96,97]. Par exemple, **Pentland et al.** [98] calculent un **sous-espace** pour **chaque orientation** et **chaque échelle** d'un visage ainsi qu'autour de certaines caractéristiques détectées (voir la **figure 2.6**). Un **nouveau visage** est ensuite **identifié** en le **projetant** sur **tous les sous-espaces** et en **sélectionnant** celui étant le **plus proche d'un vecteur de la galerie**.

Cette **technique** reposant sur plusieurs sous-espaces est **généralisée** dans [99] où des **tenseurs** à **quatre dimensions** correspondant à la **classe**, la **pose**, les conditions d'**illumination** et l'**expression faciale** sont **calculés** pour une base d'apprentissage donnée et permettent ainsi une **meilleure robustesse de la classification**.



Figure 2.6 Caractéristiques autour desquelles est réalisée une ACP dans [98].

Bien que ces **méthodes linéaires** soient **assez efficaces**, elles **manquent de précision** dès lors que les images de **visages subissent** des **transformations non linéaires**. Une *simple modification de la luminosité transforme celui-ci de façon non linéaire* étant donné la complexité de la forme.

2 Techniques non linéaires

Des techniques globales **non linéaires** ont été développées, souvent à partir des techniques linéaires. Ainsi l'Analyse en Composantes Principales à **Noyaux** (ou « **Kernel-PCA** ») [100,101] et l'Analyse Discriminante Linéaire à Noyaux (ou « **Kernel-LDA** ») [102] utilisent la notion mathématique des noyaux pour étendre les techniques linéaires que sont l'ACP et la LDA (voir **annexe A.3.1**).

D'autres techniques non linéaires ont également été utilisées dans le contexte de la reconnaissance faciale :

- le MultiDimensional Scaling (**MDS**) dans [103] ou [104],
- l'**Isomap** dans [105],
- les **Diffusion Maps** dans [106],
- le **Local Linear Embedding (LLE)** dans [107] ou [108],
- les **Laplacian EigenMaps** dans [109], [110] ou [111],
- le **Hessian LLE** dans [112],
- le Local Tangent Space Analysis (**LTSA**) dans [113],
- les approches **neuronales** dans [114] ou [115] (**autoencodeurs**), dans [116] (cartes de **Kohonen**), et dans [117] (réseaux de neurones **convolutionnels**).

L'utilisation de ces méthodes de **projection** de l'**espace des images** sur l'**espace de caractéristiques est non linéaire** et permet ainsi dans une certaine mesure de **réduire** la **dimension** des images de **meilleure façon**. Cependant, bien que ces méthodes **permettent** souvent l'**amélioration des taux de reconnaissance** sur des jeux de tests donnés, elles sont **trop flexibles pour être robustes** à de **nouvelles données**, contrairement aux méthodes linéaires.

2.2.1..3 Approches hybrides

Les méthodes **hybrides** résultent de l'**association** des **méthodes locales** et des **méthodes globales**. Elles combinent la détection de caractéristiques locales avec l'extraction de caractéristiques globales. Ces techniques essaient finalement de **tirer partie** des **avantages des deux types** de méthodes citées plus haut.

L'approche appelée Analyse en Composantes Locales (**LCA** pour Local Component Analysis) a été proposée par **Penev et Atick** [118]. Plusieurs **ACP** (Analyses en Composantes Principales) sont réalisées pour **extraire** différentes **caractéristiques locales** voir la **figure 2.7**. Celles-ci sont **ensuite combinées** et une procédure **minimisant l'erreur de reconstruction** avec une **contrainte parcimonieuse** permet de **réaliser la reconnaissance**. L'approche dite de l'Elastic Bunch Graph Matching (**EBGM**) a été proposée par **Wiskott et al.** [119].

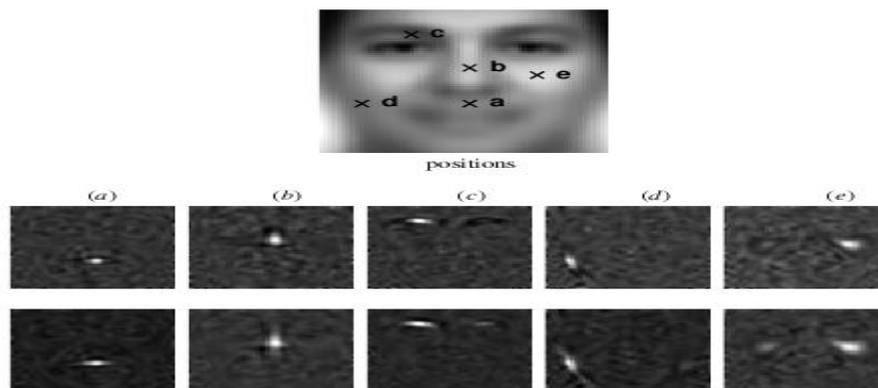


Figure 2.7 Approche Local Component Analysis. [118]

Les visages sont représentés par des Face Bunch Graph (**FBG**), où **chaque nœud** du graphe correspond à une certaine **caractéristique faciale** (l'œil droit ou gauche par exemple), voir la **figure 2.8**. A chaque **nœud** du graphe est **associée l'apparence du voisinage** de la caractéristique **via un jet**. Les **jets** représentent l'ensemble des **40 coefficients** issus de la **convolution** du **voisinage** de la caractéristique par un **filtre de Gabor** spécifique voir la **figure 2.9**. Les **arêtes** du graphe sont pondérées par la **distance relative** des **caractéristiques adjacentes**. Une fois que le **graphe est créé pour chaque personne** de la **base d'apprentissage**, un **algorithme spécifique** de **mise en correspondance** permet d'**identifier une personne inconnue**. L'algorithme essaie itérativement de faire correspondre le graphe créé à chaque graphe de la base d'apprentissage en **minimisant** une **fonction de coût** prenant en compte à la fois une **mesure de similarité géométrique** ainsi qu'une **mesure de similarité de l'apparence modélisée via les jets**.

Perlibakas présente plus récemment l'algorithme **LogGabor PCA** dans [120].

Une **convolution** par des ondelettes de **Gabor orientées** est réalisée autour de **certain points caractéristiques** du visage. Les **vecteurs** ainsi créés contiennent à la fois la **localisation** ainsi que les **amplitudes des énergies locales**. Une **Analyse en Composantes Principales** est ensuite **réalisée** afin de réduire la dimension de ces vecteurs.

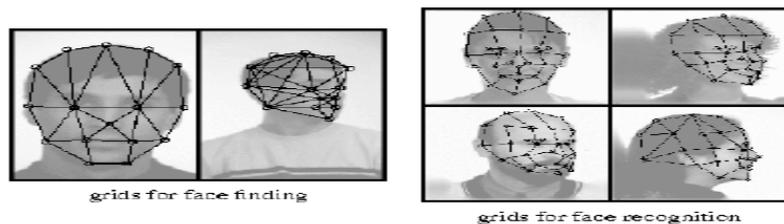


Figure 2.8 Graphe appliqué aux visages pour l'approche EBM. [119]

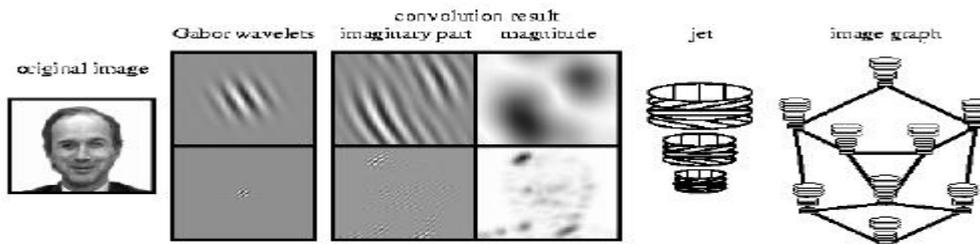


Figure 2.9 Création des jets pour l'approche EBM. [120]

Pentland et al. présentent dans [98] l'approche dite des **espaces propres modulaires** (Modular Eigenspaces) Cette technique **réalise** une Analyse en Composantes Principales et une classification sur des **régions distinctes** du visage, comme les yeux, le nez, la bouche ou encore le visage entier. La zone de la **bouche** subit de **grosses déformations** dues aux **expressions faciales**, ainsi l'ajout de cette région au processus entier fait **décroître** les **taux de reconnaissance**.

L'approche proposée par Cootes et al. [121] met en oeuvre les **Modèles Actifs d'Apparence** (AAM pour Active Appearance Models). Cette méthode **modélise** indépendamment la **forme** et la **texture** d'un visage en appliquant une Analyse en Composantes Principales. Les **vecteurs obtenus** pour la forme et la texture sont ensuite utilisés pour la **reconnaissance**. Un **nouveau visage** qui doit être identifié est **adapté** au modèle par un **processus d'optimisation** itératif. Les paramètres de forme et de texture obtenus sont ensuite comparés à ceux de la base. **Latinis et al.** [122] appliquent cette méthode pour la première fois à la reconnaissance de visages. **Edwards et al.** [123] proposent des méthodes basées sur les **AAMs** pour la reconnaissance faciale.

2.2.1.4 Récapitulatif 2D

Nous avons présenté les principales approches utilisées dans la littérature pour la reconnaissance faciale automatique. Les méthodes peuvent principalement se classer en deux catégories : les **méthodes globales** et les **méthodes locales** (les méthodes hybrides essayant de faire un lien entre ces types d'approches).

Les **méthodes globales** présentent un certain nombre d'**avantages** :

- Le problème de la reconnaissance faciale automatique est transformé en un problème d'analyse de sous-espaces de visages, pour lequel de nombreuses méthodes statistiques existent.
- Les méthodes globales sont souvent applicables à des images basses résolutions ou de mauvaises qualités.

Certains **inconvénients** se posent cependant avec les méthodes globales :

- Il est nécessaire de disposer de suffisamment de données représentatives des visages.
- Il n'y a pas d'a priori sur le physique d'un visage.
- Ces méthodes ne sont robustes qu'à des variations limitées (pose, illumination, expression).

De la même manière les **méthodes locales** présentent certains **avantages** :

- Le modèle créé possède des relations intrinsèques bien définies avec les visages réels.
- Les modèles créés peuvent prendre en compte explicitement les variations telles que la pose, l'illumination ou les expressions. La reconnaissance est ainsi plus efficace dans le cas de fortes variations.
- La connaissance a priori sur les visages peut être intégrée aux modèles afin d'améliorer leur efficacité.

Les méthodes locales présentent cependant quelques **inconvénients** :

- La construction du modèle, reposant souvent sur la détection de points caractéristiques faciaux, peut être laborieuse.
- L'extraction des points caractéristiques peut être difficile dans le cas de variations de pose, d'illumination, d'occlusion . . .
- Les images doivent être de relativement bonne qualité, et/ou être de résolution suffisante afin de pouvoir extraire les points caractéristiques.

Le **tableau 2.1** résume quelques résultats obtenus via des algorithmes classiques de reconnaissance de visages sur certaines bases de données de la littérature. Il met également en exergue la difficulté de comparer les résultats entre les différentes approches. En effet, de **nombreux paramètres** doivent être pris en compte afin de bien appréhender l'efficacité d'une méthode : la **base de données utilisée**, la **taille des images**, la **présence de variations** dans les échantillons ainsi que le **nombre d'images** utilisées pour l'enrôlement et/ou les tests.

Afin de comparer les algorithmes, certains concours ont eu lieu sur des bases de données bien définies, munies de protocoles clairs. Ainsi, le challenge FRGC [124] a permis la comparaison de nombreuses méthodes issues d'entreprises ou de laboratoires internationaux.

Tableau 2.1 Comparatif de plusieurs méthodes de l'état de l'art du SRV 2D.

Réf.	Méthodes	BDD	Taille Images	Nb. Images	Time lapse	Taux (%)	Expression	Ill.	Position
[303]	Direct LDA	ORL	112 -92	200-200	N	90:8	O	O	O
[194]	DF-LDA	ORL	112 -92	200-200	O	96		O	N
		UMIST	112 -92	160-415	N	98		N	N
[59]	DCV	Yale	126-152	150-150	N	97:33		O	N
		AR	229-299	350-350	O	99:35			
[30]	ICA	FERET	60-50	425-421	O	89	O	N	N
[190]	PDBNN	SCR	80-20	320-1280	N	100	O	O	O
		FERET	200-200		N	99	O	O	N
		ORL			N	96		O	O
[98]	RBF	ORL	160-120	300-300		98:1	O		O
[171]	HMM	FERET	128-128	500-500	N	97	O	N	N
[192]	Gabor EFM	FERET	128-128	200-100	N	99	O	N	N
		ORL	128-128	200-200	N	100	O	N	O
[296]	EBGM	FERET	256-384	250-250	N	80	O		O
[116]	WPA	MIT	480-640	155-155		80:5	O	O	
		FERET	256-384	200-400		89			
[271]	IFS	ORL	112-92	200-	N	95			
[92]	IFS	MIT	480-640	90-90		90			O
[63]	PCA	UND		166-166	N	98	O	O	N
[266]	PCA	Equinox	99 -132	770-2310	O	93	O	O	N
[50]	Th-Spectrum	Equinox		225-2500		86:8	O		O
[280]	Hyper spectral	Propriétaire		200-1200	O	92	N	O	N
[114]	LEM	Bern		40-160	N	72:09	O		O
		AR		112-336		86:03		O	N
		Yale		150-150		85:45		O	N
[157]	ICA	AR	46 -56	1685-1490		98		O	O
[186]	LDA/GSVD	CMU_PIE		68-1360	N	99:53		O	N
	LDA/QR	YaleB/Pose00		80-432	N	98:03		O	N
[117]	Cone Models	YaleB	36-42	450-4050	N	97	N	O	O

[220]	Sous-espaces	ATR		2821-804	N	98:7	N	N	O
[119]	EigenLights	CMU-PIE		5304-5304	N	36	N	O	O

La colonne **Nb. Images** indique le nombre d'images utilisées pour l'enrôlement et le nombre d'images utilisées pour les tests, la colonne **Time lapse** indique si les images d'enrôlement et de tests ont été capturées avec un intervalle significatif, les colonnes **Expr.**, **Ill.** et **Pose** indiquent si les images possèdent des variations d'expression faciale, d'illumination ou de pose (**O** pour Oui, **N** pour Non). [65]

PARTIE B : Reconnaissance de visage 3D

2.3 Reconnaissance de visage 3D

2.3.1 Systèmes d'acquisition

Nous avons passé en revue les méthodes de reconnaissance 2D de visages. Malgré les avancées réalisées ces dernières années, les techniques de reconnaissance 2D de visages **robustes** aux différents facteurs de **variabilité** (éclairage, pose, occultation) sont **loin d'être développées**. La reconnaissance **3D** de visages constitue une **alternative prometteuse** pour surmonter ces problèmes, surtout depuis l'apparition de dispositifs d'acquisition 3D performant. L'**avantage principal** des approches basées modèle 3D réside dans le fait que le **modèle 3D conserve** toutes les informations sur la **géométrie de visage**, ce qui permet d'avoir une représentation réelle de ce dernier. Dans cette section, après avoir rapidement évoqué les systèmes d'acquisition **3D**, nous présenterons les travaux récents sur la reconnaissance **3D** de visages. Cependant, les **algorithmes** de reconnaissance de visage en **3D** sont généralement plus **complexes** [19]. De plus, étant donné que les modèles de visage 3D décrivent la géométrie du visage, ils sont **plus sensibles** que les images 2D de texture aux **expressions faciales**.

Notre contribution est de réduire la dimensionnalité des données de départ en associant les modèles de visage 3D au domaine 2D à l'aide de méthodes, non rigides, d'appariement conforme. L'existence de modèles 2D représentant les visages permet alors d'utiliser les techniques précédemment développées dans le domaine de la reconnaissance de visages en 2D. Dans nos travaux, nous utilisons les cartes conformales de visages 3D.

Différents systèmes d'acquisition sont utilisés pour la reconstruction 3D de visages. De nombreux travaux de recherche visent le développement des méthodes d'acquisition 3D. 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**. [41].

Comme nous l'avons déjà cité les techniques de reconnaissance 3D de visages peuvent être regroupées en trois catégories principales : **approches basées modèle**, **approches 3D**, et **approches multimodales 2D + 3D**. C'est dans ce contexte que s'inscrivent nos travaux.

Dans ce qui suit, nous présenterons un état de l'art sur les techniques de reconnaissance de visage **3D** et nous terminerons le chapitre par une synthèse des méthodes et techniques étudiées.

2.3.2 La reconnaissance du visage 3D basée sur le profil

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 2.10**) 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 [147].



Figure 2.10 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 [147].

2.3.3 Approches basées sur les caractéristiques géométriques faciales 3D

Dans ce cas l'étude est concentrée plus particulièrement sur les approches basées sur les courbes faciales et aux approches qui utilisent les techniques de sélection des caractéristiques, pour optimiser les performances de reconnaissance:

1 Approches basées sur les courbes : les méthodes d'extraction des courbes faciales 3D. Ces courbes peuvent être explicitement analysées et comparées. Plusieurs représentations

de courbes ont été utilisées dans la littérature. **Daoudi et al.** [148] ont utilisé les courbes planes de niveaux. Les auteurs ont utilisé l'analyse de forme des courbes planes afin d'analyser les déformations du visage. En revanche, aucune étude n'a été faite pour la sélection des courbes pertinentes. Dans [149] les mêmes auteurs ont utilisé la distance géodésique pour extraire les courbes de niveaux 3D. Selon l'étude de **Bronstein et al.** [150], les changements de la distance géodésique dus aux expressions faciales est négligeable comparée aux changements de la distance euclidienne. Dans [151], les auteurs ont utilisé des points échantillons prélevés de l'intersection entre les courbes de niveaux et les profils radiaux, qui ont comme origine le bout du nez, ensuite ils ont calculé les distances euclidiennes entre les points correspondants des différents visages. Cette approche, permet de sélectionner des sous-ensembles de courbes faciales pour une mise en correspondance efficace du visage. **Drira et al.** [152] ont exploré l'utilisation de la forme de différents types de nez. Les auteurs extraient les courbes iso-géodésique, ensuite, ils ont construit les différents chemins géodésiques dans une variété Riemannienne de courbes fermées. La même approche permet de faire des calculs statistiques (moyennes) pour une recherche efficace dans des bases de visages 3D. Dans [153], les mêmes auteurs ont proposé une approche similaire afin d'analyser la forme des courbes radiales 3D. **Mpiperis et al.** [154] proposent une paramétrisation polaire, au sein de cette nouvelle paramétrisation intrinsèque, les attributs du visage sont invariants aux déformations isométriques. Quand la bouche est ouverte, cette paramétrisation n'est plus valable. Les auteurs proposent donc de déconnecter les lèvres sur toutes les surfaces faciales et changent cette paramétrisation en se basant sur trois points de références.

2 Approches basées sur la sélection des caractéristiques : Plusieurs méthodes ont été proposées pour analyser le pouvoir discriminant des différentes parties du visage. **Daniyal et al.** [155] ont proposé un algorithme dans lequel le visage est représenté comme un vecteur de distance entre les points d'intérêts (landmarks) des visages. Ils ont sélectionné les points par brute-forcing des combinaisons possibles des points utilisés/non-utilisés et ils ont comparé les taux de reconnaissance. Ils ont conclu que la meilleure sélection correspond aux points situés autour des yeux et du nez. **Kakadiaris et al.** [156] utilisent un modèle de visage annoté qui est déformé élastiquement pour s'adapter à chaque visage, ainsi lui permettre automatiquement l'annotation de ses différentes zones comme le nez, yeux, bouche, etc. Pour résoudre le problème des expressions faciales, les auteurs classent les visages en utilisant les coefficients d'ondelettes qui représentent les zones non affectées par les expressions, telles

que les yeux et le nez. Cependant, le meilleur taux de reconnaissance est atteint pour le visage entier, ce qui implique que le rejet de certaines zones, affectées par les déformations du visage, conduit à la perte de l'information discriminante. Dans [50], un appariement de plusieurs régions du nez a été proposé. Les premiers résultats montrent une amélioration par rapport à une seule grande région frontale. **Faltmier et al.** [157] proposent de recalcr, en utilisant l'algorithme **ICP**, différentes parties du visage puis de les fusionner pour améliorer les performances biométriques. Récemment, **Wang et al.** [158] calculent la carte de différence de forme signée (**SSDM**), calculée entre deux visages alignés, comme une représentation intermédiaire pour la comparaison de formes. Basé sur le **SSDM**, trois types d'opérateurs ont été utilisés pour coder à la fois la similarité locale et le changement des caractéristiques de différentes formes du visage. Les caractéristiques locales les plus discriminantes ont été sélectionnées, par **AdaBoost**, en les utilisant comme des **classifieurs faibles**. Les **caractéristiques** individuelles sont du type : **Haar-like**, **Gabor** et local binary pattern (**LBP**). En utilisant les caractéristiques similaires, **Li et al.** dans [159] proposent de concevoir un dispositif commun et un système de classement afin de **collecter** différentes **caractéristiques géométriques** bas-niveau, telles que la courbure, et les classer selon leurs sensibilités aux expressions faciales. Pour les approches basées sur les courbes, l'analyse de la forme du visage, devient l'**analyse des courbes faciales 3D**. Cependant, est ce que l'utilisation de toutes les courbes du visage donne de meilleures performances ? Y a-t-il des courbes faciales plus pertinentes que d'autres ?

3 Approche visages 3D par combinaison de courbes et techniques de sélection des caractéristiques : On propose un nouvel algorithme de reconnaissance du visage **3D** qui combine la représentation des visages **3D** par une collection de courbes et de techniques de sélection des caractéristiques. La **figure 2.11** illustre les différentes étapes de cette approche. Après l'acquisition des visages **3D** on effectue quelques prétraitements afin d'extraire la partie informative du visage et corriger certaines imperfections telles que les points aberrants, les trous ou encore la présence de parties indésirables, comme les cheveux et les habits. On extrait les **courbes faciales radiales** et de **niveaux**. Puis, selon la nature des courbes (fermées ou ouvertes), on utilise l'un des cadres géométriques pour analyser les formes des courbes faciales [44]. Cette approche géométrique est nécessaire pour comparer et analyser la forme des courbes faciales (ouvertes et fermées).

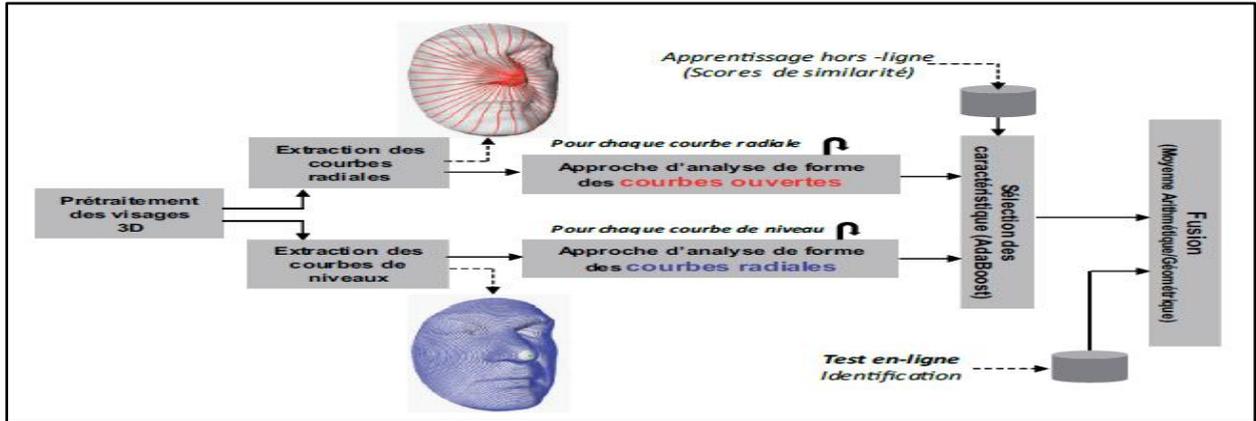


Figure 2.11 Aperçu de l'approche de reconnaissance du visage 3D. [144]

Dans l'étape d'apprentissage (hors-ligne), de sélection des caractéristiques géométriques, l'algorithme **AdaBoost** est utilisé pour sélectionner les courbes les plus pertinentes pour la reconnaissance du visage [160] sur un ensemble de scores de similarités, calculés par rapport aux courbes, pour trouver les courbes du visage les plus discriminantes. Dans l'étape de reconnaissance (en ligne), les scores de similarité des courbes sélectionnées sont cumulés pour produire un score final. Pour plus de détails sur l'extraction des courbes faciales, des courbes radiales, des courbes de niveaux voir [39].

4 Exemple de représentation utilisant les courbes faciales : Il y'a toutes les définitions nécessaires pour proposer cette représentation mathématique d'une surface faciale. On considère un chemin $\alpha : [0, L] \rightarrow (\mathbb{R}^3 \times \mathbb{C} \times \mathbb{R}_+)$, tel que $\alpha(\lambda) = (p_\lambda, v_\lambda, l_\lambda)$ pour $\lambda \in [0, L]$. On fixe le point de départ de α au point défini par le triplet $(r, v_0, 0)$. On définit H comme l'ensemble de tous les chemins, c'est à dire : $H = \{\alpha : [0, L] \rightarrow (\mathbb{R}^3 \times \mathbb{C} \times \mathbb{R}_+) \mid \alpha(0) = (r, v_0, 0)\}$ Les surfaces faciales sont représentées et analysées comme des éléments de H. La **figure 2.12** montre des exemples d'une surface faciale représentée en utilisant un grand nombre de courbes faciales. L'image supérieure montre la surface originale S tandis que la rangée inférieure montre un rendu de c_λ pour un nombre croissant de λ_s de gauche à droite. En pratique, on doit se contenter d'un nombre fini de courbes dans la représentation. [149].

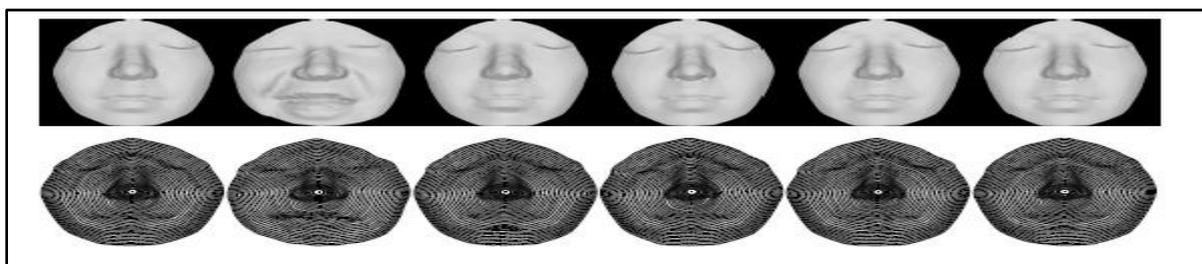


Figure 2.12 Les surfaces faciales de la même personne sous différentes expressions faciales, et leurs représentations dans l'espace H. [149]

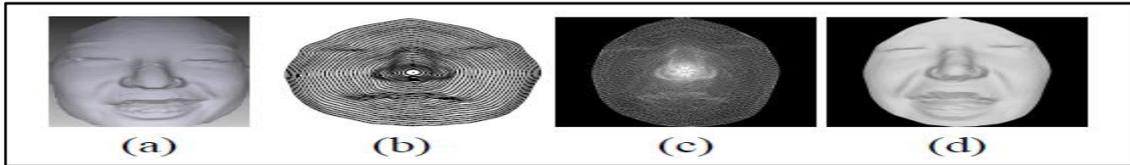


Figure 2.13 Exemple de reconstruction : (a) la surface faciale originale, (b) Sa représentation comme une collection de courbes dans l'espace H, (c) la triangulation, et (d) la surface reconstruite. [149]

2.3.4 Étapes de suppression d'occultation

Dans un contexte de reconnaissance, il s'agit de comparer un scan **3D** de test (dit probe) à un ou plusieurs scans de la base de données (dit galerie) et fournir une distance qui servira pour la décision. Comme illustré dans la **figure 2.14**.

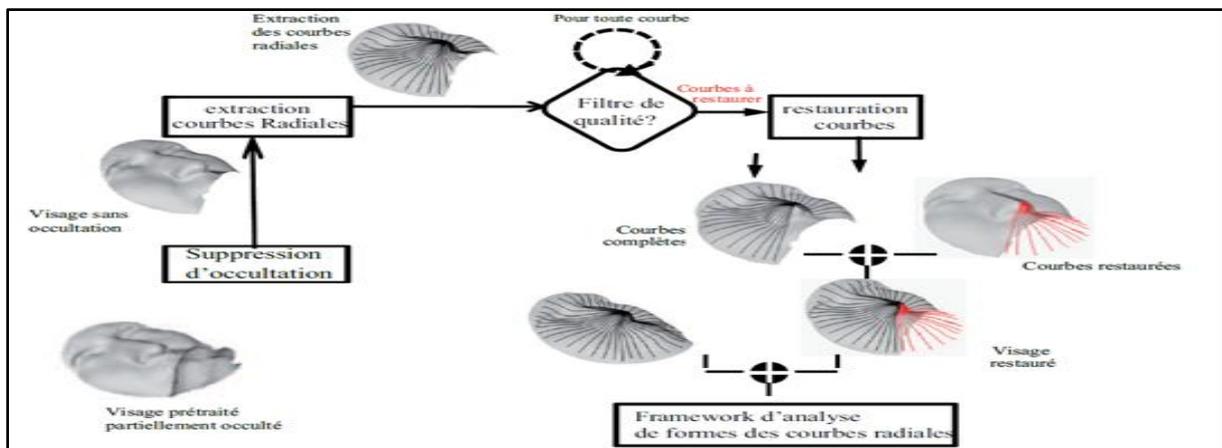


Figure 2.14 Différents étapes de l'approche : de la suppression d'occultation à la reconnaissance passant par la restauration. [153]

Cette approche est basée sur la représentation des modèles de visages **3D** par les courbes radiales émanant du bout du nez. La détection du bout du nez en présence d'occultation est faite manuellement pour le moment. Le modèle **3D** du visage est premièrement prétraité, ensuite un module de détection et de suppression d'occultation est appliqué. Les courbes radiales sont ensuite extraites tout en inspectant leur qualité d'une façon similaire. Une courbe de bonne qualité est une courbe de longueur suffisante et ayant une seule composante. La contribution de cette approche consiste à restaurer les courbes détectées par le module de qualité comme étant courbes manquantes. Enfin, les courbes restaurées et les courbes intactes (décidées complètes par le module de qualité) constituent ensemble une approximation du visage **3D** et représentent l'entrée du module de comparaison des visages à travers la comparaison des courbes du même indice [153].

1 Détection et suppression d'occultation : Ceci est accompli en alignant le scan donné avec le scan de la galerie. La procédure de mise en correspondance entre le modèle de galerie et le scan probe est effectuée en utilisant l'algorithme **ICP** récursif, implémenté comme suit.

A chaque itération, les différents points du scan sont mis en correspondance avec le scan modèle, ensuite les points jugés assez éloignés de leurs correspondants sont éliminés. La décision se fait en fixant un seuil de distance entre les points correspondants. Ce seuil a été déterminé empiriquement et fixé pour tous les visages de test. L'algorithme s'arrête lorsqu'aucun point n'est jugé extérieur à la surface faciale. Quelques exemples sont présentés dans la **Figure 2.15** de cette implémentation sur des visages 3D représentant différents types d'occultation, lunette, cheveux, main sur l'œil, main sur la bouche.

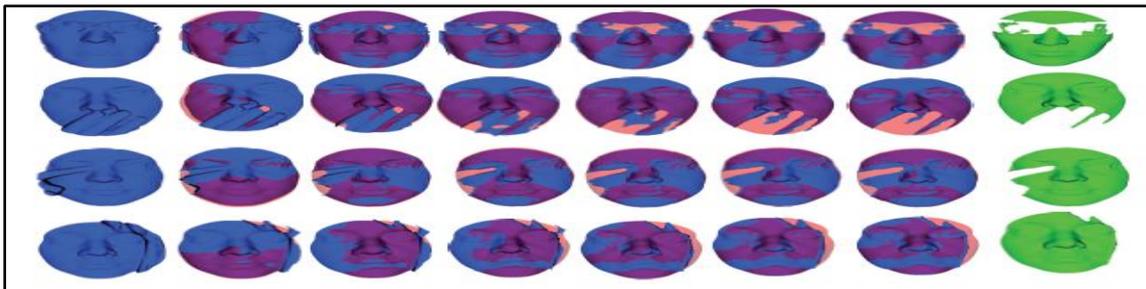


Figure 2.15 Suppression graduelle des parties occultant le scan facial 3D. [153]

De gauche à droite, chaque visage 3D illustre le modèle 3D aligné avec le modèle de galerie à différentes itérations de l'algorithme. On remarque que l'alignement s'améliore au fur et à mesure que l'on supprime des points extérieurs à la surface faciale. Ces points, labélisés comme extérieurs et colorés en rose augmentent tout au long de l'algorithme.

2 Restauration de données incomplètes et reconnaissance

La restauration de visage 3D avec des données manquantes est très importante. On a présenté dans la section précédente une étape d'apprentissage, où une base de projection est construite. Etant donné une courbe partiellement occultée (avec des données manquantes), il s'agit maintenant de prédire les données manquantes de la courbe, sachant la base de projection appropriée voir **figure 2.16**.

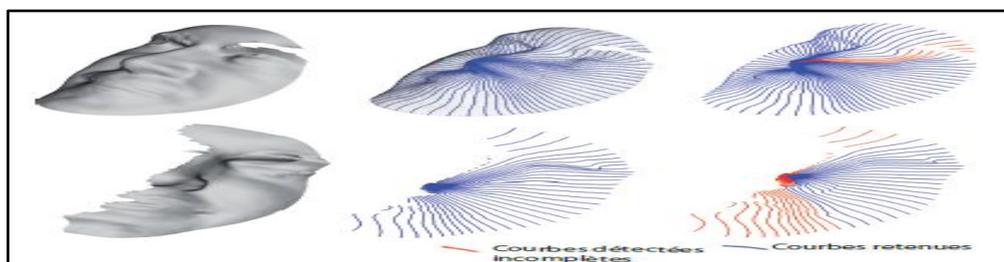


Figure 2.16 Module de qualité [153]

Exemples de détection des courbes incomplètes (en rouge) et les courbes complètes (en bleu).

3 Restauration des visages 3D

Après avoir identifié et restauré les courbes incomplètes, elles constituent avec les courbes complètes intactes la surface faciale restaurée. Trois exemples de la procédure complète sont présentés dans la **figure 2.17**, trois visages avec des occultations externes sont illustrés par **(a)**. Le résultat de la détection et la suppression des parties occultantes est illustré dans **(b)**. Finalement, le résultat de la restauration visages est illustré dans **(c)**. Afin de valider qualitativement la procédure de restauration, on applique la restauration à l'exemple illustré en haut de la **figure 2.18**.

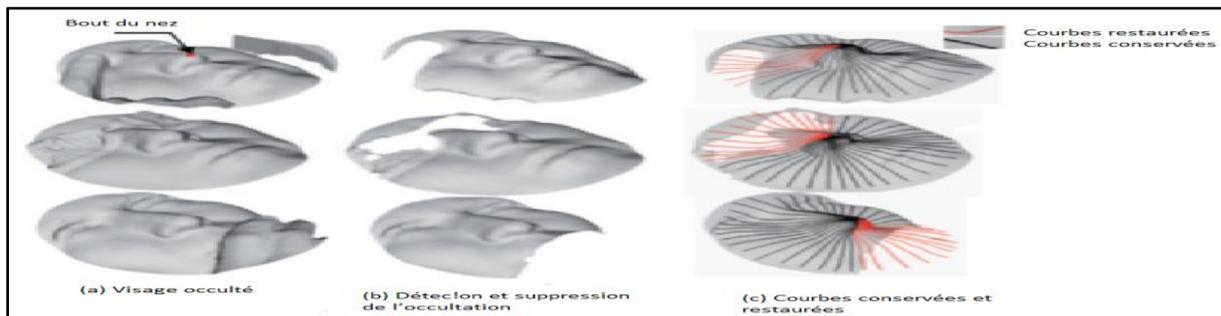


Figure 2.17 (a) Visages avec occultations externes, (b) visages après détection et suppression des parties occultantes, (c) estimation des parties occultées par modèle statistique sur l'espace de forme des courbes. [153]

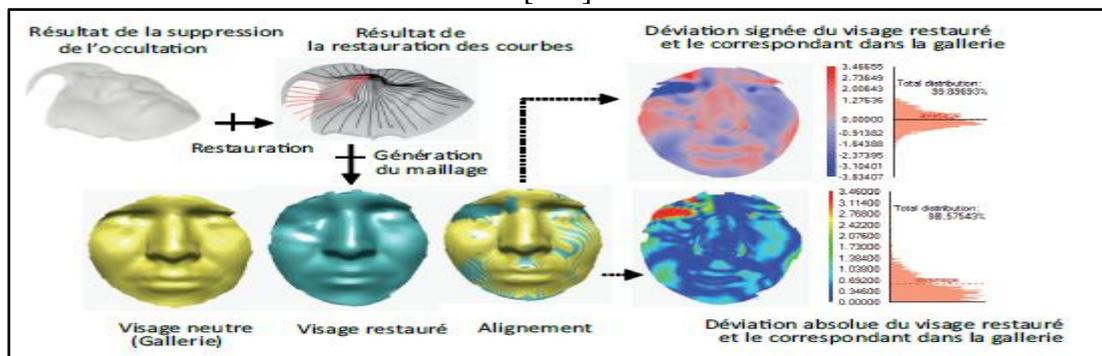


Figure 2.18 Déviation spatiale (absolue et signée) entre un visage restauré et le visage correspondant dans la galerie [153].

2.3.5 Utilisation de la couleur et profondeur pour la reconnaissance faciale

Une technique de reconnaissance faciale est développée sur la base de la profondeur et de 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 algorithmique 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 2.19**. La technique de reconnaissance est présentée basée sur la mise en œuvre de l'algorithme principal d'analyse en composantes 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** [161].

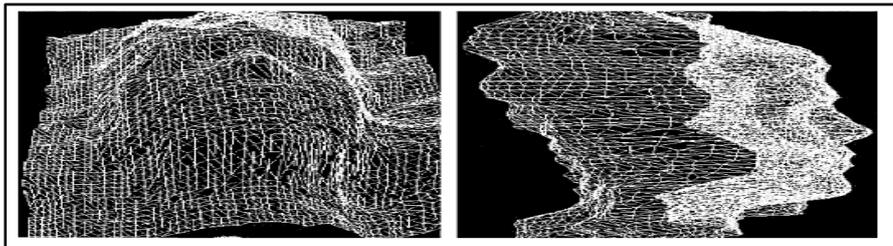


Figure 2.19 Modèle 3D de visage humain constitué de triangles. [161]

2.3.6 Modélisation 3D de surface appliquée à la reconnaissance faciale

Les systèmes de reconnaissance automatique du visage **3D** (AFR), ont attiré beaucoup d'attention ces dernières années en raison des applications intéressantes qui en découlent et à la prolifération et de performances des scanners **3D** voir la **Figure 2.20**. Ces systèmes de reconnaissance faciale en **3D** sont applicables si elles sont robustes contre les variations de position et des gestes qui peuvent produire un sujet alors qu'il / elle est en interaction avec le système.

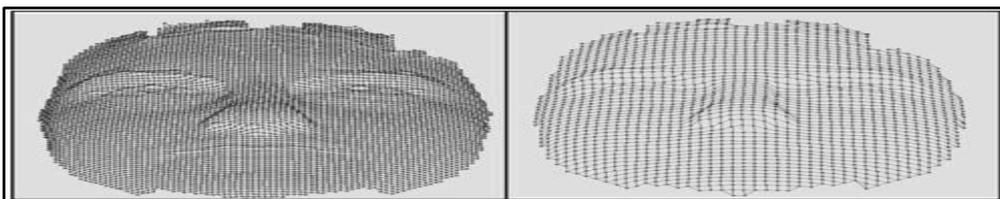


Figure 2.20 (À gauche) en maille originale produite par le scanner **3D** et (à droite) de la même maille à 1-4 la résolution originale [162].

Certains des **problèmes** et des **limites** de systèmes de reconnaissance faciale en **3D** sont les suivants:

- Il n'y a pas beaucoup de systèmes qui prennent en compte la variabilité visage en position et les expressions faciales comme la demande réelle des applications,
- Le manque de bases de données **3D** du visage avec une variabilité importante dans la pose et les gestes.

Certaines méthodes ont été proposées pour résoudre certaines variations déterminées non contrôlées, mais ils ne fonctionnent pas bien dans des conditions arbitraires. **Bowyeretal** en **2006** a travaillé avec des images tridimensionnelles trois visages (**3D**) et confirme que ça présente certains **avantages** intéressants **par rapport** aux travaux de **Hsu** et **Jain** en **2001** sur des images **2D**:

- plus géométrique dans la formation, l'information des images **2D** est formé par des projections d'objets **3D**;
- les caractéristiques mesurées à partir des données réelles en **3D** ne sont pas affectés par l'ampleur et la rotation de la face,
- si le système de reconnaissance du visage **3D** ne tient pas compte des informations de texture, la reconnaissance n'est à l'abri de l'effet des variations d'éclairage.

Deux composantes principales dans l'architecture de reconnaissance des visages ce sont:

- modelage visage
- la reconnaissance ou le régime correspondant.

La modélisation du visage transforme les maillages originaux de surface en **3D** dans un ensemble de fonctionnalités pour mieux représenter le visage dans les **AFRs**. Matching schéma implique la sélection (de la formation) d'un classificateur que l'utilisation du caractéristiques extraites à partir d'un visage de test les compare avec les traits du visage de personnes différentes dans une base de données pour déterminer l'identité du sujet de test voir la **figure 2.21**. Bien que le système d'appariement peut être efficacement mis en œuvre en utilisant des techniques d'**apprentissage** machine standard comme les **réseaux de neurones** ou les machines à vecteurs de support (**SVM**), de trouver de bons traits du visage en **3D** (modèles) pour la reconnaissance, en général, un problème complexe [162].

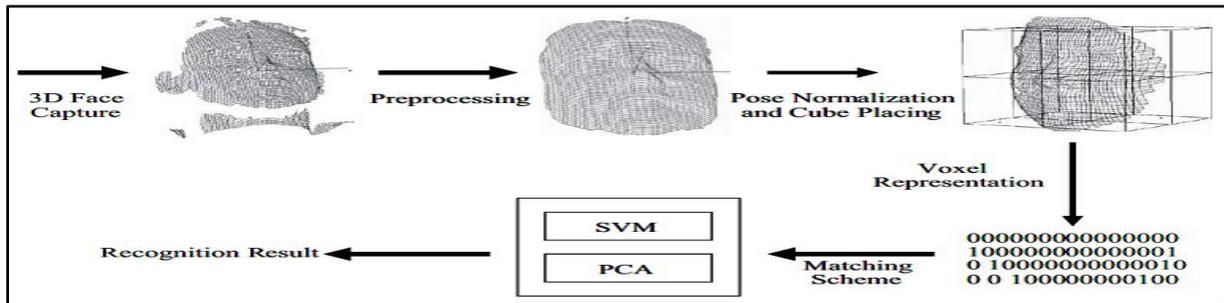


Figure 2.21 Architecture du SRV 3D avec deux génériques schémas de correspondance.

Ce travail propose et évalue plusieurs approches de **modélisation 3D** du visage. Pour valider les modèles, deux schémas de correspondance ont été utilisée:

- 1) Des Machines à Vecteurs de Support (**SVM**),
- 2) Analyse en Composantes Principales (**PCA**) en combinaison avec un classificateur

à **distance euclidienne**.

Ce travail est appliqué à une base de données **3D** appelée **Gavab DB** car il a présenté un degré plus élevé de variabilité parmi les images de chaque individu, en particulier celles liées aux expressions faciales. Les résultats de la reconnaissance obtenus avec le **profil** et la **profondeur** des modèles de cartes (à la fois sur un environnement contrôlé et sur un des environnements non contrôlés) en utilisant **SVM** et de la **PCA** que les régimes correspondants. D'après ces travaux on voit que les combinaisons de profils proposés fournissent des **résultats meilleurs** que la reconnaissance des **cartes de profondeur** correspondant à ces mêmes résolutions voir la **figure 2.22** et **figure 2.23**. [162].



Figure 2.22 Exemples de profils de combinaison.

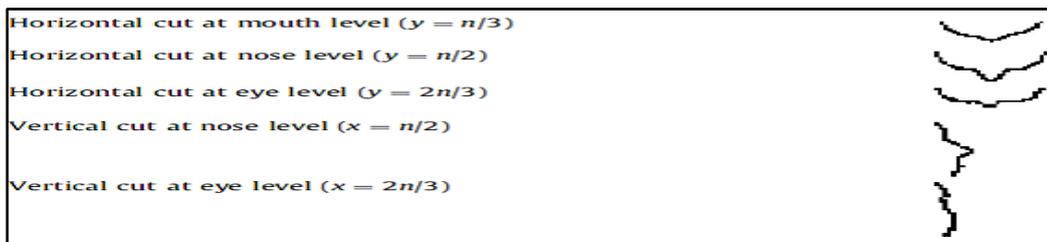


Figure 2.23 Exemples de profils définis simples.

On montre également que ce modèle est, dans une large mesure, **robuste** contre les **variations** de poser des **gestes** et du **visage**. Le système a été testé avec deux régimes correspondants: **SVM**, et de la **PCA** combinée à un classificateur à **distance euclidienne**. Les résultats expérimentaux ont démontré que, même si il n'y a pas de différence significative dans les taux de reconnaissance obtenus, et que ces différences étaient importantes en ce qui concerne le temps de formation en faveur de la **SVM**. En moyenne, lorsque vous utilisez le même type de modèle, la formation du système de reconnaissance sur le même ordinateur a pris environ **quatre fois moins de temps** lors de l'**utilisation SVM**.

Dans l'approche, maillages **3D** du visage de la même personne qui peut avoir une surface bruitée du visage voir **figure 2.24**. Ceci est causé par des occultations produites dans les frontières de la maille. On considère que ces variations dans les images d'un même individu peuvent influencer profondément sur les taux de reconnaissance. Pour cela essayer de mettre en œuvre un système de reconnaissance du visage **3D** exclusivement concentrée sur une centrale prédéterminée (vide occluse) et *veiller à ce que les mailles du visage de la même personne ont la même taille*. Cela permettra probablement d'améliorer les taux de reconnaissance des propositions de modèles **3D** de reconnaissance de visage [162].

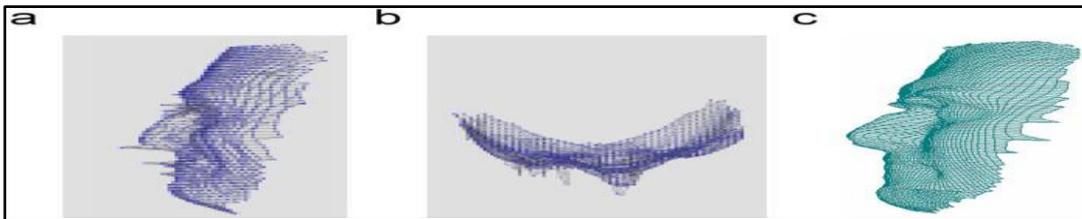


Figure 2.24 (a) et (b) maillage bruité à partir de deux points de vue différents
(c) maillage après le bruit application de filtrage et de lissage [162].

2.3.7 Tableau comparatifs des techniques de reconnaissance 3D.

3D human face description: landmarks measures and geometrical features 2012

Tableau 2.2 Comparatif de plusieurs méthodes de l'état de l'art du SRV 3D.

Table 1

Euclidean distance: applications, relative methods, data dimension, and performances obtained by the authors in their previous works.

Measure	Application	Method	Dimension	Reliability	References
Euclidean distance	Face recognition	Principal component analysis (PCA)	3D	88,8%	Gupta et al., 2007 [48]
		Linear discriminant analysis (LDA)	3D	97,60%	Gupta et al., 2007 [49]
		New algorithm (face profile recognition)	2D	90%	Liposcak et al., 1999
		PCA + LDA	3D	84,5%-96,5% depending on probe and training sets	Daniyal et al., 2009
		Invariant method (new method proposed by the authors)	3D	Accuracy: 87,7%	Cadoni et al., 2009
		Iterative closest point (ICP)	3D	Accuracy: 90,6%	
	Face registration for quantifying facial morphology	Stereophotogrammetry	3D	Reproducibility significance: P<0,05	Ras et al., 1996
	Validation of landmark localization and measurement of facial soft-tissue	Light-based techniques	3D	Mean error: 1,6+/- 1,85%; reproducibility significance: 0,05	Enciso et al., 2004
	Evaluation of laser scanner measuring performance	Laser optical surface scanning	3D	Reliability of laser scanner measurement: 12,5%-50% depending on face regions	Aung et al., 1995
	Face detection + landmark localization	Active shape models (ASM) with Kalman filtering of landmark positions	2D	Fit error mean: 6,73 pixels	Prabhu et al.
				Fit error mean: 6,84 pixels	
	Face expression recognition	Probabilistic neural network (PNN)	3D	88%	Soyel et al., 2008
		LDA + Gabor wavelets	2D	74,1%	
LDA + topographic context method		79,2%			

Tableau 2.3 Comparatif de plusieurs méthodes de l'état de l'art du SRV 3D (suite).

Table 2

Geodesic distance: applications (in this case only face recognition), relative methods, data dimension, and performances obtained by the authors in their previous works.

Measure	Application	Method	Dimension	Reliability	References
Geodesic distance	Face recognition	Principal component analysis (PCA)	3D	89,9%	Gupta et al., 2007 [48]
		Linear discriminant analysis (LDA)	3D	94,70%	Gupta et al., 2007 [49]
		New algorithm	3D	3,1% (EER for midly occluded faces); 5,5% (EER for severly occluded faces)	Bronstein et al., 2006
		PCA	3D	70%-75% for identical twins	Bronstein et al., 2005
		Rigid surface matching		82%-100% for identical twins	
		Canonical form matching (new method proposed by the authors)		100% for identical twins	
Euclidean distance + geodesic distance	Face recognition	PCA	3D	56,6%	Gupta et al., 2010
		LDA		92,6%	
		ICA		87,9%	
		Anthroface 3D (new method proposed by the authors) on landmarks		97,9%	
		Anthroface 3D on arbitrary points		87,5%	

Tableau 2.4 Comparatif de plusieurs méthodes de l'état de l'art du SRV 3D (suite). Ratios of distances

Table 3

Ratios: applications, relative methods, data dimension, and performances obtained by the authors in their previous works.

Measure	Application	Method	Dimension	Reliability	References
Ratios	Face recognition	PCA	2D	75%-98% depending on the similarity measure	Shi et al., 2006
		LDA	3D	83,6%	Tang et al., 2008
	Support vector machine (SVM)	87,1%			

Tableau 2.5 Comparatif de plusieurs méthodes de l'état de l'art du SRV 3D (suite). Curvature and shape Other features

Table 4

Curvedness and shape information: applications (i.e. face recognition and landmark detection), relative methods, data dimension, and performances obtained by the authors in their previous works.

Measure	Application	Method	Dimension	Reliability	References
Shape and curvedness info	Face recognition	Hierarchical graph matching (HGM)	2D + 3D	96,8%	Hüsken et al., 2005
		2D HGM + 3D PCA	2D + 3D	96,8%	
		PCA + nearest neighbour (NN)	3D	76,4% (correct classification rate (CCR) for global geometric features (GGF)); 82% (correct classification rate (CCR) for GGF + shape variation information (SVI))	Xu et al., 2004
		Gabor filter + PCA	2D + 3D	70%-90% depending on features and classifying theories	Wang et al., 2002
		Annotated face model (AFM)	3D	Cumulative match characteristic (CMC): 41%-87% depending on landmarking methods and datasets	Perakis et al., 2009 [79]
		ICP	3D	Matching error: 1 for neutral expressions; 37 for smiling expressions	Lu et al., 2005
		ICP + LDA	3D	Matching error: 0 for neutral expressions; 27 for smiling expressions	
		ICP	3D	88%-98% depending on rank and number of feature candidate sets	Lu et al., 2006 [67]
		Surface matching	2.5D scans and 3D models	68%-98% depending on categories of face scans	Lu et al., 2006 [68]
		Constrained LDA		69%-86% depending on categories of face scans	
Surface matching + constrained LDA	77%-99% depending on categories of face scans				
Landmark detection for face verification	ICP	3D	99% for frontal images; 86% for face images with different poses and facial expressions	Colbry et al., 2005	
Shape index + spin images	Face recognition	AFM	3D	79,4%-89,1% depending on datasets	Passalis et al., 2011
	Landmark identification for facial region retrieval	Active landmark model (ALM)	3D	Correct pose estimation: 96,25%-100%; mean distance error: 7,74-10,59 mm, depending on datasets	Perakis et al., 2009 [80]
Spin images	Face recognition	AFM	3D	94,1%-99% depending on similarity measures and datasets	Kakadiaris et al., 2007
Scale invariant feature transform (SIFT) features	Face recognition	Faceprint method (new method proposed by the authors)	3D	76,5% for comparison neutral vs. non-neutral expressions; 93,6% for comparison neutral vs. neutral	Zhang et al., 2009

Ce que l'on retient de ces tableaux :

D'après ce tableau comparatif nous constatons que la méthode **LDA** associée à différents types de métriques donne des résultats prometteurs. Dans le cas des **distances Euclidiennes** pour:

- **SRV 2D avec expressions** : **TR= 74.1%** pour **LDA+Gabor Wavelets** par **Soyel et al** en **2008**.
- **SRV 3D** : **TR=97.6%** pour **LDA** par **Gupta et al** en **2007**; **TR=84.5%** à **96.5%** pour **PCA+LDA** par **Daniyal et al** en **2009**.

Dans le cas des **distances géodésiques** pour un **SRV 3D** :

- **TR=94.70%** pour la **LDA+distance géodésique** par **Gupta et al** en **2007** ;
- **TR=92.6%** pour la **LDA+distance géodésique+distance Euclidienne** par **Gupta et al** en **2010**.

Dans le cas du **ratios** de distance toujours dans le cas de la **LDA** **Tang et al** en **2008** arrivent à un **TR= 83.6%** pour les images de faces **3D avec expressions**.

Lu et al en **2006** utilisent **LDA** avec les caractéristiques **courbes** et **shape** (forme) sur des images **2.5D** et **3D** les **TR** varient dans les intervalles suivants (tout dépend du scan utilisé):

- Cas du **constrained LDA** **TR= [69% ; 86%]**,
- Cas du **matching+constrained LDA** **TR= [77% ; 99%]**.

D'après ces travaux nous pouvons dire que la reconnaissance 3D et 2D+3D des visages reste sensibles aux expressions faciales. Pour cela nous nous sommes intéressés à l'identification de visage en présence d'expressions. Ce qui représente la dernière partie de notre chapitre.

PARTIE C : Reconnaissance de visage 3D expression

2.4 Reconnaissance de visages 3D expressions et état de l'art

La reconnaissance des expressions faciales ont débutées depuis vingtaine années parce que les expressions faciales jouent un rôle important dans la conversation humaine. Donc, l'interface homme machine du futur utilisera de nouvelles sources d'informations telles que les expressions faciales de six cas (colère, dégoût, joie, peur, surprise, tristesse), la posture humaine ou parole.

2.4.1 Qu'est-ce qu'une expression faciale?

Expression faciale est un ensemble des signes du visage qui traduisent un sentiment de changement dans le visage, perceptible visuellement, dû à l'activation (volontaire ou non) de l'un ou de plusieurs des 44 muscles composant le visage, il y a 250000 expressions possibles.



Figure 2.25 Mouvements faciaux globaux.

La reconnaissance des expressions faciales a débuté depuis une vingtaine d'années. Les expressions faciales jouent un rôle important dans la conversation humaine. Donc, l'interface homme machine du futur utilisera non seulement la souris, le clavier, mais aussi de nouvelles sources d'informations telles que les six expressions faciales (colère, dégoût, joie, peur, surprise, tristesse), la posture humaine ou parole. Dans ce chapitre, nous présenterons un état de l'art sur les techniques de reconnaissance d'expressions faciales.

2.4.2 Système de reconnaissance d'expression faciale existant

Ce système consiste à évaluer la possibilité de reconnaissance des six émotions universelles (citer les six classes considérées ici) en analysant les déformations des caractéristiques faciales telles que les yeux, les sourcils, et la bouche. On considère que ces caractéristiques contiennent suffisamment les informations pour la reconnaissance.

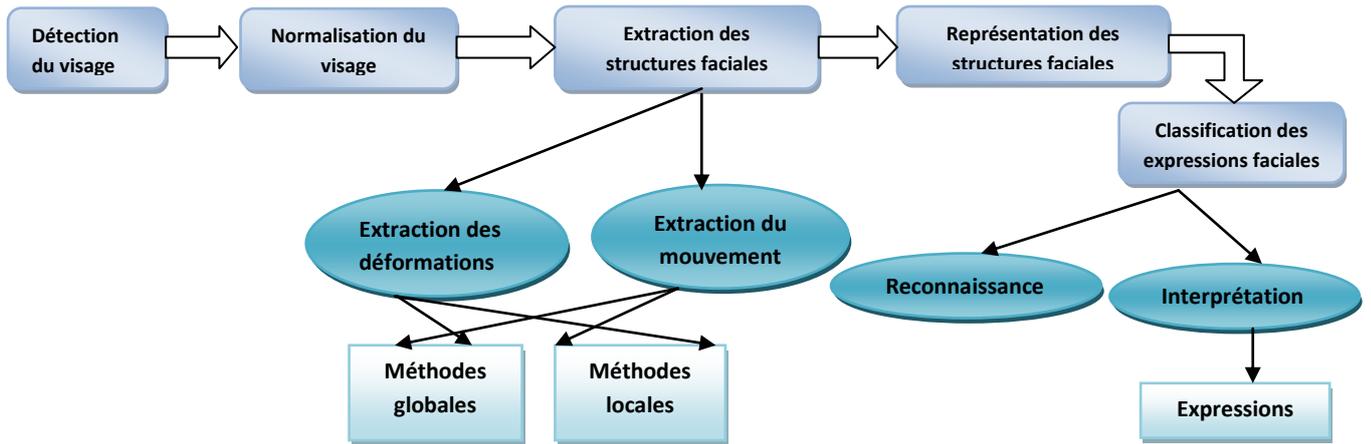


Figure 2.26 Méthodes d'analyse automatique d'un SR d'expressions faciales.

2.4.2.1 Méthode basée sur la segmentation

Ce système se présente comme l'indique la figure 2.27 sous forme de quatre phases essentielles : la segmentation, extraction de données, analyse et classification.



Figure 2.27 Les étapes dans le processus de reconnaissance des expressions faciales.

Dans l'étape de segmentation, le système va localiser les régions des yeux, des sourcils, et de la bouche. L'algorithme est présenté dans la thèse de **Hammal** [163]. Dans l'étape d'extraction de données, les squelettes sont extraits. En fait, ce sont les contours des yeux, de la bouche, des sourcils. A partir des squelettes, on va choisir cinq distances pour caractériser les déformations. Les caractéristiques considérées contiennent les contours des yeux, des sourcils et des lèvres. Finalement, on obtient les squelettes d'expressions. À partir du squelette d'expression, on peut déterminer les déformations de caractéristiques faciales. Cinq distances sont définies pour chaque squelette. À partir de ces distances caractéristiques, on va utiliser le modèle de croyance transférable pour la reconnaissance d'expressions faciales.

Après cela, chaque expression faciale est caractérisée par une combinaison d'états symboliques. Une combinaison d'états est définie pour chaque expression [164].

2.4.2.2 Méthode par masque pour le visage

L'idée c'est d'utiliser un masque pour l'enregistrement dans une reconnaissance de visage. Ce système n'est pas nouveau. Par exemple, Blanz et al.[165] utilisent un masque 3D pour enregistrer la vue frontale avec la vue de profil. Certains masques existants, sont bien conçus pour le montage surface déformable selon l'une des actions musculaires. Toutefois, lorsqu'on travaille avec des images face planes, il est difficile d'atteindre d'un bon enregistrement pour tous les sommets densément placés. En outre, bien que ces masques soient à morphologie souple. Les masques triangulaires sont avantageux dans un mappage de texture. Par conséquent, on utilise une procédure simplifiée et à masque triangulaire, comme représenté dans la **figure 2.28**. Sur ce masque, les triangles gris correspondent à des régions des sourcils, yeux, nez et bouche.

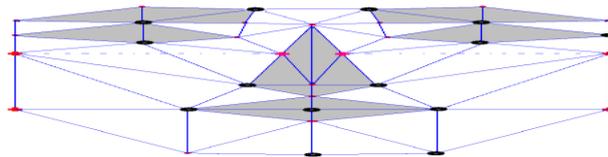


Figure 2.28 Masque pour le visage. [165]

Ils sont intentionnellement à de plus petites façon à saisir les caractéristiques plus détaillées. Ce masque contient seulement 34 sommets et 51 triangles. La **figure 2.33** ci dessous donne deux exemples d'un tel processus d'enregistrement [166].

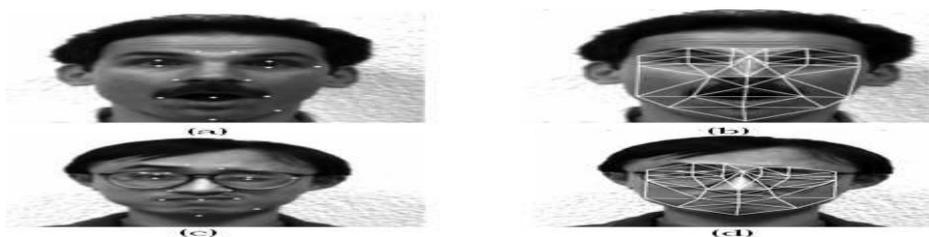


Figure 2.29 Masques de montage. (a), (c): manuellement marqueurs placés (points blancs), (b), (d): équipés masques. [166]

2.4.2.3 Méthode par le modèle d'apparence (AAM)

Franck et al [21] ont travaillé sur la reconnaissance et synthèse d'expressions faciales par modèle d'apparence. Ils traitent une application nouvelle des modèles actifs d'apparence pour l'analyse et la synthèse de visages expressifs, ainsi que pour la reconnaissance d'expressions faciales. Ils considèrent les six expressions universelles que sont les expressions de joie, colère, peur, dégoût, tristesse et surprise. Après une description de ce

type de modèle (calculé à l'aide de trois ACP ou d'une seule), ils proposent une méthode originale d'analyse et de synthèse permettant, à partir d'une seule photo, d'annuler l'expression d'un visage quelconque, puis de synthétiser une expression faciale artificielle sur ce même visage. Ils proposent pour cela deux approches permettant la modélisation des expressions faciales. Enfin, l'intérêt du modèle actif d'apparence pour la reconnaissance automatique d'expressions faciales est montré.

2.4.2.4 Autres méthodes pour la reconnaissance d'expressions faciales

D'après notre étude bibliographique nous réalisons qu'il y'a plusieurs façons d'utiliser les étapes du système de reconnaissance. Tous les systèmes de reconnaissance d'expressions faciales se composent de trois modules principaux :

- Extraction des informations,
- Reconnaissance d'expressions,
- Synthèse d'expressions.

a) Extraction des informations : **Black** et **Yacoob** [52] utilisent des modèles locaux paramétriques pour représenter le mouvement des visages. Ils estiment le mouvement relatif des traits caractéristiques dans le repère du visage. Les paramètres de ce mouvement servent par la suite à représenter l'expression faciale. De manière identique, **Cohn et al.** [167] utilisent un algorithme hiérarchique pour effectuer le suivi des traits caractéristiques par estimation du flux optique. Les vecteurs de déplacement représentent l'information sur les changements d'expression faciale. De même, **Padgett** et **Cottrell** [168] utilisent des gabarits d'œil et de bouche, calculés par Analyse en Composantes Principales d'un ensemble d'apprentissage, en conjonction avec des réseaux de neurones. D'autre part, **Hong** et al. [169] utilisent un modèle global basé sur des graphes étiquetés construits à partir de points de repère distribués sur le visage. Les nœuds de ces graphes sont formés par des vecteurs dont chaque élément est la réponse à un filtrage de Gabor extraite en un point donné de l'image. Finalement, **Cootes et al.** [170] utilise une représentation par modèle actif d'apparence (**AAM**) pour extraire automatiquement des paramètres caractérisant le visage.

b) Reconnaissance d'expressions : Après avoir détecté le visage et extrait les informations pertinentes, l'étape suivante consiste à identifier l'expression faciale affichée. Pour classer l'expression faciale dans l'une des six catégories de base citées dans [171] en plus de la catégorie neutre, **Hong et al.** [169] part du principe que deux personnes qui se ressemblent affichent la même expression faciale de manière similaire. Un graphe étiqueté est attribué à l'image de test puis la personne connue la plus proche est déterminée à l'aide d'une

méthode de mise en correspondance de graphes élastiques. La galerie personnalisée de cette personne est alors utilisée pour reconnaître l'expression faciale de l'image de test. Un graphe étiqueté par des réponses de filtres de Gabor est par ailleurs utilisé par **Lyons et al.** [172] et **Bartlett et al.** [173]. L'ensemble des graphes construits sur un ensemble d'apprentissage est ensuite soumis à une **ACP** puis analysé à l'aide d'une analyse discriminante linéaire (**ADL**) afin de séparer les vecteurs dans des classes ayant des attributs faciaux différents. Le graphe étiqueté de l'image testée sera alors projeté sur les vecteurs discriminants de chaque classe afin de déterminer son éventuelle appartenance à cette classe. Dans une finalité identique, **Essa et Pentland** [174] extraient des gabarits spatiotemporels de l'énergie du mouvement du visage pour chaque expression faciale. Le critère de similarité repose sur la distance Euclidienne entre ces gabarits et l'énergie du mouvement de l'image observée. **Heisele, Ho et Poggio** [175] utilisent des machines à vecteurs de support (**SVM**) dans le cadre de la reconnaissance de visages par des méthodes globales ainsi que par des méthodes reposant sur des traits caractéristiques. De manière identique, l'algorithme de reconnaissance de visages **FaceIt** est basé sur une technique d'analyse locale des traits caractéristiques (**LFA**) développée par **Penev et Atick** [176]. **Draper et al.** [177] compare les performances de l'analyse en composantes principales et de l'analyse en composantes indépendantes (**ICA**) pour la reconnaissance de visages et d'expressions faciales en se basant sur le système de codage **FACS**. Par contre, **Yang** [178] utilise une représentation par noyaux (**KPCA**) pour la reconnaissance de visages. Finalement, **Edwards, Cootes et Taylor** [179] utilisent le modèle actif d'apparence pour reconnaître l'identité d'un individu observé de manière robuste par rapport à l'expression faciale ainsi que l'illumination et la pose. Pour ceci, le critère de similarité utilisé repose sur la distance de **Mahalanobis**, et une **ADL** est appliqué afin de maximiser la séparation des classes.

c) Synthèse d'expressions : La synthèse d'expressions faciales est une tâche difficile compte tenu de la complexité de la forme et de la texture des visages. De plus le visage présente un grand nombre de rides et de plis ainsi que des variations subtiles de forme et de texture qui ont une importance cruciale dans la compréhension et la représentation des expressions faciales. Dans cette perspective, les techniques d'interpolation offrent une approche intuitive pour l'animation de visages. **Pighin et al.** [180] utilisent des **techniques de morphing 2D** combinées avec des transformations d'un **modèle géométrique 3D**, pour créer des modèles faciaux réalistes tridimensionnels à partir de photographies, et pour construire des transitions lisses entre les différentes expressions faciales. Dans la même optique, **Chen et**

al. [181] applique le morphing au cas **3D**. En outre, dans le cadre du "Video-Rewrite", **Bregler et al.** [182] utilisent des techniques de suivi de points **2D** sur la bouche d'un orateur dans une séquence d'apprentissage et des techniques de morphing pour combiner ces mouvements dans une vidéo finale montrant une personne différente prononçant les mêmes paroles. Dans une finalité analogue, **Ezzat et al.** [183] utilisent une représentation par modèle déformable multidimensionnel et une technique de synthèse de trajectoire pour modifier les mouvements de la bouche d'un visage parlant. Cette représentation est capable de synthétiser des configurations inconnues de lèvres parlantes "vidéo-réalistes" à partir d'une séquence initiale, en utilisant des techniques de morphing. Dans la même optique, **Kang et al.** [184] utilisent le modèle actif d'apparence en conjonction avec des techniques de régression linéaire pour annuler l'expression faciale d'un visage dans le but d'améliorer les performances de la technique de reconnaissance de visages par **AAM** décrite dans 2.2.1.1. On propose une variante au modèle **AAM** décrit dans [170] pour la représentation de visages en n'utilisant qu'une seule **ACP**. On montre que cette représentation plus directe donne des résultats sensiblement comparables au modèle **AAM** standard pour la représentation et la reconnaissance d'expressions faciales. On propose également une nouvelle approche de la reconnaissance d'expressions faciales basée sur le modèle **AAM** ainsi que sur sa variante. Finalement, on propose une extension de la méthode décrite dans [184] pour l'annulation de l'expression faciale, à une application de synthèse de nouvelles expressions faciales. De plus on introduit une nouvelle méthode d'interpolation, basée sur la représentation par le modèle actif d'apparence **AAM**, pour la synthèse et l'annulation d'expressions faciales.

La reconnaissance faciale est l'un des sujets les plus étudiés de manière intensive dans la vision par ordinateur et reconnaissance de formes. L'expression du visage, qui change la géométrie du visage, a généralement un effet négatif sur la performance d'un système de reconnaissance faciale. D'autre part, la géométrie du visage est un indice utile pour la reconnaissance. Prenant en compte ces paramètres, on utilise l'idée de séparer la géométrie de l'information de texture dans une image de visage et le modèle des deux types des informations en les projetant dans des espaces distincts de l'**APC** qui sont spécialement conçus pour capturer les traits distinctifs entre les différents individus. Par la suite, les attributs de **texture** et la **géométrie** sont recombinaés pour former un classificateur qui est capable de reconnaître des faces avec différentes expressions. Enfin, en étudiant la géométrie du visage, on est en mesure de déterminer quel type d'expression faciale a été réalisée, ainsi construire un classificateur expression.

La **sélection des fonctionnalités** est l'étape la plus importante à succès d'analyser et de reconnaître les **expressions faciales** automatiquement. Deuxièmement, il est également très important pour **extraire** les **caractéristiques** de **régions** de surface qui contient les informations **discriminantes**. Les caractéristiques optimales devraient **réduire** au minimum au sein de la classe des **variations d'expressions** tout en **maximisant** entre les **variations de la classe**. Dans la littérature, différentes méthodes sont utilisés pour extraire les traits du visage et ces méthodes peuvent être classées soit comme **méthodes en fonction apparence ou géométriques** basées sur les caractéristiques des méthodes où les formes et les emplacements de composantes faciales extraites forment un vecteur de fonctionnalité [185]. L'une des méthodes largement étudié pour extraire les informations apparence sont basées sur le Gabor ondelettes [185,186,187]. **Littlewort et al.** [188] ont montré une précision de reconnaissance élevé (97% pour **Cohn-Kanade** base de données d'expression du visage [189]) pour le visage expressions utilisant les fonctions de Gabor. Ils ont proposé d'extraire les caractéristiques de **Gabor** sur l'ensemble du visage. **AdaBoost** a été utilisé pour sélectionner les sous-ensembles de fonctionnalités. **Tian** [186] a utilisé les ondelettes de **Gabor de multi-échelles et multi-orientation** à la «différence» des images. Les images de différence ont été obtenues en soustrayant une trame expression neutre du reste des trames de la séquence. **Donato et al.** [190] ont utilisé la **technique de la division de l'image faciale** dans deux: face supérieure et inférieure pour extraire une représentation plus fine de Gabor pour la classification. En règle générale, l'**inconvenient** d'utiliser les **filtres de Gabor** est qu'il produit de **très grand nombre** de **caractéristiques** et nécessite à la fois du temps et beaucoup de mémoire pour convoluer les images de visage avec une banque de filtres de Gabor pour extraire les coefficients multi-échelles et multi-orientations. Récemment, les méthodes de description **texture** et de classification à savoir **Local Motif Binary (LBP)** [191] et de la **Phase Locale de Quantification(PLQ)** [192] sont également étudié pour extraire l'apparence du visage à base de caractéristiques. **Zhao et al.** [193] ont proposé l'utilisation du modèle de **texture** de modèles de **Volume Binaires Locales (VLBP)** une **extension** de **LBP**, pour la reconnaissance des expressions. Les auteurs ont proposé d'utiliser uniquement les **co-occurrences** de motifs binaires locales sur trois plans orthogonaux (**LBP-TOP**), afin d'améliorer l'applicabilité de la méthode de calcul en réduisant la complexité. Moyenne précision de reconnaissance faciale expression de **96,26%** a été atteinte de six expressions universelles avec leur modèle proposé utilisant **Cohn-Kanade** visage base de données d'expression. **Liao et al.** [25] ont proposé d'utiliser deux ensembles de caractéristiques pour la classification d'expression. Le premier

ensemble a été obtenu par **LBP** et le second ensemble de caractéristiques est obtenu par l'analyse discriminante linéaire (**LDA**). Ils ont testé leur modèle sur la base de données **JAFFE** [194] et la moyenne précision atteinte de la reconnaissance de **94,59%** pour les sept expressions faciales (Six universelle et une expression neutre). **Jiang et al.** [63] a étendu l'idée de **LBP-TOP** au **PLQ-TOP** et ont montré que la performance de système basé sur **LPQ** vaut mieux que le système basé sur **LBP**. Pour les méthodes géométriques basées sur les caractéristiques [195,196,197], des formes et emplacements des composants faciaux sont extraites pour former un vecteuriel de caractéristiques. Pour la reconnaissance des expressions, **Zhang et al.** [195] a mesuré et suivi le mouvement du visage à l'aide de **Filtres Kalman**. Pour atteindre la tâche de reconnaissance, ils ont également modélisé les comportements temporels des expressions faciales à l'aide **Réseaux Bayésiens Dynamiques (DBN)**. Dans [196] les auteurs ont présenté l'unité d'action de détection régime (UA) en classant les caractéristiques, calculées à partir de "filtre à particules" suivi de repère points de visage. Ils ont formé le système MMI sur l'expression du visage-base de données [198] et testés sur la base de données **Cohn-Kanade** et atteint un taux de reconnaissance de **84%**. **Bai et al.** [197] extrait des informations de forme que l'aide **Histogramme Pyramidal des Gradients d'Orientation (Phog)** et a montré l'exactitude "sourire" de détection plus élevé que **96,7%** à l'aide **Cohn-Kanade** base de données.

2.4.3 État de l'art des techniques récentes de reconnaissance d'expression faciale

Hong-Viet [164] ont travaillé en **2006** sur la **reconnaissance multimodale de gestes de communication non verbale**. Le contexte général est de présenter leurs contributions en reconnaissance automatique des expressions faciales à partir d'information vidéo. Premièrement, on extrait les caractéristiques de l'image du visage, puis on les utilise comme entrée du système de classification. La performance attendue est d'environ **87%**. Mais ils restent des limitations. Toutes les méthodes nécessitent une intervention manuelle pour la détection et la normalisation. Et toutes les méthodes vont mettre en correspondance les expressions faciales avec six expressions basiques. Le problème est qu'il existe des déformations faciales qui ne correspondent pas à ces six expressions basiques. Donc, le système doit tenir compte d'une nouvelle classe d'expressions à savoir la classe des expressions inconnues. Dans le travail de **Hammal** [163], la reconnaissance des expressions faciales utilise le Modèle des Croyances Transférable.

Ligang et al. [199] ont étudié en 2011 la **Géométrie vs Apparence pour discrimination posée par rapport version spontanée d'une émotion**. Les émotions spontanées des expressions faciales diffèrent de celles qui se posent dans l'apparence (séquences et mouvements de tête). Les images fixes ne peuvent pas fournir des informations de mouvement directement. Cependant, indirectement, les distances entre les points clés sur un visage extraites d'une image fixe à l'aide de modèles à forme active peuvent capter une partie en mouvement et changements de pose. Cette information est superposée sur l'information non-rigide du mouvement du visage qui fait également partie de l'expression. L'information géométrique permet d'améliorer la discrimination entre spontanées et les expressions posées du visage. On étudie les performances d'un système de vision artificielle pour discrimination entre les versions qui se posent et spontanées des six émotions de base qui utilise les fonctions d'apparence et les fonctionnalités relatives aux **EIPDFAP** géométriques. Des résultats expérimentaux sur la base de données **NVIE** démontrent que la **fusion de l'information géométrique** ne conduit qu'à une amélioration marginale par rapport à l'apparence caractéristique. Utilisation des fonctions de fusion, la **surprise** est l'émotion la plus facile à distinguer (83,4% de précision), tandis que le **dégoût** est le plus difficile (76,1%). Les résultats trouvés sur différentes régions importantes du visage entre discrimination posé par rapport à version spontanée d'une émotion et la classification par rapport à la même émotion et d'autres émotions. La répartition des fonctions relatives aux **EIPD** sélectionnés montre que la **bouche** est plus **importante** pour la **tristesse**, tandis que le **nez** est plus important **pour** la **surprise**. Cependant, à la fois **le nez et la bouche** sont importants pour le **dégoût, la peur et le bonheur**. **Sourcils, yeux, nez et la bouche** sont importants pour la **colère**.

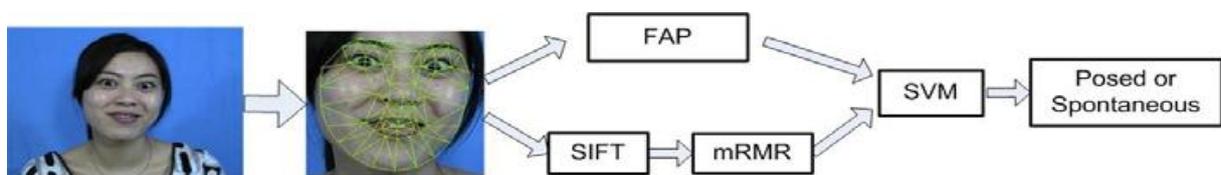


Figure 2.30 Cadre du système d'évaluation. [199]

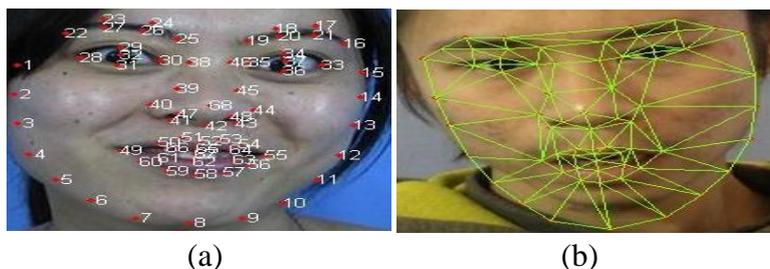


Figure 2.31 (a) 68 points de référence pour la formation, (b) les résultats de détection ASM. [199]

Amit et al. [200] ont étudiés une méthode intitulée *type-2 Intervalle modèle flou pour reconnaissance d'émotion de visage expressif*. Le document propose une nouvelle approche pour la reconnaissance des émotions de l'expression du visage d'un sujet en construisant un intervalle de type-2 modèle flou. Un type-2 intervalle flou de l'espace visage est d'abord construit avec les connaissances de base des traits du visage de sujets différents pour les émotions différentes. Le flou de l'espace visage ainsi créé comprend distributions d'adhésion primaire pour les fonctions m visage, obtenus à partir de n sujets, chacun ayant l -cas de l'expression du visage pour une émotion donnée. Deuxièmement, l'émotion d'une expression du visage inconnu est déterminée sur la base du consensus des caractéristiques mesurées du visage avec le flou face space. La précision de la classification de la méthode proposée est aussi élevé que **88,66%**. La figure **figure 2.32** explique les traits du visage sur une image sélectionnée visage.



Figure 2.32 Caractéristiques Faciales. [200]

La recherche a été effectuée avec succès ces derniers temps de **combiner** les méthodes **caractéristiques** extraites à l'aide de méthodes basées sur l'**apparence** et basées sur les **géométriques** [201,202]. Dans l'ensemble des méthodes pour la reconnaissance des expressions faciale automatique aucune d'entre elles ne peut imiter le système visuel humain. Plutôt toutes les méthodes, utilisent beaucoup de temps de calcul sur l'ensemble du visage ou divise l'image de visage basée sur une heuristique mathématique ou géométrique pour l'extraction de caractéristiques. La tâche d'analyse d'expression et la reconnaissance devrait être faite de manière favorable : seules certaines régions sont sélectionnés pour le traitement ultérieur (par exemple régions saillantes), comme il arrive dans le système visuel humain. Ainsi, on essaye de déterminer statistiquement quelles sont les régions faciales saillantes selon la vision humaine en effectuant une expérience psychovisuelle. L'expérience a été réalisée à l'aide d'eye-tracker qui enregistre les fixations et saccades des observateurs de l'homme comme ils regardent la collection des vidéos montrant six expressions faciales universelles. Les régions saillantes du visage pour les expressions spécifiques ont été déterminées par l'analyse de la fixation des données.

On propose d'extraire les caractéristiques de la région saillante du visage seulement en utilisant l'histogramme pyramidal de gradients d'orientation [209]. La précision de la reconnaissance **FER** réalisé peut atteindre **99,5%** en utilisant **Cohn-Kanade** base de données [189]. L'avantage de l'extraction de caractéristiques des régions saillantes du visage, c'est que le cadre peut être utilisé pour des applications en temps réel. Projet de cadre traite 4 fps (frames per second) en utilisant Matlab 7.6 sous Windows avec un PC 1.8 GHz et 1 Go de RAM. Les processus de la machine mêmes 2 fps pour extraire les mêmes caractéristiques à partir de l'image de visage en entier [204].

Rizwan et al. [204] ont aussi travaillé en **2012** sur une méthode de reconnaissance des expressions du visage basée sur la **perception**. Les humains peuvent reconnaître très facilement les expressions du visage en temps réel. Toutefois, la reconnaissance fiable et rapide des expressions faciales en temps réel est une tâche difficile pour un ordinateur. Ils présentent une nouvelle approche de reconnaissance de trois types d'expressions faciales qui se base sur l'idée de ne considérer que de petites régions du visage bien définies pour en extraire les caractéristiques. Cette proposition est basée sur une étude psycho-visuel expérimental menée avec un eye-tracker. Les mouvements des yeux de quinze sujets ont été enregistrés dans des conditions de visualisation libre d'une collection de 54 vidéos montrant six expressions faciales universelles. Les résultats de cette étude montrent que pour certaines expressions du visage une unique région est perceptuellement plus attractive que les autres. Les autres expressions montrent une attractivité pour deux ou trois régions du visage. Cette connaissance est utilisée pour définir une méthode de reconnaissance se concentrant uniquement sur certaines régions perceptuellement attrayantes du visage et ainsi réduire par un facteur de deux les temps de calcul. Les résultats montrent une précision de reconnaissance automatique de trois expressions de **99.5%** sur la base de données d'expression faciale **Cohn-Kanade**.

2.4.4 Tableau comparatif des travaux récents

Tableau 2.6 Comparaison des performances des algorithmes proposés sur les bases de visages.

Source	Année	Base	Méthode	Taux de reconnaissance
Zilu Ying Beihang university, Beijing	2009	JAFFE	LBP et LE	70.48%
Ligang Zhang. Queensland University of Technology.Australia	2011	Le naturel d'expression visible et infrarouge du visage NVIE	FAP SIFT+mRMS	76.1% 83.4 %
Anisha Halder. Jadavpur University	2011		GT2FS	96.67%
Amit Konar1. Jadavpur University	2012		IT2FS	88.66%
Rizwan Ahmed Khan. Université de Lyon	2012	CK (Cohn-Kanade)	SVM C4.5 Decision Tree RF	97.3% 97.6% 99.5%

2.5 Conclusion

Dans ce chapitre, nous avons passé en revue les principales techniques de reconnaissance de visages. Le problème n'est pas du même niveau de complexité selon que l'on cherche à reconnaître un visage neutre **2D** ou **3D** et un visage **3D expressions** et sous plusieurs variantes. Dans le cadre de ce travail nous nous limiterons à la reconnaissance d'un visage unique dans une image. Parmi les différentes caractéristiques invariantes du visage utilisées pour la détection, la couleur de la peau paraît parmi les plus prometteuses. En effet, les méthodes basées sur la couleur de la peau sont robustes aux occultations ainsi qu'aux variations de la pose, de l'échelle, et de l'expression. Ces méthodes restent cependant peu robustes au changement d'éclairage. Toutefois, les techniques comme le classifieur de Bayes ou bien les réseaux de neurones peuvent conduire à des performances satisfaisantes lorsqu'elles sont appliquées sur des bases de données de petites tailles. Nous avons montré que des approches qui utilisent des prétraitements préalables, comme par exemples la correction de couleur, permettent d'améliorer les performances de la reconnaissance et constituent une solution intéressante au problème de l'illumination.

Par ailleurs, nous avons classifié les techniques de reconnaissance **2D** de visages en trois grandes catégories : les approches globales, les approches locales et les approches hybrides. Nous avons souligné que l'**ACP** permet de réaliser un processus de reconnaissance même si l'on ne possède qu'une seule image d'une personne, ce qui n'est pas possible avec des méthodes telles que l'analyse discriminante linéaire ou les « Support Vector Machine ».

Toutefois, nous avons soulevé la difficulté de l'**ACP**, et de ses différentes extensions, à gérer les occultations et les variations de poses, d'illumination et d'expressions. Les approches locales ont prouvé leur efficacité dans le cas d'un seul exemple d'apprentissage, mais elles souffrent d'un manque de robustesse (la méthode **EGBM**, par exemple, n'est pas robuste face aux occultations). Enfin, les approches hybrides semblent intéressantes pour résoudre le problème de la robustesse, cependant elles ne sont pas appropriées pour l'instant pour traiter le cas d'un seul exemple d'apprentissage (tel que les **Modèles d'Apparence Flexibles** ou le **LFA** hybride).

La reconnaissance **3D** de visages constitue une alternative prometteuse pour résoudre les problèmes de robustesse que connaît la reconnaissance **2D**. Cependant, elle n'a pas encore atteint une certaine maturité, à cause notamment de la lourdeur du processus d'acquisition, et de la non-disponibilité de grandes bases de données de visages **3D**, à accès libre, afin de tester et d'évaluer les techniques élaborées. Les approches de reconnaissance **3D** de visage développées jusqu'à présents sont soit des extensions d'approches **2D**, comme l'**ACP** qui en outre donne des résultats plus précis en **2D** qu'en **3D**, soit des adaptations d'approches **3D** de reconnaissance de formes rigides, tel que l'**ICP** qui donnent des taux d'identification plus intéressants, mais dont l'inconvénient majeur reste l'initialisation du recalage des surfaces faciales. Par ailleurs, les algorithmes de reconnaissance **3D** de visage basés sur les caractéristiques faciales locales ne sont pas fondés sur aucune étude de la structure du visage. Les **points d'intérêt** du visage sont plutôt choisis arbitrairement ou bien correspondent à une extension de caractéristiques locales **2D**. En comparaison avec les images faciales **2D** qui donnent des informations sur la **texture faciale**, les modèles faciaux **3D** fournissent des informations riches sur la **structure** même du visage, ce qui rend le problème de reconnaissance **3D** de visage fondamentalement différent de celui de la reconnaissance **2D**. Nous avons donné, également, une taxonomie des approches sur les quelles sont organisées les deux parties de notre état de l'art.

Deux grandes catégories d'approches d'acquisitions tridimensionnelles ont été distinguées : les **approches actives** et les **approches passives**. Peu d'approches de reconnaissance faciales **3D** qui s'intéressent aux approches passives malgré leur contexte peu contraignant (pas de lumière projetée sur le visage). Ceci est dû au manque de précision des mesures et la production de modèles non-denses ou grossiers. C'est pour ces raisons, que la vision active (introduction de motif lumineux envoyé sur le visage) a pris un intérêt plus important. Moyennant la triangulation active, ces approches sont plus précises et produisent des modèles

plus denses couvrant la géométrie **3D** du visage, malgré l'introduction de lumière projetée sur le visage. Les systèmes axés sur le **profil** existant, les **courbes** de niveau, la **couleur** et des approches **cartes de profondeur** sur la base pour la modélisation **3D** faces sont **très sollicités**.

Dans le **chapitre 6** nous présentons notre modèle de reconnaissance faciale 3D basé sur la **couleur** et la **profondeur**.

Enfin, nous avons souligné l'avantage des techniques multi modales qui **combinent** des données **2D et 3D** de visages pour améliorer les performances et la robustesse de la reconnaissance. Nous avons présenté quelques travaux qui ont clairement rapporté des taux de reconnaissance nettement supérieurs à ceux des techniques de reconnaissance **2D et 3D** isolées.

Dans le cadre de cette thèse, nous proposons une approche de reconnaissance de visage basée sur la couleur qui traite le problème de l'illumination. Nous retenons à ce stade du travail l'utilisation d'un jet de filtres avec une ACP au préalable et les méthodes non linéaires dans les modèles sous espaces, le SVM pour la classification, l'utilisation de la couleur + profondeur pour le 3D et la fusion des régions d'intérêts pour le 3D expression.

Dans la suite de ce travail nous allons présenter les études que nous avons faites dans le cadre de la vérification d'identité. Dans ce contexte suite aux travaux que nous avons présenté dans ce **chapitre 2**, nous avons décidé d'utiliser une stratégie de reconnaissance basée sur la combinaison **2D+3D** et nous tentons d'implémenter cette approche au **chapitre 6**. La dernière partie du chapitre nous a permis de tirer un certain nombre d'enseignements du processus de reconnaissance d'expressions faciales. Après le développement des principales difficultés inhérentes, nous avons présenté le principal du système de reconnaissance d'expressions faciales et un état de l'art des techniques proposées jusqu'à ce jour. Dans le chapitre 4 nous détaillerons le traitement réalisé permettant d'extraire les caractéristiques d'un visage, puis des régions de ce visage à travers une implémentation de l'algorithme de filtres de **Gabor**. Nous mettons en avant l'intérêt de l'**ACP** modulaire sur les images **2D**, puis nous proposons une méthode de vérification d'identité à partir de données **3D**. Nous présentons l'essentiel de notre application en reconnaissance monomodale. Enfin dans le **chapitre 6**, nous explorerons les possibilités de combinaison de ces deux types d'information. Mais avant nous allons présenter les données utilisées pour nos évaluations et les techniques de notre approche. C'est ce qui fera l'objet du **chapitre 3 et 4**.

Chapitre 3

Méthodes de réduction et classification.

3.1 Introduction

Tout système de reconnaissance biométrique comporte une phase très importante basée sur la réduction d'espace. Pour cela nous consacrons un chapitre pour présenter les différentes techniques de projection d'espace. La classification n'est pas à négliger. Nous essayons de présenter l'essentiel des approches utilisées.

Dans les différents domaines de recherches scientifiques, le développement technologique et le besoin de superviser des systèmes de plus en plus complexes nécessitent l'analyse de bases de données de taille importante (signaux, images, documents, ...). Toutefois, si dans cette accumulation de données, on est sûr d'avoir une information complète et utile, celle-ci risque d'être "noyée" dans la masse. Ceci pose les problèmes de la structuration des données et de l'extraction des connaissances. En effet, les bases de données sont en général définies par des tableaux à deux dimensions correspondant aux données et aux attributs les caractérisent. Ces deux dimensions peuvent prendre des valeurs très élevées, ce qui peut poser un problème lors du stockage, de l'exploration et de l'analyse. Pour cela, il est important de mettre en place des outils de traitement permettant l'extraction des connaissances sous-jacentes. L'extraction des connaissances s'effectue selon deux directions, la catégorisation des données (par regroupement en classes) et/ou la réduction de la dimension de l'espace de représentation de ces données (par sélection ou extraction d'attributs). La classification vise à découvrir la structure intrinsèque d'un ensemble de données en formant des groupements qui partagent des caractéristiques similaires.

La réduction de la dimension se pose comme une étape primordiale dans le processus de prétraitement des données (filtrage, nettoyage, élimination des points aberrants, etc.). En effet, pour des données appartenant à un espace de grande dimension, certains attributs n'apportent aucune information voire expriment du bruit, d'autres sont redondants ou corrélés. Ceci rend les algorithmes de décision complexes, inefficaces, moins généralisables et d'interprétation délicate. Les méthodes de réduction de la dimension de l'espace de représentation peuvent être divisées en méthodes d'extraction d'attributs et méthodes de sélection d'attributs. L'extraction d'attributs transforme l'espace d'attributs de départ en un nouvel espace formé de la combinaison linéaire ou non linéaire des attributs initiaux. La sélection d'attributs

choisit les attributs les plus pertinents selon un critère donné. Les données sont alors analysées après projection dans un espace de représentation composé des attributs les plus pertinents. Toutefois, l'interprétation des attributs extraits est plus délicate que l'interprétation des attributs sélectionnés. Le point clé de la sélection d'attributs est la définition d'un score mesurant la pertinence de chacun des attributs. Cette sélection s'appuie sur la connaissance explicite et implicite sur les données. Quand on ne dispose d'aucune information à priori sur le regroupement des données en classes, le contexte d'apprentissage est dit non supervisé. La pertinence d'un attribut est alors mesurée en évaluant ses capacités à préserver la structure des données. Pour de nombreuses applications, on dispose des informations à priori sur la répartition des données en classes. Ainsi, pour ces données, les labels des classes ont été fournis. Dans ce cas, la sélection supervisée consiste à mesurer la corrélation entre l'attribut et les labels des classes des données. Nous n'aborderons que ce dernier cas de classification car tous système biométrique est fondé sur une phase d'apprentissage où la création de signature des données est effectuée d'où la création d'une base de données indexée.

Notre **objectif** c'est l'**identification** des personnes dans un premier temps par la modalité visage en se basant sur une analyse globale en **2D**, puis par la couleur et profondeur en **3D** et finalement une **fusion** des différentes régions d'intérêts en présence **d'expressions faciales**. Donc la fusion est prise en considération et nous lui consacrons le **chapitre 4** pour son étude.

3.2 Méthodes de réduction de dimension

L'Analyse en Composantes Principales (**PCA**) est l'une des méthodes les plus utilisées dans la reconnaissance de visages, elle a été proposée par **M.A.Turk et M.P. Pentland** [4]. Dans l'identification de visage basée sur la **PCA**, les images de visage **2D** sont transformées en vecteurs colonnes **1D**. Le calcul de la matrice de covariance à base de ces vecteurs est difficile à cause de la grande taille des vecteurs **1D** et le nombre important d'échantillons d'apprentissage. En général, le calcul des vecteurs propres d'une grande matrice de covariance prend beaucoup de temps. L'analyse discriminante linéaire (**LDA**) est née des travaux de **Belhumeur et al. en 1997**. La **LDA** effectue une véritable séparation de classe et cela en **minimisant** les **variations** entre les images d'un **même individu** tout en **maximisant** les **variations** entre les **images d'individus différents**. Néanmoins, lorsque le nombre d'individus est inférieur à la résolution de l'image, il est difficile d'appliquer la **LDA** qui peut faire apparaître des matrices de dispersions singulières. Afin de contourner ce problème, certains algorithmes basés sur la **LDA** ont été proposées, le plus connu est la **RLDA** (**Regularized LDA**).

Roli en 2002 [205], ont remarqué que la **LDA** et la **PCA** ne sont pas corrélées car la **LDA** génère un espace propre significativement différent de la **PCA**. Les expériences effectuées dans [205] montrent que la *fusion de la LDA et de la PCA ont donné de bons résultats*.

Dans [206], il a été prouvé expérimentalement que la **PCA** et la **LDA** peuvent être appliquées sur un nombre réduits de coefficients **DCT** pour réaliser une meilleure reconnaissance avec un gain en temps de calcul et en espace mémoire. Afin d'améliorer le taux de reconnaissance de la **PCA** et **LDA**, leur fusion est proposée par **G. L. Marcialis et al.** La réduction de dimensionnalité employant la **PCA** ou la **LDA** nécessite un temps prohibitif lorsque la dimension et le nombre d'échantillons d'apprentissage sont importants. Pour cette raison, la réduction de la complexité informatique est fortement exigée. Pour cela, la transformée en cosinus discrète (**DCT**) a été utilisée dans l'identification de visages pour la réduction de dimension.

Dans ce chapitre nous nous attachons à décrire plusieurs méthodes de réduction de dimension. La réduction de dimension consiste à transformer des données représentées dans un espace de grande dimension en une représentation dans un espace de dimension plus faible. Idéalement, la nouvelle représentation a une dimension égale au nombre de paramètres nécessaires pour décrire les données observées [207]. La réduction de dimension est importante dans de nombreux domaines étant donné qu'elle facilite la classification, la visualisation ou encore la compression de données de grande dimension. Elle permet également souvent de limiter l'effet de la malédiction de la dimension et d'autres propriétés non désirées des espaces de grande dimension [208].

Récemment, un grand nombre de méthodes de réduction de dimension ont été proposées [208, 209,210,211,212,100,213,214,215]. Ces techniques sont capables de traiter des problèmes complexes non linéaires et ont souvent été proposées comme une alternative aux techniques linéaires classiques telles l'analyse en composantes principales (**ACP**) ou l'analyse discriminante linéaire (**LDA**).

De précédentes études ont en effet montré que les *approches non linéaires surpassent les méthodes linéaires* sur des jeux de *données artificiels* hautement non linéaires. Cependant, les succès de réduction de dimension avec les *méthodes non linéaires sur des jeux de données naturelles sont plutôt rares*. Dans la suite, nous décrivons celles qui sont les plus proches de notre modèle des techniques linéaires classiques telles :

l'Analyse en Composantes Principales (**ACP**) [216], la Factorisation Non négative de Matrices (**NMF**) [48], l'Analyse en Composantes Indépendantes (**ICA**) [217] et l'Analyse

Discriminante Linéaire (LDA) [217], ainsi que dix méthodes non linéaires (le nom de chaque méthode n'a volontairement pas été traduit) :

Multi Dimensional Scaling (MDS) [218,219], **Isomap**[200,214], **Kernel PCA (KPCA)** [221,100], **Diffusion Maps** [211,222], **Multi Layer Auto Encoders**[223,210], **Locally Linear Embedding (LLE)** [212], **Laplacian Eigenmaps** [208], **Hessian LLE** [209], **Local Tangent Space Analysis (LTSA)** [215] et **Locally Linear Coordination (LLC)** [213].

D'autres techniques non linéaires ont été proposées, telles que :

Principal Curves [224], **Generalized Discriminant Analysis** [225], **Kernel Maps** [226], **Maximum Variance Unfolding** [227], **Conformal EigenMaps** [257], **Locality Preserving Projections** [229], **Linear Local Tangent Space Alignment** [230], **Stochastic Proximity Embedding** [231], **FastMap** [232], **Geodesic NullSpace Analysis** [234].

La plupart d'entre elles sont des variantes des dix méthodes énoncées plus haut, et ne seront donc pas décrites ici. Nous nous intéressons à celles qui ont fait leurs preuves dans le domaine de la reconnaissance de visages et d'après l'état de l'art établi dans le **chapitre 2**. Il en sort que la **PCA**, **DPCA**, **ICA**, **KPCA**, **EFM** sont les plus sollicitées. Nous essayons de les présenter dans ce qui suit.

3.2.1 La réduction de dimension

Supposons qu'un jeu de données soit décrit par la matrice X de taille $(n \times D)$ où n est le nombre de vecteurs x_i de dimension D . Ce jeu de données possède une dimension propre (ou intrinsèque) d , où $d < D$ voire $d \ll D$. En termes mathématiques, la dimension intrinsèque signifie que le jeu de données repose sur une variété de dimension d , contenu dans un espace de plus grande dimension D . Une technique de réduction de dimension transforme le jeu de données X en un nouvel ensemble Y de dimension d , en gardant au maximum l'essentiel de l'information de l'ensemble de départ. Généralement, ni la géométrie de la variété, ni la dimension d sont connus. Les techniques de réduction de dimension peuvent être classées en plusieurs groupes (voir la figure **3.1**). Le principal critère de classement est l'aspect linéaire ou non des méthodes. Les méthodes linéaires supposent que les données reposent sur une variété linéaire de l'espace de grande dimension. Les méthodes non linéaires ne reposent pas sur cette hypothèse et sont capables de caractériser des variétés plus complexes. Cette hypothèse et sont capables de caractériser des variétés plus complexes.

Figure 3.1 Taxonomie des techniques de réduction de dimension.

3.2.2 Méthodes linéaires de réduction de dimension

Nous décrivons ici quatre des méthodes linéaires les plus couramment utilisées : l'Analyse en Composantes Principales (ACP), la Factorisation de Matrices Non négatives (NMF), l'Analyse en Composantes Indépendantes (ICA) et l'Analyse Discriminante Linéaire (LDA ou FLD : Fisher Linear Discriminant) et l'EFM (Enhanced Fisher Linear Discriminant Model ou Modèle de Fisher Amélioré).

3.2.2.1 L'Analyse en Composantes Principales

L'Analyse en Composantes Principales (ACP) [235], aussi connue sous le nom de transformée de **Karhunen–Loève** [236] est une méthode très utilisée en statistique. Introduite par **Pearson** [237] puis plus tard par **Hotelling** [216], sa principale idée est de **réduire** la **dimension** d'un jeu de données tout en gardant un **maximum d'informations**. Cela est réalisé grâce à une **projection** qui **maximise la variance** tout en **minimisant l'erreur quadratique moyenne** de la reconstruction pour plus de détails, voir [238,235,239,240].

Pour la dérivation **Hotelling** définit l'ACP comme une projection orthogonale maximisant la variance dans l'espace projeté. Étant donné n échantillons $x_i \in R^D$ et $u \in R^D$ tel que :

$$\|u\| = u^T u = 1 \quad (3.1)$$

soit un vecteur ortho normal de projection. Un échantillon x_i est projeté sur u par : $a_i = u^T x_i$

La variance de l'échantillon peut donc être estimée :

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}) \quad (3.2)$$

où \bar{x} est la moyenne des projetés des échantillons de la base :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{d'où} \quad \bar{a} = u^T \bar{x} \quad (3.3)$$

Ainsi la variance du projeté est donnée par :

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (a_i - \bar{a}) \quad (3.4)$$

$$= \frac{1}{n-1} \sum_{i=1}^n (u^T x_i - u^T \bar{x}) \quad (3.5)$$

$$= u^T C u \quad (3.6)$$

Où $C \in R^{D \times D} = \frac{1}{n-1} \sum_{i=1}^n ((x_i - \bar{x})(x_i - \bar{x})^T)$ (3.7)

est la matrice de covariance de $X = [x_1; \dots; x_n] \in R^{D \times n}$. Le problème de maximisation de la variance dans l'espace projeté peut donc s'écrire : $\max u^T C u$ avec $u^T u = 1$

Le calcul de la solution optimale peut être réalisé grâce au multiplicateur de Lagrange :

$$f(u, \lambda) = u^T C u + \lambda(1 - u^T u) \quad (3.8)$$

Par dérivation partielle selon u : $\frac{\partial f(u, \lambda)}{\partial u} = 2Cu - 2\lambda u = 0$ (3.9)

on obtient : $Cu = \lambda u$ (3.10)

Ainsi, le maximum pour le multiplicateur de Lagrange est obtenu si λ est une valeur propre et u un vecteur propre de C . Ainsi la variance décrite par le vecteur de projection u est donnée par λ . D'autres méthodes de **dérivation** de l'ACP sont données dans [241,235]. Pour une vue probabiliste de la dérivation de l'ACP, voir [84,85].

Calcul de l'ACP pour la mise en œuvre de méthodes : il est supposé que le jeu de données d'entraînement est disponible en entier. Ainsi nous avons un ensemble de n observations $x_i \in R^D$ organisés sous forme matricielle $X = [x_1; \dots; x_n] \in R^{D \times n}$. L'estimation de la base de projection de l'ACP revient donc à estimer les éléments propres de la matrice de covariance C de X . Le calcul requiert d'abord l'échantillon moyen : $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ (3.11)

Puis les échantillons sont normalisés par rapport à la moyenne x_{moy} : $\hat{x} = \bar{x}_i = x_i - \bar{x}$ pour former la nouvelle matrice $\hat{x} = [\hat{x}_1; \dots; \hat{x}_n]$. La matrice de covariance $C \in R^{D \times D}$ est ensuite calculée par : $C = \frac{1}{n-1} \hat{x} \hat{x}^T$ (3.12)

La recherche des éléments propres de C conduit à l'obtention de la base de vecteurs propres $u_i \in R^D$, pour lesquels, à chacun d'eux, est associée une valeur propre λ_i . Généralement triés par ordre décroissant de valeur propre associée, les premiers vecteurs propres forment alors une base dans laquelle la plupart de l'information du jeu de données d'entraînement est gardée.

ACP pour des données de grande dimension.: La dimension de la matrice de covariance dépend de la dimension D des vecteurs du jeu de données, qui peut être relativement grande pour certains types de données (typiquement des images). La méthode décrite plus haut devient alors difficile à appliquer, essentiellement à cause de la recherche des éléments

propres de la matrice de covariance C . En effet, pour des images de taille 100x100 par exemple, la matrice de covariance C à inverser est de taille 10000x10000. Cependant, il est connu que pour toute matrice X , les produits matriciels XX^T et $X^T X$ partagent les mêmes valeurs propres différentes de zéro.

Ainsi, le calcul des éléments propres de $C = XX^T$ $C \in R^{D \times D}$ peut se ramener au calcul des éléments propres de la matrice $M \in R^{n \times n}$ où $M = X^T X$. Soit e_i les vecteurs propres de M associés aux valeurs propres δ_i . On a donc :

$$X^T X e_i = \delta_i e_i \quad (3.13)$$

En multipliant à gauche par X les deux côtés de l'équation, on obtient ainsi :

$$X (X^T X e_i) = X (\delta_i e_i) \quad (3.14)$$

$$X X^T (X e_i) = \delta_i (X e_i) \quad (3.15)$$

On voit donc que $X e_i$ est vecteur propre de XX^T et que δ_i est la valeur propre associée, d'où

$$\begin{cases} u_i = X e_i \\ \lambda_i = \delta_i \end{cases} \quad (3.16)$$

La matrice M étant beaucoup plus petite que la matrice C (typiquement, on passe d'une complexité de l'ordre de la dimension des échantillons à une complexité de l'ordre du nombre d'échantillons d'apprentissage), les calculs sont donc plus efficaces. L'algorithme de l'Analyse en Composantes Principales est résumé à l'Algorithme ci-dessous.

Des variantes de l'ACP ont été proposées. Ainsi plusieurs méthodes ont été proposées pour extraire des axes principaux robustes notamment au bruit contenu dans les images d'apprentissage [242,243], ou des méthodes basées sur une formulation Espérance–Maximisation de l'ACP [84,244,85]. Dans le cas où les données d'apprentissage ne sont pas toutes disponibles au départ (cas de vidéos par exemple), des versions incrémentales de l'ACP ont été mises au point [245,246,247,248]. Des méthodes combinant l'aspect incrémental et robuste ont également été proposées dans [249,250].

Algorithme : Calcul de l'ACP**Entrées** : matrice X **Sorties** : vecteur moyen \bar{x} , base de vecteurs propres U , valeurs propres associées λ_i

Calcul du vecteur moyen :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Normalisation des images d'entrées :

$$\hat{x} = x_i - \bar{x}$$

$$\hat{x} = [\hat{x}_1; \dots; \hat{x}_n].$$

si Données de grande dimension **alors**

$$M = \frac{1}{n-1} \hat{x} \hat{x}^T$$

Calcul des éléments propres de M :

$$E = [e_1; \dots; e_n]$$

$$\delta = [\delta_1; \dots; \delta_n]$$

Calcul des éléments finaux :

$$u_i = X e_i \quad U = [u_1; \dots; u_n]$$

$$\lambda_i = \delta_i \quad \lambda = [\lambda_1; \dots; \lambda_n]$$

sinon

$$C = \frac{1}{n-1} \hat{x} \hat{x}^T$$

Calcul des éléments propres de C :

$$U = [u_1; \dots; u_n]$$

$$\lambda = [\lambda_1; \dots; \lambda_n]$$

retourner \bar{x}, U, λ **3.2.2.2 Factorisation de Matrice Non Négative**

La factorisation de matrice non-négative (ou NMF pour Non Negative Matrix Factorization) a été proposée dans [251] et [252]. Introduite dans le cadre de la vision par ordinateur dans [253], cette technique, contrairement à l'ACP, *n'autorise pas de valeurs négatives* dans les *vecteurs de base* ni dans les *vecteurs de projection*. Les vecteurs de base sont donc additifs et représentent des structures locales. Plus formellement, la méthode peut être décrite ainsi : Étant donnée une matrice $V \in R^{n \times m}$ positive contenant les images vectorisées, le but est de trouver les matrices non-négatives $W \in R^{n \times r}$ et $H \in R^{r \times m}$ qui approximent la matrice V :

$$V \approx WH \quad (3.17)$$

Les deux matrices W et H doivent être estimées itérativement en considérant le problème d'optimisation suivant :

$$\min \|V - WH\|_2^2 \quad \text{avec } W; H > 0 \quad (3.18)$$

Les règles de mise à jour pour les matrices W et H sont alors :

$$H_{i,j} \leftarrow H_{i,j} \frac{(W^T V)_{i,j}}{(W^T W H)_{i,j}} \quad (3.19)$$

$$W_{i,j} \leftarrow W_{i,j} \frac{(V H^T)_{i,j}}{(W H H^T)_{i,j}} \quad (3.20)$$

Plus de détails sur la dérivation de la méthode ainsi que sur des descriptions de l'algorithme peuvent être trouvées dans [254] et [255]. De plus, pour améliorer la rapidité de l'algorithme ainsi que pour s'assurer que la solution trouvée soit le minimum global (le problème d'optimisation n'est en effet pas convexe en W ni en H), plusieurs extensions ont été proposées [256,257]. Elles considèrent une contrainte additionnelle de parcimonie et reformulent le problème en un problème convexe.

3.2.2.3 Analyse en Composantes Indépendantes

L'Analyse en Composantes Indépendantes (ou **ICA** pour **I**ndependant **C**omponent **A**nalysis) a été introduite par **Hérault, Jutten** et **Ans** dans [258,259] et [260] dans le contexte de la neurophysiologie. Elle devint populaire lors de son utilisation dans le domaine du traitement du signal pour la séparation de sources aveugles dans [261] et [262]. Le but est d'exprimer un ensemble de n variables aléatoires x_1, \dots, x_n comme une combinaison linéaire de n variables aléatoires statistiquement indépendantes s_j :

$$x_j = a_{j;1}s_1 + \dots + a_{j;n}s_n \quad \forall j \quad (3.21)$$

ou sous forme matricielle : $x = As$ (3.22)

où $x = [x_1; \dots; x_n]^T$, $s = [s_1; \dots; s_n]^T$ et A est une matrice contenant les coefficients a_{ij} . Le but de l'Analyse en Composantes Indépendantes est l'estimation des composantes originales s_i , ou de manière équivalente des coefficients a_{ij} . Par définition, les variables aléatoires s_i sont mutuellement indépendantes et la matrice de mélange est donc inversible.

Ainsi le problème de l'ICA peut être formulé [263] : $u = Wx = WAs$ (3.23)

Plusieurs fonctions objectives ont été proposées, ainsi que des méthodes efficaces de résolution : **InfoMax** [264] ou **FastICA** [265]. Pour plus de détails sur la théorie et les applications possibles de l'Analyse en Composantes Indépendantes, voir [266]. Pour l'application de l'ICA à la **reconnaissance de visages** [89] et [267] proposent **deux architectures**. Dans la **première**, les images sont considérées comme un **mélange linéaire d'images** de base statistiquement indépendantes. Dans la **seconde**, le but est de trouver **des coefficients statistiquement indépendants représentant l'image d'entrée**.

Pour ces deux architectures, une Analyse en Composantes Principales est **appliquée en prétraitement**.

3.2.2.4 Analyse Discriminante Linéaire

Si les données d'apprentissage sont labélisées, ces informations peuvent être utilisées pour l'apprentissage du sous-espace. Ainsi, pour assurer une classification plus efficace,

l'Analyse Discriminante **Linéaire** de **Fisher** (**LDA** pour **Linear Discriminant Analysis**) a pour but de maximiser la distance entre les classes tout en minimisant la variance intra-classe. Plus formellement, soient $\{x_1; \dots; x_n\}$ n échantillons appartenant à une classe parmi c $\{X_1; \dots; X_c\}$. L'Analyse Discriminante Linéaire calcule une fonction de classification $g(x) = W^T x$, où la matrice W est choisie comme la projection linéaire minimisant la variance intra-classe :

$$S_B = \sum_{j=1}^c n_j (\bar{x}_j - \bar{x})(\bar{x}_j - \bar{x})^T \quad (3.24)$$

tandis que la variance inter-classe est maximisée :

$$S_W = \sum_{j=1}^c \sum_{x_k \in X_j} (x_k - \bar{x}_j)(x_k - \bar{x}_j)^T \quad (3.25)$$

où \bar{x} est le vecteur moyen de tous les échantillons, \bar{x}_j est le vecteur moyen des échantillons appartenant à la classe j , et n_j est le nombre d'échantillons de la classe j . Le calcul de la projection est ainsi obtenu en **maximisant** le **critère de Fisher** :

$$W_{\text{opt}} = \arg \max \frac{|(W^T S_B W)|}{|(W^T S_W W)|} \quad (3.26)$$

La solution optimale à ce problème d'optimisation est donnée par la résolution du problème généralisé des valeurs propres :

$$S_B W = \lambda S_W W \quad (3.27)$$

ou en calculant directement les vecteurs propres de $S_W^{-1} S_B$. Le rang de $S_W^{-1} S_B$ est au plus $c - 1$. Ainsi, pour de nombreuses applications, cette matrice est singulière et le problème des valeurs propres ne peut être résolu. Ce problème est souvent appelé le problème des échantillons de petite taille (small sample size problem). Pour surmonter ce problème, plusieurs solutions ont été proposées [83,268,269]. De plus, de nombreuses variantes de la **LDA** ont été introduites telles la classification robuste [270], ou la **LDA** incrémentale [271]. La **LDA** est étroitement liée à l'**ACP**, du fait que toutes les deux recherchent des combinaisons linéaires des variables qui représentent au mieux les données. La **LDA** essaye explicitement de modéliser la différence entre les classes des données. L'**ACP** quand à elle, ne tient pas compte des différences entre les classes.

Chaque visage, qui se compose d'un grand nombre de pixels, est réduit à un plus petit ensemble de combinaisons linéaires avant la classification. Chacune des nouvelles dimensions est une combinaison linéaire des valeurs de pixels. Les combinaisons linéaires obtenues en utilisant **FLD** s'appellent les Fisherfaces, en analogie avec les visages propres (EigenFaces) [272,55]. La **LDA** est une technique qui cherche les directions qui sont efficaces pour la discrimination entre les données. La **figure 3.2** représente un exemple de classification de deux nuages de points. L'axe principal de la **LDA** est l'axe de projection qui maximise la

séparation entre les deux classes. Il est clair que cette projection est optimale pour la séparation des deux classes par rapport à la projection sur l'axe principal calculé par **ACP**.

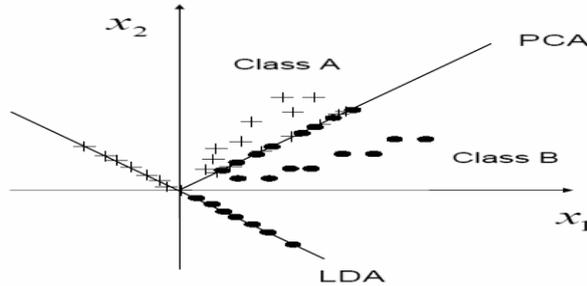


Figure 3.2 Comparaison entre les projections de deux classes de points ("class 1" et "class 2") sur les axes principaux construits par **ACP** et par **LDA**.

Les étapes de l'analyse discriminante linéaire **LDA** sont :

1) Calcul des moyennes

Nous calculons la moyenne des images dans chaque classe (m_i) et la moyenne de toutes les images m

$$m_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_j^i ; \quad i = 1, 2, \dots, C \quad (3.28)$$

$$m = \frac{1}{P} \sum_{i=1}^P x_i ; \quad (3.29)$$

P : Le nombre d'images d'apprentissage.

n_i : Le nombre d'images dans chaque classe.

C : Le nombre de classes.

2) Centrer les images dans chaque classe

Soustraire la moyenne de chaque classe de l'image de cette classe

$$\forall x \in X_i, X_i \in X, \hat{x} = x - m_i \quad (3.30)$$

3) Centrer les moyennes de chaque classe

Soustraire la moyenne totale des moyennes de classes.

$$\hat{m}_i = m_i - m \quad (3.31)$$

4) Calculer la matrice de dispersion intra-classes S_W (within class scatter matrix)

La matrice de dispersion intra-classes (S_W) est la somme des matrices de dispersion des classes. Pour la $i^{\text{ème}}$ classe la matrice de dispersion S_i est calculée par la somme des matrices de covariance des images centrées.

$$S_i = \sum_{x \in X_i} \hat{x}^T \hat{x} \quad (3.32)$$

La matrice de dispersion intra-classes (S_W) est la somme de toutes les matrices de dispersion.

$$S_W = \sum_{i=0}^C S_i \quad (3.33)$$

Où C est le nombre de classe.

5) **Calculer de la matrice de dispersion inter-classes S_B** (between class scatter matrix)

La matrice de dispersion inter-classes (S_B) est la somme de dispersion entre classes.

$$S_B = \sum_{i=1}^C n_i \hat{m}_i \hat{m}_i^T \quad (3.34)$$

Où n_i : Le nombre d'images dans la classe.

\hat{m}_i : La moyenne des classes.

Le but est de maximiser les distances inter-classes tout en minimisant les distances intra-classes, ce qui revient à retrouver la matrice de transformation W qui maximise le critère :

$$J(W) = \frac{W^T S_b W}{W^T S_w W} \quad (3.35)$$

donc W est optimale pour :

$$W_{opt} = \arg \max_W \left(\frac{|W^T S_b W|}{|W^T S_w W|} \right) = [w_1, w_2, \dots, w_m] \quad (3.36)$$

6) **Résoudre le problème de valeurs propres généralisé**

- Résoudre le problème généralisé de vecteurs propres (V) et des valeurs propres (Λ) de la matrice de dispersion S_w et la matrice de dispersion S_B .

$$S_B V = \Lambda S_w V \quad (3.37)$$

La solution est rendue par calcul des vecteurs propres et des valeurs propres de la matrice $S_w^{-1} * S_B$.

- Ordonner les vecteurs propres par ordre décroissant des valeurs propres correspondantes. La matrice de transformation de la **LDA** est constituée par les premiers vecteurs propres.

Implémentation de l'Analyse Discriminante Linéaire (LDA) :

1. Calcul la matrice de dispersion intra-classes S_w (équation (3.33)).
 2. Calcul de la matrice de dispersion inter-classes S_B (équation (3.34)).
 3. Calcul des valeurs et vecteurs propres de la matrice $S_w^{-1} * S_B$.
 4. Ordonner les vecteurs propres par ordre décroissant des valeurs propres correspondantes.
 5. La matrice de transformation de la **LDA** est les m premiers vecteurs propres (U_{LDA}).
- **Inconvénient**

La **FLD** exige un grand nombre d'échantillons de l'ensemble d'apprentissage pour la bonne généralisation. Quand un tel besoin n'est pas répondu, la **FLD** crée un problème de sur-ajustement aux données d'apprentissage et ceci s'apprête mal aux nouvelles données de test. [273,274,275]

- **Solution** [276]

Le modèle **Discriminant Linéaire Amélioré de Fisher (Enhanced Fisher Model EFM)**.

3.2.2.5 Le modèle discriminant linéaire amélioré de Fisher (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 inter-classe [279]. La diagonalisation simultanée est une étape sagement équivalente à deux opérations comme l'a souligné **Fukunaga** [278]. Blanchiment de la matrice de dispersion intra-classe et l'application de l'**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 [279]. 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 une meilleure généralisation.

Le choix de rang des composantes principales (m) pour la réduction de la dimension, prend en compte de l'ordre de grandeur de l'énergie spectrale. Les valeurs propres de la matrice de covariance fournissent un bon indicateur pour répondre au critère de l'énergie. Il faut ensuite calculer les valeurs propres de la matrices de dispersion intra-classe dans l'espace **ACP** réduit pour faciliter le choix du rang des composantes principales de sorte que l'exigence de grandeur est respectée. A cette fin, on effectue la **FLD** par des étapes comme décrit ci-dessous. En particulier, ces étapes **FLD** permettent de trouver les valeurs propres et les vecteurs propres de $S_W^{-1}S_b$ comme résultat de la diagonalisation simultanée de S_W et S_b .

Les étapes de l'**EFM** sont présentées comme suit :

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

$$S_W V = V A \quad \text{et} \quad V^T V = I \quad (3.38)$$

$$A^{-1/2} V^T S_W V A^{-1/2} = I \quad (3.39)$$

Où V , $A \in \mathbf{R}^{m \times m}$ 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 (3.37)).

Donc, **EFM** diagonalise en premier lieu la matrice de dispersion intra-classe S_W (3.38) et (3.39). Notez que V et A 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:

$$A^{-1/2} V^T S_b V A^{-1/2} = K_b \quad (3.40)$$

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

$$K_b V_b = V_b A_b \quad \text{et} \quad V_b^t V_b = I \quad (3.41)$$

Où $V_b, A_b \in \mathbf{R}^{m \times m}$ 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 :

$$U = V A^{-1/2} V_b \quad (3.42)$$

Implémentation du modèle Discriminant Linéaire Amélioré de Fisher (EFM) :

1. Calcul la matrice de dispersion intra-classes S_W (équation (3.33)).
2. Calcul de la matrice de dispersion inter-classes S_B (équation (3.34)).
3. Calcul des valeurs (A) et vecteurs (V) propres de la matrice S_W .
4. Calculer la nouvelle matrice de dispersion inter-classe $K_b = A^{-1/2} V^T S_b V A^{-1/2}$
5. Calcul des valeurs (A_b) et vecteurs (V_b) propres de la matrice K_b .
6. Calcul de la matrice $U = V A^{-1/2} V_b$.
7. La matrice de transformation de l'EFM est les m premiers vecteurs de U (U_{EFM}).

3.2.3 Méthodes non linéaires de réduction de dimension

Nous décrivons ici les méthodes non linéaires de réduction de dimension. Les techniques non linéaires peuvent être catégorisées en trois principaux types : les techniques essayant de préserver les propriétés globales des données d'apprentissage dans l'espace de faible dimension, les techniques s'attachant à préserver les propriétés locales des données d'apprentissage, et les techniques réalisant un alignement global de modèles linéaires.

3.2.3.1 Méthodes globales

Les méthodes globales de réduction non linéaires de dimension essaient de préserver les propriétés globales des données d'apprentissage dans le nouvel espace de faible dimension. On peut citer les techniques : **MDS**, **Isomap**, **Diffusion Maps** et les **AutoEncoders Multi-Couches**. Nous nous limitons à la description de la **Kernel PCA** qui fait l'objet de notre application au système de reconnaissance de visages monomodale et multi algorithmiques. Notre choix est justifié par le fait que la **KPCA** est une extension de la **PCA** qui est à la base de tous nos travaux. Nous voulons par l'introduction de noyaux améliorer les performances de notre système et aussi soulever les limites de la **PCA**.

Kernel PCA L'Analyse en Composantes Principales à Noyaux (ou **KPCA** pour **Kernel Principal Component Analysis**) est la reformulation non linéaire de la technique linéaire classique qu'est l'Analyse en Composantes Principales en utilisant des fonctions à noyaux

[280]. Depuis plusieurs années, la reformulation de techniques classiques à l'aide de l'astuce du noyau a permis l'émergence de nombreuses techniques comme les machines à support de vecteurs (ou **SVM** pour **Support Vector Machine**) [281]. L'**ACP** à noyaux calcule les principaux vecteurs propres de la matrice de noyaux plutôt que la matrice de covariance. Cette reformulation de l'**ACP** classique peut être vue comme une réalisation de l'**ACP** sur l'espace de grande dimension transformée par la fonction noyau associée. L'**ACP** à noyaux permet ainsi de construire des **Mappings Non Linéaires**.

L'**ACP** à noyaux calcule d'abord la **matrice de noyaux** K des points x_i dont les entrées sont définies par :

$$k_{ij} = k(x_i, x_j) \quad (3.43)$$

où k est la fonction noyau [258]. Ensuite, la matrice de noyaux K est centrée :

$$k_{ij} = k_{ij} - \frac{1}{n} \sum_l k_{il} - \frac{1}{n} \sum_l k_{jl} + \frac{1}{n^2} \sum_{lm} k_{lm} \quad (3.44)$$

Cette opération correspond à la soustraction de la moyenne des vecteurs caractéristiques dans l'**ACP** linéaire classique.

Les d principaux vecteurs propres v_i de la matrice de noyaux centrée sont ensuite calculés. Il peut être montré que les vecteurs propres α_i de la matrice de covariance (dans l'espace de grande dimension) sont des versions mises à l'échelle des vecteurs propres v_i de la matrice de noyaux.

$$\alpha_i = \frac{1}{\sqrt{\lambda_i}} v_i \quad (3.45)$$

La technique **KPCA** est une méthode basée sur les noyaux et ses performances dépendent alors grandement du choix de la fonction noyau k . Les noyaux classiquement utilisés sont le noyau linéaire (cela revient alors à effectuer une **AC** classique), le noyau polynomial ou encore le noyau gaussien [281].

L'**Analyse en Composantes Principales à Noyaux** a été appliquée avec succès à plusieurs problèmes comme la reconnaissance de la parole [282], ou la détection de nouveaux éléments d'un ensemble [283]. Un gros **défaut** de l'**Analyse en Composantes Principales à noyaux** est que la **taille de la matrice de noyaux est le carré du nombre d'échantillons** de l'ensemble d'**apprentissage** ce qui peut rapidement être prohibitif. Une approche permettant de résoudre ce problème peut être trouvée dans [284].

3.2.3.2 Méthodes locales

Les méthodes dites locales de réduction de la dimension essaient de préserver les propriétés dans le voisinage des points. Ce type de technique repose sur la supposition qu'en préservant les propriétés locales des données, les propriétés globales de la variété le seront tout autant. La plupart de ces techniques peuvent se ramener à une définition valide dans le

cadre de l'ACP à Noyaux à l'aide de noyaux locaux spécifiques [285,287]. Sont présentés dans [65] les méthodes LLE, Laplacian Eigenmaps, Hessian LLE et LTSA.

3.3 Classification

3.3.1 Classification par mesure de similarités

3.3.1.1 Comparaisons entre deux vecteurs

Lorsqu'on souhaite comparer deux vecteurs de caractéristiques issus du module d'extraction de caractéristiques d'un système biométrique, on peut soit effectuer une mesure de similarité (ressemblance), soit une mesure de distance (divergence).

La première catégorie de distances est constituée de distances Euclidiennes et sont définies à partir de la *distance de Minkowski d'ordre p* dans un espace euclidien \mathbf{R}^N (N déterminant la dimension de l'espace euclidien).

Considérons deux vecteurs $X = (x_1, x_2, \dots, x_N)$ et $Y = (y_1, y_2, \dots, y_N)$, la *distance de Minkowski d'ordre p* notée L_p est définie par :

$$L_p = \left(\sum_{i=1}^N |x_i - y_i|^p \right)^{1/p} \quad (3.46)$$

Nous allons présenter quelques mesures de distance dans l'espace original des images puis dans l'espace de Mahalanobis.

1) Distances Euclidiennes

- Distance City Block (L_1)

Pour $p = 1$, on a :

$$L_1(x, y) = \sum_{i=1}^N |x_i - y_i| \quad (3.47)$$

- Distance Euclidienne (L_2)

Pour $p = 2$, on a :

$$L_2(x, y) = \sqrt{\sum_{i=1}^N |x_i - y_i|^2} \quad (3.48)$$

2) Distances dans l'Espace de Mahalanobis

Soit u et v deux vecteurs propres de J_m , issus de l'algorithme PCA, et m et n deux vecteurs de E_{Mah} . Soit λ_i les valeurs propres associées aux vecteurs u et v , et σ_i l'écart type, alors on définit $\lambda_i = \sigma_i^2$. Les vecteurs u et v sont reliés aux vecteurs m et n à partir des relations suivantes :

$$m_i = \frac{u_i}{\sigma_i} = \frac{u_i}{\sqrt{\lambda_i}} \quad \text{et} \quad n_i = \frac{v_i}{\sigma_i} = \frac{v_i}{\sqrt{\lambda_i}} \quad (3.49)$$

- **Mahalanobis L_1 (Mah L_1)**

L_1 est définie par :

$$Mah_{L_1}(u, v) = \sum_{i=1}^N |m_i - n_i| \quad (3.50)$$

- **Mahalanobis L_2 (Mah L_2)**

L_2 est définie par :

$$Mah_{L_2}(u, v) = \sqrt{\sum_{i=1}^N |m_i - n_i|^2} \quad (3.51)$$

Par défaut, lorsqu'on parle de distance de Mahalanobis, c'est à cette distance que l'on doit se référer.

3.3.1.2 Comparaisons entre deux matrices

Dans **Yang** a proposé une nouvelle mesure de similarité au plus proche voisin pour la reconnaissance de visages. La distance de **Yang** se base sur la classification de matrices caractéristiques obtenues par l'**ACP2D**. Cette distance a été, également, adoptée par **Visani et al.** [Vis04] et **Bengherabi** [Ben08] elle est définie pour deux matrices caractéristiques réduites $Y_i = [y_1^{(i)} y_2^{(i)} \dots y_{d_1}^{(i)}]$ et $Y_j = [y_1^{(j)} y_2^{(j)} \dots y_{d_1}^{(j)}]$ comme suit :

$$d(Y_i, Y_j) = \sum_{k=1}^{d_1} \|y_k^{(i)} - y_k^{(j)}\|_2 \quad (3.52)$$

Où $\|y_k^{(i)} - y_k^{(j)}\|_2$ désigne la distance Euclidienne entre les deux vecteurs $y_k^{(i)}$ et $y_k^{(j)}$ d'où :

$$d(Y_i, Y_j) = \sum_{k=1}^{d_1} \left(\sum_{h=1}^n (y_{hk}^{(i)} - y_{hk}^{(j)})^2 \right)^{1/2} \quad (3.53)$$

Avec $y_k^{(i)} = [y_{1k}^{(i)} y_{21}^{(i)} \dots y_{d_1 k}^{(i)}]$ et $y_k^{(j)} = [y_{1k}^{(j)} y_{21}^{(j)} \dots y_{d_1 k}^{(j)}]$.

Dans [Zuo05], **Zuo** a proposé l'**Assembled Matrix Distance: AMD** définie par :

$$d(Y_i, Y_j) = \left(\sum_{k=1}^{d_1} \left(\sum_{h=1}^n (y_{hk}^{(i)} - y_{hk}^{(j)})^2 \right)^{p/2} \right)^{1/p} \quad \text{avec } p > 0 \quad (3.54)$$

pour laquelle la distance de Yang est obtenue pour $p = 1$ et la distance de Frobenius pour $p = 2$.

3.3.2 Classification par Machine à Vecteurs de Support (SVM)

Le **SVM** (Support Vector Machine) est une nouvelle technique d'apprentissage statistique, proposée par **V. Vapnik** en 1995. Elle permet d'aborder des problèmes très divers comme le classement, la régression, la fusion, etc...

Depuis son introduction dans le domaine de la Reconnaissance de Formes (**RdF**), plusieurs travaux ont pu montrer l'efficacité de cette technique principalement en traitement d'image. L'idée essentielle consiste à projeter les données de l'espace d'entrée (appartenant à des

classes différentes) non linéairement séparables, dans un espace de plus grande dimension appelé espace de caractéristiques, de façon à ce que les données deviennent linéairement séparables. Dans cet espace, la technique de construction de l'hyperplan optimal est utilisée pour calculer la fonction de classement séparant les classes. Tels que les vecteurs appartenant aux différentes classes se trouvent de différents côtés de l'hyperplan.

Le but du SVM est donc de trouver une séparatrice qui minimise l'erreur de classification sur l'ensemble d'apprentissage mais qui sera également performante en généralisation sur des données non utilisées en apprentissage. Pour cela le concept utilisé est celui de marge (d'où le nom de séparateurs à vaste marge). La marge est la distance quadratique moyenne entre la séparatrice et les éléments d'apprentissage les plus proches de celle-ci appelés vecteurs de support (figure 3.3). Ces éléments sont appelés vecteurs de support car c'est uniquement sur ces éléments de l'ensemble d'apprentissage qu'est optimisée la séparatrice.

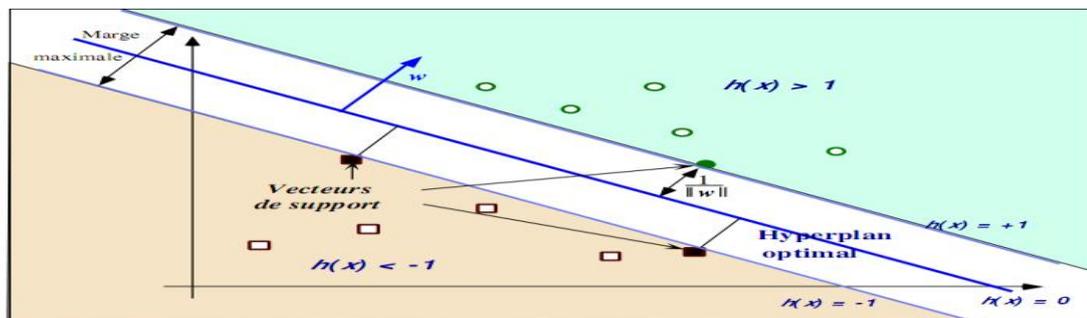


Figure 3.3 Principe de la technique SVM

(hyperplan optimal, vecteurs de supports, marge maximale).

Hyperplan optimal : est un Hyperplan qui classe correctement les données (lorsque c'est possible) et qui se trouve le plus loin possible de tous les exemples, on peut dire aussi que cet hyperplan maximise la marge.

Vecteurs de support : ce sont les points les plus proches, qui seuls sont utilisés pour la détermination de l'hyperplan.

La marge : est la distance entre l'hyperplan et les exemples. La marge est calculée à partir du produit scalaire entre les vecteurs situés la frontière de chaque classe et le vecteur unitaire normal de l'hyperplan séparateur. [287]

Tout classifieur a pour but de classifier un élément x , ici $x = (s_1, \dots, s_N)$ est un vecteur de scores de dimension N , dans l'une des classes possibles. Dans notre problème il y a deux classes, Client ou Imposteur, dont l'étiquette sera noté y avec $y = -1, 1$, -1 correspondant à la classe des Imposteurs et 1 à la classe des Clients. Le classifieur a donc pour but de déterminer f telle que

$$Y = f(x) \quad (3.55)$$

Le SVM a pour but de trouver la meilleure séparatrice linéaire (en terme d'émarge maximale, c'est à dire la meilleure généralisation) dans l'espace transformée par la fonction de noyau \mathbf{K} , c'est à dire de déterminer le vecteur \mathbf{w} et la constante b tels que la séparatrice ait pour équation :

$$\mathbf{w} \cdot \mathbf{k}(x) + b = 0 \quad (3.56)$$

La distance entre un point de l'espace x_i et l'hyperplan d'équation $w.K(x)+b = 0$ est égal à :

$$h(x_i) = \frac{w \cdot K(x_i) + b}{\|w\|} \quad (3.57)$$

Pour maximiser la marge, il faut donc minimiser $\|w\|$ en maximisant $\mathbf{w} \cdot \mathbf{K}(x_i) + b$ pour les x_i définis comme vecteurs de support. Ces **vecteurs de supports** sont les x_i pour $i = 1 : m$ de la base d'**apprentissage** tels que :

$$\mathbf{w} \cdot \mathbf{K}(x_i) + b = \pm 1. \quad (3.58)$$

La résolution de ce problème d'optimisation est faite par l'utilisation des multiplicateurs de Lagrange où le Lagrangien est donné par :

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^m \alpha_i (y_i (w \cdot K(x_i) + b) - 1) \quad (3.59)$$

Avec les coefficients α_i appelés multiplicateurs de Lagrange. Pour résoudre ce problème d'optimisation, il faut minimiser le Lagrangien par rapport à \mathbf{w} et b et le maximiser par rapport à α . Dans la pratique, il est souvent impossible de trouver un séparateur linéaire (même dans l'espace transformé par la fonction noyau) car il y a toujours des erreurs de classification. Il a donc été introduit par **Vapnik** [288] la technique de marge souple. Ce principe de marge souple tolère les mauvais classements par l'introduction de variables ressorts ξ_i qui permettent de relâcher les contraintes sur les éléments d'apprentissage qui ne doivent plus être à une distance supérieure ou égale à 1 de la marge (l'égalité correspondant aux vecteurs de support), mais à une distance supérieure ou égale à $1 - \xi_i$, c'est à dire :

$$y_i (w \cdot K(x_i) + b) \geq 1 - \xi_i \quad (3.60)$$

Avec $\xi_i \geq 0$ pour $i = 1 : M$; M étant le nombre d'éléments de la base d'apprentissage.

Le problème d'optimisation est donc modifié et le Lagrangien devient :

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^M \xi_i - \sum_{i=1}^M \alpha_i (y_i (w \cdot K(x_i) + b) - 1) \quad (3.61)$$

Où C est une constante strictement positive qui permet de régler le compromis entre le nombre d'erreurs de classification et la largeur de la marge. Cette constante est en général déterminée empiriquement par validation croisée sur l'ensemble d'apprentissage. [58]

3.4 La Décision

Pour estimer la différence entre deux images, il faut introduire une mesure de similarité. Il est important de noter que le système de vérification automatique de visage se base en sa totalité sur la méthode de localisation.

3.5 Conclusion

Dans ce chapitre, nous avons présenté un certain nombre de méthodes utilisées dans les cinq étapes d'un système de reconnaissance de visage. Nous nous sommes limités aux méthodes proches du modèle de notre approche. Toutes ces méthodes ont des avantages et des inconvénients selon la complexité, le besoin de données d'apprentissage ou les paramètres à optimiser. Le choix d'une méthode va donc dépendre de l'application visée selon le compromis entre performance, simplicité, confort et quantité de données d'apprentissage disponible. Toutes ces méthodes seront implémentées puis comparées sur la base de données **XM2VTS**, **Casia 3D** et **Casia (3D expression)** dans les **chapitres 5 et 6**. L'utilisation de systèmes biométriques multimodaux a été encouragée par la menace d'usurpation d'identité, où l'on estime qu'un système monomodal est insuffisant pour authentifier les individus. La multi modalité repose sur des techniques de fusion. La multi modalité est l'utilisation de plusieurs systèmes biométriques. La combinaison de plusieurs systèmes a pour objectif d'en diminuer les limitations. En effet, l'utilisation de plusieurs systèmes a pour premier but d'améliorer les performances de reconnaissance. En augmentant la quantité d'information discriminante de chaque personne, on souhaite augmenter le pouvoir de reconnaissance du système.

Plusieurs méthodes non linéaires de reconnaissance ont été présentées tout au long de ce chapitre tel que la théorie des noyaux reproduisant a permis le développement fulgurant d'une classe d'algorithmes de reconnaissance des visages dont la formulation ne dépend pas de la nature des données traitées, ni de l'espace de représentation adopté pour résoudre les problèmes. Au delà de ce caractère universel, celles que l'on range désormais sous le qualificatif de méthodes à noyau doivent également leur succès à l'essor de la théorie statistique de l'apprentissage, au sein de laquelle la prédiction de leurs performances en généralisation fait aujourd'hui encore l'objet d'études approfondies.

Chapitre 4

Méthodes de Fusion et Normalisation.

4.1 Introduction

Ajouter une modalité à un système biométrique, c'est ajouter une nouvelle source d'information. C'est pourquoi les systèmes multimodaux permettent d'obtenir de meilleurs résultats que les systèmes mono modaux correspondants. Ajouter des modalités permet également d'augmenter l'universalité du système. Car si l'utilisateur est réfractaire à une caractéristique biométrique donnée, l'information récoltée sur les autres modalités peut compenser la faible qualité de l'acquisition. La fusion biométrique est possible à chacun de ces niveaux : aux deux premiers niveaux, elle consiste souvent en une normalisation des données suivie d'une concaténation [289]; au niveau de la comparaison, elle se traduit par une combinaison des scores; enfin, au niveau de la décision, elle consiste en une confrontation des résultats de classification qui aboutit à une décision finale [290]. Nous traitons la question de la fusion et ses différents niveaux. Particulièrement la **fusion** des **caractéristiques** et des **scores** qui font objet de notre travail. Les principales méthodes de **normalisation des scores**, de fusion par **combinaison** et **classification** sont étudiées.

4.2 Les types de fusion

La fusion d'éléments biométriques peut se référer à de nombreux scénarios différents.

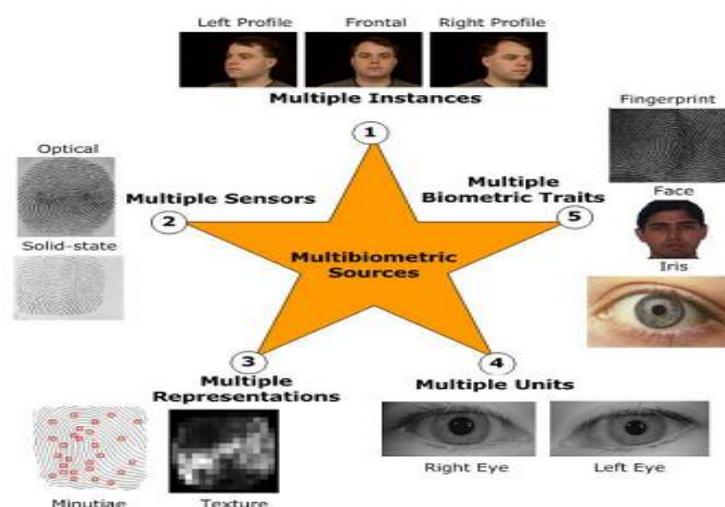


Figure 4.1 Sources de différents types de fusion de traits biométriques [3].

- **Systèmes multi algorithmes:** C'est le type de système le plus classique implicitement utilisé par de nombreuses approches. Les caractéristiques sont extraites via différents

algorithmes puis fusionnées. La fusion de caractéristiques extraites via un algorithme analysant les textures et un autre la forme d'un caractère biométrique entre dans ce cadre.

- **Systèmes multi échantillons:** Un capteur unique peut capturer plusieurs instances du même caractère biométrique dans le but de rendre plus robuste l'extraction des caractéristiques ou d'enrichir le modèle biométrique d'une personne. C'est le cas, par exemple, de plusieurs captures de visage d'une personne sous différents angles. L'utilisation de vidéos entre également dans ce cadre.
- **Systèmes multi capteurs:** Plusieurs capteurs permettent de capturer le même caractère biométrique sous différents angles. Ainsi la capture d'un visage à l'aide d'une caméra classique et d'une caméra infrarouge entre dans ce scénario. Ce type de système permet notamment la fusion au niveau capteur, ce que ne permettent pas d'autres systèmes comme les systèmes multi caractères.
- **Systèmes multi instances:** Ce type de système permet de capturer plusieurs instances du même caractère biométrique. L'acquisition de plusieurs empreintes digitales via le même capteur est l'exemple typique de ce type de système. Ces systèmes n'entraînent pas de surcoût de capteurs, ni le développement de nouveaux algorithmes. À ne pas confondre avec les systèmes multi échantillons.
- **Systèmes multi caractères:** Ce type de système combine différents traits biométriques d'un individu. Les fusions visage iris, ou visage empreinte digitale font partie de ce type d'approche. Ces systèmes nécessitent différents capteurs ainsi que des algorithmes dédiés à chaque caractère biométrique. Ce type de système a comme principale caractéristique que les caractères biométriques considérés peuvent être plus décorrélés que pour les systèmes multi capteurs.

La fusion de données issues de visages capturés via une caméra en lumière visible et une autre en lumière infrarouge entre dans le cadre des systèmes multi capteurs, où il est considéré que les deux captures sont issues de modalités différentes. Même si les deux captures sont sensiblement décorrélées (la chaleur émise par un visage n'est pas visible en lumière visible), la fermeture des yeux d'un individu est visible sur les deux modalités. À noter la présence de systèmes hybrides combinant plusieurs scénarios. Une revue de nombreux systèmes biométriques multimodaux développés peut être trouvée dans [291].

4.3 L'architecture des systèmes multimodaux.

Les systèmes multimodaux associent plusieurs systèmes biométriques et nécessitent donc l'acquisition et le traitement de plusieurs données. L'acquisition et le traitement peuvent se faire successivement, on parle alors **d'architecture en série**, ou simultanément, on parle alors **d'architecture en parallèle**.

L'architecture est en réalité surtout liée au traitement. En effet, l'acquisition des données biométriques est en général séquentielle pour des raisons pratiques. Il est difficile d'acquérir en même temps une empreinte digitale et une image d'iris dans de bonnes conditions. Il existe cependant certains cas où les acquisitions peuvent être faites simultanément lorsque les différentes données utilisent le même capteur par exemple les capteurs d'empreintes multi-doigts qui permettent d'acquérir plusieurs doigts simultanément ou même les empreintes palmaires. L'architecture est donc en général liée au traitement et en particulier à la décision. En effet la différence entre un système multimodal en série et un système multimodal en parallèle réside dans le fait d'obtenir un score de similarité à l'issue de chaque acquisition (fusion en série) ou de procéder à l'ensemble des acquisitions avant de prendre une décision (fusion en parallèle).

L'**architecture en parallèle (figure 4.3)** est la plus utilisée car elle permet d'utiliser toutes les informations disponibles et donc d'améliorer les performances du système. En revanche, l'acquisition et le traitement d'un grand nombre de données biométriques est coûteux en temps et en matériel, et réduit le confort d'utilisation. C'est pour cela que l'**architecture en série (figure 4.3)** peut être privilégiée dans certaines applications ; par exemple si la multimodalité est utilisée pour donner une alternative pour les personnes ne pouvant pas utiliser l'empreinte digitale. Pour la majorité des individus seule l'empreinte est acquise et traitée mais pour ceux qui ne peuvent pas être ainsi authentifiés on utilise un système à base d'iris alternativement.

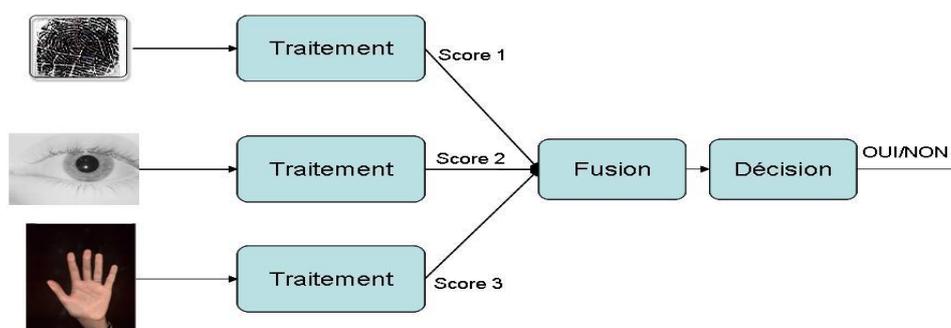


Figure 4.2 Architecture de fusion en parallèle

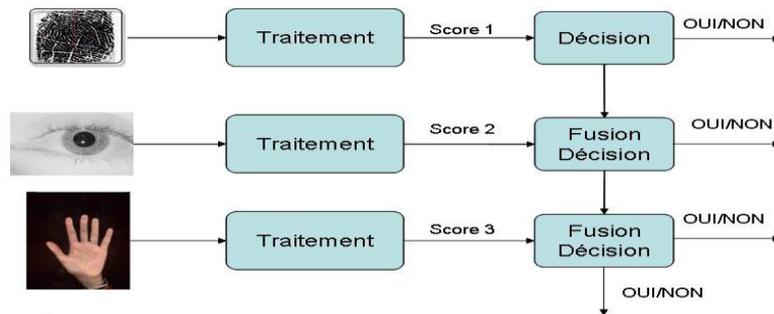


Figure 4.3 Architecture de fusion en série (incrémentale ou séquentielle)

4.4 Les niveaux de fusion [55]

La combinaison de plusieurs systèmes biométriques peut se faire à quatre niveaux différents : au niveau des **données**, au niveau des **caractéristiques** extraites, au niveau des **scores** issus du module de comparaison ou au niveau des **décisions** du module de décision (figure 4.4).

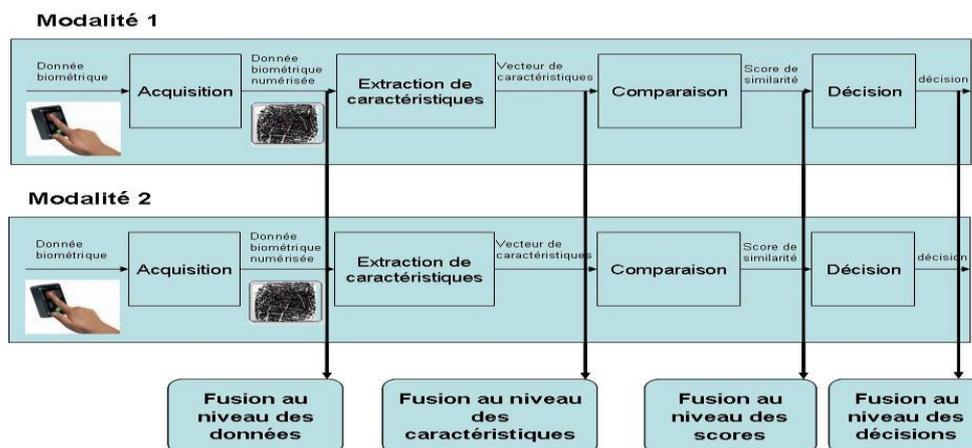


Figure 4.4 Les différents niveaux de fusion

Ces quatre niveaux de fusion peuvent être classés en deux sous-ensembles :

- la fusion pré-classification (avant comparaison),
- la fusion post-classification (après la comparaison).

4.4.1 La fusion pré-classification

La fusion pré-classification correspond à la fusion des informations issues de plusieurs données biométriques au niveau du capteur (images brutes) ou au niveau des caractéristiques extraites par le module d'extraction de caractéristiques.

1) Niveau du capteur (*Sensor Level*)

La fusion au niveau **capteur** est relativement peu utilisée car se faire uniquement si les diverses captures sont des instances du même trait biométrique obtenu à partir de **plusieurs capteurs** compatibles entre eux ou **plusieurs instances** du même trait biométrique obtenu à

partir d'un seul capteur. De plus, les captures doivent être compatibles entre elles et la correspondance entre les points dans les données brutes doit être connue par avance. Par exemple, les images de visage obtenues à partir de plusieurs caméras peuvent être combinées pour former un modèle **3D** du visage. Un autre exemple de fusion au niveau capteur consiste à mettre en mosaïque plusieurs images d'empreintes digitales afin de former une image d'empreinte digitale finale plus complexe. La fusion au niveau capteur n'est généralement pas possible si les instances des données sont incompatibles (par exemple, il est peut être difficile de fusionner des images de visages provenant de caméras ayant des résolutions différentes).

2) Niveau Caractéristiques (*Feature Level*)

La fusion au niveau des caractéristiques est moins limitée par la nature des données biométriques. Cependant une certaine homogénéité est nécessaire pour la plupart des méthodes de fusion au niveau des caractéristiques comme par exemple la moyenne de plusieurs "templates" d'empreintes ou de visage. Un exemple de fusion au niveau des caractéristiques qui ne nécessitent pas vraiment d'homogénéité est la concaténation de plusieurs vecteurs de caractéristiques avant le traitement par l'algorithme de comparaison. Par exemple, dans [292], **Jing et al** proposent une méthode de fusion de caractéristiques pour de la fusion de visage et d'empreinte palmaire. La fusion est effectuée par concaténation d'images obtenues par transformée de Gabor sur les images de visage et d'empreinte de la main. Mais la concaténation pose le problème de la dimension de l'espace de classification qui lorsqu'il augmente, rend plus difficile la tâche de classification.

Les méthodes de fusion pré-classification **sont assez peu utilisées** car elles posent un certain nombre de contraintes qui ne peuvent être remplies que dans certaines applications très spécifiques. En revanche, la fusion post-classification est très étudiée par les chercheurs.

4.4.2 La fusion post-classification

La fusion post-classification peut se faire au niveau des scores issus des modules de **comparaison** ou au niveau des **décisions**. Dans les deux cas, la fusion est en fait un problème bien connu de la littérature sous le nom de "Multiple Classifier systems".

1) Niveau Décision (*Decision Level*)

La fusion au niveau des décisions est souvent utilisée pour sa simplicité. En effet, chaque système fournit une décision binaire sous la forme **OUI** ou **NON** que l'on peut représenter par **0** et **1**, et le système de fusion de décisions consiste à prendre une décision finale en fonction de cette série de **0** et de **1**. Les méthodes les plus utilisées sont des méthodes

à base de **votes** telles que le **OR** (si un système a décidé 1 alors OUI), le **AND** (si tous les systèmes ont décidé 1 alors OUI) ou le vote à la **majorité** (si la majorité des systèmes ont décidé 1 alors OUI). On peut également utiliser des méthodes plus complexes qui pondèrent les décisions de chaque sous-système ou qui utilisent des classifieurs dans l'espace de décisions telles que **BKS** (Behaviour Knowledge Space). Dans [293], **Verlinde** présente un grand nombre de méthodes de fusion de décision. Ces méthodes de fusion au niveau des décisions sont très simples mais utilisent très peu d'information (0 ou 1).

2) Niveau Score (*Score Level*)

La fusion au niveau des scores est le type de fusion le plus utilisé car elle peut être appliquée à tous les types de systèmes (contrairement à la fusion pré-classification), dans un espace de dimension limité (un vecteur de scores dont la dimension est égale au nombre de sous-systèmes), avec des méthodes relativement simples et efficaces mais traitant plus d'information que la fusion de décisions. La fusion de scores consiste donc à la classification : **OUI** ou **NON** pour la décision finale, d'un vecteur de nombres réels dont la dimension est égale au nombre de sous-systèmes. Il existe un grand nombre de méthodes de fusion de scores qui seront présentée dans le paragraphe 4.3.5.

4.5 Les approches de fusion [55]

Nous allons maintenant nous intéresser aux méthodes de **fusion de scores**. Les méthodes de fusion de scores combinent les informations au niveau des scores issus des modules de comparaison comme indiqué sur la **figure 4.5**.

Un système de fusion est constitué de deux modules, un module de **fusion** et un module de **décision** (voir **figure 4.5**). 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 de sous-systèmes du système **multi-algorithmes** .

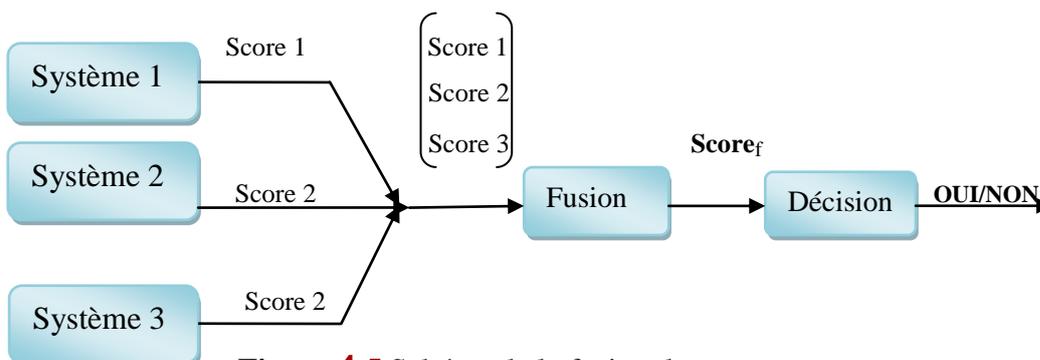


Figure 4.5 Schéma de la fusion de scores.

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

La première approche est de traiter le sujet comme un problème de **combinaison**, tandis que l'autre approche est de voir cela comme un problème de **classification**. Il est important de noter que **Jain et al.** ont montrés que les approches par combinaison sont plus performantes que la plupart des méthodes de classification [294].

1) **Approche par combinaison de scores** : consiste à traiter le sujet comme un problème de **combinaison de scores** par des méthodes mathématiques de combinaison. Dans **l'approche par combinaison**, les scores individuels sont combinés de manière à former un unique score qui est ensuite utilisé pour prendre la décision finale. Afin de s'assurer que la combinaison de scores provenant de différents systèmes soit cohérente, les scores doivent d'abord être transformés dans un domaine commun : on parle alors de **normalisation de score**. [291]

On peut par exemple citer:

Kittler et al. [296] qui ont développés un cadre théorique pour combiner les informations d'identification obtenues à partir de plusieurs classifieurs en utilisant des méthodes simples de combinaison telles que la règle somme ("**sum rule**"), la règle "produit" ("**product rule**"), la règle maximum ("**max rule**"), la règle minimum ("**min rule**") et la règle médiane ("**median rule**"). Afin d'employer ces schémas, les scores de correspondance doivent être convertis en probabilités à posteriori conformément à un utilisateur authentique ou imposteur. Ils considèrent le problème de classer un modèle d'entrée X dans l'une des m classes possibles (dans un système de vérification, $m = 2$) d'après l'information d'identification fournie par R classifieurs ou matchers différents.

Kwak et al [297] réalisent l'intégration floue de **Choquet** pour deux systèmes multi classifieurs. Le premier système combine des images traits du visage (yeux, nez et la bouche) et image globale de visage. Le deuxième système est basé sur des classifieurs opérant sur quatre ensembles de sous-image générée par décomposition en ondelettes.

Rasheed et al [298] utilisent l'intégrale floue de **Sugeno** pour la combinaison de scores pour un système multi classifieurs pour la décomposition d'un signal électromyographique (EMG).

Chia et al [299] utilisent une méthode hybride de calcul du minimum, maximum ou somme pour la combinaison des scores de deux systèmes d'authentification de visages et voix. **Yong**

Li et al [300] utilisent la somme pondérée pour la fusion par combinaison des scores. **Shukla et al** [301] proposent un calcul adaptatif et hybride utilisant la combinaison de scores par **logique floue** à base de l'intégrale de **Sugeno** ou de **Choquet**.

2) **L’approche par classification de scores** : consiste à voir cela comme un problème de *classification des scores*. Plusieurs classifieurs ont été utilisés pour classifier les scores de correspondance afin d’arriver à une décision. Dans **l’approche par 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 : Client ou Imposteur. En général, le classifieur utilisé pour cette opération est capable d’apprendre la frontière de décision sans tenir compte de la manière dont le vecteur de caractéristiques a été généré. Ainsi, les scores en sortie de différents systèmes peuvent être non-homogènes (mesure de distance ou de similarité, différents intervalles de valeurs prises, etc...) et aucun traitement n’est requis avant de les envoyer dans le classifieur. On peut citer :

Verlinde et Chollet [302] combinent les scores provenant de deux modules de reconnaissance faciale et d’un module de reconnaissance de la parole avec l’aide de trois classifieurs : un premier classifieur utilisant la méthode des "k plus proches voisins" ("knearest neighbor algorithm", "**k-NN**") avec une quantification vectorielle, un deuxième classifieur basé sur **un arbre décisionnel** et un dernier classifieur basé sur un modèle de **régression logistique**.

Chatzis et al. [303] utilisent une méthode de regroupement ("clustering") appelée "**fuzzy k-means**" et une quantification vectorielle floue ("**fuzzy vector quantization**"), couplée à un classifieur de réseau de neurones de **RBF** médiane pour fusionner les scores obtenus à partir de systèmes biométriques basés sur des caractéristiques visuelles (visage) et acoustiques (voix).

Sanderson et al. [304] utilisent un classifieur basé sur une **Machine à Vecteurs de Support (SVM)** pour combiner les scores issus d’un module de reconnaissance faciale et d’un module de reconnaissance de la parole. Ils montrent que la performance d’un tel classifieur se détériore en la présence de conditions de bruit en entrée. Afin de surmonter ce problème, ils implémentent des classifieurs résistants au bruit structurel comme un classifieur linéaire défini par morceau ("**piece-wise linear classifier**") et un **classifieur Bayésien modifié**.

Wang et al. [305] qui considèrent que les scores provenant de modules de reconnaissance faciale et de reconnaissance par l’iris comme un vecteur de caractéristiques à deux dimensions. Une analyse discriminante linéaire de Fisher (**LDA**) et un classifieur par réseau de neurones combiné à une fonction de base radiale (**RBF**) sont ensuite utilisés pour la classification.

Ross et Jain [306] utilisent un **arbre décisionnel** et des **classifieurs discriminants linéaires** pour combiner les scores des modalités du visage, d'empreinte digitale et de géométrie de la main.

4.6 Normalisation de scores [55]

Les méthodes de normalisation de scores ont pour objectif de transformer individuellement chacun des scores issus des systèmes pour les rendre **homogènes** avant de les combiner. En effet, les scores issus de chaque système peuvent être de nature différente. Certains systèmes produisent des scores de **similarité** (plus le score est grand, plus la référence ressemble au test, donc l'utilisateur est un Client), d'autres produisent des **distances** (plus la distance est faible, plus la référence et le test sont proches, plus l'utilisateur est un Client). De plus chaque système peut avoir des intervalles de variations des scores différents, par exemple pour un système les scores varient entre 0 et 1 et pour un autre les scores varient entre 0 et 1000. On comprend bien la nécessité de normaliser les scores avant de les combiner. Les méthodes de normalisation présentées dans la suite, traitent des scores qui varient déjà tous dans le même sens (en général on considère tous les scores sous forme de similarité). Pour transformer des distances en similarité il existe deux solutions : l'inverse ou l'opposé. Dans toute la suite, nous considérerons que tous les scores à fusionner ont été transformés en scores de similarité (**scores Client > scores Imposteur**).

Les différentes techniques de normalisation de scores sont :

- Normalisation par la méthode **Min-Max** ;
- Normalisation par une fonction quadratique-linéaire-quadratique (**QLQ**) ;
- Normalisation par la méthode **Z-Score** ;
- Normalisation par la médiane et l'écart absolu médian (**MAD**) ;
- Normalisation par la méthode **tangente hyperbolique "Tanh"** ;
- Normalisation par une fonction **double sigmoïde**.

1 Normalisation par la méthode Min-Max

Cette méthode n'est pas robuste (insensible aux valeurs aberrantes). La normalisation **Min-Max** conserve la distribution de scores originale à un facteur d'échelle près et transforme tous les scores dans l'intervalle [0,1]. Le score normalisé Min-Max pour le score de test s_{ik} est donné par (4.1):

$$S'_{ik} = \frac{S_{ik} - \min(\{S_i\})}{\max(\{S_i\}) - \min(\{S_i\})} \quad (4.1)$$

s_{ik} le $k^{\text{ème}}$ score de sortie du $i^{\text{ème}}$ système, où $i = 1, 2, \dots, N$ et $k = 1, 2, \dots, M$

2 Normalisation par la méthode Z-Score

Cette méthode n'est pas robuste. De plus, la normalisation *Z-Score* ne garantit pas un intervalle commun pour les scores normalisés provenant de différents systèmes. Pour une distribution arbitraire, la moyenne et l'écart-type sont respectivement des estimateurs raisonnables de position et d'échelle, mais ne sont pas optimaux.

$$S'_{ik} = \frac{S_{ik} - \mu}{\sigma} \quad (4.2)$$

Où μ est la moyenne arithmétique et σ l'écart-type des données.

3 Normalisation par la médiane et l'écart absolu médian (MAD)

C'est une méthode robuste. Cependant, les estimateurs issus de la médiane et de la MAD ont une faible efficacité comparée aux estimateurs issus de la moyenne et de l'écart-type, c'est-à-dire que lorsque la distribution de score n'est pas gaussienne, la médiane et la MAD sont de pauvres estimateurs des paramètres de position et d'échelle. Ainsi, cette technique de normalisation ne conserve pas la distribution d'entrée et ne transforme pas les scores dans un intervalle commun.

$$S'_{ik} = \frac{S_{ik} - \text{median}}{\text{MAD}} \quad (4.3)$$

$$\text{MAD} = \text{median}(\{|s_i - \text{median}(\{s_i\})|\})$$

4 Normalisation par une fonction Quadratique-Linéaire-Quadratique (QLQ)

Snelick et al. [307] utilisent une fonction Quadratique-Linéaire-Quadratique (QLQ) pour normaliser des scores au préalable transformés dans l'intervalle $[0, 1]$ (s_{MM}) grâce à une normalisation *MinMax* (s_{MM}). Cette normalisation QLQ prend comme paramètres le centre c et la largeur w de la zone de recouvrement des distributions des scores Imposteurs et Client.

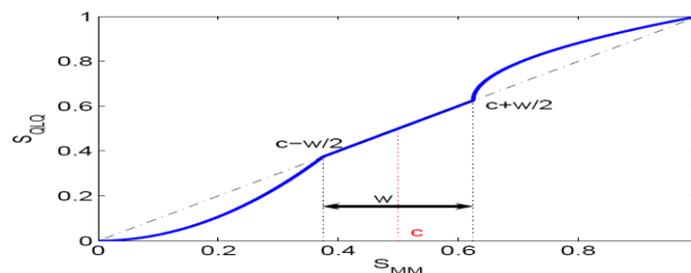


Figure 4.6 Normalisation QLQ.

La zone de recouvrement reste inchangée tandis que les autres régions sont transformées à l'aide de deux fonctions quadratiques par segment. Le score normalisé est donné par (4.4):

$$s_{QLQ} \begin{cases} \frac{1}{c-\frac{\omega}{2}} S_{MM}^2, & \text{si } s_{MM} \leq \left(c - \frac{\omega}{2}\right) \\ S_{MM}, & \text{si } \left(c - \frac{\omega}{2}\right) < s_{MM} \leq \left(c + \frac{\omega}{2}\right) \\ \left(c + \frac{\omega}{2}\right) + \sqrt{\left(1 - c - \frac{\omega}{2}\right) \left(s_{MM} - c - \frac{\omega}{2}\right)}, & \text{sinon} \end{cases} \quad (4.4)$$

5 Normalisation par la méthode tangente hyperbolique "Tanh"

Les scores normalisés sont donnés par (4.5):

$$s'_{ik} = \frac{1}{2} \left\{ \tanh\left(0.001 \frac{s_{ik} - \mu}{\sigma}\right) + 1 \right\}, \quad (4.5)$$

Où μ est la moyenne arithmétique, σ l'écart-type des données et \tanh la tangente hyperbolique. La méthode tangente hyperbolique met chaque score normalisé dans l'intervalle $[0, 1]$.

6 Normalisation par une fonction double sigmoïde

Cappelli et al. [308] ont utilisé une **fonction double sigmoïde** pour la normalisation de score dans un système biométrique multimodal qui combine différents systèmes d'empreinte digitale. Le score normalisé est donné par (4.6) :

$$s'_{ik} = \begin{cases} \frac{1}{1 + \exp\left(-2\left(\frac{s_{ik}-t}{r_1}\right)\right)} & \text{si } s_k < t, \\ \frac{1}{1 + \exp\left(-2\left(\frac{s_{ik}-t}{r_2}\right)\right)} & \text{sinon} \end{cases} \quad (4.6)$$

Où t est le point de fonctionnement de référence et r_1 et r_2 sont respectivement les bords gauche et droit, la **figure 4.7** montre un exemple d'une normalisation double sigmoïde, où les scores étant dans l'intervalle $[0, 300]$ sont transformés dans l'intervalle $[0, 1]$, avec $t = 200$, $r_1 = 20$ et $r_2 = 30$.

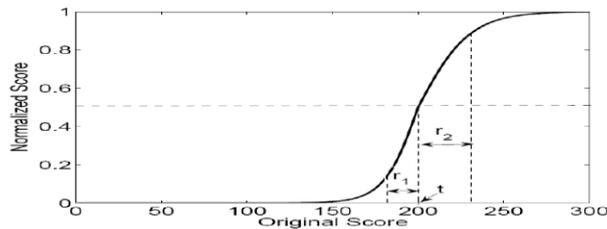


Figure 4.7 Normalisation double sigmoïde.

En général, t est choisi de façon à être une valeur appartenant à la région de recouvrement entre les distributions des scores client et imposteurs, et r_1 et r_2 sont fixés de telle sorte qu'ils correspondent respectivement aux extrémités de recouvrement entre les deux distributions vers la gauche et vers la droite de t (on peut également voir cela comme les valeurs correspondant respectivement au minimum des scores de similarité des clients et au

maximum des scores de similarité des imposteurs). Ce schéma de normalisation fournit donc une transformation linéaire des scores dans la région de recouvrement, tandis que les scores à l'extérieur de cette région sont transformés de manière non linéaire.

4.7 Approche par combinaison de scores

4.7.1 Méthode de combinaisons simples

Les méthodes de combinaisons de scores simples sont des méthodes très simples dont l'objectif est d'obtenir un score final s à partir des N scores disponibles s_i pour $i = 1$ à N issus de N systèmes.

Les méthodes les plus utilisées sont la moyenne, le produit, le minimum, le maximum ou la médiane.

- Combiner les scores par la moyenne consiste à calculer s tel que

$$s = \frac{1}{N} \sum_{i=1}^N s_i \quad (4.7)$$

- Combiner les scores par le produit consiste à calculer s tel que

$$s = \prod_{i=1}^N s_i \quad (4.8)$$

- Combiner les scores par le minimum consiste à calculer s tel que

$$s = \min (s_i) \quad (4.9)$$

- Combiner les scores par le maximum consiste à calculer s tel que

$$s = \max (s_i) \quad (4.10)$$

- Combiner les scores par la médiane consiste à calculer s tel que

$$s = \text{med} (s_i) \quad (4.11)$$

- La somme pondérée c'est une méthode un peu plus évoluée qui nécessite une adaptation par le réglage de paramètres.

$$s = \sum_{i=1}^N \omega_i s_i \quad (4.12)$$

La somme pondérée permet de donner des poids différents ω_i à chacun des systèmes en fonction de leur performance individuelle ou de leur intérêt dans le système multi-algorithmes.

4.7.2 Combinaison de scores par logique floue

La théorie de la **logique floue** (des sous-ensembles flous) a été introduite par Zadeh en 1965 [309] comme une extension de la logique binaire d'une part et une amélioration de la

logique multivaluée (admettant plusieurs valeurs de vérité) d'autre part. L'importance de la logique floue réside dans le fait qu'elle s'approche du **raisonnement humain** par l'intégration et le traitement du caractère approximatif, vague, imprécis ou flou de la connaissance humaine. Les termes linguistiques tels que **environ**, **moyenne**, **approximativement** sont de nature à donner un caractère flou aux phrases énoncées. Par exemple, la règle « si le prix est inférieur à 6000 DA, j'achète » sera intuitivement utilisable si le prix est de 6002 DA, mais elle ne pourrait être exploitée en logique classique puisque le prix indiqué ne satisferait pas la prémisse.

Dans [22] on réalise l'**intégration floue de Choquet** pour deux approches de fusion :

- La **première approche** est de combiner des images des traits du visage (yeux, nez et la bouche) et images globales de visage.
- La **deuxième approche** est à l'agrégation des classificateurs opérant sur quatre ensembles de sous-image générée par décomposition en ondelettes.

Mesure floue

Un jeu de fonction $g : P(Y) \rightarrow [0, 1]$ est appelée une **mesure floue** si les conditions suivantes sont remplies :

1. conditions aux limites: $g(\phi) = 0$, $g(Y) = 1$
2. monotonie : $g(A) \leq g(B)$, si $A \subset B$ et $A, B \in P(Y)$
3. continuité : $\lim_{i \rightarrow \infty} g(A_i) = g(\lim_{i \rightarrow \infty} A_i)$, si $\{A_i\}_{i=1}^{\infty}$ est une suite croissante d'ensembles mesurables.

A partir de cette définition, **Sugeno** [310] a introduit un soi-disant g_λ **mesure floue** qui est livré avec une propriété supplémentaire

$$g(A \dot{\cup} B) = g(A) + g(B) + \lambda g(A)g(B) \quad (4.13)$$

Pour tous les $A, B \subset Y$ et $A \cap B = \phi$, et pour certains $\lambda > -1$. Évidemment quand $\lambda=0$, le g_λ mesure floue devient une mesure de probabilité standard.

En général, la valeur de λ peut être déterminée en raison de l'état limite de la mesure floue g_λ . Cette condition pour $g(Y) = 1$. Par conséquent, la valeur de λ est déterminée par la résolution de ce qui suit:

$$g_\lambda(Y) = \frac{1}{\lambda} (\prod_{i=1}^n (1 + \lambda g^i) - 1) , \lambda \neq 0 \quad (4.14)$$

est l'équivalent de :

$$\lambda + 1 = \prod_{i=1}^n (1 + \lambda g^i) \quad (4.15)$$

Où $\lambda \in (-1, +\infty)$, $\lambda \neq 0$, et g^i est la valeur de **la fonction de densité floue**. La solution peut être facilement obtenue; évidemment on s'intéresse à la racine supérieure à -1.

2 Intégrale floue

L'intégrale floue de la fonction **h** calculée sur **Y** par rapport à une mesure floue **g** est définie sous la forme

$$\int_Y h(y) \circ g(\cdot) = \sup_{\alpha \in [0,1]} [\min[\alpha, g(\{y | h(y) \geq \alpha\})]] \quad (4.16)$$

a) Intégrale floue de Sugeno

Lorsque les valeurs des $h(\cdot)$ sont classés dans l'ordre décroissant,

$$h(y_1) \geq h(y_2) \geq \dots \geq h(y_n)$$

L'intégrale floue de **Sugeno** est calculée comme suit:

$$\int_Y h(y) \circ g(\cdot) = \max_{i=1:n} [\min(h(y_i), g(A_i))] \quad (4.17)$$

Où $A_i = \{y_1, y_2, \dots, y_i\}$ désigne un sous-ensemble d'éléments. Les valeurs de $g(A_i)$ pris en charge par la mesure floue sur les sous-ensembles correspondant d'éléments peut être déterminée de manière récursive sous la forme :

$$g(A_1) = g(y_1) = g^1 \quad (4.18)$$

$$g(A_i) = g^i + g(A_{i-1}) + \lambda g^i g(A_{i-1}), \quad (4.19)$$

Le calcul de la fonction de densité floue g^i sur la base des données sont assurées par la manière suivante

$$\begin{cases} g^i = \beta p_i, & i = 1 \\ g^i = (1 - \beta) p_i, & i = 2, 3, 4 \end{cases} \quad (4.20)$$

Où p_i est le taux de classification dans l'intervalle $[0, 1]$ pour chaque système.

$B \in [0, 1]$ est un facteur qui met en place un certain équilibre entre les résultats de la classification.

b) Intégrale floue de Choquet

Il a été démontré que (4.16) n'est pas une extension correcte de l'intégration de **Lebesgue** habituelle. En d'autres termes, lorsque la mesure est additive l'expression ci-dessus ne retourne pas l'intégrale au sens de **Lebesgue**. Afin de remédier à cet inconvénient, Murofushi et Sugeno [311] ont proposés un soi-disant intégrante floue de Choquet calculé de la manière suivante:

$$\int_Y h(y) d g(\cdot) = \sum_{i=1}^n [h(y_i) - h(y_{i+1})] g(A_i) h(y_{n+1}) = 0 \quad (4.21)$$

4.8 Approche par classification de scores

L'approche par classification de scores, cherche à prendre une décision à partir du vecteur de scores de dimension N . Le but des méthodes de fusion basées sur des classifieurs de chercher une séparatrice entre les 2 classes, **Client** et **Imposteur**, dans l'espace à N dimensions des scores. La séparatrice peut être plus ou moins complexe selon le classifieur utilisé. Toutes les méthodes classiques de reconnaissance des formes et d'apprentissage automatique peuvent être utilisées pour la fusion de scores qui est un problème de classification comme un autre. Dans le domaine des systèmes multi-algorithmes, dont l'objectif est d'utiliser un classifieur pour combiner plusieurs systèmes sources (ici ce sont des systèmes biométriques que l'on veut fusionner) afin de consolider la décision de classification, un grand nombre de méthodes ont été utilisées. Les méthodes utilisées dans notre travail sont :

- la méthode de **statistique de Fisher**,
- les machines à vecteurs de support (**SVM** : Support Vector Machine)
- les réseaux de neurones **RN**.

4.8.1 Fusion par méthode statistique de Fisher

La méthode **statistique** introduite ici, se base sur les travaux de **Fischer** [312] et fait usage d'une frontière de décision linéaire pour séparer deux populations données, à savoir les clients et les imposteurs dans notre cas. Dans le contexte particulier de l'authentification de personnes, **Duc et al** [289] utilisent eux aussi une règle de décision linéaire, mais l'optimisent en se fixant un critère dérivé de la théorie de **Bayes**. [313]

Envisageons à présent la règle de décision développée par **Fisher**. Elle se base sur le rapport de vraisemblance repris ci-dessous:

$$\frac{T(z|c)}{T(z|i)} > k \quad (4.22)$$

où k représente un seuil d'acceptation dont la valeur dépend du compromis **TFA/TFR** que l'on veut obtenir. Dans le problème qui nous préoccupe, $T(z|c)$ et $T(z|i)$ sont inconnus et doivent être estimés à partir des données d'apprentissage. Une hypothèse courante consiste à approcher les distributions réelles par des distributions normales à p variables $N_p(\mu_A, \Sigma)$, où $A = \{c, i\}$ représente la classe d'individus, μ_A le vecteur des scores moyens et Σ la matrice de covariance entre experts. En un premier temps, on supposera la matrice Σ indépendante de la classe d'individus. Sous de telles hypothèses, les fonctions de densité de probabilité s'écrivent sous la forme:

$$f_A(z) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (z - \mu_A)' \Sigma^{-1} (z - \mu_A) \right\} \quad (4.23)$$

Les paramètres μ_c , μ_i et Σ sont inconnus, mais peuvent être estimés à partir des données d'apprentissage, soit x les n_c données relatives aux accès clients et y , les n_i données relatives aux accès imposteurs (simulés). On a :

$$\hat{\mu}_c = \sum_{q=1}^{n_x} x_q / n_c \quad (4.24)$$

$$\hat{\mu}_i = \sum_{q=1}^{n_u} y_q / n_i \quad (4.25)$$

$$\hat{\Sigma}_c = \sum_{q=1}^{n_x} (x_q - \hat{\mu}_c)(x_q - \hat{\mu}_c)' / (n_c - 1) \quad (4.26)$$

$$\hat{\Sigma}_i = \sum_{q=1}^{n_y} (y_q - \hat{\mu}_i)(y_q - \hat{\mu}_i)' / (n_i - 1) \quad (4.27)$$

$$\hat{\Sigma} = [(n_c - 1)\hat{\Sigma}_c + (n_i - 1)\hat{\Sigma}_i] / (n_c + n_i - 2) \quad (4.28)$$

Notons que l'on tient compte ici, par l'intermédiaire de Σ , de $\hat{f}_c(z)/\hat{f}_i(z) \geq k$ qui peut exister entre experts. En combinant les équations (4.22) à (4.28), on peut réécrire sous la forme de $D_L(z) \geq \ln(k) = k^*$ où :

$$D_L(z) = (z - \frac{1}{2}(\hat{\mu}_c + \hat{\mu}_i))' \hat{\Sigma}^{-1} (\hat{\mu}_c - \hat{\mu}_i) \quad (4.29)$$

Fisher fut le premier à utiliser cette fonction à des fins de classification.

Comme $D_L(z)$ est linéaire en z , elle fut communément appelée **fonction linéaire discriminante** ou **LDF** (**L**inear **D**iscriminant **F**unction). Ainsi, la procédure à suivre pour vérifier l'identité d'un candidat, consiste à calculer $\hat{\mu}_c$, $\hat{\mu}_i$ et $\hat{\Sigma}$ à partir des données d'entraînement (ce qui est effectué une fois pour toutes), puis $D_L(z)$ et comparer celui-ci au seuil k^* donné. Si $D_L(z) \geq k^*$, le candidat est accepté comme client.

Dans le cas de distributions de scores clients et imposteurs ne satisfaisant pas l'hypothèse d'une covariance Σ unique, la règle de décision (4.29) peut se réécrire sous la forme $D_Q(z) \geq 2k^*$ où :

$$\begin{aligned} D_Q(z) = & (z - \hat{\mu}_i)' \hat{\Sigma}_i^{-1} (z - \hat{\mu}_i) \\ & - (z - \hat{\mu}_c)' \hat{\Sigma}_c^{-1} (z - \hat{\mu}_c) \\ & + \ln(|\hat{\Sigma}_i|/|\hat{\Sigma}_c|) \end{aligned} \quad (4.30)$$

$D_Q(z)$ est appelée **fonction quadratique discriminante** ou **QDF** (**Q**uadratic **D**iscriminant **F**unction). Cette règle permet d'obtenir de meilleurs résultats que ceux obtenus dans le cadre d'une fusion linéaire, mais requiert un ensemble d'apprentissage suffisamment étendu pour pouvoir évaluer Σ_c et Σ_i avec précision.

4.8.2 Fusion par méthode SVM

Les machines à vecteurs de support (Support Vector Machine : SVM), également appelés Séparateurs à Vaste Marge, introduits par Vapnik [288] est utilisée dans la fusion de données dans les références [16,314,315,316]. Ils ont pour but de définir un hyperplan de séparation entre les deux classes, hyperplan qui minimise l'erreur de classification sur un ensemble d'apprentissage. Cette méthode d'apprentissage supervisée peut apprendre une séparatrice plus ou moins complexe selon la nature du noyau choisi. Le noyau le plus simple est le noyau linéaire qui correspond à chercher un séparateur linéaire dans l'espace à N dimensions des scores. Le but des fonctions noyau est de transformer l'espace initial (des scores à N dimensions) en un espace de dimension **plus grande** dans lequel les données pourraient être linéairement séparables. La séparatrice est donc toujours linéaire dans l'espace transformé par la fonction noyau, mais ne l'est plus dans l'espace des scores.

Le but du SVM est donc de trouver une séparatrice qui minimise l'erreur de classification sur l'ensemble d'apprentissage mais qui sera également performante en généralisation sur des données non utilisées en apprentissage. Pour cela le concept utilisé est celui de marge (d'où le nom de séparateurs à vaste marge). La marge est la distance quadratique moyenne entre la séparatrice et les éléments d'apprentissage les plus proches de celle-ci appelés vecteurs de

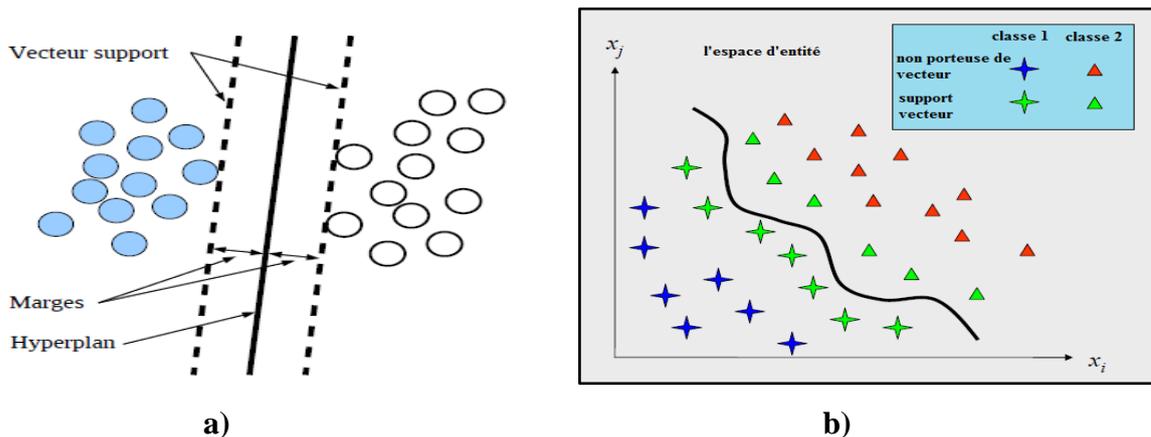


Figure 4.8 Séparation dans un espace à deux dimensions a) linéaire et b) non linéaire.

support (**figure 4.8**). Ces éléments sont appelés vecteurs de support car c'est uniquement sur ces éléments de l'ensemble d'apprentissage qu'est optimisée la séparatrice.

Le principe des SVM est présenté dans la section 3.3.2.

4.8.3 Fusion par Réseaux de Neurones Artificiels

Le principe général des **Réseaux de Neurones Artificiels (RNA)** est à l'origine inspiré de certaines fonctions de base des neurones naturels du cerveau.

Un réseau de neurones artificiel est généralement organisé en **plusieurs couches** :

- une couche d'entrée,
- une couche de sortie,
- des couches intermédiaires appelées couches cachées.

La présence de couches cachées permet de **discriminer** des classes d'objets non linéairement séparables. En général, un réseau de neurones est fondamentalement un classifieur. Il réalise un travail de classification pendant la phase d'apprentissage, et de classement lors de la reconnaissance. Mais on peut l'utiliser pour réaliser la fusion de données [317] pour séparer deux populations données, à savoir les clients et les imposteurs dans notre cas.

Les grands avantages des réseaux de neurones résident dans leur capacité **d'apprentissage automatique** (approximation universelle (**Cybenko, Hornik**)), ce qui permet de résoudre des problèmes sans nécessiter l'écriture de règles complexes, tout en étant tolérant aux erreurs. Ils résident aussi dans leur capacité à prendre une décision à partir de critères **non formalisables explicitement**.

Structure du réseau perceptron multicouche MLP (Multi-Layer perceptron)

Les réseaux monocouches ont rapidement montré des limites car ils ne permettaient de traiter ni le cas des classes non linéairement séparables, ni le cas multi-classes. Le perceptron multicouche s'est imposé par la suite comme une réponse à ces problèmes, grâce à sa propriété d'approximateur universel [318], qui le rend capable d'approcher une fonction quelconque avec une précision arbitraire. Comme son nom l'indique, sa structure est organisée en couches (voir **figure 4.9**).

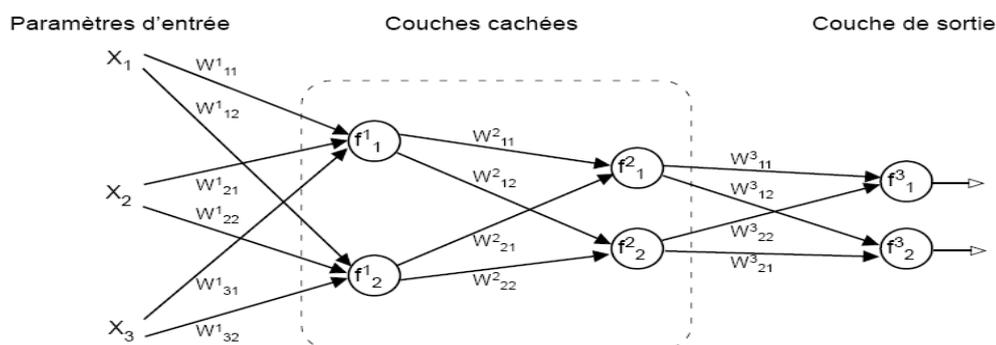


Figure 4.9 Exemple de perceptron multicouches

La **figure 4.9** illustre un exemple de perceptron multicouches avec 3 entrées, 2 couches cachées comportant 2 neurones chacune et une couche de sortie comportant 2 neurones.

Chaque neurone élémentaire est connecté à tous les neurones de la couche qui suit celle à laquelle il appartient. Les fonctions d'activation peuvent être quelconques, et le nombre de neurones en sortie dépend généralement du codage adopté. Par exemple, dans un problème de discrimination, chacun d'eux est généralement dédié à une classe donnée.

4.9 Conclusion

Ce chapitre nous a permis de connaître un certain nombre de méthodes de fusion de scores. Nous avons présenté les deux approches essentielles de fusion des scores la première est l'approche par combinaison. Cette dernière traite les scores séparément avant de les combiner et elle est basée sur des méthodes simples (les plus utilisées) telles que la moyenne, le produit, le minimum, le maximum ou la médiane ou bien par la logique floue.

Nous avons également montré qu'il y a différents niveaux de "complexité" pour les méthodes de combinaison.

La deuxième approche étudiée est basée sur la classification de scores qui cherche également à modéliser la séparatrice des classes dans l'espace multidimensionnel. Les méthodes utilisées dans notre travail sont :

- la méthode de **statistique de Fisher**,
- les machines à vecteurs de support (**SVM** : Support Vector Machine)
- les réseaux de neurones **RN**.

Une phase de normalisation préalable aux deux approches est réalisée c'est ce qui fait l'originalité de notre travail. Car dans les travaux antérieurs la normalisation se limitait à la première approche seulement (la fusion par combinaison). Cependant, il y a toujours de nombreuses questions ouvertes qui sont laissées sans réponse. Toutes ces méthodes ont des avantages et des inconvénients selon la complexité, le besoin de données d'apprentissage ou les paramètres à optimiser. Le choix d'une méthode va donc dépendre de l'application visée selon le compromis entre performance, simplicité, confort et quantité de données d'apprentissage disponible. Toutes ces méthodes ont été réalisées et comparées dans nos travaux sur la base de données **XM2VTS** [319].

Chapitre 5

Conception de l'Approche RV Monomodale.

5.1 Introduction

Dans ce chapitre, nous détaillons les principales expériences que nous avons menées durant la thèse ainsi que les résultats monomodaux (sur une seule modalité à la fois) obtenus via les trois principales approches testées :

- l'approche 2D pour laquelle des investigations sur la meilleure orientation du filtre de Gabor et une comparaison des méthodes de réduction sont réalisés. Dans cette approche nous étudions les méthodes de réduction linéaires et non linéaires,
- l'approche 3D dans laquelle la génération de cartes de profondeur est réalisée,
- l'approche 3D expression basée sur la combinaison par concaténation des régions d'intérêts et du visage.
- et enfin l'approche 3D expression basée sur la fusion des caractéristiques du visage neutre et des six expressions (dégout, joie, peur, surprise, colère,

Les deux principaux modes de test d'un système biométrique sont l'authentification (ou vérification) et l'identification. L'utilisation d'un système biométrique en authentification nécessite le calcul d'un score de similarité entre deux modalités biométriques, la décision étant ensuite prise en fonction de ce score et d'un seuil. L'utilisation d'un système d'identification nécessite le calcul de n scores de similarité. Le calcul du rang pour une image test, c'est à dire le moment à partir duquel la bonne identité est retrouvée, permet d'évaluer qualitativement le système biométrique.

Toutes les expériences menées sont en identification, c'est-à-dire que le système doit retrouver l'identité de la personne testée. Celles-ci ont été réalisées sous l'hypothèse forte que la personne testée a préalablement été enrôlée.

Dans la suite, nous utilisons parfois le terme galerie pour désigner les images utilisées lors de l'enrôlement des personnes, et le terme probe pour désigner les images utilisées pour tester la méthode d'identification.

5.2 Système de reconnaissance de visage

- a) Première expérimentation, dans un premier temps, nous avons mis en œuvre un système de reconnaissance monomodale sur le visage 2D neutre.

- b) Deuxième expérimentation, nous essayons de contourner le problème de la première expérimentation, à savoir la limite des images 2D face aux changements d'expressions. Nous tentons de faire la fusion des différentes régions d'intérêts afin de mieux identifier les personnes.
- a) Troisième expérimentation, dans cette dernière application, nous essayons d'améliorer les performances du système de reconnaissance 3D qui a des difficultés à reconnaître les visages avec expressions. Nous essayons de faire la fusion du visage neutre avec les six expressions et des résultats encourageants sont obtenus.

5.2.1 Détection de visage

Une multitude de techniques ont été développées allant de la simple détection du visage, à la localisation précise des régions caractéristiques du visage, tels que les yeux, le nez, les narines, les sourcils, la bouche, les lèvres, les oreilles, etc...



Figure 5.1 Détection de visage

Les performances globales de tout système automatique de reconnaissance dépendent amplement des performances de la détection de visages[320].

5.2.2 Prétraitement

Où les données biométriques sont traitées pour enlever le bruit lié à l'environnement ou au dispositif de capture. Pour cela nous faisons une photo normalisation qui a un double effet : d'une part elle supprime pour tout vecteur image x un éventuel décalage par rapport à l'origine et ensuite elle supprime tout effet d'amplification (multiplication par un scalaire). Pour chaque image on effectue l'opération suivante :

$$photo\ normalisation(x) = \frac{x - mean(x)}{std(x)}$$

$mean(x)$: le moyen de x et $std(x)$: l'écart type de x .

5.2.3 Extraction de caractéristiques

Dans toutes les méthodes de reconnaissance faciale, le point le plus délicat concerne la définition et l'extraction des caractéristiques faciales les plus pertinentes, à savoir les caractéristiques qui représentent le mieux les informations portées par un visage. Le choix de ces informations utiles revient à établir un modèle pour le visage, elles doivent être discriminantes et non redondantes. Le choix de caractéristiques locales présente plusieurs

avantages par rapport à des caractéristiques globales. C'est pour cette raison que les systèmes les plus récents s'appuient sur des caractéristiques faciales locales. Les plus performantes dans le contexte de reconnaissance de visages sont les ondelettes de **Gabor** et les indices **LBP Local Binary Patterns**. Ici on se limitera à la première pour les raisons suivantes :

- Les ondelettes de Gabor sont connues comme le moyen d'analyse espace-fréquence qui minimise l'incertitude de Heisenberg à savoir si on est précis dans l'analyse de l'information dans l'espace du pixel on le sera moins dans l'espace fréquentiel et vice versa.
- Cette caractéristique fait des ondelettes de Gabor un moyen puissant pour l'analyse de la texture [321].
- D'après **Morizet** qui a consacré un chapitre à la **Perception Visuelle** et les **Neurosciences** (**Chapitre 3** [3]) en se basant sur les expériences de **Pollen et Ronner** [322] et **Jones et Palmer** [323]. Le filtre de Gabor est le filtre le plus proche des **champs réceptifs linéaires (RFs)** [324]

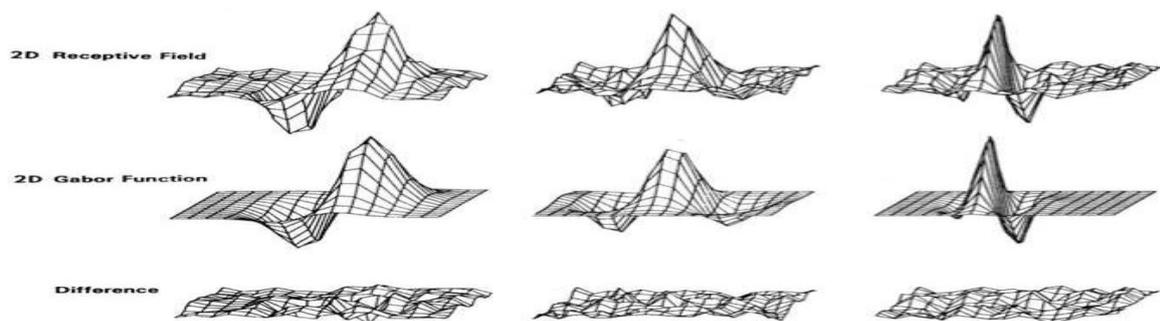


Figure 5.2 Adaptation d'un filtre de Gabor aux données correspondant aux mesures du comportement de la réponse d'une certaine cellule simple. [323]

Sur la première ligne de la **figure 5.2** on peut voir le **RF expérimental** et sur la deuxième ligne, le **RF modélisé par des ondelettes 2D de Gabor**. On peut apprécier la **très grande similitude** entre les deux entités par leur **très faible différence** représentée sur la dernière ligne. Pour chaque type de caractéristique, il y'a plusieurs façons de les utiliser. Nous présentons dans la suite seulement le principe de base. Il convient aussi de noter qu'il existe des variantes de ces caractéristiques originales.

Les ondelettes de Gabor [325]

Le filtre de **Gabor**, défini par **Dennis Gabor**, est largement utilisé en traitement d'images car les ondelettes de **Gabor** présentent deux propriétés intéressantes : la localisation fréquentielle et la sélectivité en orientation. Les représentations en fréquence et en orientation du filtre de

Gabor s'apparentent à celles du système visuel humain [323]. Les articles [326,327,328] (le premier est dans Nature) indiquent que la représentation par ondelettes de **Gabor** des images faciales est robuste aux changements causés par des variations d'éclairage ou par des modifications d'expressions faciales.

Les ondelettes en dimension deux de **Gabor** ont été introduites dans le domaine de la recherche biométrique par **Daugman** [329] pour la reconnaissance d'iris. **Ladesetal.** [131] ont utilisé les premières ondelettes de **Gabor** en reconnaissance faciale en utilisant la **Dynamic Link Architecture**. Un noyau de filtre de **Gabor** est le produit d'une onde complexe sinusoidale avec une enveloppe gaussienne. Une ondelette de **Gabor** généralement utilisée dans la reconnaissance faciale est définie comme suit :

$$\Psi_{\theta,\lambda}(z) = \frac{\|k_{\theta,\lambda}\|}{\sigma^2} e^{-\frac{\|k_{\theta,\lambda}\|^2 \|z\|^2}{2\sigma^2}} [e^{ik_{\theta,\lambda}z} - e^{-\frac{\sigma^2}{2}}] \quad (5.1)$$

où $\mathbf{z} = (\mathbf{x}, \mathbf{y})$ est le point de coordonnées (\mathbf{x}, \mathbf{y}) dans le plan image. Les paramètres θ et λ définissent l'orientation et la fréquence des noyaux de **Gabor**. $\|\cdot\|$ est l'opérateur norme, et σ l'écart-type de l'enveloppe gaussienne. Le paramètre d'orientation θ (ou θ) fait pivoter l'ondelette autour de son centre. Il détermine l'angle des contours ou des lignes de l'image auxquelles l'ondelette va être sensible. Le paramètre λ (ou r) spécifie la longueur d'onde du cosinus ou inversement la fréquence centrale de l'ondelette. Les ondelettes avec une grande longueur d'onde seront sensibles à des changements progressifs d'intensité dans une image. Les ondelettes avec une petite longueur d'onde seront sensibles à des contours et des lignes abruptes. La représentation en ondelettes de **Gabor** d'une image résulte du produit de convolution de l'image avec une famille de noyaux de **Gabor** de fréquences et d'orientations différentes. La convolution de l'image I et d'un noyau de **Gabor** $\Psi_{\theta,\lambda}(z)$ est définie par :

$$G_{\theta,\lambda}(z) = I(z) * \Psi_{\theta,\lambda}(z) \quad (5.2)$$

Comme on peut le remarquer dans l'équation 5.1, les filtres de **Gabor** ont une forme complexe, il est important d'utiliser les informations données par la partie réelle et la partie imaginaire des coefficients de **Gabor**. Deux choix triviaux s'offrent à nous l'étude de l'amplitude et l'étude de la phase de **Gabor**. L'amplitude et la phase seront notées respectivement par :

$$M(IG_{r,\theta}(x, y)) = \sqrt{\text{Im}(IG_{r,\theta}(x, y))^2 + \text{Real}(IG_{r,\theta}(x, y))^2}$$

$$P(IG_{r,\theta}(x, y)) = \arctan\left(\frac{\text{Im}(IG_{r,\theta}(x, y))}{\text{Real}(IG_{r,\theta}(x, y))}\right)$$

L'intérêt d'utiliser les filtres de **Gabor** pour extraire des caractéristiques faciales est qu'ils permettent de capturer les informations de visage dans des orientations et des résolutions différentes. De plus, ils sont robustes aux changements d'illumination, aux distorsions et aux variations d'échelle. En effet, la convolution d'une image avec une banque de **40 noyaux de Gabor** (5 échelles et 8 orientations) conduit à 40 cartes d'amplitude et 40 cartes de phase qui sont de même taille que l'image originale, comme illustrée sur la **figure 5.6**. Par conséquent, si on ne considère que la réponse en amplitude, chaque pixel est décrit par un vecteur de dimension 40. Ce vecteur de dimension 40 est également appelé « jet » et il a été utilisé largement dans les systèmes les plus anciens, tels que le **DLA** et les plus récents **EGBM**[330]. Notons que ce sont des méthodes basées sur des points caractéristiques qui doivent être détectés très précisément. Dans les méthodes telles que [279], les auteurs ont utilisé directement une image de caractéristiques contenant les 40 cartes d'amplitude comme entrée de reconnaissance. Dans [57], **Mellakh** utilise les réponses en phase comme indice facial. Bien entendu, des méthodes de réduction de dimension ont été utilisées, telles que le sous-échantillonnage ou l'**ACP** avec ses variantes. Plusieurs métriques ont été testées pour les caractéristiques basées sur les filtres de **Gabor** et celle qui est la plus souvent utilisée est la distance cosinus.

5.3 Evaluation expérimentale de la reconnaissance monomodale 2D

5.3.1 Base de données

Nos expériences sont exécutées sur des images frontales de visage de la base de données **XM2VTS**. C'est une base multimodale développée au sein du projet européen **ACTS**, elle est utilisée pour la vérification d'identité, elle contient 8 images par visage de 295 personnes. La base de données **XM2VTS**[331] a été réalisée sur une longue période ce qui a permis d'obtenir plusieurs images d'une même personne et donc une grande variabilité d'apparence (changement de coiffure, présence et/ou absence de lunettes, etc...). Par contre, seules les expressions neutres ont été considérées. Pour la tâche de vérification, un protocole standard d'estimation de performances a été développé (appelé «Lausanne Protocol Splits Randomly»). Il existe deux configurations différentes, la configuration I et la configuration II. Nous n'utiliserons que la **configuration I** puisqu'elle est la plus complexe. Elle consiste à séparer les individus en deux classes, Client et Imposteur. Le groupe Client contient 200 sujets, alors que le groupe Imposteur est divisé en 25 imposteurs pour l'évaluation et 70 imposteurs pour le test, huit images des quatre sessions sont utilisées.



Figure 5.3 Exemples des images de la base de données XM2VTS
 (a) personnes différents, (b) même personne.

La répartition des photos dans les différents ensembles est présentée dans le **tableau 5.1**.

Tableau 5.1 Répartition des photos dans les différents ensembles.

Ensemble	Clients	Imposteurs
Apprentissage	600 images (3 par personne)	0 images
Evaluation	600 images (3 par personne)	200 images (8 par personne)
Test	200 images (2 par personne)	560 images (8 par personne)

Notre travail est exécuté avec un ordinateur : Intel Pentium Dual CPU 2.2 GHz , 1.49 Go de RAM et le langage de programmation utilisé est le **MATLAB R2009b**.

5.3.2 Architecture de l'approche monomodale 2D

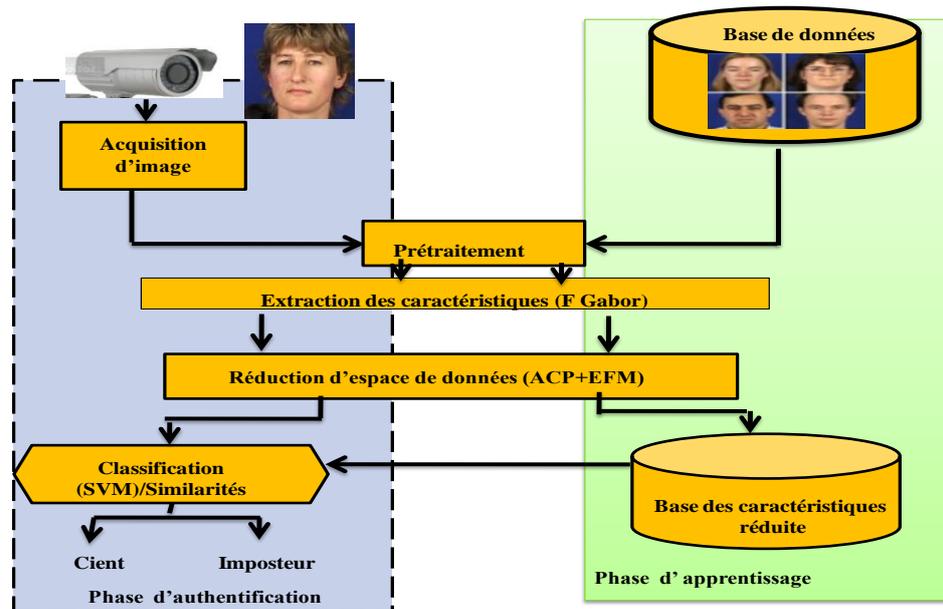


Figure 5.4 Schéma de conception de notre approche monomodale.

5.4 Représentation du visage et choix des filtres de Gabor

5.4.1 Représentation du visage : Chaque image est constituée de plusieurs informations comme la couleur, l'arrière plan, les cheveux, les cols de chemise, les oreilles etc... Pour cela, la première étape nécessaire est de découper l'image par une fenêtre rectangulaire de taille **161x121** centrée autour des caractéristiques les plus stables liées aux yeux, aux sourcils, au nez et à la bouche. Une décimation de facteur 2 pour réduire par 4 la taille de l'image

découpée puis nous avons utilisé l'espace de couleur **HSV** (Hue, Saturation, Value). Notre choix est justifié par les résultats performants dans la littérature [57,332,333].



Figure 5.5(a) Image originale, (b) Image découpée, (c) Image décimée, (d) Image en système HSV, (e) La composante S de HSV.

Nous considérons la composante *S* "Saturation" comme caractéristique selon [17] puis nous faisons la comparaison des performances des méthodes de réduction d'espace par calcul de distance.

5.4.2 Choix des filtres de Gabor : La méthode de reconnaissance que nous étudions dans cette partie repose principalement sur l'extraction des caractéristiques du visage par filtrage de **Gabor 2D**. La robustesse des filtres de Gabor a fait d'eux un moyen puissant d'analyse de textures et de classification. Les filtres de Gabor analysent la texture d'un objet suivant différentes résolutions et différents angles. Dans le domaine spatial, un filtre de Gabor 2D est une fonction à noyau gaussien modulée par une onde sinusoïdale plane complexe. L'équation générale d'une ondelette de Gabor en 2D est :

$$W(x, y, \theta, \lambda, \varphi, \sigma, \gamma) = e^{-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}} e^{i(\frac{x'}{\lambda} + \varphi)} \quad (5.1),$$

Où : $x' = x \cos \theta + y \sin \theta$ et $y' = -x \sin \theta + y \cos \theta$.

Il y a donc cinq paramètres qui contrôlent l'analyse par ondelettes [320,330]. Nous choisissons les cinq paramètres comme suit :

- θ spécifie l'orientation du filtre. Nous utilisons dans ce cas huit orientations : $\theta = \{0, \pi/8, \pi/4, 3\pi/8, \pi/2, 5\pi/8, 3\pi/4, 7\pi/8\}$.
- λ spécifie la longueur d'onde et donc la fréquence de la sinusoïde. Cette variable peut prendre cinq valeurs comprises entre 4 et 16 : $\lambda = \{4, 4\sqrt{2}, 8, 8\sqrt{2}, 16\}$.
- φ spécifie la phase de la sinusoïde. Elle vaut 0 ou $\pi/2$ selon que l'on veut la partie réelle ou imaginaire.
- σ spécifie la variance de la gaussienne (sa taille ou résolution). Elle est proportionnelle à la longueur d'onde de la sinusoïde. Dans notre cas $\sigma = \lambda$.
- γ spécifie l'aspect de la gaussienne. Ici les gaussiennes sont de forme circulaire $\gamma = 1$.

Rappelons que la représentation de **Gabor** d'une image de visage est le résultat de la convolution de l'image avec la famille des filtres de **Gabor**, définie par $IG(r, o) = I * G(r, o)$ où $IG(r, o)$ à une certaine résolution r et à une orientation o .

Nous commençons donc par étudier l'influence des caractéristiques des familles des filtres de **Gabor** sur les performances de l'authentification pour en déduire le choix optimal. Pour cela nous utilisons huit orientations $\{0, \pi/8, \pi/4, 3\pi/8, \pi/2, 5\pi/8, 3\pi/4, 7\pi/8\}$ et cinq longueurs d'onde $\{4, 4\sqrt{2}, 8, 8\sqrt{2}, 16\}$. La représentation de l'amplitude l'image considérée est comme le montre **figure 5.6**.

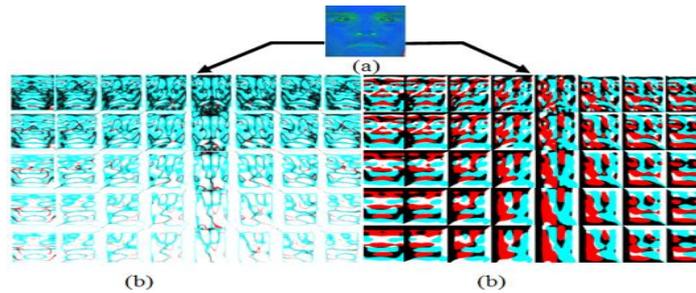


Figure 5.6 Résultats de la convolution d'une image de visage avec une famille de 40 filtres de Gabor (8 orientations (horizontales) et 5 résolutions (verticales)). (a) Image en espace de couleur HSV, l'ensemble (b) les amplitudes et (c) les phases de cette convolution.

- Influence des caractéristiques de la famille des filtres de Gabor sur les performances de la reconnaissance* : Nous nous intéressons d'abord à l'étude de l'amplitude des réponses des filtres de Gabor de l'image considérée **figure 5.6** (b). L'image d'entrée est une image de visage découpée et décimée. Nous considérons la composante **S** de l'image couleur dans l'espace **HSV**. L'algorithme de reconnaissance utilisé dans ce cas est l'*EFM* et la mesure de distance utilisée est la *corrélation*. Le **tableau 5.2** présente les résultats de reconnaissance en terme TEE dans l'ensemble d'évaluation pour différent niveau de résolution et d'orientations des filtres.

Tableau 5.2 TEE pour différents niveaux de résolution et d'orientations des amplitudes des filtres.

Résolution ($\sigma = \lambda$)	Orientations des filtres (θ)							
	0	$\pi/8$	$\pi/4$	$3\pi/8$	$\pi/2$	$5\pi/8$	$3\pi/4$	$7\pi/8$
4	9.28	10.1	8.13	8.02	8.01	8.04	8.63	7.3
$4\sqrt{2}$	8.33	9.01	7.54	9.61	5.35	7.95	7.2	8.5
8	9.31	7.34	8.7	5.7	7.85	5.13	8.02	8.17
$8\sqrt{2}$	9.54	8.64	7.31	9.36	10.19	8.3	7.07	7.54
16	9.17	8.48	8.65	9.18	9.18	8.64	7.84	7.62

Dans ce tableau nous constatons que le meilleur résultat est obtenu avec un **TEE= 5.13%** ce qui n'est pas satisfaisant.

- Problème de l'utilisation de la phase de Gabor pour les visages* : Lorsque nous considérons une image du visage, certaines parties du visage n'ont aucune texture informative qui pourrait être analysée par les basses résolutions des filtres de Gabor. Pour ces régions, l'analyse par

filtrage de Gabor donne $Real(IG_{s,o}) \cong 0$ et $Im(IG_{s,o}) \cong 0$. Même si ces valeurs sont très proches de 0, l'amplitude de la convolution n'est pas affectée par ce problème, alors que la phase devient une forme indéterminée pour ces régions spécifiques. Pour éviter les formes indéterminées, nous sélectionnons les régions informatives par seuillage de l'amplitude à chaque point d'analyse.

$$P(IG_{s,o}(x,y)) = \begin{cases} \arctan\left(\frac{Im(IG_{s,o}(x,y))}{Real(IG_{s,o}(x,y))}\right) & if M(IG_{s,o})(x,y) > Th \\ 0 & if M(IG_{s,o})(x,y) < Th \end{cases} \quad (5.3)$$

Où $(x; y)$ sont les coordonnées du point analysé et Th est le seuil de sélection de la phase.

- *Optimisation du seuil pour la sélection de la phase*

Afin d'étudier l'influence du seuillage de la phase en fonction des performances. La **figure 5.7** montre l'évolution des taux d'égale d'erreur *TEE* en fonction du seuil Th par un filtre de Gabor avec résolution $\sigma = \lambda = 4$ et Orientation $\theta = \pi/2$.

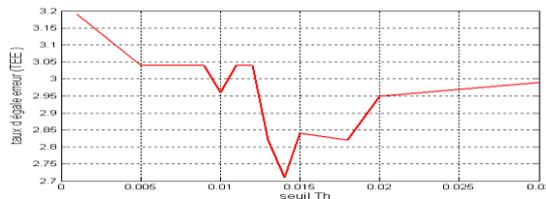


Figure 5.7 Taux d'égale erreur dans l'ensemble d'évaluation en fonction du seuil Th .

D'après la courbe de la **figure 5.7** nous constatons que la variation de *TEE* en utilisant la phase de Gabor est bien liée aux seuils du filtrage. Donc, notre choix est porté sur le seuil du filtrage $Th = 0.014$.

Dans cette deuxième série de mesures, ce seuil est maintenu et nous nous intéressons donc aux phases de Gabor. Nous cherchons dans ce cas les filtres de Gabor optimaux ce que nous présentons dans le **tableau 5.3**.

Tableau 5.3 TEE pour différents niveaux de résolutions et d'orientations des phases des filtres.

λ	Orientations des filtres (θ)							
	0	$\pi/8$	$\pi/4$	$3\pi/8$	$\pi/2$	$5\pi/8$	$3\pi/4$	$7\pi/8$
4	4.79	5.14	4.12	4.96	2.69	3.3	3.79	4.64
4	4.8	5.29	6	5.28	4.15	4.88	4.87	5.3
$\sqrt{2}$								
8	6.03	6.53	7.16	6.79	6.04	6.85	6.29	7.04
$8\sqrt{2}$	6.64	7.47	7.29	8.21	8.52	8.14	7.66	7.8
16	6.5	7.01	8.16	8.45	9.01	8.61	7.99	7.84

Nous remarquons que la première résolution et les orientations : $\theta = \pi/2, 5\pi/8, 3\pi/4$ donne le meilleur TEE. Les résultats obtenus par les phases de **Gabor** sont satisfaisants et encourageants. Ce qui confirme les travaux d'**Oppenheim** ainsi que ceux de

Mellakh dans [334,57], qui ont démontré sur d'autres bases de données que la phase des analyses multi-résolutions est plus informative que les amplitudes qui peuvent souffrir des problèmes d'illumination.

Nos meilleurs systèmes d'authentification de visages sont présentés dans le **tableau 5.4**.

Tableau 5.4 Résultats du système d'authentification de visages pour les quatre meilleurs systèmes retenus dans l'ensemble d'évaluation et test (intervalle de confiance paramétrique à 95%).

Méthodes	Ensemble évaluation		Ensemble de test	
	TEE %	TFR %	TFA %	TR %
Système 1	2.66 ± 0.72	2 ± 1.37	3.66 ± 0.11	94.33 ± 1.48
Système 2	2.69 ± 0.72	0.5 ± 0.69	4.07 ± 0.12	95.43 ± 0.81
Système 3	3.3 ± 0.8	2 ± 1.37	4.41 ± 0.12	93.59 ± 1.49
Système 4	3.79 ± 0.85	0.5 ± 0.69	4.47 ± 0.12	95.03 ± 0.82

- Le **système 1** qui utilise l'image couleur originale prétraitée.

Les trois autres systèmes utilisent la phase filtrée de la convolution de la **composante S** de l'espace de couleur TSV avec le filtre de **Gabor** pour la première résolution ($\sigma = \lambda = 4$) pour diverses orientations.

- Le **système 2** utilise l'orientation ($\theta = \pi/2$).
- Le **système 3** utilise l'orientation ($\theta = 5\pi/8$).
- Le **système 4** est l'orientation ($\theta = 3\pi/4$).

Nous utiliserons donc ces trois phases des filtres de **Gabor** pour la conception de notre Système Multi-Classifieurs (MCS en **chapitre 6**).

5.5 Choix des méthodes de réduction

Première expérience : Cas de méthodes linéaires

5.5.1 Comparaison des méthodes de réduction associée au calcul de distance [336]

Dans cette partie, nous étudions les algorithmes de réduction d'espace : l'ACP, la LDA, la DLDA, LDA2D et EFM et surtout leur application au domaine du visage. Une présentation théorique de ces approches, est faite dans la première partie du **chapitre 3**. Une étude comparative entre ces algorithmes est présentée dans ce chapitre. Nous nous focaliserons ensuite sur plusieurs particularités que rencontrent ce type d'algorithmes comme le choix des conditions adéquates d'apprentissage. L'objectif de cette partie du chapitre est de comparer les différentes méthodes afin de pouvoir choisir le meilleur algorithme pour la suite de nos études.

Mesure de distance

Cette étape consiste à calculer la distance euclidienne minimale entre le visage test et l'ensemble des visages A_i représentant soit la classe de personnes à vérifier (2D) ou la totalité des cartes de profondeur (3D) en mode identification. Lorsqu'on souhaite comparer deux

vecteurs de caractéristiques issus du module d'extraction de caractéristiques d'un système biométrique, on peut soit effectuer une mesure de similarité (ressemblance), soit une mesure de distance (divergence). La première catégorie de distances est constituée de distances Euclidiennes et sont définies à partir de la *distance de Minkowski d'ordre p* dans un espace euclidien R^N (N déterminant la dimension de l'espace euclidien). [24,69,103]

$$d_{Euclidienne} = \|A - B\|^2 \quad (5.4)$$

La distance employée dans notre expérience pour la mesure similarité est l'angle ou (corrélation normalisée) qui est définis comme suit: $Angle(A, B) = -\frac{A^T B}{\|A\| \cdot \|B\|}$ (5.5)

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.

$$d_k = \min(d_{euclidienne_i}) \text{ pour } i = 1 \dots k. \quad (5.6)$$

Dans les systèmes de reconnaissance nous calculons deux types de distances :

- **La distance_intra** : la distance entre deux images de visage de même individu.
- **La distance_extra** : la distance entre deux images de visage de deux individus différents.

Seuillage : Cette étape consiste à évaluer le résultat de l'étape précédente avec un ensemble de seuil et produit un résultat selon le mode de vérification ou le mode d'identification.

1 Mode de vérification

Lors de la reconnaissance en mode vérification on compare la distance minimale d_k à un ensemble de seuils, l'opération se déroule comme suit :

- Si d_k est inférieure à un seuil s_1 ; alors l'image du visage est considérée comme un **visage connu**.
- Si d_k est supérieure au seuil s_1 , mais inférieure au deuxième seuil s_2 , l'image peut être déterminée comme un **visage inconnu**.
- Si le visage d'entrée est supérieur aux deux seuils, le visage déterminé **n'est pas un visage**.

Dans le problème de vérification d'identité, nous cherchons à définir, pour chaque personne, ou de manière globale un seuil. Ce seuil va déterminer le minimum de ressemblance entre deux images pour admettre qu'il s'agit de la même personne. Ce **minimum de ressemblance** va s'exprimer comme un **distance maximale** entre les caractéristiques des deux images. Pour fixer ce seuil nous allons utiliser l'**ensemble d'évaluation**. Théoriquement $\max(\text{distance intra}) < \min(\text{distance extra})$ dans cette condition, il est possible de déterminer un seuil S qui détermine une classification parfaite.

Il suffit de prendre :
$$S = \frac{\max(\text{distance intra}) + \min(\text{distance extra})}{2} \quad (5.7)$$

Malheureusement cette situation ne se rencontre que très rarement en pratique. Par conséquent, un compromis doit être fait dans le choix du seuil, et ce dernier est très dépendant de l'application: **TFA** doit avoir une valeur très faible dans les applications bancaires afin de garantir plus de sécurité. Généralement, un système d'authentification opère entre les deux extrêmes pour la majorité des applications: on ajuste le seuil de décision pour atteindre une valeur désirée et prédéfinie de **TFR** ou de **TFA**.

Donc nous devons choisir un critère à minimiser pour fixer le seuil de manière à atteindre le **TEE** (**TFA=TFR =TEE**, Taux d'erreur égale). Dans notre expérience nous choisissons le seuil de manière à atteindre **TEE** tel que :

$$(TFA - TFR) < 10^{-2} \quad (5.8)$$

Le programme ainsi conçu se présente sous plusieurs fonctions, l'essentiel est présenté sous la forme suivante :

- *V_{caractéristique}()* pour préparer les images ainsi lues à la phase de traitement.
- *PCA ()* cette fonction renvoie une base de l'ACP normalisée dans laquelle on projette les deux matrices moyennes centrées de chaque classe et de chaque images
- *Fisherface ()* cette fonction renvoie une base LDA dans laquelle on projette toutes les images d'apprentissage et du test.
- *Dis₁ ()* cette fonction renvoie les valeurs des distances entre les images des clients évaluation ou des imposteurs et les images d'apprentissage

2 Mode d'identification

A chaque visage de référence I_k est associée une identité, sous la forme d'un numéro $id(k)$. Nous chercherons à cette étape l'identification d'un visage test à partir des visages de référence. La méthode consiste à comparer le visage test J avec chaque image de référence I_k . La dissimilitude entre les deux est quantifiée par la **distance Euclidienne**: En évaluant cette distance pour chaque visage de référence, on peut déterminer le visage de référence I_k le plus proche du **visage test** J . Si l'erreur d'identification, (distance correspondante) est inférieure à un seuil θ , on peut estimer qu'il s'agit d'un visage reconnu d'identité $id(J)=id(k)$, sinon le visage est classifié comme inconnu :
$$id(J) = \begin{cases} \text{connu} & \min d_k(J) \leq \theta \\ \text{inconnu} & \text{sinon} \end{cases} \quad (5.9)$$

Finalement, si le visage est un visage inconnu, on peut décider si on souhaite ajouter à l'avenir le visage pour la reconnaissance. Nous utilisons dans nos expériences plusieurs algorithmes de réduction d'espace, on peut légitimement se poser la question :

Quel algorithme choisir?

Pour répondre à cette question nous effectuons nos mesures sur plusieurs algorithmes et nous retenons le meilleur au sens performance du système. Nous utilisons entre autre les deux méthodes d'extraction des caractéristiques l'Analyse Discriminante Linéaire Bidimensionnelle Orientée en Lignes (ADL2DOL) et en Colonnes (ADL2DOC). Pour ce type d'analyse nous utilisons la distance proposée dans [103] l'AssembledMatrixDistance: AMD (équation (5.10) pour la mesure de similarité.

$$d(Y_i, Y_j) = \left(\sum_{k=1}^{d_1} \left(\sum_{h=1}^n (y_{hk}^{(i)} - y_{hk}^{(j)})^2 \right)^{\frac{p}{2}} \right)^{\frac{1}{p}} \text{ avec } p > 0 \quad (5.10)$$

Avec : $y_k^{(i)} = [y_{1k}^{(i)} y_{2k}^{(i)} \dots y_{d_1k}^{(i)}]$ et $y_k^{(j)} = [y_{1k}^{(j)} y_{2k}^{(j)} \dots y_{d_1k}^{(j)}]$ les deux matrices caractéristiques. Les figures ci-dessous illustrent les taux d'égale erreur TEE obtenus sur l'ensemble d'évaluation pour les méthodes (ADL2DOL, ADL2DOC) en appliquant la AMD pour la mesure similarité avec $p=0.125, 0.5, 1, 2$. Le meilleur résultat est obtenu par $p=0.125$ pour les deux méthodes d'extraction des caractéristiques par un nombre restreint des vecteurs propres.

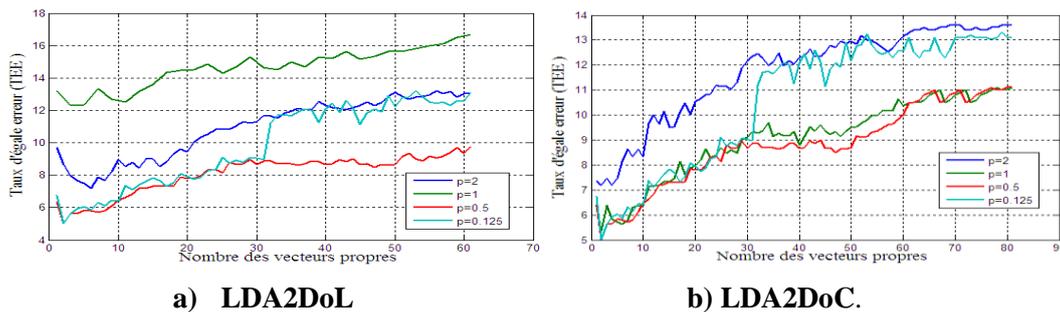


Figure 5.8 TEE dans l'ensemble d'évaluation par différentes fonctions de mesure de distance

Tableau 5.5 Comparaison des performances des méthodes LDA2DOL et LDA2DOC dans l'ensemble d'évaluation et test.

Méthodes	Ensemble évaluation	Ensemble de test			N _p
	TEE%	TFR%	TFA%	TR%	
LDA2DOL	5.1 ±0.98	2±1.37	6.3±0.15	91.7±1.52	2*81
LDA2DOC	5.01±0.97	2±1.37	6.19±0.15	91.81±1.5	2*61

Nous remarquons que le meilleur résultat est obtenu par $p=0.125$ pour les deux méthodes d'extraction des caractéristiques par un nombre restreint des vecteurs propres. Le **tableau 5.5** montre les résultats obtenus dans l'ensemble d'évaluation et l'ensemble de test pour la meilleure distance avec $p=0.125$. La méthode LDA2DOC nous donne le même résultat avec un nombre de caractéristiques $2*61=122$ pour un **TR=91.81±1.5%**.

Le **tableau 5.6** montre les résultats obtenus dans l'ensemble d'évaluation et l'ensemble de test des différentes méthodes de réduction d'espace.

Tableau 5.6 Comparaison des performances des méthodes ACP, ADL, EFM et ADLD dans l'ensemble d'évaluation et de test (intervalle de confiance paramétrique à 95%).

Méthodes	Ensemble évaluation		Ensemble de test		Np
	TEE %	TFR %	TFA %	TR %	
	ACP	5.7 ±1.04	2.7 ± 1.58	8.1 ±0.17	
ACP+LDA	2.83 ±0.74	2 ±0.08	4.12 ±1.94	93.88 ±2.03	90
ACP+EFM	2.66 ±0.13	2 ± 1.37	3.66 ±0.12	94.33 ±1.49	80
ACP+DLDA	3 ±0.72	2.1 ± 1.4	4.3 ±0.12	93.6 ±1.53	195
LDA2DOL	5.1 ±0.98	2 ± 1.37	6.3 ± 0.15	91.7 ±1.52	2*81
LDA2DOC	5.01 ±0.97	2 ± 1.37	6.19 ± 0.15	91.81 ±1.52	2*61

Les taux d'égale erreur TEE obtenus sur l'ensemble d'évaluation des quatre méthodes (ACP, ADL, EFM et ADLD) pour l'authentification des visages en appliquant corrélation normalisée pour la mesure similarité sont représentés sur la **figure 5.9**.

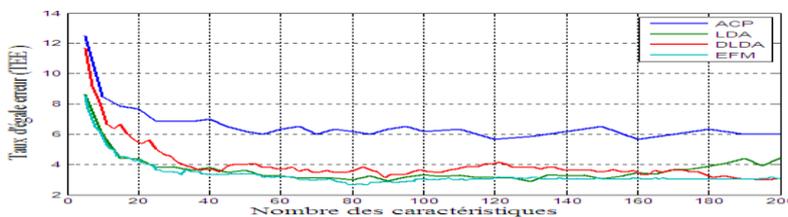


Figure 5.9 Comparaison des performances des méthodes ACP, ADL, EFM et ADLD dans l'ensemble d'évaluation.

La méthode EFM est mieux performante que l'ADL suivie de l'ADLD suivie de l'ACP en employant un nombre restreint de caractéristiques (vecteurs propres) dans notre base de données XM2VTS. Nous constatons que le résultat obtenu avec ACP+EFM est le meilleur avec un TEE=2.66±0.13% et un TR=94.33±1.49% en utilisant seulement un nombre de paramètres caractéristiques Np = 80. Nous remarquons aussi que la DLDA ne dépasse pas de trop les performances de la LDA et le mauvais résultat est présenté par la PCA. Les deux méthodes LDA2DOL et LDA2DOC donnent presque le même résultat avec un nombre de paramètres caractéristiques restreint. Nous remarquons aussi que ces méthodes à deux dimensions n'améliorent pas la performance d'authentification de visage par rapport aux ADL, EFM et ADLD car la base de données XM2VTS a été réalisée sur une longue période

ce qui a permis d'obtenir plusieurs images d'une même personne et donc une grande variabilité d'apparence (changement de coiffure, présence et/ou absence de lunettes, etc...). Par contre, seules les expressions neutres ont été considérées, donc *nous proposons d'utiliser ces méthodes (LDA2DOL et LDA2DOC) pour les bases de données qui contiennent plusieurs expressions.*

L'algorithme de réduction d'espace retenu est donc l'ACP+EFM et la mesure de similarité utilisée est la corrélation.

5.5.2 Méthode retenue ACP+EFM associée au SVM [335]

Dans cette partie de nos travaux nous retenons que la méthode ACP+EFM est la meilleure méthode de réduction d'espace. Nous l'utiliserons donc dans ce qui suit.

La méthode à vaste marge SVM est utilisé comme méthode de classification à deux classes. Pour cela, nous proposons d'utiliser un nombre de SVMs égale au nombre des Clients (200 Clients dans le cas de la XM2VTS). La figure 5.10 montre l'apprentissage des SVMs. On voit bien que les deux entrées du premier SVM₁ sont composées par :

- trois vecteurs caractéristiques réduits associés aux trois poses du Client₁,
- deux cents vecteurs caractéristiques réduits des huit poses des vingt cinq Imposteurs.

Pour les deux cent SVMs, il n'y a qu'une seule entrée qui varie c'est celle correspondant aux clients : SVM₁ est associée au Client₁, SVM₂ au Client₂,..... SVM₂₀₀ au Client₂₀₀. Les 200 vecteurs caractéristiques réduits des 8 poses des 25 imposteurs sont maintenus pour les 200 SVMs. Les vecteurs étiquettes sont les mêmes pour chaque SVM avec, -1 correspondant à la classe des Imposteurs et +1 à la classe des Clients.



Figure 5.10 Apprentissage des SVM

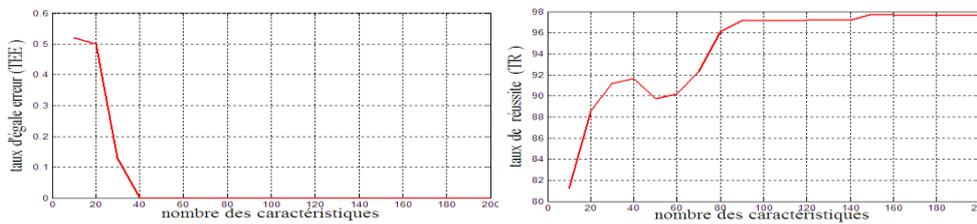
Chaque client est caractérisé par trois descripteurs correspondants aux trois poses (600 vecteurs). Alors que les imposteurs sont représentés par 25x8 descripteurs (25x8 images).

Nous choisissons d'utiliser une machine à vecteurs de support (SVM) avec un noyau RBF (Radial Basis Function ou fonction à base radiale) (voir équation 5.11).

Le SVM a été implémenté en utilisant la librairie **LibSVM** disponible gratuitement (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>). Le noyau **RBF** utilisé est de la forme :

$$K_{RBF}(u, v) = e^{-\gamma \|u-v\|^2} \quad (5.11)$$

Où γ est un paramètre qui règle la largeur de la marge. Le SVM nécessite également le réglage du paramètre C qui règle le compromis entre la largeur de la marge et le nombre d'erreurs admissibles. Les figures ci-dessous présentent les taux d'égale erreur (**TEE**) dans l'ensemble d'évaluation et taux de réussite (**TR**) dans l'ensemble de test en fonction du nombre des caractéristiques de la méthode de réduction d'espace **ACP+EFM**. La méthode de classification est celle qui utilise un nombre des **SVMs** égale au nombre des Clients.



a) **TEE** dans l'ensemble d'évaluation b) **TR** dans l'ensemble de test.

Figure 5.11 Taux en fonction des caractéristiques de l'ACP+EFM associée au SVM.

Nous remarquons que le **TEE** s'annule à partir d'un un nombre de caractéristiques égale à 40 c'est un **TEE** parfait et cela dans l'ensemble d'évaluation. Mais d'après la figure **5.11** nous remarquons que dans l'ensemble de test le **TR** prend une valeur intéressante et stable depuis un nombre de caractéristiques égale 90.

Nous présentons dans le **tableau 5.7** les résultats dans l'ensemble d'évaluation et l'ensemble de test, le nombre des caractéristiques et le temps de calcul pour la méthode proposée. Notre travail est exécuté avec un ordinateur : Intel Pentium Dual CPU 2.2 GHz, 1.49 Go de RAM et le langage de programmation utilisé est le **MATLAB R2009b**.

Tableau 5.7 Performance de notre système d'authentification de visages.

Ensemble d'évaluation	Ensemble de test			Np	Tps de calcul
	TFA%	TFR %	TR %		
0±0	1.5±0.09	0.78±0.01	97.72±1.46	150	0.32 s

Le descripteur donne un **TEE= 0±0 %** dans l'ensembled'évaluation. La méthode proposée est parfaite dans la phase d'évaluation. L'erreur augmente dans l'ensemble de test.

Nous pouvons affirmer à ce stade que dans ce cas de **classification par SVM**, nous obtenons un parfait taux d'égale erreur avec **TEE=0±0%** dans l'ensemble d'évaluation et un bon taux de réussite **TR=97.72±1.46%** dans l'ensemble de test avec un temps de calcul Tc satisfaisant **t=0.32s**. Cette méthode est plus performante que celle qui utilise la **classification** basée sur la **distance** de corrélation normalisée avec un : **TEE=2.66 ±0.13% et TR=94.33 ±1.49%**.

1) Tableau comparatif avec notre approche

D'après cette série de résultats, un tableau de synthèse (**tableau 5.8**) permet de comparer la performance de notre système de reconnaissance de visages par rapport à la performance obtenue dans d'autres travaux sur la base de données **XM2VTS** selon le protocole de **Lausanne I**.

Tableau 5.8 Tableau de synthèse de performances, obtenues dans l'authentification de visages.

Dates	Auteurs	Méthodes	Évaluation(%)	Test (%)
2006	Saigaa[337-1]	couleur ACP+EFM+RN	TEE=1.51	TFA=1.49,TFR=1.5,TR=97.01
2007	Yan[337-2]	Weighted Variant for Effectiveness	TEE=2.37	/
2007	Kotsia[337-3]	Discriminant Non-negative Matrix Factorization Algorithms	TEE=0.7	/
2009	Ejarque[337-4]	Fisherfaces +SVM	TEE=1.25	/
2009	Salimi[337-5]	KPCA+LDA	TFA=1.08, TFR=1.17	TFA=1.23,TFR=1,TR=97.77
2011	Nos systèmes	ACP+EFM+corrélation PCA+EFM+200 SVMs	TEE=2.66±0.13 TEE=0	TR=94.33±1.49 TFA=1.5,TFR=0.78,TR=97.72

Nous pouvons dire que la performance de notre système d'authentification de visages est presque équivalente à celle obtenue en **2009** par **Salimi** dans l'ensemble de test. En plus notre système est parfait dans l'ensemble d'évaluation. Notre objectif était l'amélioration du système de reconnaissance de visages nous affirmons en nous basant sur les résultats obtenus que notre objectif est bien atteint. De cette étude comparative de six méthodes de réduction d'espace, il en ressort que les méthodes d'extraction des caractéristiques : **ADL**, **EFM** et **ADLD** sont plus performantes que l'**ADL2DO** et la **meilleure méthode de réduction d'espace** reste l'**ACP+EFM** [336]. En utilisant un nombre de **SVMs** égale au nombre des Clients nous obtenons une bonne efficacité du système justifiée par un **TEE** parfait dans l'ensemble d'évaluation, un bon taux de réussite avec **TR=97.72±1.46%** dans l'ensemble de test et un temps de calcul égal **0.32**seconde.

Deuxième expérience : Cas de méthode de réduction non linéaire KPCA

Notre système ainsi conçu est associée à la **KPCA**. Nous utilisons pour ce cas deux classifieurs. Le premier basé sur une mesure métrique et le deuxième est assuré par **SVM**.

1 Architecture de l'approche à base de KPCA

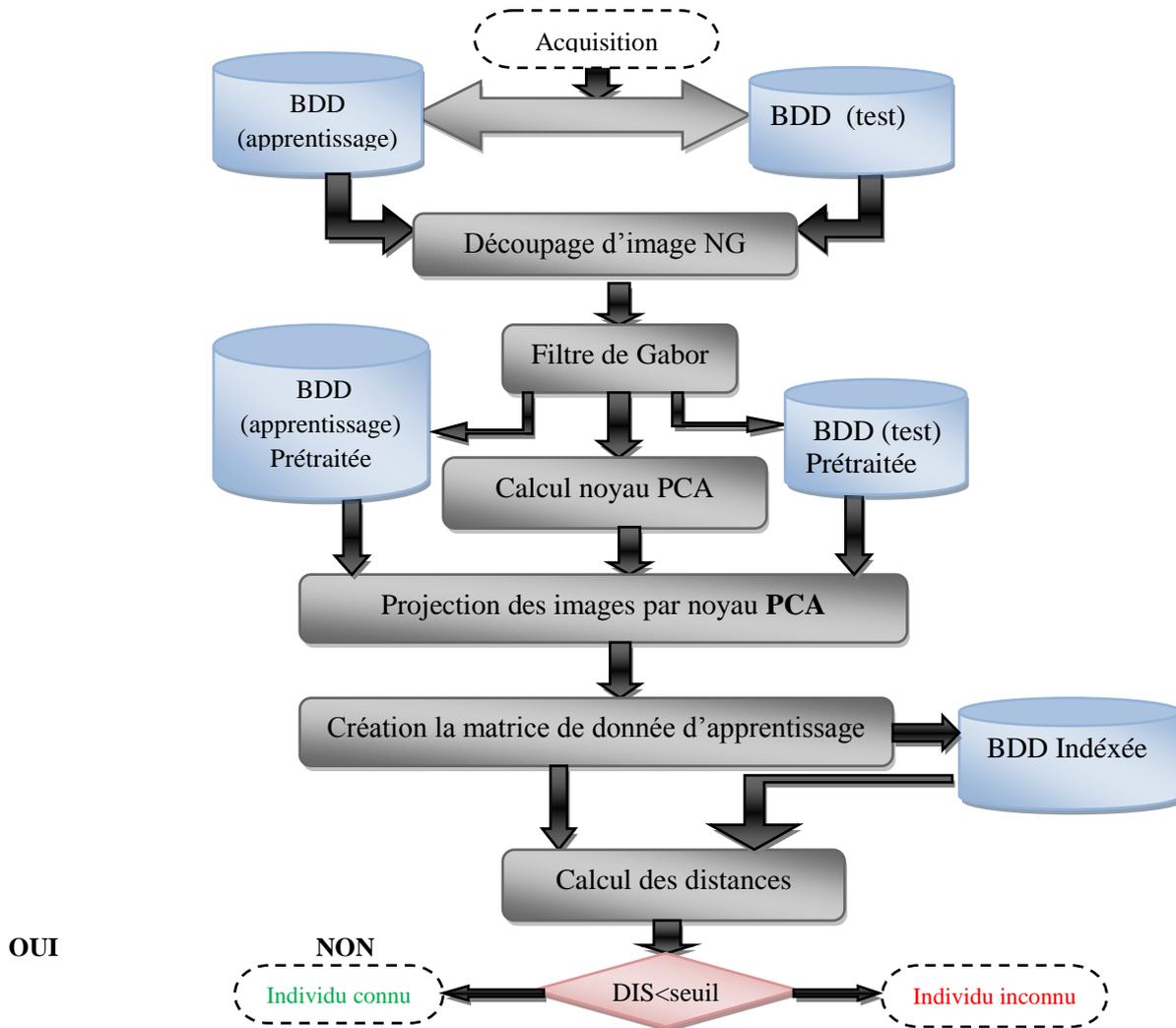


Figure 5.12 Organigramme du système de vérification d'identité de visage

1.1 L'algorithme de KPCA

Procédure de calcul KPCA /* Algorithme de KPCA*/

Fonction noyau Gaussien

x, y : deux coord.points

$Param = 0.5 / \sigma * \sigma$ /*paramètre de la fonction kernel, σ : la largeur du noyau Gaussien*/

Fonction $k = \text{kernel}(x, y, param)$

$Diff = x - y$;

$K = \exp(-(dif * dif) * param)$;

fonction KPCA ($data, \sigma, \text{numev}, \text{outlier}$)*

$[n, d] = \text{size}(data)$; /*matrice de n lignes et m colonnes

*centrer les données de la matrice noyau *

$K = \text{zero}(n, n)$;

Calcul de la matrice de covariance

Calculer la matrice de covariance c 'est le produit de la matrice centrée et sa transposée.

Calcul des valeurs et vecteurs propres

Calculer les valeurs propres et les vecteurs propres par la fonction $\text{eig}()$;

*Tri des valeurs propres par ordre décroissant *

Trier les valeurs propres λ dans l'ordre décroissant avec la fonction

$\text{sort}()$;

fin

Mesures de similarités

Une fois que les images sont projetées dans un sous espace il reste à déterminer quelles sont les images, semblables...Il y a beaucoup de mesures possibles de distance et desimilitude. Dans ce cas la distance L_2 est retenue. Connue aussi sous le nom norme euclidienne, elle est définie comme suit :c'est la somme de la différence carré entre les composantes des deux vecteurs A et B. Elle est donnée par l'équation suivante:

$$L_2 = \sum_{i=1}^N (A_i - B_i)^2 \quad (5.12)$$

1.2 Etude comparative

a) Comparaison entre PCA et KPCA(avec normalisation et classification par mesure de métrique L_2)

Tableau 5.9 La comparaison entre PCA et KPCA classification métrique L_2 .

N_p		10	100	400	500	600
TR(%)	PCA	85.9	86.2	86.01	86	86
	KPCA	89.92	91.7	91.51	92.33	93.51

D'après les résultats illustrés dans le **tableau 5.9** le **TR** obtenu pour la **KPCA** est nettement mieux que le celui donné par la **PCA**. La courbe suivante nous montre cette différence qui est loin d'être négligeable. Elle est accentuée par le nombre de paramètres caractéristiques N_p .

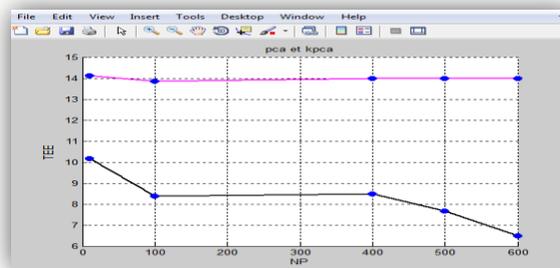


Figure 5.13 Les courbes du TEE pour la PCA et la KPCA en fonction N_p . (avec normalisation et mesure de métrique L_2).

Les courbes de la figure **figure 5.13** montre le grand écart entre les deux méthodes du taux d'égale erreur le TEE. Cet écart atteint **8.51%** ce qui est considérable dans le domaine de reconnaissance en biométrie.

b) Classification par SVM

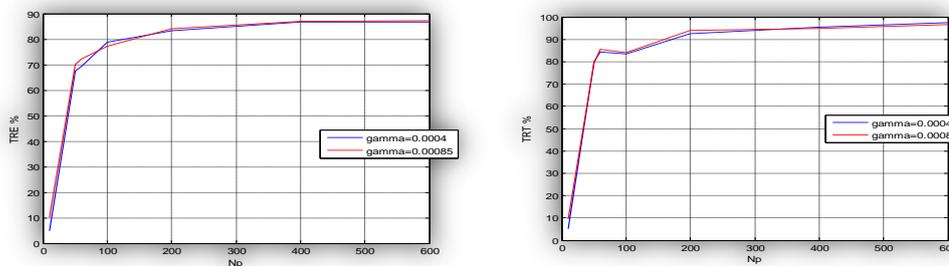
Pour $N_p = 600$ et σ variable dans les deux ensembles :

Sigma (σ) : Largeur du noyau **Gausséen**.

Tableau 5.10 Taux de reconnaissance classification SVM

Résolution(σ)	0.0004	0.0007	0.00085	0.0009	0.003	0.01
N_p	600	600	600	600	400	400
TR évaluation%	86.83	87	87.33	87.33	86.83	82.83
TR test%	97.5	97	96.5	96.5	95.5	94

Dans ce tableau nous constatons que le meilleur $TR_{test} = 97.5\%$ avec résolution $\sigma = 0.0004$ et $N_p = 600$ c'est un bon résultat.



a) Ensemble d'évaluation

b) Ensemble de test

Figure 5.14 Taux de reconnaissance en fonction de N_p pour diverses σ . (Classification par SVM)

Tableau de KPCA et SVM

Comparaison entre le KPCA et la PCA avec différentes classifications

Tableau 5.11 Comparaison entre le **KPCA** et **PCA** pour différentes classifications

		Classifieurs	
		Distances	SVM
TR(%)	PCA	86.2	90.18
	KPCA	93.51	97.5

Nous remarquons que le TEEs'annule à partir d'un un nombre de caractéristiques égale à $N_p = 40$ c'est un TEEparfait et cela dans l'ensemble d'évaluation. Mais d'après la **figure 5.14** nous remarquons que dans l'ensemble de test le **TR** prend une valeur intéressante et stable depuis un nombre de caractéristiques égale $N_p = 90$. En se basant sur les résultats de Salimi en 2009 voir **tableau 5.8** qui arrive à $TR=97.77\%$ en utilisant la KPCA +LDA, nous pouvons affirmer que la KPCA est une approche stable et prometteuse. Notre approche est plus simple et performante. Nous arrivons à un $TR=97.55\%$ en utilisant seulement la KPCA, en plus l'outil SVM est très efficace et facilement implémentable.

5.6 Approche monomodale 3D

5.6.1 Description de la BDD CASIA 3D expression

Entre Août 2004 et Septembre 2004, on a recueilli une base de données de visages 3D constituée de 4624 scans de 123 personnes qui utilisent numériseur 3D, Minolta Vivid 910, comme indiqué dans la figure 5.15 a). Au cours de la construction de la base de données, on considère non seulement les variations individuelles de poses, les expressions et les illuminations, mais aussi les variations combinées des expressions sous illumination et poses sous des expressions, comme montré sur la figure 5.15 b) et 5.15 c). Pour les sujets avec des lunettes, on a recueilli un scan supplémentaire avec des lunettes. Ainsi, chaque personne dispose de 37 ou 38 balayages. Et à partir de chaque balayage, une image couleur 2D et une surface 3D du visage triangulaire sont également générés. On vise à construire une base de données complète du visage en 3D, qui est en outre amenés à être une plate-forme publique pour tester les algorithmes de reconnaissance faciale 3D ou autres[9].

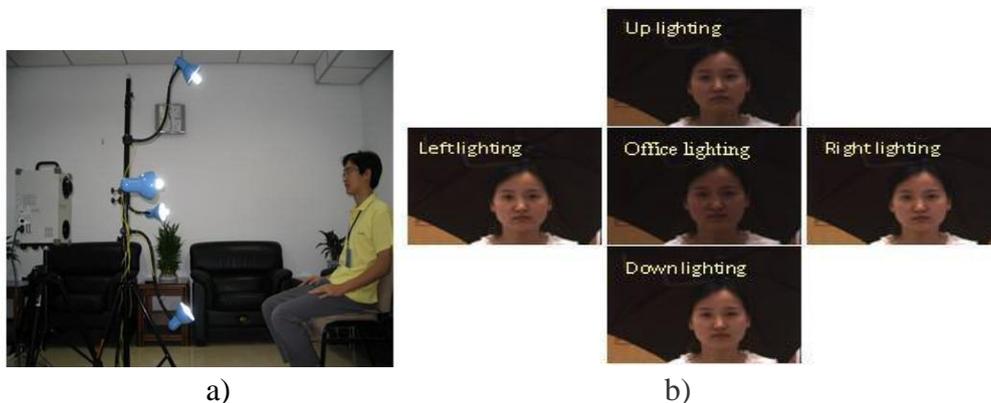
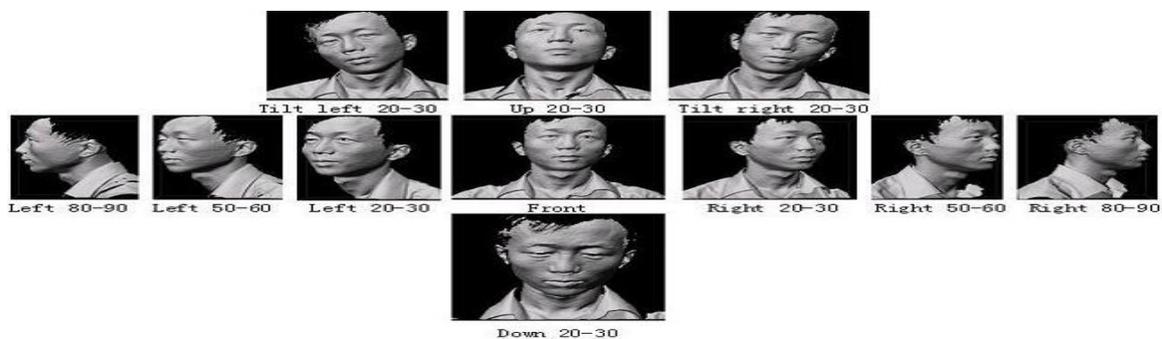


Figure 5.15 a) Capture des scènes de la base de données CASIA.

b) Variations d'éclairage de la base de données CASIA visage 3D.



c) Variation des poses de la base de données CASIA visage 3D.

- **Base d'apprentissage** : contient l'information concernant les personnes connues par le système (seulement des clients). Pour cela on prend les M premières images de 1

jusqu'à M de chaque personne. La **BDD CASIA 3D** expressions contient des images de visages présentant des expressions faciales dans l'ensemble d'apprentissage. Ce sous-ensemble contient des variations de pose des visages, des variations d'éclairage ainsi que des expressions faciales.

- **Base de test** : permet de tester le système en lui présentant des images de personnes imposteurs et clients les N dernières images de x à y de chaque individu nous ont servies pour la réalisation des différents tests.
- **Stockage et nomenclature** : Toutes les images sont stockées sous la nomenclature « $xx-yy.wrl$ ».

' xx ' : représente le numéro de la personne (de «00 » à «123 » pour les deux sexes).

' yy ' : représente numéro de l'image de «01 » à «37».

5.6.2 Système de reconnaissance de visage 3D

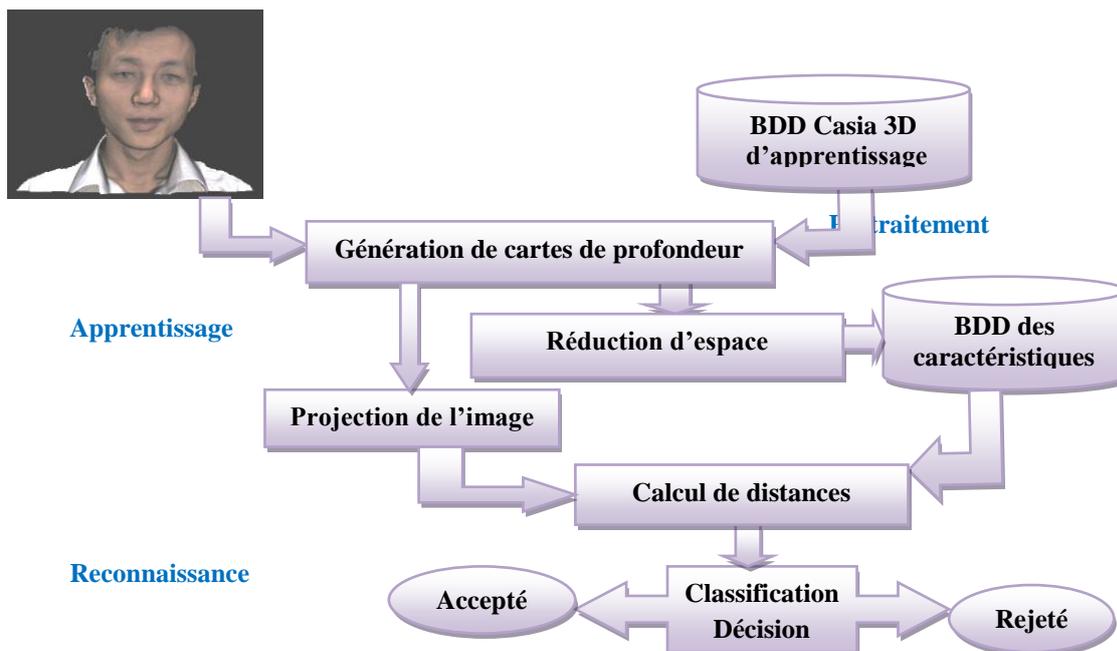


Figure 5.16 Organigramme du SRV 3D (basé sur la génération de cartes de profondeur).

1 Phase de prétraitement

1.1 Génération des cartes de profondeur

Comme nous l'avons déjà vu, la base de données utilisée est un ensemble de fichiers VRML contenant des maillages triangulaires. La façon d'exploiter l'information 3D, des caractéristiques géométriques d'un visage humain, est la construction de la carte de profondeur. La carte de profondeur est la fonction donnant pour chaque pixel de l'image la

profondeur du point 3D correspondant. Les étapes de génération de cartes de profondeur sont les suivantes :

- *Préparation de la carte de profondeur ;*
- *Normalisation des Z-coordonnées ;*
- *Chargement de la carte par les coordonnées Z normalisées ;*
- *Rotation ;*
- *Découpage ;*
- *Affichage de la carte de profondeur.*

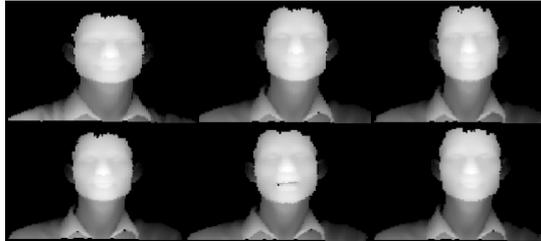


Figure 5.17 Cartes de profondeur avant le découpage

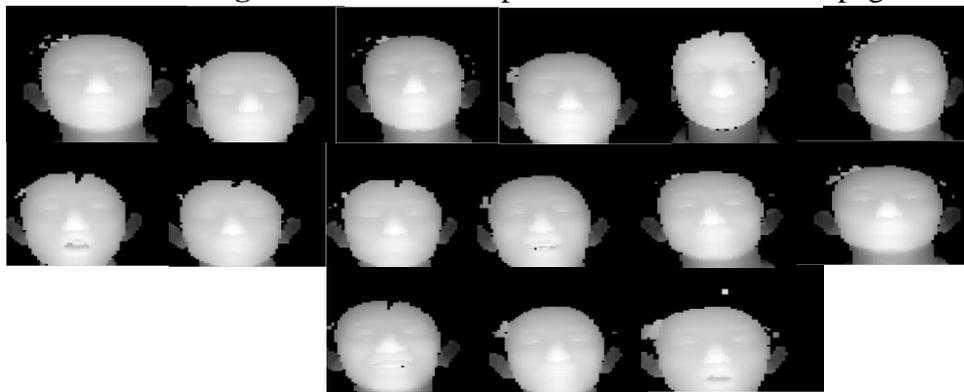


Figure 5.18 Cartes de profondeur après le découpage

2 Phase d'apprentissage

Le processus d'apprentissage a pour but de recueillir des informations biométriques sur les personnes à identifier. Il prend comme entrée les cartes de profondeur filtrées générées à partir de la base de données d'apprentissage et produit comme sortie les k premiers grands vecteurs propres en se basant sur l'ACP.

3 Phase de reconnaissance

Le processus de reconnaissance se subdivise en trois étapes : l'image d'entrée $I(N \times N)$ contenant le visage à identifier est transformé en un vecteur Γ de dimension $(N^2 \times 1)$ puis comparé avec le vecteur moyen Ψ . La différence entre le vecteur Γ et le vecteur moyen est ensuite multipliée par l'ensemble des vecteurs propres e_k pour générer le vecteur de poids Ω^T , tel que :

$$\omega_k = e_k^T (\Gamma_i - \Psi), \quad k = 1, \dots, M \quad (5.13)$$

$$\Omega^T = [\omega_1, \omega_2, \dots, \omega_M]^T \quad (5.14)$$

5.6.3 Résultats et discussions

Afin de mieux étudier l'apport de notre méthode, nous allons valider les résultats obtenus avec une méthode de base et ces résultats serviront par la suite de comparaison.

Pour cela nous faisons varier les paramètres suivants:

- Coefficients : les coefficients de projection des vecteurs propres (triés suivant les valeurs propres décroissantes).
- Les nombres des paramètres de l'ACP et de LDA : pour l'ACP le nombre de paramètres maximum est égal à $N_{eapp} * N_c$ et pour la LDA la dimension maximale est égale à N_c (N_c est le nombre des personnes dans la base d'apprentissage et N_{eapp} c'est le nombre des images (poses) de chaque personne dans la base d'apprentissage).
- Mesure de score (similarité) : corrélation normalisée (l'angle).
- Classification globale : en utilisant l'algorithme de calcul de TFA et TFR.

Nous testons notre approche sur différents aspects de la BDD CASIA 3D. Nous menons trois expériences essentielles sur une personne sous trois variantes: changement d'illumination, changement d'expressions et changement d'illumination et expression.

1 Changement d'illumination

Pour un nombre de personnes égal à **50** et un nombre d'images pour chacune égal à **5** ; **3** images pour la base d'apprentissage et **2** images pour la base du test.

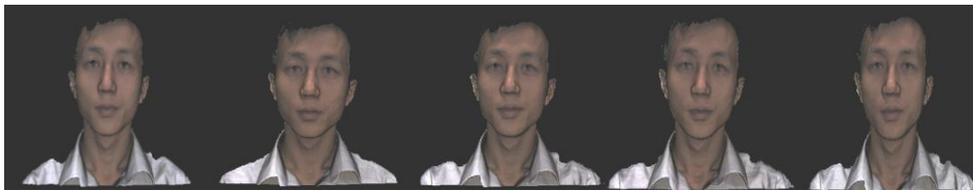


Figure 5.19 Les cinq visages de changement d'illumination.

Meilleur résultat d'identification : TR= 88 % pour $N_p(ACP)=20$ et $N_p(LDA)=45$

Les meilleurs résultats du taux de reconnaissance sont obtenus dans le cas de l'intervalle N_p de l'ACP [25,50] quelque soit $N_p(LDA)$. Pour $N_p(ACP) > 50$ le taux chute considérablement. Alors que le système reste stable concernant le nombre de paramètres $N_p(LDA)$. Nous pouvons affirmer à ce stade que *les paramètres de l'ACP ont une grande influence sur la performance du système et le grand nombre de paramètres caractéristiques est non sollicité.*

2 Changement d'expression faciale

Pour un nombre de personnes égal à **50** et un nombre d'images de chacun d'elles égal à **10** ; **6** images dans la base d'apprentissage et **4** images pour la base du **test**. Nous notons que:

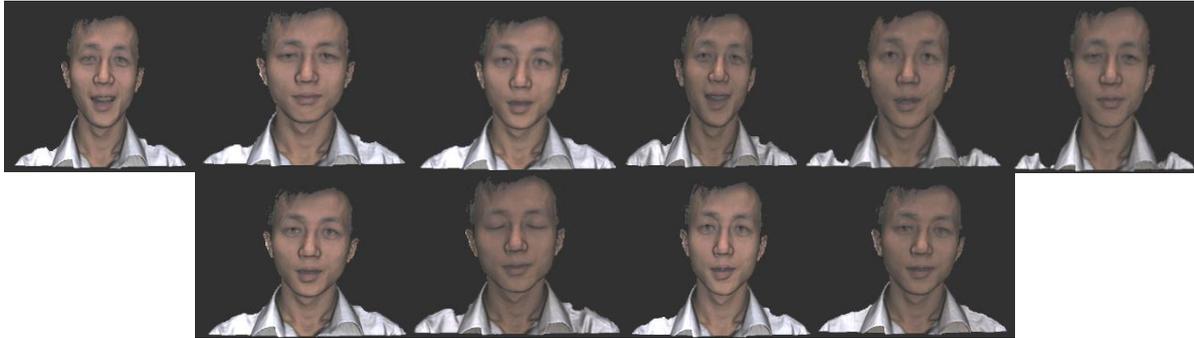


Figure 5.20 Les dix visages de changement d'expressions faciales

Résultats d'identification : $TR = 84,5\%$ pour $N_p(ACP) = 45$ et $N_p(LDA) = 40$. Notre approche est **performante** pour l'intervalle $N_p(ACP)[25,50]$. Nous remarquons une **stabilité** du système pour $N_p(LDA) > 30$ et un $N_p(ACP) < 100$. Pour $N_p = 100$ le taux de réussite (**TR**) est **trop faible** et le **système** devient **instable** et ceci **quelque soit** $N_p(LDA)$. *L'expression faciale influe beaucoup sur le taux de réussite relativement au changement d'illumination.* Nous constatons que **notre approche est vulnérable pour ce cas précis**. Cette différence réside à notre avis dans le fait de normalisation dans l'étape génération de cartes de profondeur. Et ceci malgré un échantillon d'images apprentissage et test important. Pour cette raison nous jugeons bon de faire des tests sur la présence d'illumination et expressions faciales en augmentant la taille de l'échantillon images soumis dans les deux phases.

3 Changement d'illumination avec changement d'expressions faciales

Pour un nombre de personnes égal à **50** et un nombre d'images pour chacune d'elles égal à **15** ; **10** images dans la base d'apprentissage et **5** images pour la base du **test**.

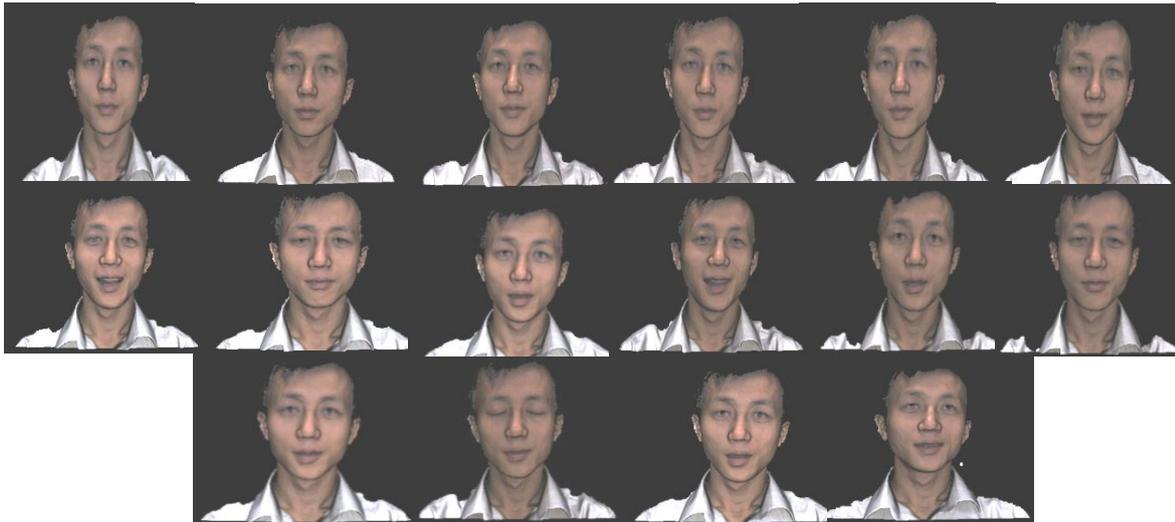


Figure 5.21 Visages avec changement d'illumination et d'expressions faciales.

Meilleur résultat d'identification : TR = 91,5% pour $N_p(\text{ACP}) = 30$ et $N_p(\text{LDA}) = 20$

Notre approche répond mieux à cette dernière expérience. Nous pensons que ceci peut être dû à l'échantillon plus important des images soumis à l'apprentissage et au test. Nous remarquons que dans l'intervalle de nombre de paramètres $N_p(\text{ACP})$ [25,50], le système est performant alors que pour $N_p(\text{ACP})$ supérieur à 50 il devient instable avec des taux de réussite très faible. Le taux de réussite obtenu reste non satisfaisant pour cela nous tentons dans ce qui suit étendre notre recherche et nous intéresser aux visages avec expressions. Nous prenons en compte les régions d'intérêts (yeux, nez et bouche) et ceci dans le souci d'améliorer la performance du système de reconnaissance.

5.7 Approche monomodale 3D expressions (Fusion des régions d'intérêts)

Un échantillon de visages de la **BDD CASIA 3D expressions** et le protocole utilisé dans notre approche est présenté dans la **figure 5.23**.

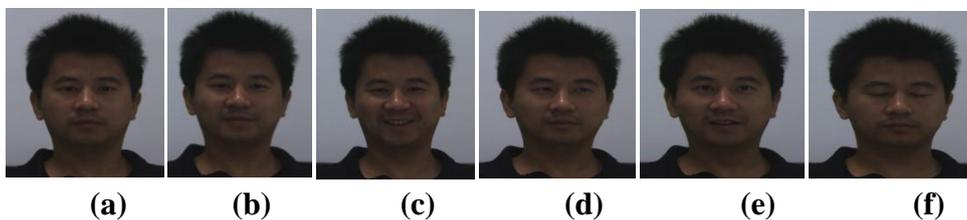


Figure 5.22 Variations d'expression de la base de données CASIA:

(a) visage neutre, (b) sourire, (c) joie, (d) colère, (e) surprise, (f) près des yeux.

5.7.1 Protocole de la BDD CASIA 3D expressions

La **CASIA 3D expressions** est partagée en trois sous ensembles selon le protocole suivant.

- *Apprentissage* de l'image (a),(c),(d) dans la base CASIA l'image 1,7,8.

- *Evaluation* de l'image (b),(e) dans la base CASIA l'image 6,9.
- *Test l'image*(f) dans la base CASIA l'image 10.

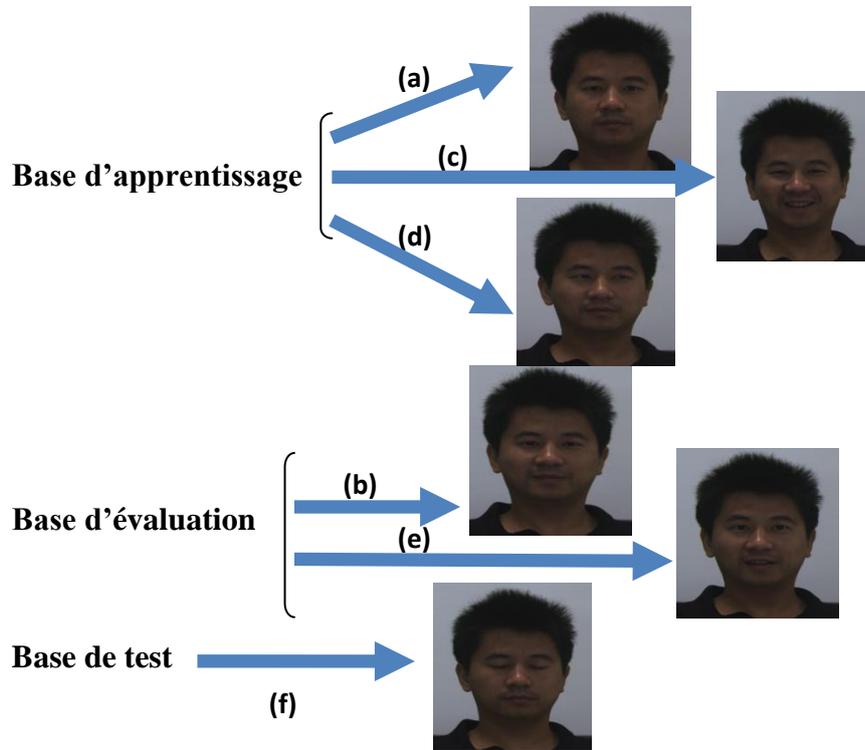


Figure 5.23 Protocole choisi pour notre SRV

5.7.2 Architecture structurelle du système de conception

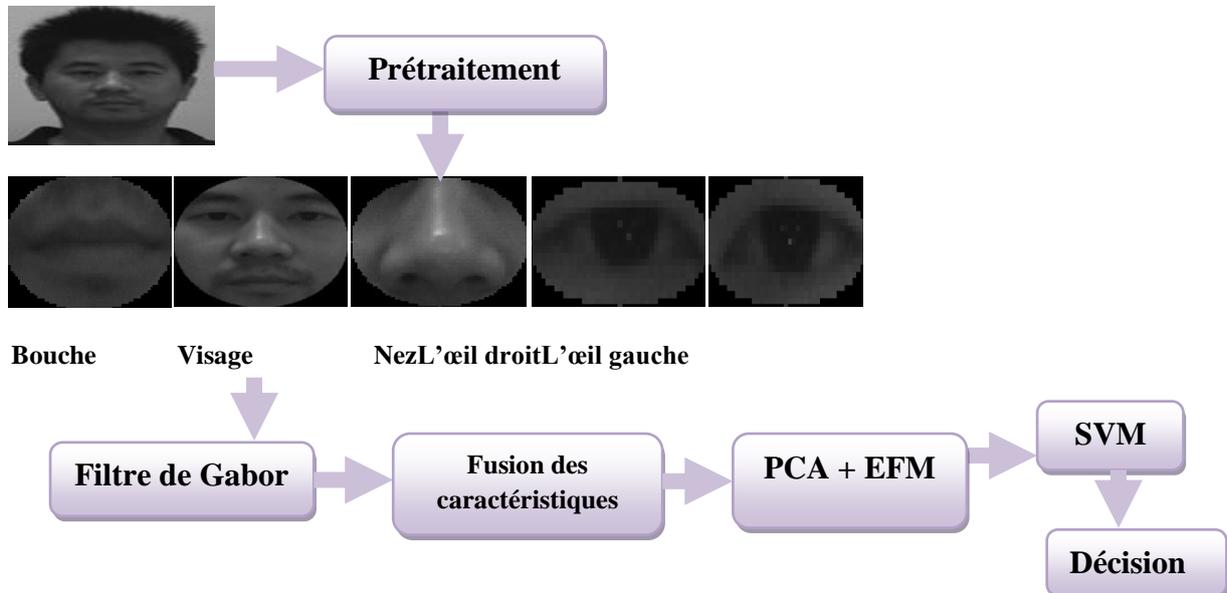


Figure 5.24 Schéma de l'architecture structurelle du système de conception SRV 3D expressions (fusion des régions d'intérêts).

1 Prétraitement d'image

Pour notre travail nous avons besoin de trois phases de prétraitement : la décimation, la conversion de la couleur en l'espace de couleur TSV puis la détection de visage et des régions d'intérêts.

1.1 La composante V de l'espace de couleur TSV de l'image

Les systèmes d'identification de visages utilisent souvent la représentation en niveaux de gris de l'image de visage comme caractéristique d'entrée. Mais lorsque les images sont en couleurs, il est avantageux d'utiliser cette information supplémentaire pour améliorer les performances de ce système d'identification. Nous rappelons que nous utilisons en général l'espace de couleur TSV (Teinte, Saturation, Valeur) ou HSV (Hue, Saturation, Value) ce sont les deux systèmes les plus utilisés dans la littérature [Mel09]. La quantification de la couleur de notre base de données CASIA est faite dans le système RVB (Rouge, Vert et Bleu), le passage du système de couleurs RVB au système TSV se fait par une fonction de transfert non linéaire. Pour plus de détails sur cette transformation, voir [Raf02]. Dans ce cas de base de données CASIA nous utilisons les espaces de couleurs TSV et nous trouvons que la composante V a donné le meilleur résultat. Partant de ce résultat nous utilisons donc la composante V de l'espace de couleur TSV comme caractéristique de l'image. La figure 5.25 présente la conversion de couleur et l'extraction de la composante V de visage.



(a) Image couleur (b) Image en système TSV (c) Composante V

Figure 5.25 Extraction de la composante V du système couleur TSV

1.2 Détection de visage et de points et régions d'intérêts

1.2.1 Détecteurs de visages et points d'intérêts

Avant d'entamer toute opération d'extraction de paramètres caractéristiques de notre image. Nous procédons tout d'abord à la détection de ce qui est significatif dans l'image du visage captée c'est-à-dire nous détectons le visage proprement dit. Car les cheveux, fond, la peau présentent peu d'intérêt pour l'analyse et l'identification. La détection du visage est la première phase de prétraitement et la figure nous montre le résultat de notre détection sur une image de visage neutre de la base CASIA 3Dexpression utilisée.

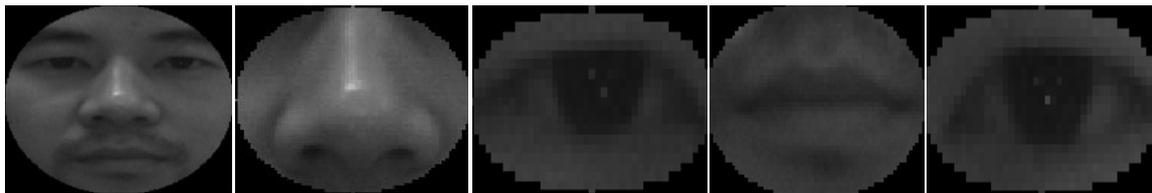


(a) Composante V de TSV (b) Détection manuelle des yeux (c) Détection de visage

Figure 5.26 Détection du visage.

1.2.2 Détecteurs des régions d'intérêts

Les régions d'intérêts sont illustrées par la **figure 5.27**.



(a) Détection de visage (b) Détection de nez (c) Détection de bouche (d) Détection d'œil droit (e) Détection d'œil gauche

Figure 5.27 Détection des régions d'intérêts du visage.

1.3 Extraction des caractéristiques par filtre de Gabor

Cette partie repose principalement sur l'extraction des caractéristiques du visage par filtrage de Gabor 2D.

1.4 Réduction d'espace

La réduction d'espace est réalisée par l'ACP+EFM. Finalement nous réalisons notre base de données indexée dans la phase enrôlement selon la **figure 5.28**.



Figure 5.28 Schéma de la phase enrôlement de notre modèle RV.

1.5 Classification de scores avec les machines à vecteurs de support (SVM)

Les images originales des personnes de la base de données **CASIA 3D expressions** subissent la phase de prétraitement, puis des descripteurs sont extraits et ensuite projetés sur des espaces réduits et stockés dans une nouvelle base de données indexée. Celle-ci est répartie sous le protocole décrit précédemment. La base d'apprentissage nous servira dans le calcul de la méthode de réduction et à l'entraînement de nos **SVMs** dans leur phase d'apprentissage. Une fois cette phase est effectuée nous procédons à l'étape classification par **SVM** en utilisant la base d'évaluation et test de la **BDD CASIA 3D expressions**. Le schéma de ce processus est présenté dans la **figure 5.29**.

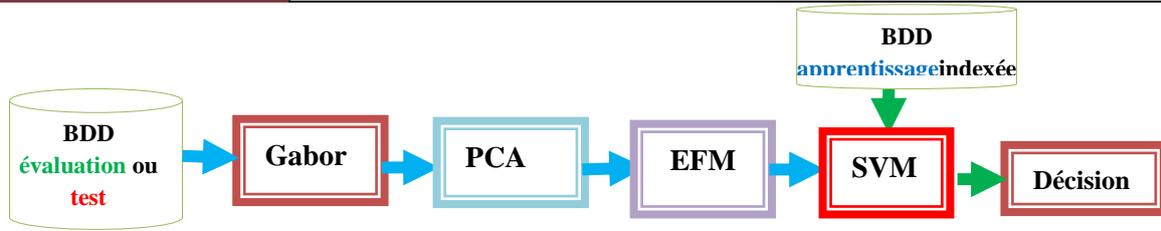


Figure 5.29 Schéma de principe de la classification par SVM.

Ainsi notre étude sur la classification des visages par SVM est menée sur deux phases :

1. **Phase d'apprentissage** : Pour notre protocole de la base de données CASIA 3D expressions, nous avons 123 images de personnes.

D'où : $N_c=123$ classes $\Rightarrow 123$ SVMs puisque le Nombre de classes = Nombre de SVMs. Chaque personne représente 3 échantillons $\Rightarrow N_e=3$. Nous choisissons le numéro d'échantillons d'apprentissage égal 3. Alors $N_{e_{app}}=3$.

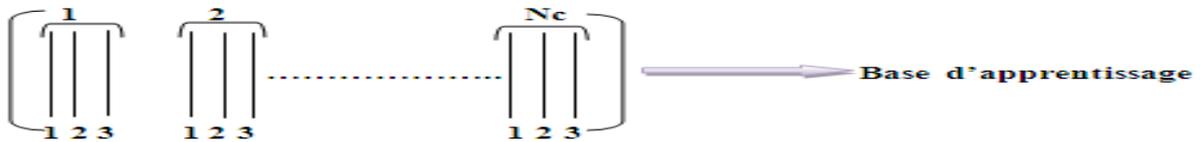


Figure 5.30 Schéma représentatif de la base d'apprentissage.

Le schéma de la phase de test est comme suit :

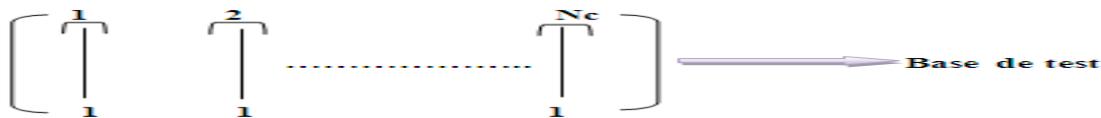


Figure 5.31 Schéma représentatif de la base de test.

Donc la phase d'apprentissage par SVMs sera :

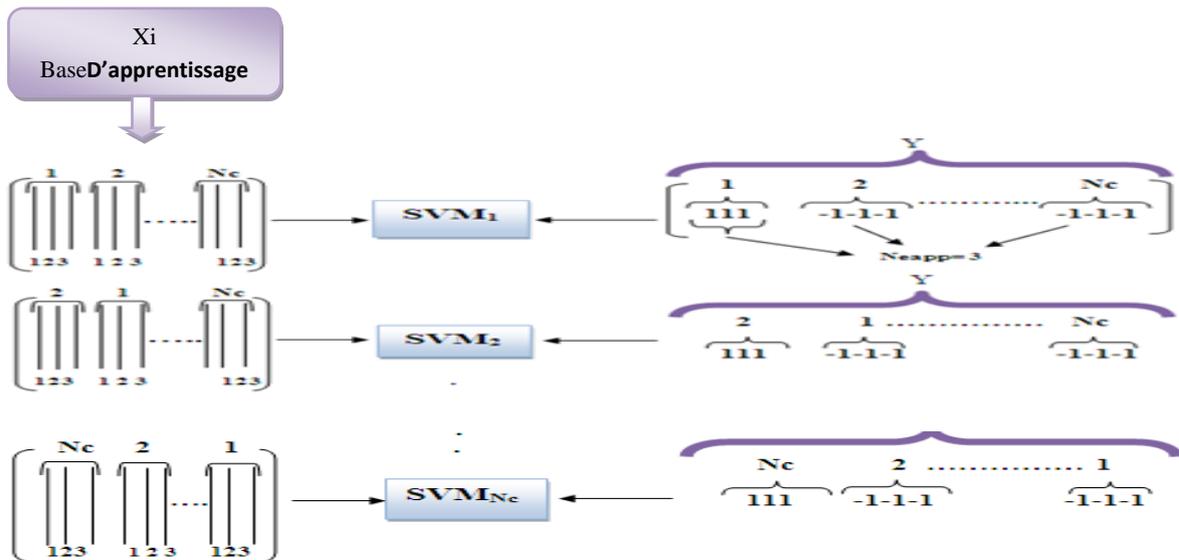


Figure 5.32 Schéma de la phase d'apprentissage par SVMs.

D'après la **figure 5.32** nous constatons que **123 SVMs** sont utilisés dans la phase apprentissage un SVM pour chaque classe (personne).

2. Phase d'évaluation : Dans cette phase nous présentons à l'entrée des **123 SVMs** le vecteur caractéristique réduit et nous sauvegardons l'indice de max des scores. Ainsi nous obtenons **2x123** scores/personnes d'où deux indices max/personne. Ces indices nous permettent de calculer notre taux de réussite **TR**.

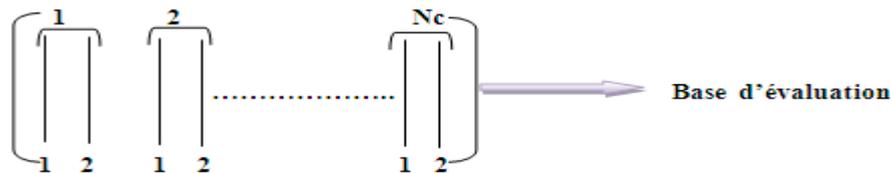


Figure 5.33 Schéma représentatif de la base d'évaluation.

Le schéma de la phase d'évaluation est :

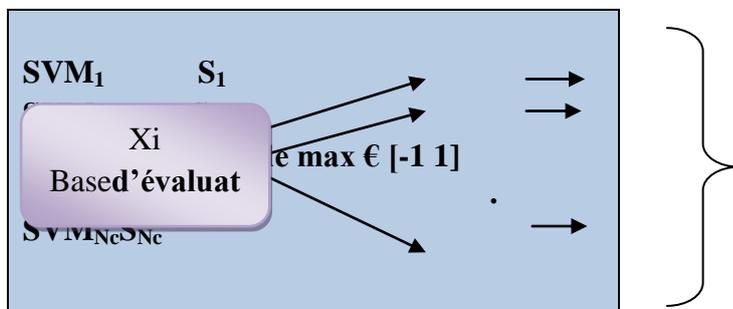


Figure 5.34 Schéma de la phase d'évaluation par SVMs

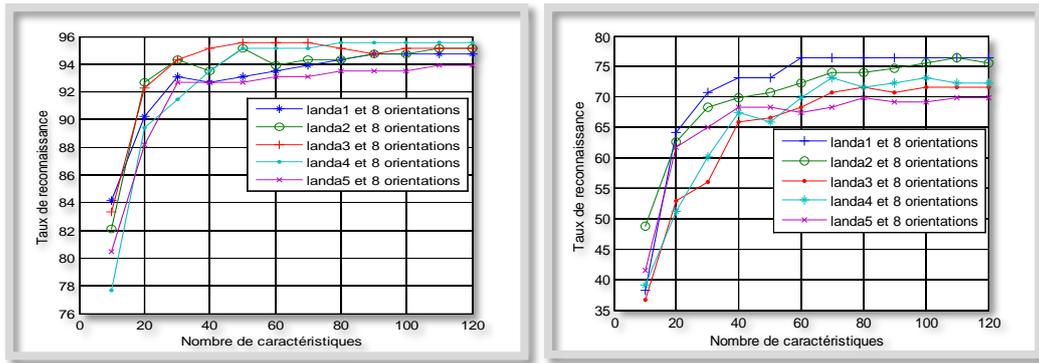
5.7.3 Résultats obtenus

5.7.3.1 Recherche des filtres optimaux

Nous nous intéressons au calcul du taux de reconnaissance dans l'ensemble d'évaluation et l'ensemble de test pour $\lambda = \{4, 4\sqrt{2}, 8, 8\sqrt{2}, 16\}$ et $\theta = \{0, \pi/8, \pi/4, 3\pi/8, \pi/2, 5\pi/8, 3\pi/4, 7\pi/8\}$.

- **1^{er} expérience λ variable avec 8 orientations**

Dans les expériences menées nous fixons λ et nous calculons le **TR** dans les 8 orientations pour divers nombre de paramètres caractéristiques puis, nous varions λ et nous reprenons les mêmes expériences. Comparaison des taux de reconnaissance dans l'ensemble d'évaluation et de test pour **5** les résolutions et **8** orientations:



(a) l'ensemble d'évaluation (b) l'ensemble de test

Figure 5.35 Taux de reconnaissance pour 5 résolutions et 8 orientations.

Nous remarquons que le meilleur TR dans la base d'évaluation est obtenu pour $\lambda=8\sqrt{2}$ avec un TR = 95.52%. Dans la base de test le meilleur taux de réussite est obtenu pour $\lambda=4$ et $\lambda=4\sqrt{2}$ avec un TR=76.42%. C'est le résultat retenu dans cette première expérience.

Tableau 5.12 Tableau récapitulatif.

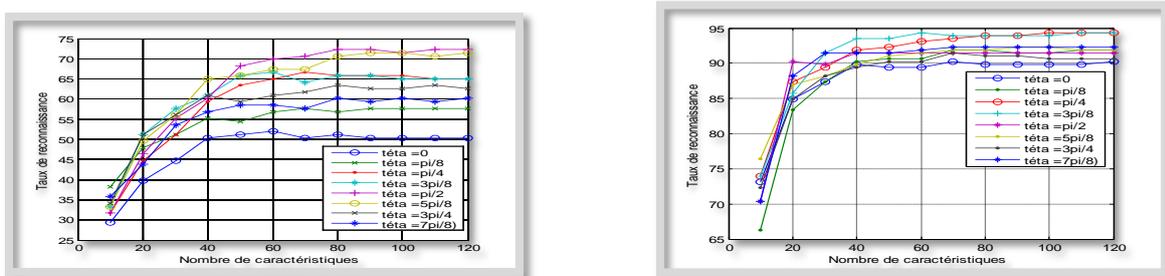
Résolution(λ)	4	$4\sqrt{2}$	8	$8\sqrt{2}$	16
TR % évaluation	94.71	95.12	95.12	95.52	93.90
TR% test	76.42	76.42	76.42	73.17	69.91
Np	100	110	100	100	110

Dans ce tableau nous constatons que le meilleur obtenu est $TR_{test} = 76.42\%$ avec résolution $\lambda = 4$ et 8 orientations θ et c'est un bon résultat concernant la reconnaissance de visages présentant des expressions qui reste à améliorer.

• 2^{ème} expérience $\lambda=4$ avec 8 orientations

Le meilleur résultat de TR obtenu sur la base de test est $TR=76.42\%$ pour la première résolution $\lambda=4$. Pour cela, nous essayons dans cette deuxième expérience de fixer $\lambda=4$ et changer les orientations $\{\theta = 0, \pi/8, \pi/4, 3\pi/8, \pi/2, 5\pi/8, 3\pi/4, 7\pi/8\}$ en appliquant ceci sur les deux BDD test et évaluation.

Comparaison du taux de reconnaissance dans l'ensemble d'évaluation et de test pour $\lambda = 4$ dans les 8 orientations:



(a) l'ensemble de test (b) l'ensemble d'évaluation

Figure 5.36 Comparaison du taux de reconnaissance pour $\lambda = 4$ dans les 8 orientations.

Nous pouvons dire que le meilleur résultat dans la base d'évaluation est obtenu pour les orientations $\theta = 3\pi/8$ et $\theta = \pi/4$. Dans la base de test les orientations $\theta = \pi/2$ et $5\pi/8$ s'avèrent les plus favorables.

Tableau 5.13 Résultats du système d'identification pour les quatre meilleurs systèmes retenus dans l'ensemble d'évaluation et test.

Orientations des filtres (θ)	0	$\pi/8$	$\pi/4$	$3\pi/8$	$\pi/2$	$5\pi/8$	$3\pi/4$	$7\pi/8$
TR % évaluation	89.43	91.46	94.30	94.30	91.46	92.27	91.05	92.27
TR% test	52.03	57.72	65.85	66.66	72.35	71.54	63.41	60.16
Nombre de caractéristiques	60	90	100	60	80	90	80	80

5.7.2 Fusion par concaténation des filtres de Gabor associés aux orientations optimales

D'après les résultats obtenus précédemment nous retenons quatre orientations $\theta = \{\pi/4, 3\pi/8, \pi/2, 5\pi/8\}$, $\lambda = 4$ et pour diverses N_p . Ces descripteurs ainsi obtenus sont concaténés et de nouveaux TR calculés en fonction de N_p .

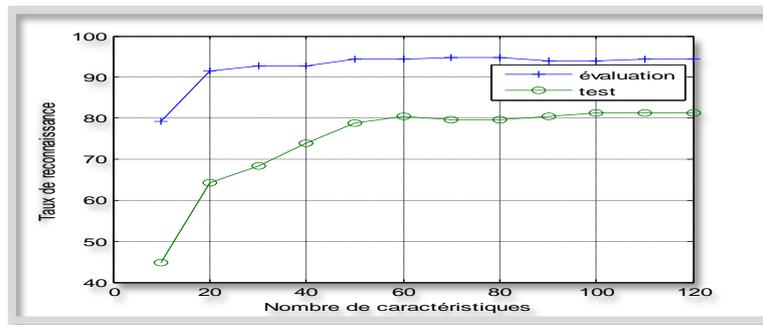


Figure 5.37 Valeurs de TR pour fusion par concaténation des descripteurs optimaux.

Les résultats de cette nouvelle expérience sont représentés dans le tableau suivant :

Tableau 5.14 Valeurs de TR pour fusion par concaténation des descripteurs optimaux.

	Nombre de caractéristiques											
	10	20	30	40	50	60	70	80	90	100	110	120
TR% évaluation	79.26	91.46	92.68	92.68	94.30	94.30	94.71	94.71	93.90	93.90	94.30	94.3
TR% test	44.71	64.22	68.29	73.98	78.86	80.48	79.67	79.67	80.48	81.30	81.30	81.3

Le meilleur résultat est obtenu avec un $TR_{test} = 81.30\%$ avec un $TR_{éval} = 94.30\%$. Nous pouvons affirmer que c'est un bon résultat relativement à ce que l'on espère avoir dans les applications 3D expressions en reconnaissance de visage monomodale. Une autre question se pose à ce stade du travail : Pourquoi ne pas simplifier notre système de reconnaissance en fusionnant l'image neutre avec ses six expressions et tenter d'identifier (ou authentifier) l'individu dans ce cas. C'est l'objet de notre application dans la section suivante.

5.8 Fusion du visage neutre et des visages avec les six expressions

La conception du système est illustrée par l'organigramme de la figure 5.38. Nous consacrons cette partie à l'implémentation du modèle étudié, la mise en œuvre du système et les résultats obtenus. Les outils utilisés, les algorithmes développés et l'approche réalisée sont mis en évidence. Dans cette étape nous fusionnons les scores du visage neutre et des visages de différentes expressions. Ces scores sont représentés par les distances calculées précédemment, la méthode de fusion utilisée est la **fusion par Ondelette**. Nous utilisons l'instruction **wfusmat** disponible dans **Matlab**, cette instruction fusionne deux matrices ou deux vecteurs. Dans un processus d'authentification, la stratégie de décision nous permet de choisir entre les deux alternatives suivantes: l'identité de l'utilisateur correspond bien à l'identité de la personne proclamée (ou recherchée) ou elle ne correspond pas. En comparant le modèle de l'utilisateur à celui qui existe dans la base de données. Cette phase est basée sur le seuil prédéfini.

$V2D_{test}$: image de visage 2D $_{test}$ (couleur). (Image de l'utilisateur)

$V3D_{test}$: image de visage 3D expressions test. (Image de l'utilisateur)

$XV2D \setminus XV3D$: modèle de la personne proclamée stocké dans la base de données

DIS_V2D : distance entre $XV2D$ et $V2D_{test}$. (projetées dans le même sous espace)

DIS_V3D : distance entre $X3D$ et $V3D_{test}$. (projetées dans le même sous espace)

DIS : distance fusionnée de DIS_V2D et DIS_V3D .

Phase test

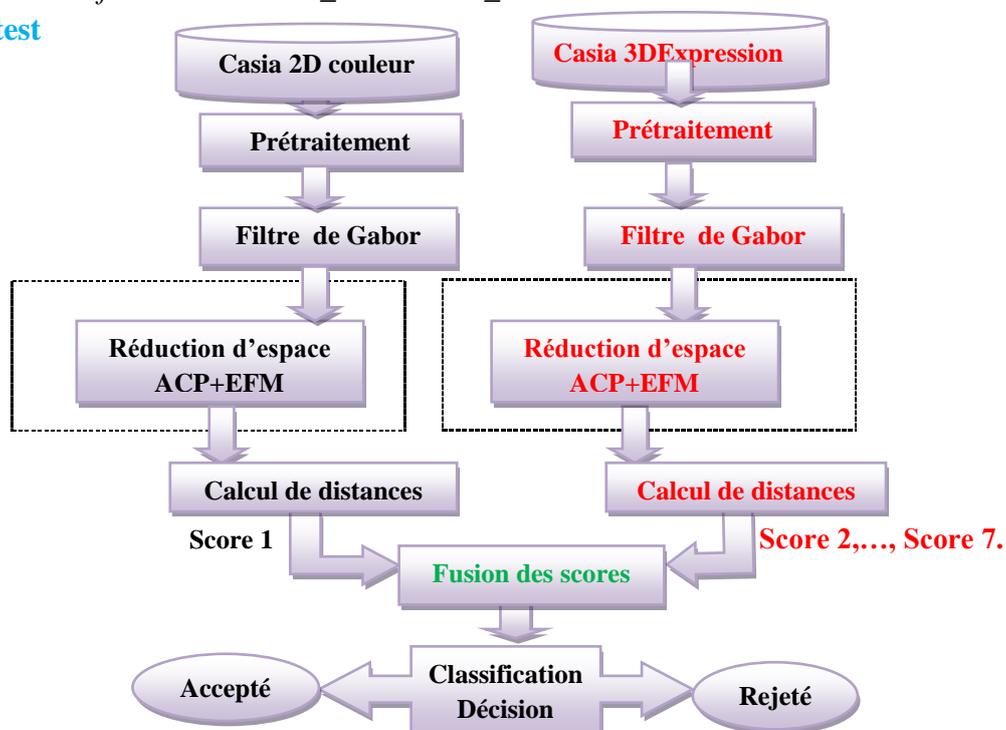


Figure 5.38 Organigramme du SRV 3D en présence d'expressions (Fusion des expressions)

Le programme se présente sous plusieurs fonctions, l'essentiel est présenté sous la forme suivante :

- *Lire()* pour lire l'ensemble des images d'apprentissage Visages **2D** et **3D** expressions , d'évaluation et de test.
- *Pretraitement()* pour préparer les images Visages **2D** et **3D** expressions , à la phase de traitement.
- *PCA()* *EFM()* cette fonctions renvoie une base dans laquelle on projette toutes les images Visages **2D** et **3D** expressions , d'apprentissage, d'évaluation et de test.
- *distance_intra()* renvoie les valeurs des distances entres les images des clients évaluation et les images d'apprentissage.
- *distance_extra()* renvoie les valeurs des distances entres les images des imposteurs évaluation et les images d'apprentissage.
- *Fusion_dis()* fusionne les distance intra et extra des visages **2D** avec les distances intra et extra visage **3D** expressions.
- *seuillage ()* cette fonction sert à estimer les valeurs de **TEE** et le seuil.

Nous utilisons le filtre de **Gaboret** la fusion des scores (Mean, Max, Min).

- Nous étudions d'abord le cas où le nombre de paramètres de $N_p(\text{EFM})$ variant de **10** à **200** et **PCA = 200**. Nous remarquons que la différence entre les erreurs pour chaque fusion n'est pas très grande. Ceci est une caractéristique très importante, car on peut dire que le système est stable.
- Nous étudions maintenant le cas où le nombre de paramètres de $N_p(\text{EFM})$ qui variant de **10** à **100** et **PCA = 100**.

Les résultats obtenus sont présentées sur le tableau suivant où nous présentons la comparaison entre $N_p(\text{PCA})=200$ et $N_p(\text{PCA})=100$ pour TR :

Tableau 5.15 Comparaison du TR pour fusion des expressions à base d'ACP+EFM.

PCA+EFM	Mean	Min	Max
100	92	92.25	78.92
200	77	82	69

Comme le montre ce tableau le meilleur **TR = 92.25 %** est obtenu pour $N_p(\text{PCA})=100$ pour la fusion par combinaison **Min** et classification par mesure de la norme Euclidienne L_1 .

5.9 Conclusion

Ce chapitre est structuré en deux grandes parties :

1 Approche monomodale 2D

Dans cette partie nous traitons la reconnaissance de visages 2D à base de techniques globales, divers classifieurs et à l'aide de techniques de réduction d'espace linéaires et non linéaires :

1.1 Méthodes de réduction d'espace linéaires :

Nous présentons les résultats obtenus en effectuant en premier lieu une étude comparative entre des méthodes de réduction d'espace à base d'ondelettes de **Gabor** comme méthodes d'extraction de caractéristiques pour la reconnaissance de **visages 2D** sur la base de données **XM2VTS**. Ces résultats montrent :

- L'efficacité de l'approche **ACP+EFM** par rapport **ACP, LDA, DLDA, ADL2DOC(ADLDOL)**. La méthode **EFM** est **mieux performante** que l'**ADL** suivie de l'**ADLD** suivie de l'**ACP** en employant un nombre restreint de caractéristiques (vecteurs propres) dans notre base de données **XM2VTS**. Nous constatons que le résultat obtenu avec **ACP+EFM** est le meilleur avec un **TEE=2.66±0.13%** et un **TR=94.33±1.49%** en utilisant seulement en utilisant seulement un nombre de caractéristiques égale **80**. Nos expériences sont menées avec la distance **corrélacion normalisée** ou Angle.
- Les phases filtrées de la convolution des images de visages par les filtres de **Gabor** donnent des bons résultats pour quelques filtres. Nous remarquons que les meilleurs résultats sont obtenus avec la première résolution et les orientations : $\theta = \pi/2, 5\pi/8, 5\pi/8$.

Les résultats obtenus par les phases de **Gabor** sont donc satisfaisants. Pour cela nous utilisons ces trois phases des filtres de **Gabor** pour la conception de notre système multi-algorithmes dans le chapitre suivant. Ce résultat nous a motivé à poursuivre la suite de nos travaux avec des méthodes linéaires améliorées telles la **FLD (LDA)** et l'**EFM**. En réalisant la fusion multi algorithmique des quatre meilleurs systèmes résultats des investigations précédentes est réalisée et nous pouvons affirmer à ce stade que dans ce cas de **classification par SVM**, nous obtenons un parfait taux d'égale erreur avec **TEE=0±0%** dans l'ensemble d'évaluation et un bon taux de réussite **TR=97.72±1.46%** dans l'ensemble de test avec un temps de calcul T_c satisfaisant **t=0.32s**. Notre méthode est plus performante que la méthode qui utilise la **classification** basée sur la **distance de corrélacion normalisée** avec un : **TEE=2.66 ±0.13%** et **TR=94.33 ±1.49%**.

1.2 Méthodes de réduction non linéaires

Ensuite nous avons étudié la méthode non linéaire **KPCA** pour une meilleure identification. L'association de la **KPCA** avec une classification par mesure de similarité basée sur la mesure de la norme **Euclidienne L2** nous donne dans la base d'évaluation un **TR=93.51%**.

Alors que l'association de la **KPCA** avec le classifieur **SVM** donne les meilleurs résultats avec un **TR_{test} = 97.5 %** pour un noyau **Gausséen** avec la résolution **$\sigma = 0.0004$** et **N_p = 600** et c'est un bon résultat.

2 Approche monomodale 3D

Pour nos travaux en **3D** nous avons utilisé la **BDD CASIA 3D** version **2004**. Nous avons établi un protocole pour notre **BDD CASIA 3D** et assuré une bonne phase d'entraînement. Nos expériences ainsi menées nous conduisent à des résultats encourageants surtout que le domaine de l'identification des visages en présence d'expressions est très délicat d'autant plus que la base **CASIA 3D** est une collecte de visages Asiatiques où la ressemblance est très accentuée.

Notre étude comporte trois approches : l'approche monomodale **3D** et l'approche monomodale **3D expressions avec concaténation des régions d'intérêts** et l'approche monomodale **3D avec fusion des visages avec expressions faciales**.

2.1 Dans l'approche 3D c'est le prétraitement qui prime, pour cette raison la génération de cartes de profondeur est réalisée, et pour la facilité d'implémentation dans l'application l'**ACP+LDA** est utilisée. Nous avons testé notre approche sur trois variantes : **changement d'illumination**, **changement de poses** et **changement d'illumination + expression**. Nous obtenons pour un nombre de paramètres réduit respectivement **TR=88%**, **TR=84,5%** et **TR=91,5%**. On en déduit que dans ce cas que l'**ACP** en amont de la **LDA** améliore les performances du système d'identification. Les meilleures performances de l'**ACP** sont obtenues pour un nombre de paramètres réduits [**25,50**] en utilisant des tailles proches de l'**ACP** et de la **LDA**. Les performances du système sont dépendantes de la taille de l'**ACP** pour l'identification. La distance **corrélacion normalisée** pour des petites dimensions de l'**ACP** donne les meilleurs résultats pour la reconnaissance. Ce résultats reste insuffisant pour cela nous pensons que la fusion des caractéristiques ou scores des images de visages **2D** et **3D** pourrait améliorer la performance de notre SRV. Ce qui sera réalisé dans le chapitre 6.

2.2 Dans notre deuxième expérience pour la reconnaissance des visages **3D**, nous nous sommes intéressés aux expressions faciales qui posent grand problème en reconnaissance biométriques. Dans ce premier cas nous essayons d'extraire l'information contenue dans les régions d'intérêts en détectant ces derniers et en les soumettant au filtrage de **Gabor**. Puis les vecteurs descripteurs ainsi obtenus sont concaténés, puis réduits par **ACP+EFM** et soumis au classifieur **SVM**. D'après les résultats obtenus et présentés dans ce chapitre

nous remarquons que le meilleur TR dans la base d'évaluation est obtenu pour le filtre de résolution $\lambda = 8\sqrt{2}$ avec un TR = 95.52%. Dans la base de test le meilleur taux de réussite est obtenu pour $\lambda=4$ et $\lambda=4\sqrt{2}$ avec un TR = 76.42%. Les orientations présentant les meilleurs TR sont $\theta = 3\pi/8$ et $\theta = \pi/4$ dans l'ensemble d'évaluation $\theta = \pi/2$ et $5\pi/8$ dans l'ensemble de test (pour $\lambda = 4$). Une fois l'étude faite sur les différentes orientations et résolutions du filtre un choix est établi. Les meilleurs filtres sont concaténés et ce nouveau descripteur est soumis à l'étape de réduction d'espace. Et finalement, le meilleur résultat dans ce dernier cas est obtenu avec un **TR_{test} = 81.30 %** avec un **TR_{éval} = 94.30 %**. Nous pouvons affirmer que c'est un bon résultat relativement à ce que l'on espère avoir dans les applications **3D expressions**. Nous pouvons déduire aussi que le classifieur **SVM** donne toujours les meilleurs résultats pour toutes les techniques utilisées et ceci pour les visages **2D** et **3D**.

2.3 Notre étude s'est étendue à la fusion des caractéristiques du visage neutre+les visages présentant les six expressions faciales. Les résultats obtenus sont prometteurs

TR = 92.25% est obtenu pour une réduction d'espace **ACP+EFM** pour **N_p(PCA)=100** et **N_p(EFM) = 100** pour la fusion par combinaison **Min** et la classification par mesure de la norme **Euclidienne L₁**.

Nous pouvons dire que la concaténation des descripteurs basée sur les régions d'intérêts associés au SVM donne de meilleurs résultats que la fusion des expressions faciales. Ceci est dû en partie au SVM qui reste un classifieur de premier choix et aussi à la phase de prétraitement qui prend en compte l'information des régions d'intérêts mise en jeu dans les expressions faciales. Et à tous cela s'ajoute le temps d'exécution qui est plus rapide pour l'algorithme concaténation de descripteurs et au SVM.

Enfin pour améliorer davantage les performances de notre système d'authentification des visages. Nous proposons d'utiliser les filtres de Gabor au lieu d'utiliser l'information des pixels comme caractéristiques des visages et utiliser la fusion des scores issus de plusieurs experts (Multi algorithmes).

Malgré des résultats satisfaisants pour les systèmes d'authentification de visages obtenus dans ce chapitre les taux d'erreur associés à des systèmes biométriques monomodaux restent relativement élevés, ce qui les rend inacceptables pour un déploiement d'applications critiques de sécurité. Pour pallier ces inconvénients, une solution est l'utilisation de *plusieurs modalités biométriques* au sein d'un même système, on parle alors de système biométrique multimodal qui est étudié dans le reste de notre mémoire.

Chapitre 6

Conception de l'Approche RV Multi Modales.

6.1 Introduction

Plusieurs expériences sont menées en multi algorithmes et multimodales où les deux approches par combinaison et par classification de la fusion sont étudiées en association à différentes méthodes de normalisation sur un système multi algorithmes appliqué aux images de la **XM2VTS**. Les techniques retenues sont appliquées aux **BDD** Bimodales dépendantes des scores comme (Voix, Visage) et aussi à une application bimodales indépendantes (Empreintes, Visage). Nos travaux se sont étendus au domaine **3D** pour contourner le problème d'illumination et rotation de visages. Seulement dans ce dernier cas le système présente des limites ce qui nous a mené à orienter nos recherches sur le **3D** expressions. Toutes nos expériences en **3D** sont réalisées à base de la **BDD CASIA 3D** et dans le cas des empreintes nous utilisons la **BDD FRVC**. Toutes ces bases de données sont universelles et agréées.

Notre objectif c'est l'identification des personnes en se basant sur une fusion multi algorithmique et multi biométrique.

- La fusion multi algorithmique est basée dans un premier temps sur quatre systèmes issus des meilleurs filtres de Gabor retenus par nos diverses expériences. Puis dans un deuxième temps une fusion des caractéristiques codées d'une part par un codage entropique et d'autre part par le calcul de la moyenne. Ensuite nous réalisons la réduction des caractéristiques par l'**EFM** et la **KPCA**. Les scores obtenus par mesure de similarité par des métriques sont combinés et une décision est prise en fonction d'un seuil prédéfini.
- La fusion multi biométrique est assurée par le filtre de Gabor optimal retenu à travers nos expériences sur la base de données **XM2VTS**. Une réduction d'espace est choisie selon les meilleures performances de notre système de reconnaissance. *Dans le cas de nos travaux l'EFM est bonne candidate*. La phase de classification est réalisée par calcul de distances ou par **SVM**. Les différentes modalités étudiées sont les expressions faciales, les empreintes, la voix et le visage **2D** et **3D**.

6.2 Bases de données

Les expériences ont été menées sur certaines bases spécifiques (**chapitre 5**) :

- La base de données **XM2VTS** [331] est composée de 395 personnes, organisée selon le protocole de Lausanne. Les images présentent des variations d'illumination, de pose et des artefacts (lunettes, barbes, . . .).
- La base de données **FRVC** [331-1] pour les empreintes.
- La base de données **CASIA 3D**[338] est composée de près de $n= 123$ personnes en 3D pour un total de plus de $m= 37 \times 123$ images. Celles-ci présentent de nombreuses variations de luminosité, de pose, rotation de la tête ou d'expressions faciales.
- La base principalement utilisée pour les tests multimodale est la base de données des **scoresXM2VTS**. Elle dispose d'un protocole bien défini permettant une comparaison précise avec des résultats publiés précédemment sur cette base.

Des expérimentations ont été conçues pour tester respectivement les problèmes de luminosité avec expressions ou visages neutres et la reconnaissance à travers le temps. Pour chacune de ces expérimentations, des protocoles sont effectués.

Plusieurs variantes sont prises en compte. Celles-ci dénotent les particularités des images associées et donc les caractéristiques que l'on souhaite tester. Elles sont organisées selon les cas suivants :

- les visages ont une expression faciale neutre,
- les visages présentant des expressions faciales,
- les visages sont capturés avec l'illumination,
- les visages sont capturés avec l'illumination et expressions faciales.

Ces jeux de tests permettent donc de tester différentes combinaisons, comme par exemple :

- un enrôlement où les visages ont une expression faciale neutre avec deux types d'expression, et les visages d'évaluation et de tests présentant des expressions faciales différentes.
- les visages d'enrôlement et de test ont la même expression faciale, mais la luminosité est changée.

La base de données XM2VTS a été largement utilisée car elle possède un protocole bien défini, ce qui est un atout certain pour la comparaison à de précédents travaux, pour peu que ceux-ci suivent le protocole de test. Certains travaux [33] [260] [261] utilisent cette base de données et proposent des algorithmes pour la reconnaissance ainsi que pour la fusion des modalités voix et visage. Dans [33], les fusions au niveau pixels et au niveau caractéristiques

issues d'une ACP sont considérées via des algorithmes génétiques. Les auteurs de [260] et [261] considèrent l'utilisation de multiples Machines à Support de Vecteurs (SVM) [260], ou de transformées en ondelettes [261] pour réaliser la fusion des images visible et infrarouge.

6.2.1 Résultats sur la XM2VTS

Les tests sur la base XM2VTS ont été réalisés en trois temps, ce qui conduit aux deux approches détaillées ci-après :

- a) Première expérimentation, dans un premier temps, nous avons mis en œuvre un système de reconnaissance multi algorithmes sur la modalité visage. Nous avons étudié les cas suivants :
 - la fusion des trois systèmes conçus par les meilleures orientations du filtre de Gabor et la composante **S** du système de couleurs **HSV**.
 - la fusion des caractéristiques issues des algorithmes entropie et moyenne.
- b) Deuxième expérimentation, nous avons essayé de tester notre approche sur deux modalités (bimodales) dépendantes (visage, voix) et sur deux modalités indépendantes (visage, empreintes). Les tests sont réalisés sur la BDD des **scores** XM2VTS pour les deux modalités (visage, voix), sur la BDD XM2VTS (visages) et la BDD FRVC pour les empreintes.

6.2.2 Résultats sur la CASIA 3D

Dans ce qui suit nous avons essayé de contourner le problème de la première expérimentation, à savoir la limite des images 2D face aux changements d'illumination.

Les tests sur la base CASIA 3D ont été réalisés en trois temps, ce qui conduit aux trois approches détaillées ci-après. Nous avons étudié les cas suivants :

- la fusion des caractéristiques 2D et 3D,
 - la fusion des scores 2D et 3D,
 - la fusion de la couleur et de la profondeur des images 3D.
- a) Première expérimentation, dans un premier temps, nous avons mis en œuvre un système de reconnaissance multimodale sur le visage 3D couleur. Il s'agit d'une fusion entre les caractéristiques ou scores de l'image couleur 2D (projection de l'image couleur RVB 3D sur le plan 2D) et une image profondeur nommée 2.5D (cartes de profondeur) issue de l'image couleur RVB 3D. Dans ce premier cas nous utilisons un classifieur basé sur la mesure de similarité pour sa simplicité.
 - b) Dans la deuxième expérimentation nous nous intéressons à la composante **V** du système HSV et nous faisons la fusion des caractéristiques ou scores des images 2D (composante

V) et 3D (profondeur). Dans ce deuxième cas le classifieur SVM est choisi pour sa rapidité et son efficacité. Nous tentons de faire l'identification du visage sous diverses variantes (variation de lumière(IV), visage avec expressions faciales (EV) et visage avec expression faciale et variation de lumière(EVI)).

6.2 Les systèmes d'authentification de visage retenus

Le principe de fonctionnement de base de notre système faciale (**figure 6.1**) est toujours le même que dans les applications monomodales. Il peut être résumé en **quatre étapes**.

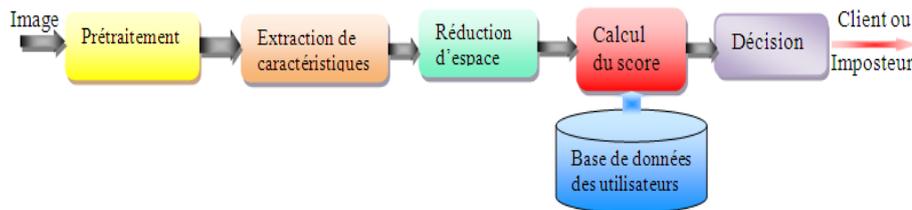


Figure 6.1 Principe de fonctionnement d'un système de reconnaissance de visage.

L'image d'entrée subit le même processus que dans les travaux réalisés au **chapitre 4**.

A.Fusion Multi Systèmes

Au cours des dix dernières années les systèmes **multi-systèmes** ou **multi classifieurs** (**Multiple Classifier Systems MCS**) sont devenus une approche établie pour concevoir des systèmes de classification. Un grand nombre de travaux à la fois théoriques et expérimentaux montre que les systèmes multi-classifieurs peuvent surpasser un classifieur unique dans plusieurs applications réelles, en termes de précision de la classification [339,296]). En particulier, plusieurs auteurs ont montré que les **MCS** peuvent permettre d'améliorer l'authentification biométrique de visages [339]. Toutefois, il n'est pas évident de savoir comment on fusionne les classifieurs. Le mécanisme de **fusion** peut se faire à différents niveaux de la classification [58] : au niveau des **données**, au niveau des **caractéristiques extraites**, au niveau des **scores** et au niveau des **décisions**. Notre **travail** dans ce cas se concentre sur la **fusion des scores** car c'est le type de fusion le plus utilisé. Elle peut être appliquée à tous les types de systèmes (contrairement à la fusion au niveau des données et au niveau des caractéristiques extraites). La fusion des scores est réalisée dans un espace de dimension limité représenté par un vecteur de scores dont la dimension est égale au nombre de sous systèmes, avec des méthodes relativement simples et efficaces mais traitant plus d'information que la fusion de décisions.

Les meilleurs systèmes de vérification de visages retenus (section 4.3.2.1) sont utilisés finalement pour étudier les méthodes de **fusion de scores** associés aux méthodes de **normalisation des scores**.

6.3 Principe de la fusion de scores MCS (Multi Systèmes)

Nous rappelons qu'un système de fusion de scores est constitué de deux modules, un module de fusion et un module de décision (**figure 6.2**).

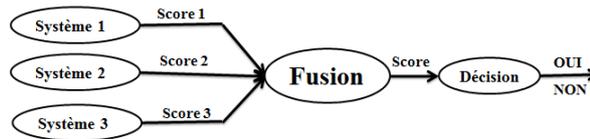


Figure 6.2 Schéma de la fusion de scores.

La fusion de scores consiste donc à la classification : **OUI** ou **NON** pour la décision finale, d'un vecteur de nombres réels. Il existe deux approches pour fusionner les scores obtenus par différents classifieurs :

- (i) Approche par combinaison de scores
- (ii) Approche par classification de scores.

Plus récemment **Morizet** et **Allano** ont essayé d'utiliser les deux approches précédentes et arrivent à des taux de réussite très intéressants.

Allano[58] utilise les deux approches basées sur le **SVM** pour l'approche par classification de scores et les méthodes simples de combinaison citées plus haut et les statistiques pour l'approche par combinaison de scores.

Morizet [3] a travaillé seulement avec l'approche basée sur la combinaison de scores en utilisant les mêmes méthodes que Allano et une nouvelle technique de fusion adaptative appelée "*WaveletScore DenoisingStatisticalFusion*" (**WSDSF**),

Nous nous sommes inspirés des travaux de Morizet et Allano et nous utilisons les deux approches citées précédemment pour l'étude des stratégies de fusion en ajoutant plus de méthodes comme :

La combinaison de scores par **logique floue** à base de l'intégrale de **Sugeno** ou de **Choquet**. Et les statistiques de **Fisher** et les réseaux de neurones artificiels (**MLP**) dans le cas de la première approche. Nous proposons aussi de faire la normalisation des scores dans ce cas bien précis de la fusion par classification de scores et c'est ce qui fait l'originalité de notre travail. Notre objectif est de réduire le coût et le temps d'utilisation des systèmes multimodaux

et améliorer les performances du système biométrique. Puis nous faisons une étude comparative de ces différentes méthodes.

Dans la **première approche** (fusion par combinaison des scores) nous utilisons donc des méthodes simples comme la **moyenne**, le **produit**, la **médiane** et **somme pondérée**[300] et en plus la combinaison de scores par **logique floue** à base de l'intégrale de **Sugeno** ou de **Choquet** [298,301]. Une autre question importante à laquelle nous apportons notre réponse est la normalisation des scores qui est une phase nécessaire avant de faire la combinaison.

Dans le cas de la **deuxième approche** (fusion par classification des scores) en plus de la méthode des machines à vecteurs de support **SVM** utilisée par Allano [58] et dans [287] nous utilisons les statistiques de **Fisher** [180] et les réseaux de neurones artificiels[317,318] (**MLP**). En ajoutant en plus des travaux antérieurs la phase de normalisation dans ce cas bien précis et c'est ce qui fait l'originalité de notre travail.

Nous faisons tout d'abord le choix de la meilleure méthode de réduction d'espace au sens des performances du système biométrique en se basant sur l'image brute (comme au **chapitre 5**). Une fois la méthode de réduction sélectionnée, nous étudions les amplitudes et les phases des filtres de **Gabor** et nous retenons le meilleur filtre présentant les meilleures performances biométriques (**TFA, TFR, TEE**) (voir paragraphe précédent). Ensuite les scores pour chaque système sont calculés puis soumis à la normalisation des scores. Cette dernière représente une étape très importante. Pour cela nous essayons d'étudier à ce stade plusieurs méthodes de normalisation. Nous les associons aux méthodes de fusion dans le cas des approches combinaison des scores et classification des scores et nous retenons la meilleure normalisation.

Finalement la meilleure approche de fusion est appliquée sur une base de données de scores **XM2VTS** utilisant deux modalités visage et parole. Et nous testons ainsi la validité de notre système de fusion retenu.

La solution pour les méthodes de réduction dans notre cas c'est le modèle discriminant Linéaire amélioré de Fisher (Enhanced Fisher Model **EFM**) [336]. Le modèle discriminant Linéaire amélioré de Fisher améliore la capacité de généralisation de la FLD en décomposant la procédure FLD.

La mesure de similitude entre les visages est assurée par un calcul de métriques classiques. Bien que la distance Euclidienne soit optimale dans la théorie. Les diverses expériences ont trouvé que la distance Euclidienne est surpassée par d'autres.

Une d'entre elles est la corrélation normalisée (normalized correlation) [3,65] que nous rappelons est définie par :

$$S(A, B) = \frac{A^T B}{\|A\| \|B\|} \quad (6.1)$$

Cette fonction calcule simplement le cosinus de l'angle entre les deux vecteurs caractéristiques.

Nous étudions un multi systèmes de vérification faciale par deux approches de fusion de scores:

- **combinaison de scores** par des **méthodes simples** (somme, produit, min-max et médiane) et la **logique floue** (intégrales de Sugeno et Choquet),
- **classification de scores** par les méthodes statistiques de **Fisher**, la méthode des machines à vecteurs de support **SVM** et les réseaux de neurones artificiels (**MLP**).

Une normalisation est réalisée au préalable. Différentes méthodes de normalisation sont étudiées et la **Znorm**, **QLQ** (Quadratique Linéaire Quadratique) et la **fonction double sigmoïde** sont retenues. L'étude de la fusion pour les deux approches combinaison et classification des scores des quatre systèmes d'authentification de visages est présentée par la **figure 6.3**. Nous commençons par comparer les méthodes de normalisation de scores associées aux méthodes de combinaison. Ensuite nous comparons les méthodes de classification de scores sans et avec la phase de normalisation. Toutes les méthodes sont testées sur la base de données de visages **XM2VTS** et selon le protocole de **Lausanne I**.

6.3.1 Conception de la fusion de scores Multi Systèmes

Après une étude des performances biométriques du système de reconnaissance de visages. Les systèmes qui présentent les meilleurs **TEE** (calculé sur l'ensemble d'évaluation) et **TR** (calculé sur l'ensemble de test) sont retenus.

Nous proposons un système multi-classifieurs de vérification faciale par l'étude de fusion de scores des quatre systèmes d'authentification de visages. Qui sont construits par l'extraction des caractéristiques du visage utilisant les phases filtrées des ondelettes de **Gabor** et l'Analyse en Composantes Principales (**ACP**) plus le modèle discriminant linéaire amélioré de Fisher (**EFM**) comme méthode de réduction d'espace de données.

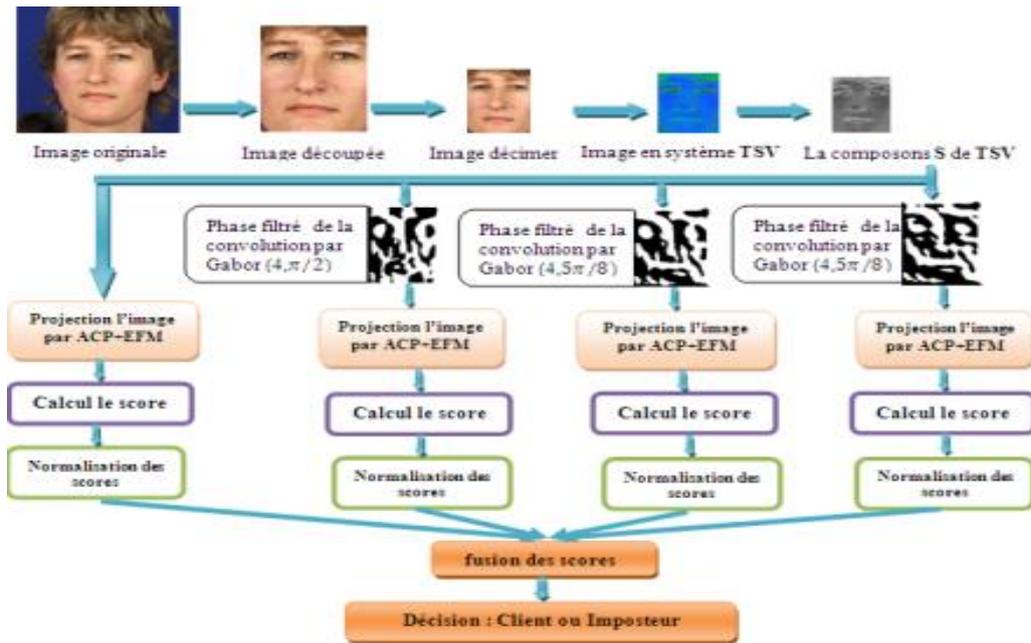


Figure 6.3 Schéma du multi systèmes par fusion de scores.

La représentation spatio-fréquentielle du visage a largement été utilisée et étudiée dans la littérature. Dans la majorité de ces études, seule l'amplitude de la réponse des filtres de Gabor a été employée alors que la phase est omise. Nous avons expliqué les raisons de la limitation de l'utilisation de cette phase et nous avons fourni une solution simple pour pallier cette limitation par une seuillage de la phase qui donne les meilleurs résultats par rapport aux amplitudes (section 4.3.2.1 chapitre 5). La phase de prétraitement, de réduction d'espace et de mesure de score sont commune aux quatre systèmes.

Les meilleurs résultats de TEE et TR obtenus sont présentés dans le tableau 6.1.

Tableau 6.1 Les quatre meilleurs systèmes retenus dans l'ensemble d'évaluation (TEE) et de test. (Intervalle de confiance paramétrique à 95%)

Performance	Système 1	Système 2	Système 3	Système 4
TEE %	2.66 ± 0.72	2.69 ± 0.72	3.3 ± 0.8	3.79 ± 0.85
TR %	94.33 ± 1.48	95.43 ± 0.81	93.59 ± 1.49	95.03 ± 0.82

Pour l'étude de fusion de scores nous utilisons deux approches :

- i) la **classification de scores** par : *méthode statistique de Fisher*, *Support Vector Machine (SVM)* et les *réseaux de neurones artificiels (MLP)*
- ii) les **combinaisons de scores** par : *la somme pondérée* et *la logique floue*.

6.3.2 Choix de la meilleure méthode de normalisation

6.3.2.1 Les distributions et normalisation de scores

Les distributions de scores pour les quatre systèmes d'authentification de visages sont représentées sur la **figure 6.4**. Nous remarquons que les quatre systèmes donnent des distributions Client et Imposteur différentes. Les distributions sont différentes en termes de plage de variation, ce qui rendra nécessaire l'étape de normalisation des scores. Elles sont également différentes dans leurs formes et leurs recouvrements entre les deux classes même si toutes les distributions ont un seul mode (mais pas forcément symétrique).

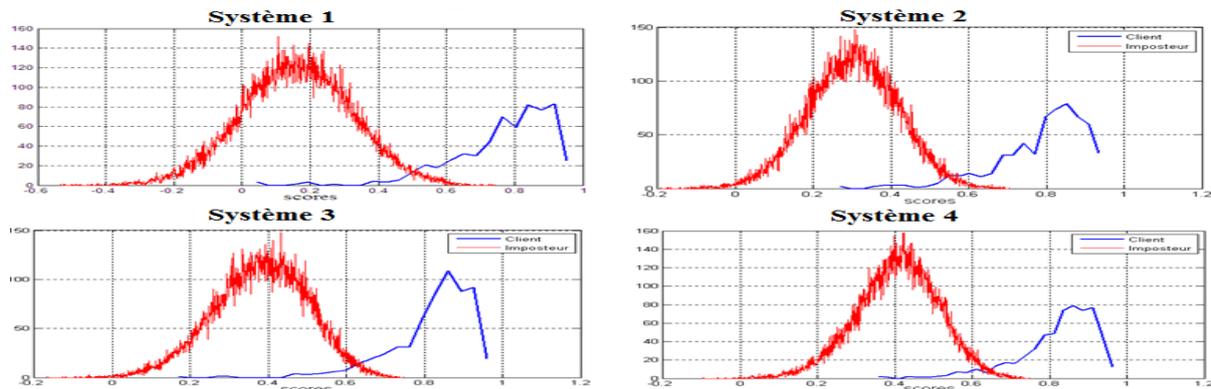


Figure 6.4 Les distributions de scores des quatre systèmes d'authentification de visages

6.3.2.2 Méthodes de normalisation de scores [295]

Les méthodes de normalisation de scores ont pour objectif de transformer individuellement chacun des scores issus des systèmes pour les rendre homogènes avant de les combiner. Les méthodes de normalisation de scores utilisés sont: **Znorm**, la **double sigmoïde** et la fonction quadratique-linéaire-quadratique (**QLQ**). Dans la **figure 6.5** la transformation des scores du premier système est présentée pour les deux normalisations (**Znorm** et la **fonction double sigmoïde**).

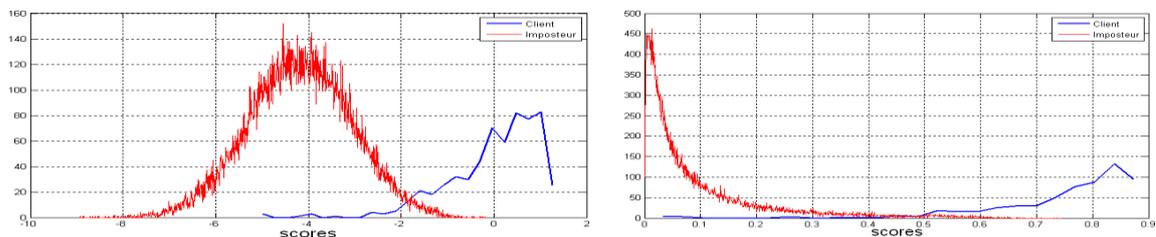


Figure 6.5 Normalisation de scores

Nous remarquons que la méthode de normalisation **Znorm**, ne modifie pas la forme des distributions mais leur différence se situe dans la façon dont chaque distribution sera répartie dans l'intervalle et donc dans la façon dont les systèmes à combiner se juxtaposeront dans

l'intervalle défini (avec Z_{norm} la moyenne de distribution des Clients de chaque système sera en 0). De plus, la méthode de normalisation **fonction double sigmoïde** modifie la forme des distributions.

Choix de la méthode de fusion

Comme dans la section **6.A** nous nous intéressons aux deux approches de fusion de scores la *combinaison* et la *classification*.

6.3.2.3 Fusion par combinaison des scores [340] :Pour ce type de fusion nous utilisons des méthodes simples [341] (moyenne, produit, minimum, maximum, médiane et somme pondérée) et deux intégrales floues [310,311] (**Sugeno** et **Choquet**).La somme pondérée est la meilleure méthode de combinaisons simples. Elle consiste en la définition de poids associés à chaque système, poids calculés de façon à minimiser le taux d'erreur sur la base de développement. On constate que la somme pondérée améliore les performances comparée à la deuxième meilleure méthode de combinaison simples qui est la somme simple (avec des poids identiques pour toutes les systèmes, ici $1/4$). Pour cette raison nous nous limiterons à cette combinaison simple.

Dans ce paragraphe, nous sélectionnons trois méthodes de normalisation de scores associées aux méthodes de combinaisons par logique floue à base de l'intégrale floue de **Sugeno** et de **Choquet**, ainsi que la combinaison simple par somme pondérée:

- une méthode qui *ne modifie pas la forme des distributions* est la **Znorm** puisqu'elle donne un meilleur résultat que la **minmax**, la **médiane** et l'**écart absolu médian** qui est plus simple que la **tangente hyperbolique**,
- deux méthodes qui *modifient la forme des distributions* les fonctions quadratique-linéaire-quadratique (**QLQ**) et **fonction double sigmoïde**.

Les différents taux d'erreur et de succès dans l'ensemble d'évaluation et de test en utilisant la fusion par combinaison de scores se trouvent dans le **tableau 6.2**.

Tableau 6.2 Performances des méthodes de normalisation associées aux méthodes de combinaisons pour MCS (intervalle de confiance paramétrique à 95%).

Méthodes de combinaison	Taux de Performance	Méthode de normalisation		
		Znorm	QLQ	Double sigmoïde
Intégrale floue de Sugeno	TEE	2.08 ± 0.64	1.22 ± 0.49	1.14 ± 0.47
	TR	96.28 ± 0.8	98.05 ± 0.76	98.36 ± 0.75
Intégrale floue de Choquet	TEE	2.16 ± 0.65	1.89 ± 0.61	2.4 ± 0.68
	TR	96.31 ± 0.79	97.08 ± 0.78	95.68 ± 0.81
Somme pondérée	TEE	1.95 ± 0.62	2.13 ± 0.64	2.17 ± 0.65
	TR	96.73 ± 0.79	96.6 ± 0.79	96.36 ± 0.79

Nous pouvons dire que dans le cas des méthodes de normalisation de scores qui modifient la forme des distributions la *fonction double sigmoïde* est la mieux performante associée à différentes méthodes de fusion par combinaison.

La méthode de fusion par combinaison de scores l'intégrale floue de *Sugeno* donne le meilleur résultat avec un **TEE=1.14±0.47%** et **TR=98.36 ± 0.75%** avec un temps de calcul très réduit **t=0.94 seconde**.

6.3.2.4 Fusion par classification des scores [345]: Dans cette section nous étudions la fusion de scores par l'approche de classification de scores (voir paragraphe **chapitre 4**). Normalement les méthodes de fusion par classification de scores cherchent à prendre une décision à partir du vecteur de scores de dimension N sans normalisation de scores. Mais dans notre travail nous proposons d'utiliser les méthodes de normalisation de scores pour transformer individuellement chacun des scores issus des systèmes afin de les rendre homogènes avant la classification, et ceci dans le souci d'augmenter la performance d'authentification. Nous utilisons trois méthodes de normalisation: **Znorm**, fonction quadratique-linéaire-quadratique (**QLQ**) et normalisation par une **fonction double sigmoïde**. Les différents taux d'erreur et de succès dans l'ensemble d'évaluation et de test en utilisant la fusion par classification de scores sans et avec les méthodes de normalisation se trouvent dans le tableau **6.3**. Nous utilisons la classification de Fisher (en utilisant la fusion par fonction linéaire discriminante (**LDF**) et fonction quadratique discriminante (**QDF**), une machine à vecteurs de support (**SVM**) avec un noyau **RBF** (Radial Basis Function) et une classification par MLP. Le **SVM** a été implémenté en utilisant la librairie **LibSVM** disponible gratuitement au site (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>).

Le noyau **RBF** utilisé est de la forme :

$$K_{RBF}(u, v) = e^{-\gamma \|u-v\|^2} \quad (6.2)$$

Où γ est un paramètre qui règle la largeur de la marge.

Les paramètres choisis par notre expérience pour la classification par **MLP** sont :

- Deux entrées sont les scores pour chaque système.
- Une couche cachée avec dix neurones et la fonction d'activation sigmoïde.
- Deux neurones dans la couche de sortie et la fonction d'activation tangente hyperbolique.

Tableau 6.3 Performances des méthodes de normalisation associées aux méthodes de classification pour MCS (intervalle de confiance paramétrique à 95%).

Méthodes de classification	Taux de performance	Méthode de normalisation			
		Sans normalisation	Znorm	QLQ	Double sigmoïde
SVM	TEE	2±0.62	1.5±0.54	1.96±0.62	1.66±0.62
	TR	96.53±0.79	97.44±0.77	97.36±0.78	97.11±0.78
MLP	TEE	1.83±0.6	1.69±0.57	1.66±0.57	1.33±0.51
	TR	96.25±1.07	96.69±1.27	97.84±0.77	97.49±0.77
LDF	TEE	2.14 ± 0.64	1.98 ± 0.62	1.97 ± 0.62	2.17 ± 0.65
	TR	96.32 ± 0.79	96.66 ± 0.79	96.82 ± 0.79	96.35 ± 0.79

Nous pouvons dire que la méthode proposée qui utilise la normalisation de scores avant la fusion par classification (et dans la majorité des cas les méthodes de normalisation qui modifient la forme des distributions) améliore la performance d'authentification de visages comparativement à celle obtenue sans normalisation. La meilleure performance est obtenue par la normalisation QLQ associée au MLP avec des taux **TEE=1.66±0.57%** et **TR=97.84±0.77%**. Seulement nous notons au passage que le SVM est plus sollicité dans les SRV pour sa rapidité, son efficacité et sa facilité d'implémentation malgré les taux obtenus dans notre cas **TEE=1.5±0.54 %** et un **TR=97.44± 0.77%** (pour la normalisation double sigmoïde).

6.3.2.5 Discussion

Nous avons montré comment l'utilisation d'un **système multi classifieurs (MCS)** permet d'améliorer notablement les performances d'un système de vérification d'identité monomodale de visage et nous affirmons que les méthodes de normalisation de scores améliorent la performance en générale de l'authentification de visages pour toutes les méthodes de classification de scores utilisées. Le meilleur résultat est obtenu dans le cas de la fusion par combinaison (intégrale floue de **Sugeno**) avec un **TEE=1.14±0.47%** et **TR=98.36 ± 0.75%** pour la méthode de normalisation de scores la fonction double sigmoïde (**tableau 6.2**). Les deux méthodes de classification de scores **SVM** (TEE=1.5±0.54 % et un TR=97.44± 0.77%) et **MLP** (TEE=1.66±0.57% et TR=97.84±0.77%.) nous donnent satisfaction pour les normalisations de scores respectives **Znorm** et **QLQ** (**tableau 6.3**). Rappelons que dans le cas de *l'approche monomodale* appliquée à la même BDD XM2VTS, pour une classification **SVM** nous sommes arrivés au meilleur des cas à des taux de performance **TEE=0±0%** et **TR=97.72±1.46%** dans l'ensemble de test avec un temps de calcul T_c satisfaisant $t=0.32s$ dans le cas de méthode de réduction linéaire choisie

(ACP+EFM+200SVMs). Alors que pour le cas non linéaire (KPCA) nous obtenons $TR_{\text{test}} = 97.5\%$ pour un noyau **Gausséen** avec la résolution $\sigma = 0.0004$ et $N_p = 600$.

Le meilleur système de fusion est donc construit par la normalisation de scores :

- 1) Normalisation **double sigmoïde** pour la combinaison par l'intégrale floue de Sugeno [342].
- 2) Normalisation **Znorm** pour la fusion par classification SVM [345].
- 3) Normalisation **QLQ** pour la fusion par classification MLP [345].

B. Fusion Multi Biométries [344]

6.4 Application du meilleur système de fusion sur la base de données de scores XM2VTS

6.4.1 Base des scores XM2VTS

Pour les deux protocoles de Lausanne (LP1 et LP2), Norman Poh a mis à disposition un ensemble de scores issus de **5 classifieurs de visage** et **3 classifieurs de voix** présentés dans [39]. Ces **scores** donnent 21 configurations différentes pour 21 couples de classifieurs **voix** et **visage** (*15 couples pour le protocole I et 6 couples pour le protocole II*). Chaque configuration correspond à un couple de systèmes: un système de voix et un système de visage. Chaque système est composé de :

- un module d'extraction de caractéristiques,
- un classifieur.

Un seul classifieur est utilisé pour la voix : le **GMM** (mélange de gaussiennes). Deux classifieurs sont utilisés pour le visage : le **GMM** (mélange de gaussiennes) et le **MLP** (réseaux de neurones appelé "Multi Layer Perceptron"). Trois modules d'extraction de caractéristiques sont disponibles pour le visage (**FH** : Face Histogram, **DCTs** : Discrete Cosinus Transform Small Size et **DCTb** : Discrete Cosinus Transform big size) et trois modules d'extractions de caractéristiques pour la voix (**LFCC** : Linear Filter-bank Cepstral Coefficient, **PAC-MFCC** : Phase Auto-Correlation Mel Filter-bank Cepstral Coefficient et **SSC** : Spectral Subband Centroid). Tous les détails des classifieurs et des modules d'extraction de caractéristiques sont donnés dans [39]. Rappelons qu'un système de fusion de scores 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. Afin de s'assurer que la combinaison de scores provenant de différents

si les systèmes sont cohérents, les scores doivent d'abord être transformés dans un domaine commun : on parle alors de *normalisation de score*.

6.4.1.1 Application à la base de données des scores XM2VTS

C'est une BDD multimodale prenant en compte trois systèmes de voix et les cinq systèmes de visage et délivrant les scores obtenus par des systèmes d'authentification biométriques particuliers. Nous nous intéressons uniquement au Protocole I (LP_1) pour lequel nous allons tester notre méthode de fusion. Le **tableau 6.4** présente la performance en terme de **TEE** dans l'ensemble d'évaluation et en terme TR dans l'ensemble de test pour les quinze couples de systèmes voix et visage (voix,visage). La méthode de normalisation de scores utilisée dans ce cas est la fonction **double sigmoïde**.

Tableau 6.4 Fusion de scores des visages et voix.

Méthodes de fusion	Systèmes de Voix ,Visage	Performance	
		TEE%	TR %
Intégrale floue de Sugeno	(LFCC,GMM),(DCTs, GMM)	0.001±0.01	99.99±0.04
SVM	(LFCC, GMM), (DCTs, MLP)	0.03±0.07	99.30±0.41
MLP	(LFCC, GMM),(FH, MLP)	0.23±0.21	99.57±0.32

La méthode de normalisation et la méthode de fusion de scores sont : la fonction **double sigmoïde** et l'intégrale de floue de **Sugeno** respectivement. Le tableau montre clairement que la fusion de scores appliquée à la multimodalité est toujours mieux que le meilleur système monomodal pour tous les couple des systèmes (voix ,visage). Le meilleur résultat est obtenu avec la fusion par l'intégrale de **Sugeno** dans le cas de la première approche pour le couple voix-visage (**LFCC,GMM),(DCTs,GMM)**. Une performance d'authentification presque parfaite est réalisée avec un **TEE=0.001±0.01%** et un **TR=99.99±0.04%**. Pour la deuxième approche la meilleure performance est atteinte dans le cas de la fusion par **MLP** pour le couple voix-visage ((LFCC, GMM),(FH,MLP)) avec TEE=0.23±0.21 et un TR=99.57±0.32. Nous nous intéressons uniquement au Protocole I (LP_1) pour lequel nous testons notre méthode de fusion. Le **tableau 6.5** présente la performance en terme de TEE dans l'ensemble d'évaluation et en terme TR dans l'ensemble de test pour les 15 couples des systèmes voix et visage (voix,visage). La méthode de normalisation et la méthode de fusion de scores sont : la **fonction double sigmoïde** et l'intégrale floue de **Sugeno** respectivement.

Tableau 6.5 Performances de fusion de scores pour les 15 couples des systèmes (voix, visage) dans d'ensemble d'évaluation et test (intervalle de confiance paramétrique à 95%).

Méthodes Voix ,Visage	Intégrale floue de Sugeno		Le meilleur système	
	TEE%	TR %	TEE%	TR %
(LFCC, GMM),(FH, MLP)	0.002±0.02	99.99±0.04	0.84±0.4	97.7± 1.15
(LFCC, GMM),(DCTs, GMM)	0.001±0.01	99.99±0.04	0.84±0.4	97.7± 1.15
(LFCC, GMM),(DCTb, GMM)	0.007±0.03	99.93±0.12	0.84±0.4	97.7± 1.15
(LFCC, GMM), (DCTs, MLP)	0.03±0.07	99.91±0.14	0.84±0.4	97.7± 1.15
(LFCC, GMM), (DCTb, MLP)	0.04±0.08	99.88±0.17	0.84±0.4	97.7± 1.15
(PAC, GMM),(FH, MLP)	0.17±0.18	99.49±0.35	1.66±0.57	96.25±1.45
(PAC, GMM),(DCTs, GMM)	0.24±0.21	98.69±0.56	4.37±0.91	91.55± 2.04
(PAC, GMM),(DCTb, GMM)	0.03±0.07	99.39±0.28	1.53±0.55	96.62±1.27
(PAC, GMM), (DCTs, MLP)	0.04±0.08	99.66±0.28	2.96±0.76	93.32±1.91
(PAC, GMM), (DCTb, MLP)	0.63±0.53	99.21±0.43	4.36±0.19	91.07±2.15
(SSC, GMM),(FH, MLP)	0.09±0.13	99.72±0.25	1.66±0.57	96.25±1.45
(SSC, GMM),(DCTs, GMM)	0.28±0.23	99.48±0.35	4.36±0.19	91.07±2.15
(SSC, GMM),(DCTb, GMM)	0.03±0.07	99.64±0.29	1.53±0.55	96.62±1.27
(SSC, GMM), (DCTs, MLP)	0.2±0.2	99.40±0.37	2.96±0.76	93.32±1.91
(SSC, GMM), (DCTb, MLP)	0.45±0.3	97.65±0.74	4.36±0.19	91.07±2.15

Le **tableau 6.5** montre clairement que la fusion de scores est toujours mieux que le meilleur système pour tous les couples des systèmes (voix,visage), et le meilleur résultat est obtenu avec le **couple voix visage (LFCC,GMM),(DCTs,GMM)** qui nous donne une performance d'authentification presque parfaite avec un **TEE=0.001±0.01%** et un **TR=99.99±0.04%**.

6.4.1.2 Travaux récents sur la multimodalité appliquée à la BDD XM2VTS

La série de résultats obtenue, est résumée par un tableau de synthèse (**tableau 6.6**) qui nous permet de comparer notre systèmes d'authentification multi-algorithmes et multi-biométries (visage,voix) à d'autres travaux connus en fusion multimodale, et aussi de faire la comparaison entre les systèmes multi-algorithmes et multi-biométries sur la base de données XM2VTS et la base de scores **XM2VTS** selon le protocole de Lausanne I.

Tableau 6.6 Tableau de synthèse des résultats obtenus en fusion multimodale

Date	Auteurs	Multimodalités	Niveau de fusion	Méthodes de fusion	BDD	Meilleurs résultats
2007	Sadeghi et al [85]	Multi algorithmes visage	Scores	Confidence Based Gating of Scores	XM2VTS	TRA=0.48%, TFR= 0.5% TR=98.45%
2009	Villegas et al [346]	Multi biométries visage,voix	Scores	Maximisation de AUC	Base de scores XM2VTS	TEE= 0.28%
2009	Li et al [42]	Multi biométries visage,voix	Scores	Nouvelle méthode basée sur TFR et TFA	Base de scores XM2VTS	TEE=0.27%
2009	E.Jarque et al [314]	Multi biométries visage ,voix	Caractéristiques	SVM	XM2VTS	TEE=0.008%
2009	Chia et al [299]	Multi biométries visage,voix	Scores	Hybride par min ou max ou somme	Base de scores XM2VTS	TEE=0.5%
2010	Lobrano et al [347]	Multi biométries visage,voix	Scores	DynamicLinearCombination(DLC)	Base de scores XM2VTS	TEE= 0.0045%
2012	Notre système [341,342]	Multi systèmes visages	Scores	Intégrales de floue de Sugeno	XM2VTS	TEE=1.14% TR=98.36%
2010	Notre système [340]	Multi biométries visage,voix	Scores	Intégrales de floue de Sugeno	Base des scores XM2VTS	TEE=0.001% TR=99.99%

Nous pouvons remarquer que la performance de notre système multi-algorithmes d'authentification de visages et presque équivalent à celui obtenu en 2007 par **Sadeghi et Khoshrou** avec un **TEE =1.14%** et un **TR=98.36%**.

Notre système multi biométries d'authentification de visage et voix (appliqué à la base des **scores XM2VTS**) est le meilleur comparé aux travaux 2009 et 2010 avec un **TEE=0.001%** et un **TR=99.99%** qui est justifié par l'efficacité de la méthode de normalisation de scores la fonction double sigmoïde et la fusion de scores par l'intégrale floue de **Sugeno**.

Pour la comparaison entre les deux approches multi systèmes et multi biométries, nous affirmons que le système multi-biométries a donné une performance presque parfaite avec un **TR=99.99%** et **TEE=0.001%**, mais avec beaucoup d'inconvénients : temps de capture de visage et voix très lent [5], coûteux, confort d'utilisateurs etc....

Par contre nous avons obtenu une performance satisfaisante par un système multi-algorithmes d'authentification de visages avec un **TR = 98,36%** et **TEE=1.14%** avec beaucoup d'avantages : temps de calcul réduit puisque le temps de capture est très supérieur au temps de traitement [5], confort des utilisateurs etc

C. Fusion Multi Algorithmes

6.5 Fusion basée sur l'entropie, moyenne, EFM et KPCA

Dans cette partie du travail nous essayons de faire la fusion des caractéristiques à la source afin de tirer le maximum d'informations. Et par souci d'encombrements de données nous codons nos images par deux algorithmes statistiques très simples qui sont la moyenne et l'entropie. Ces deux derniers sont ensuite fusionner, une réduction d'espace réalisée par l'EFM ou KPCA est assurée puis une classification par calcul de similarités nous permet d'obtenir nos scores qui à leurs tour sont fusionnés pour faciliter la décision de notre système. Notre approche est validée et testée sur des images frontales de la base de données XM2VTS et nous discutons les résultats en fonction de différents paramètres.

6.5.1 Architecture structurelle du système

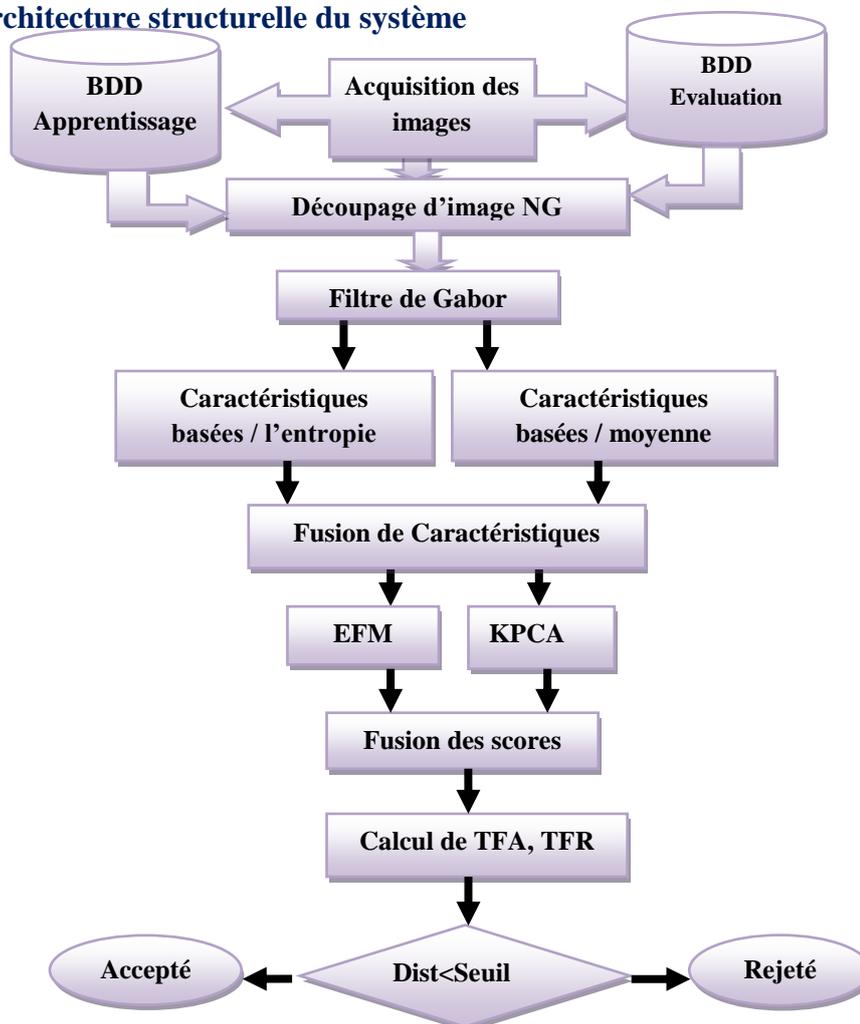


Figure 6.6 Architecture structurelle du système.

Comme le montre si bien la figure 6.6 notre système est structuré par plusieurs composantes fonctionnelles. Le prétraitement représente une fonction fondamentale du processus SRV. Nous préparons notre image pour être traitée dans cette phase. Ensuite, nous codons l'image pour en extraire des paramètres caractéristiques réduits. Cette caractérisation est réalisée par

deux algorithmes en parallèle : l'algorithme *moyenne* et l'algorithme *entropie*. Les matrices ainsi obtenues sont fusionnées par des méthodes de combinaison simples et le résultat est ainsi soumis à l'analyse. Il s'agit là d'une fusion multi algorithmes des caractéristiques (niveau caractéristiques). Cette fusion a l'avantage de traiter les caractéristiques brutes de l'image, donc nous agissons sur l'information à l'entrée. Et nous codons l'image pour contourner le traitement de taille de données énorme dû au nombre d'images issues de la base de données. Maintenant pour l'analyse nous procédons tout d'abord à la réduction de la taille des paramètres caractéristiques. Les méthodes que nous utilisons dans notre approche sont finement sélectionnées suite à notre étude bibliographique et l'état de l'art des travaux récents présenté au **chapitre 2**. Donc nous choisissons de faire une réduction d'espace par méthode le modèle de Fisher Amélioré **EFM** (Enhanced Fischer Model) et par Analyse en Composantes Principales à Kernel (Noyau) **KPCA** (Kernel Principle Component Analysis). Ces deux transformations sont réalisées en parallèle. L'**EFM** nous renforce la discrimination des classes et leur généralisation et la **KPCA** est là pour nous faciliter la tâche dans la classification non linéaire. Ces deux algorithmes sont une fois encore fusionnés et cette fois ci il s'agit d'une fusion des scores de même nature que la première c'est-à-dire une fusion multi algorithmes. Ensuite les paramètres de performances du **SRV** sont calculés. Finalement la classification est réalisée par mesure de similarité et une décision est prise se basant sur un test sur un seuil.

Organigramme de test

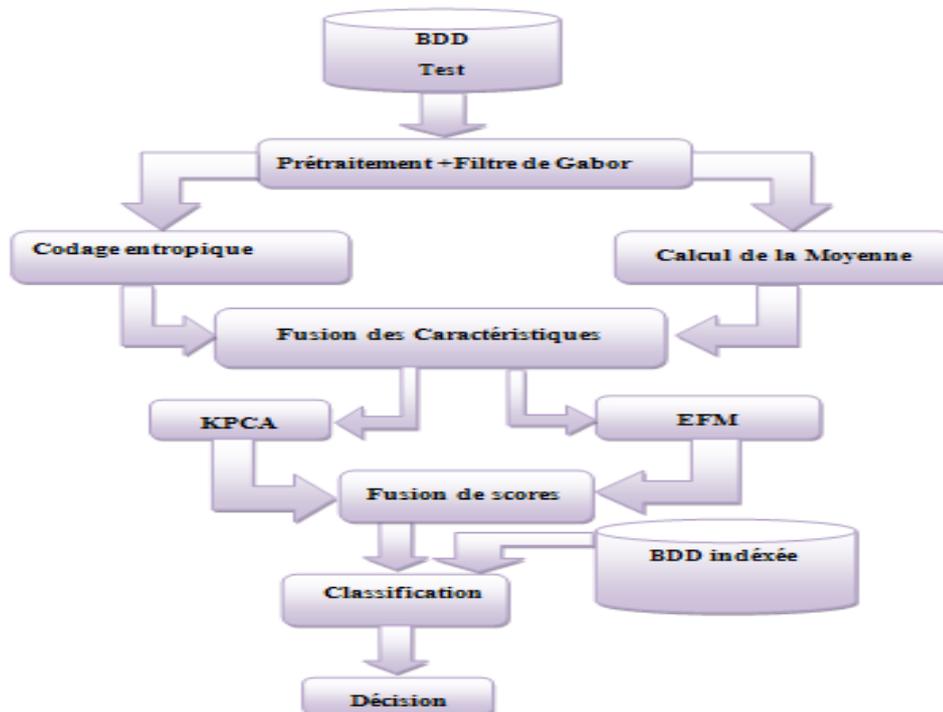


Figure 6.7 Organigramme de test

6.5.2 Implémentation du modèle

La mise en œuvre et essais de notre approche se présente en deux variantes :

- 1) La première basée sur les différents noyaux de la **KPCA**.
 - i) La prise en compte de plusieurs fonctions de noyaux (noyaux Gauss, Linear et Poly existant dans le toolbox Matlab 2009).
 - ii) La prise en compte des fonctions réalisant la fusion par combinaison. Dans notre travail nous nous limitons aux algorithmes de fusion par min, max et mean pour leurs simplicités. (Ces opérateurs sont utilisés dans la fonction **wfusmat** du Matlab 2009).
- 2) La deuxième sur le nombre de paramètres de l'**EFM**.
 - i) La variation du nombre de paramètres caractéristiques,
 - ii) Les différentes méthodes de fusion par combinaison citées précédemment.

6.5.3 Comparaison des performances

Nous effectuons des mesures pour diverses fonctions de noyau (Gauss, Linear, Poly) afin de retenir la meilleure au sens performances en changeant aussi les méthodes de fusion. Nous menons nos expériences pour $N_p(\text{PCA}) = 100$ avec la variation de $N_p(\text{EFM})$ et différentes fusions par combinaison (**mean, min, max**) dans le cas des trois noyaux du **Noyau Gausséen**, **linéaire** et **polynomial** sur la **BDD** évaluation et test. Les meilleurs résultats sont répertoriés dans le tableau suivant :

Tableau 6.7 Valeurs **TEE** et **TR** pour différents noyaux de la **KPCA**(fusion multialgorithmes).

Kernels	Gausséen	Linéaire	Polynomial
TR	Min=76.75%	Min =76.20%	Min=76.20%
TEE	Min=23.24%	Mean= 23.79%	Mean=25.78%

D'après ces résultats nous pouvons dire que le **Kernel Gausséen** reste celui qui présente le meilleur taux de réussite pour la fusion **min**. Dans le cas du **Kernel Gausséen**, nous constatons que la fusion par **min** donne le meilleur **TR = 76.75%** dans l'ensemble d'évaluation. Les valeurs de **TR** obtenus dans le cas de la fusion par **max** et **mean** sont semblables et stables autour de **TR = 64.75%**. Ce qui confirme l'importance et l'apport de la fusion basée sur **min**. Nous constatons une stationnarité dans les valeurs de **TR** pour la fusion min. Ce qui prouve la **stabilité** de ce type de fusion pour **divers Kernels**.

D. Fusion des informations profondeur et couleur

6.6 Fusion de la couleur RVB et profondeur

6.6.1 Fusion des caractéristiques par combinaison (couleur RVB + profondeur)

La reconnaissance faciale **3D**, combinant la forme et l'intensité / couleur **2D** dans la formation, est un sujet en développement. La combinaison d'information **2D** et **3D** fournit une occasion d'améliorer les performances de reconnaissance des visages. Wang et al. [158] décrivent les points caractéristiques du visage à l'aide de réponses des filtres de Gabor dans un domaine **2D** et les signatures de points dans un domaine **3D**. La classification a été faite ensuite par des machines à vecteurs de support à une décision "graphe orienté acyclique". Tsalakanidou et al. [161] ont construit un classifieur "embarqué modèles de Markov cachés (EHMM)" basé sur la profondeur et l'intensité, puis utilisé des règles de fusion pour combiner les scores. Dans [35], Chang et al. ont évalué le système de reconnaissance avec différentes combinaisons de l'information **2D** et **3D** et ont montré que la combinaison des informations **2D** et **3D** était plus efficace dans la caractérisation d'un individu. Dans les méthodes existantes susmentionnées, certains problèmes demeurent distincts et en suspens:

- Certains travaux [348,349,350] ont été effectués pour comparer la reconnaissance faciale **2D** et **3D**. Cependant, certains détails sont ignorés sur la façon dont la profondeur et l'intensité d'information contribue à la reconnaissance des visages avec une expression des variations d'éclairage.
- Seule la fusion au niveau de décision est considérée.

Nous consacrons cette partie à l'implémentation du modèle d'identification basé sur la fusion des caractéristiques et des scores *profondeur* et *couleur RVB* basée sur une classification métrique. Les différentes étapes de la conception sont :

- Projection de l'image **3D** en **2D** : Pour des facilités d'implémentation nous projetons nos images **3D** sur un espace **2D**. Nous réalisons deux projections en parallèle :
 - Génération des cartes de profondeur **2.5**.
 - Projection de l'image **3D** dans l'espace des couleurs **2D** couleur (**RGB**).
- Fusion de données (image **2D RGB** et **2.5D** profondeur).
- Extraction de paramètres (Filtrage **GABOR**).
- Réduction d'espace (**LDA** et **EFM**).
- Classification : Les descripteurs extraits par le filtre de **Gabor** sont projetés d'une façon parallèle sur deux espaces. La première est calculée par **LDA** et la deuxième obtenue par l'**EFM**. Les vecteurs projetés dans chacune des espaces séparément sont

comparés aux vecteurs de la base d'apprentissage indexée et un calcul de distances intra et extra est effectué. Ce calcul est réalisé par des normes Euclidiennes (Euclidienne norme 1 (L_1)).

- Décision

6.6.1.1 Architecture fonctionnelle du SRV à base de fusion des caractéristiques

Cesystème de reconnaissance de visages 3D se présente comme suit :

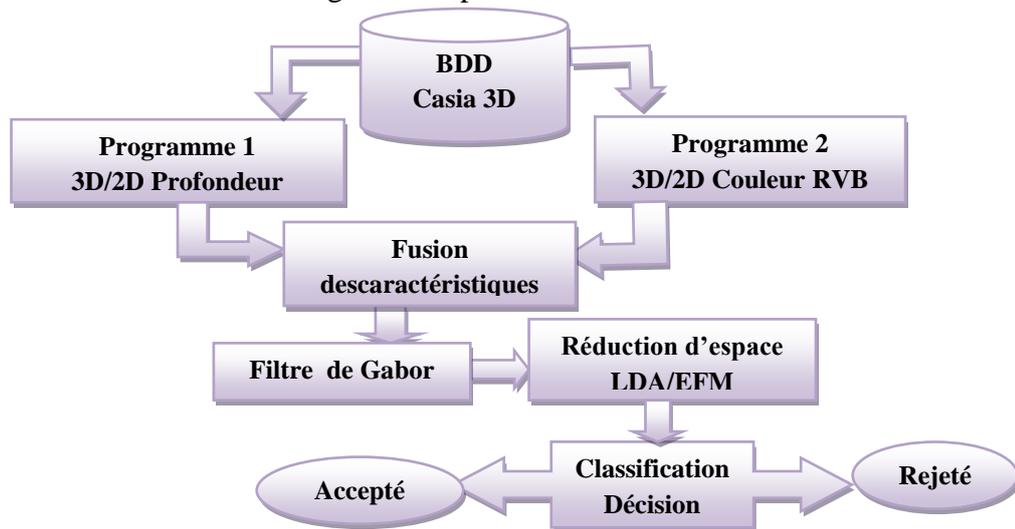


Figure 6.8 SRV 3D à base de fusion des caractéristiques : *profondeur* et *couleur RVB*

Nous utilisons une fusion des caractéristiques. Cette fusion présente l'avantage de prendre en compte l'information contenue dans l'image. Mais elle a l'inconvénient de mettre en jeu une taille importante de données. Cet inconvénient est remédié par le filtre de **Gabor** qui nous permet de retirer que les détails dans nos images. Ces détails représentent des contours dans l'image qui sont très représentatifs puisqu'il s'agit de points et régions d'intérêt dans le visage (exemple : contour du nez (narines), des yeux, la bouche, sourcils et tout autre détails dans le visage). Cette étape de filtrage est suivie d'une méthode de réduction d'espace à pouvoir discriminatoire c'est l'**EFM** qui nous prépare à la phase classification. Le choix de l'**EFM** est justifié par les résultats obtenus dans les travaux de M.Belahcène et A.M.Ouamane en 2011[55].

6.6.1.2 Implémentation de l'Approche SRV 3D (couleur RVB+profondeur)

Pour augmenter la lisibilité de notre programme et faciliter les différents tests nécessaires, nous avons essayé de le partitionner en un ensemble de modules.

1 Prétraitement de visage

En projetant l'image **3D** couleur sur un espace plan nous obtenons notre image couleur **2D**. Par un programme de prétraitement nous obtenons la carte profondeur appelée

aussi l'image **2.5D**(figure 6.9). Ces deux images après avoir subies des traitements séparés seront fusionnées.

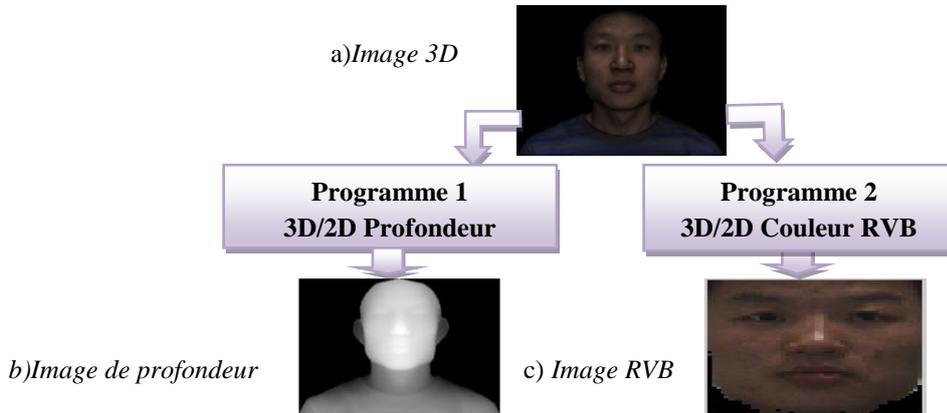


Figure 6.9 Phase de prétraitement de l'image **3D** couleur

2 Fusion des caractéristiques

Nous fusionnons deux images originales : l'image **3D** couleur projetée sur un espace **2D** et l'image **2.5D** qui représente les cartes de profondeur. Pour cela nous utilisons des méthodes de fusions appliquées aux coefficients d'approximations et les coefficients de détails de l'ondelette. Les deux images doivent être de la même taille et sont censées être associées à des images indexées sur une palette de couleurs commune. Les méthodes de fusion par combinaison des coefficients de l'ondelette sont ainsi calculées. Nous changeons le mode d'extension de rembourrage symétrique, en utilisant la commande: **dwtmode** ("**symétrique**"), et la fonction (**wfusing**) pour la fusion d'images. L'algorithme suivant nous présente une méthode de fusion de données réalisée sur les images profondeur et couleur transformées par ondelettes. Cette dernière transformation nous permet de réduire les données considérablement à traiter en tenant compte des détails contenus dans l'image. En effet, il existe différentes méthodes d'extraction de paramètres, et plusieurs types de fusion de données. Nous effectuons nos expériences sur plusieurs types de fusion et nous gardons le type qui convient le mieux à notre étude.

3 Extraction des paramètres et classification basée sur LDA/EFM

Il existe des différents filtres de **Gabor** pour diverses résolutions et orientations. Celui qui nous donne le meilleur score est maintenu. La classification est réalisée par la norme Euclidienne L_1 . Nous calculons les paramètres de performances du **SRV** pour différentes paramètres caractéristiques (N_p variable).

4 Résultats

Tableau 6.8 Tableau des résultats de fusion couleur RVB+profondeur

Approche	Combinaison		
Types de fusion	Fusion des caractéristiques		
Méthodes de fusion	Mean	Max	Min
Méthodes de réduction	LDA	EFM	EFM
N_p	10	40	70
Distance	L1		
TR(%)	54.76	75.75	75.25

Les résultats obtenus avec la classification par **LDA** ne sont pas satisfaisants. Car le meilleur **TEE** est de l'ordre de **45.24%** avec un taux de réussite **TR=54.76%** en utilisant seulement **10 paramètres** de **LDA**. Après le choix du filtre optimal ainsi que la mesure de distance, nous calculons les taux d'erreur et les taux de réussite dans le cas où nous faisons une projection sur l'espace **LDA/EFM** du vecteur caractéristique. Nous appliquons la fusion de données par trois méthodes (**max**, **min**, **mean**). La méthode de fusion par la moyenne (**mean**) donne un bon résultat par rapport aux méthodes de fusion par max et min. Les meilleures performances du système se trouvent dans l'intervalle de N_p entre **[70,100]**. La méthode **EFM** est une fois de plus la meilleure réduction dans le SRV. Pour N_p (**EFM**) entre **[10,70]** nous remarquons que les différences entre les erreurs dans les trois ensembles (**max**, **min**, **mean**) sont faibles. Ceci est une caractéristique très importante, car on peut dire que le système de reconnaissance de visages est stable. D'après ces résultats nous pouvons dire que la méthode de fusion par la moyenne (**mean**) donne le meilleur taux de réussite **75.75%** avec le nombre de paramètres égal à $N_p = 40$. Et la fusion par **min** donne un taux égal à **75.25%** avec un nombre de paramètres $N_p = 70$. Malgré que le taux d'erreur reste élevé pour dire que notre système est stable, mais nous pouvons affirmer que c'est un résultat qui nous encourage à faire davantage d'investigations.

6.6.2 Fusion des scores par combinaison (RVB + profondeur)

Notre but est d'étudier la fusion des scores entre les images 3D couleur et 3D profondeur à partir de la base de données **CASIA 3D**, et de classifier des personnes particulières pour les identifier. Nous présentons les différentes étapes de la conception de notre système. Nous effectuons nos expériences sur plusieurs types de fusion et nous gardons le type qui convient à notre étude.

Notre programme principal se présente comme suit :

- 1) Prétraitement de l'image
 - Extraction de la couleur du visage 3D et sa projection sur 2D.
 - Génération des cartes de profondeur du visage 3D.
- 2) Projection des images sur l'espace **PCA** (visages 2D et visages 2D de profondeur).
- 3) Mesure des distances (pour visages 2D et visages 2.5D de profondeur): Euclidienne norme 1 (L_1), Euclidienne Norme 2 (L_2), **covariance**.
- 4) Fusion des scores.
- 5) Classification et décision.

6.6.2.1 Architecture du système de conception

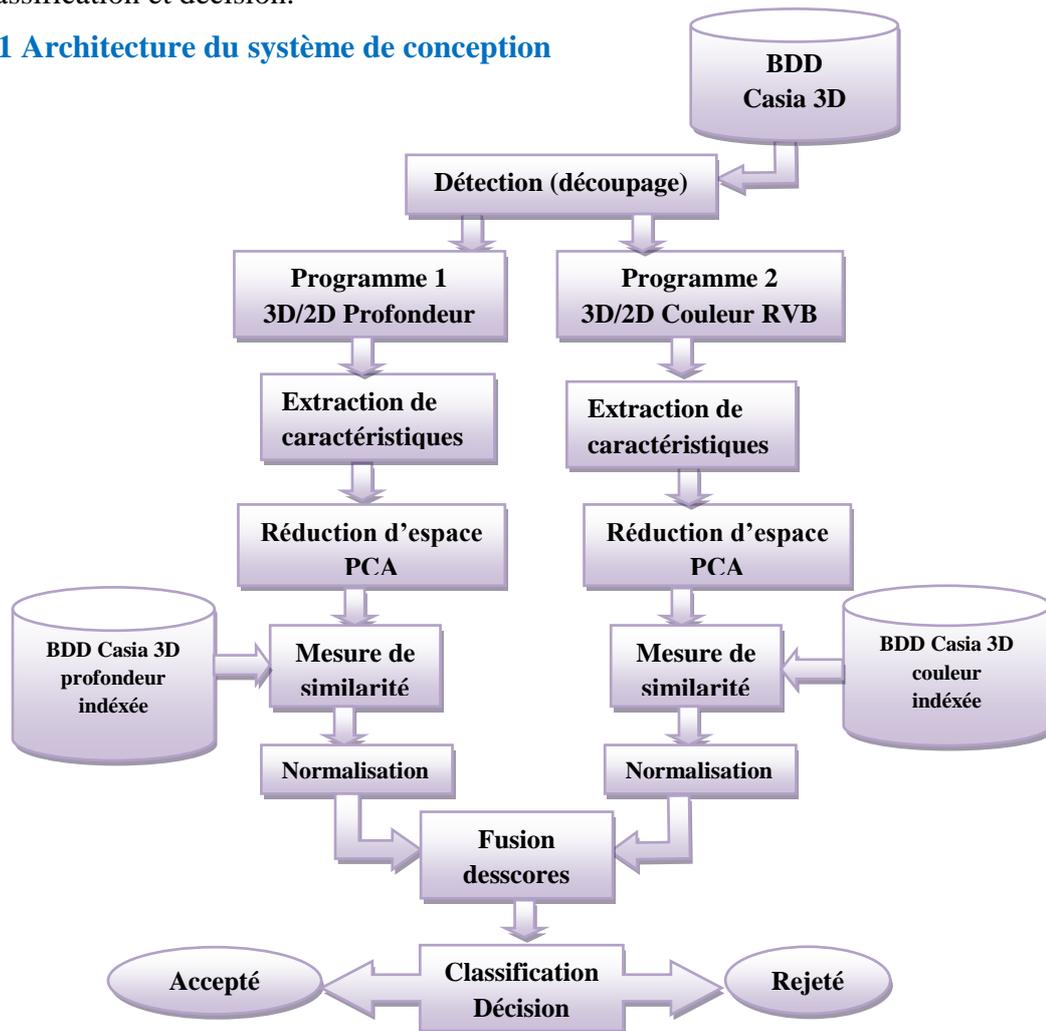


Figure 6.10 Architecture structurelle du système de conception.

Dans l'étape de fusion nous réalisons la fusion des scores, ces derniers sont représentés par les distances calculées précédemment, la méthode de fusion que nous utilisons est la fusion par combinaisons simples des scores basés sur les calculs de moyenne, max et min. Nous utilisons des instructions disponibles dans Matlab, cette instruction fusionne deux matrices ou deux vecteurs.

Dans un processus d'authentification, la stratégie de décision nous permet de choisir entre les deux alternatives suivantes: l'identité de l'utilisateur correspond bien à l'identité de la personne proclamée (ou recherchée) ou elle ne correspond pas. En comparant le modèle de l'utilisateur à celui qui existe dans la base de données. Cette phase est basée sur le seuil prédéfini. L'algorithme suivant explique le processus de la phase de décision :

Algorithme de processus de la phase de décision

Lire V_{test}

$Y_1 = \text{prétraitement 1}(V_{test})$

$Y_2 = \text{prétraitement 2}(V_{test})$

Extraction des paramètres ($Y_1|Y_2$)

$DIS_V = \text{mesurer distance entre } XV, Y_1$

$DIS_P = \text{mesurer distance entre } XP, Y_2$

$DIS = \text{Fusion entre } DIS_V, DIS_P$

Seuillage

Si $DIS < \text{seuil}$

Personne connue (identité parfaite)

Sinon

Personne inconnue (identité non parfaite)

Fin

V_{test} : image de visage test.(image de l'utilisateur)

$XV \setminus XP$: modèle (couleur,profondeur) de la personne proclamée stocké dans la base des données

DIS_V2D couleur : distance entre XV et Y_1 (Projetées dans le même sous espace)

DIS_V2D profondeur : distance entre XP et Y_2 (Projetées dans le même sous espace)

6.6.2.2 Résultats

Nous réalisons plusieurs types de fusion par combinaison des scores (mean, max, min).Nous étudions d'abord le cas où le nombre de paramètres $N_p(\text{PCA})$ varie de 20 à 200. Nous présentons la comparaison de TR entre les normes : L_1 , L_2 et covariance L_3 :

Tableau 6.9 TR de la fusion des scores (couleur RVB et profondeur) pour L_1 , L_2 et L_3

	Mean	Min	Max
L_1	79.50	78.85	74.74
L_2	65.79	78.75	63.33
L_3	76.45	76.75	67.48

D'après ces résultats nous pouvons dire que dans le cas de la norme L_1 la méthode de fusion des scores par moyenne pour les images de visages 2D couleur RVB et 2.5D profondeur donne le meilleur taux de précision 79.50% avec un nombre de paramètre égal à $N_p= 60$. Par contre la fusion par min donne un taux égal à 78.85% avec $N_p= 20$.

Malgré que le taux d'égale erreur reste élevé pour dire que notre système est stable. Mais nous pouvons affirmer que c'est un bon résultat dans le cas de la reconnaissance de visage en 3D.

Tableau 6.10 Tableau comparaison des résultats de la **fusion des caractéristiques** et des **scores** de la couleur RVB et profondeur.

Approche	Combinaison					
	Fusion des caractéristiques			Fusion des scores		
Types de fusion	Mean	Max	Min	Mean	Max	Min
Méthodes de fusion	LDA	EFM	EFM	ACP		
Méthodes de réduction	10	40	70	60	20	
Np	L1			L1		
Distance	54.76	75.75	75.25	79.50	78.85	74.74
TR(%)						

Ce tableau nous montre que la meilleure performance **SRV** est le cas de fusion des scores malgré une réduction d'espace primaire. Ceci est dû au fait que l'information traitée est prise sur l'image de visage directement sans aucun filtrage. En plus le traitement des scores est plus rapide que celui des caractéristiques vu la taille de données mises en jeu.

6.7 Fusion 2D couleur HSV et 3D profondeur

Dans cette partie du travail, nous tentons de résoudre ces problèmes restants dans la reconnaissance du visage et nous examinons comment l'information de profondeur et de couleur HSV contribue à la reconnaissance faciale lorsque la variation d'éclairage et l'expression sont prises en compte. La classification utilisée est basée sur **SVM** et ceci d'après les travaux de M.Belahcène en 2011[351].

Les principales contributions dans cette partie du travail sont les suivantes:

- Faire l'étude comparative des méthodes de réduction d'espace de données suivantes : L'Analyse Discriminante Linéaire (ADL), le Modèle Discriminant Linéaire Amélioré de Fisher (EFM) et la ADL Directe (ADLD) pour choisir le meilleur espace de projection des données de profondeur et de couleur HSV.
- Etudier la fusion au niveau des caractéristiques.
- Etudier la fusion au niveau de scores par: l'intégrale floue de **Sugeno**, de **Choquet** et les machines à vecteurs de support (**SVMs**).

6.7.1 Identification de visages à base de fusion par combinaison

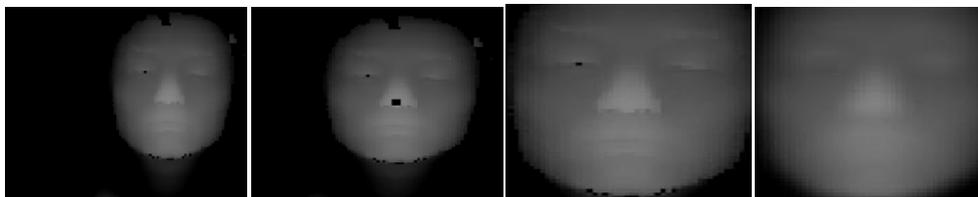
(couleur HSV+ profondeur)

Idéalement, un système de reconnaissance faciale doit pouvoir identifier des visages grâce à l'information profondeur et/ou couleur. 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** correspondantes. Dans cette section, nous décrivons comment les données originales en **3D** sont prétraitées.

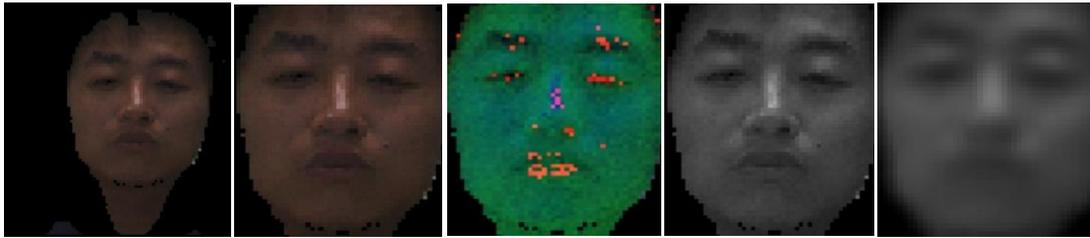
Les données **3D** sont converties en une image en profondeur (voir **Figure 6.11(a)**) et une image de couleur (voir **Figure 6.12(a)**). Dans la plupart des images, le nez est la plus proche partie du visage au scanner **3D**, c'est à dire, il a la valeur la plus élevée en profondeur entre tous les 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 **6.11(b)**). Après avoir détecté le nez, toutes les images dans la base de données sont découpées par une fenêtre rectangulaire de taille **57x47** centrée autour du centre du nez (voir **Figure 6.11(c)**, **6.12(b)**). Pour l'image de couleur **RGB**, nous avons utilisé l'espace de couleur **HSV** (Hue, Saturation, Value) (voir **Figure 6.12(c)**) et nous considérons la composante **V** " Value " parce qu'elle est moins sensible à la variation d'éclairage (voir **Figure 6.12(d)**) [20].

Toutefois, en raison de la qualité des données originales en **3D**, les images de profondeur et de 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 par la sous-fenêtre de taille **5 × 5** autour de lui. Le résultat est montré dans les **Figures 6.11 (d)**, **6.11(e)**.



(a) Image de profondeur (b) Détection du bout de nez (c) Image découpée (d) Image après suppression de bruit et remplissage de trous.

Figure 6.11 Prétraitement d'image de profondeur



(a) Image de couleur (b) Image découpée (c) Image en système HSV (d) Composante V de HSV (e) Image après suppression de bruit et remplissage de trous.

Figure 6.12 Prétraitement de l'image couleur 3D.

Bien que la distance euclidienne soit optimale dans la théorie, les diverses expériences ont montré que la distance euclidienne est surpassée par d'autres distances. Une d'elles est la corrélation normalisée (normalized correlation) équation 6.1.

6.7.2 Résultats

6.7.2.1 La base de données CASIA 3D

Nous utilisons la base de données des visages CASIA 3D [338] pour tester notre système d'identification proposé. La base est construite par un numériseur 3D sans contact **Minolta VIVID 910** travaillant sur le mode rapide. Cette base de données contient 123 sujets, chaque sujet ayant 37 ou 38 images avec les variations individuelles de poses, d'expression, d'illumination, de variations combinées des expressions sous illumination et de pose sous des expressions. Cette base de données contient des variations complexes qui sont difficiles à n'importe quel algorithme.

Dans notre travail, nous avons étudié les variations d'illuminations (**Figure 6.13**), expressions (**Figure 6.14**) et les variations combinées des expressions sous illumination. Nous avons donc utilisé 15 images pour chaque sujet. La base de données de **1845 images** est divisée en deux sous-ensembles, l'ensemble de **galerie** et l'ensemble de **test**. L'ensemble de galerie contient une image pour chaque sujet (sous la condition de **vue de face, l'éclairage des bureaux et l'expression neutre**).

L'ensemble de **test** est en outre divisé en **trois sous-ensembles**:

- **IV** (400 images): variations d'éclairage, y compris l'éclairage en haut, bas, gauche et à droite.
- **EV** (500 images): variations d'expression, y compris sourire, rire, la colère, la surprise et les yeux fermés.
- **EVI** (500 images): variations d'expression en vertu de l'éclairage du côté droit.

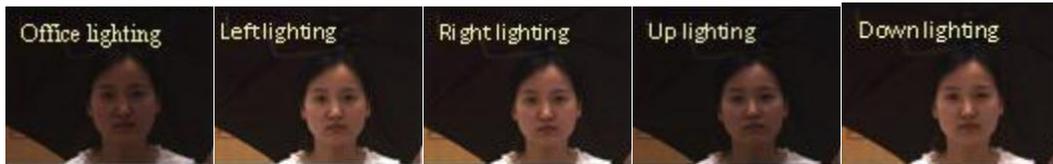


Figure 6.13 Illumination variations of the CASIA 3D Face Database.



(a) sourire (b) rire (c) colère (d) surprise (e) yeux fermés

Figure 6.14 Variations d'expression de la base de données des visages 3D CASIA

6.7.2.2 Comparaison des méthodes de réduction sur les images de profondeur et couleur HSV

Dans notre expérience, des images de profondeur et des images de la composante **V** de l'espace de couleur **HSV** sont utilisées pour caractériser les sujets. Les figures 6.15 et 6.16 présentent les performances des méthodes de réduction pour d'identification des images de profondeur et de couleur.

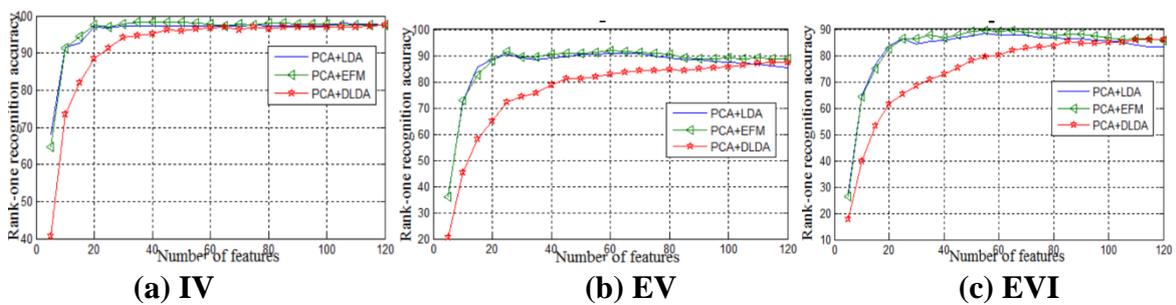


Figure 6.15 Performances des méthodes de réduction pour d'identification des images de profondeur.

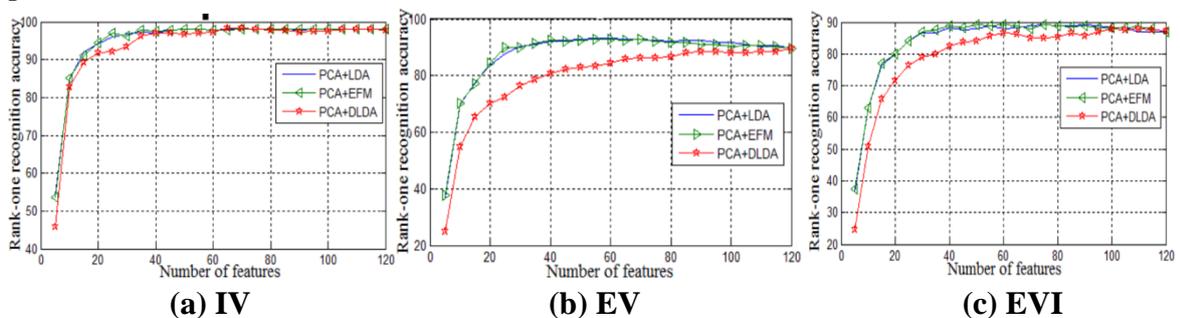


Figure 6.16 Performances des méthodes de réduction pour d'identification des images de couleur.

Nous pouvons constater dans les **figures 6.15** et **6.16** que la méthode **PCA+EFM** est mieux performante que la **PCA+LDA** suivie de la **PCA+DLDA** en employant un nombre restreint de caractéristiques dans notre base de données **CASIA 3D** pour les images de profondeur et de couleur. Le tableau 1 donne les meilleurs résultats pour la méthode de réduction d'espace **PCA+EFM**.

Tableau 6.11 Les meilleurs résultats pour la méthode de réduction d'espace PCA+EFM.

Sous-ensembles de test	Images de profondeur		Images en couleur	
	TR(%)	Nombre de caractéristiques	TR(%)	Nombre de caractéristiques
IV	98.3740	35	98.1707	70
EV	92.0325	60	93.0081	55
EVI	89.5935	55	89.1057	60

Le tableau montre que l'information de profondeur est presque équivalent à l'information de couleur, mais l'information de profondeur est moins sensible à la variation d'éclairage et l'information de couleur est plus performante pour des variations d'expression.

6.7.2.3 Fusion par combinaison au niveau caractéristiques (FeatureLevel)

Pour la **fusion** au niveau **caractéristiques** nous avons utilisé deux types :

- La fusion par la **concaténation** pour les deux vecteurs caractéristiques de profondeur et de couleur.
- La fusion des pixels de l'image de profondeur et de couleur par: la **moyenne**, le **produit**, le **minimum** et le **maximum**.

La **figure 6.17** présente les performances des méthodes de fusion au niveau caractéristiques pour l'identification multimodale des visages.

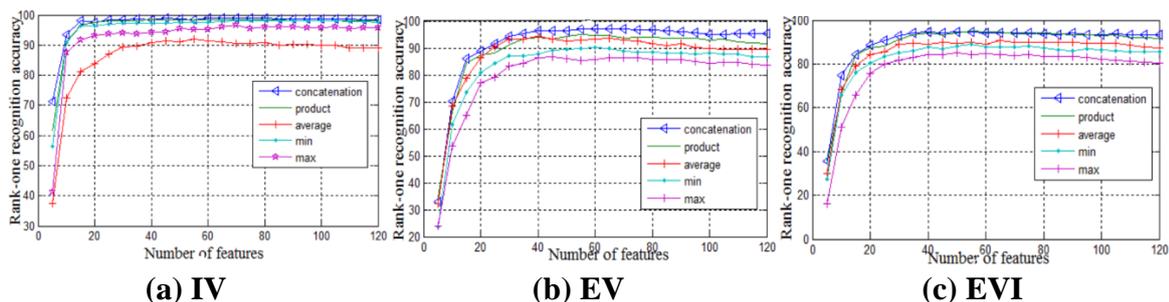


Figure 6.17 Performances des méthodes de fusion au niveau caractéristique pour l'identification multimodale des visages.

Nous avons remarqué dans la **figure 6.17** que la fusion par la concaténation et le produit ont donnés presque la même performance pour les trois sous-ensembles de test. Le **tableau 6.12** présente les meilleurs résultats.

Tableau 6.12 Les meilleurs résultats pour la fusion des caractéristiques de couleur **HSV** et **profondeur**

Sous-ensembles de test	Méthodes de fusion	TR(%)	Nombre de caractéristiques
IV	Concaténation	98.9837	60
EV	Concaténation	97.0732	60
EVI	Produit	94.9593	55

6.7.2.4 Fusion par combinaison et classification au niveau scores (Score Level)

La fusion au niveau des scores est le type de fusion le plus utilisé car elle peut être appliquée à tous les types de systèmes (contrairement à la fusion des caractéristiques), dans un espace de dimension, avec des méthodes relativement simples et efficaces mais traitant plus d'information que la fusion de décisions. La fusion de scores consiste donc à la classification pour la décision finale.

Il existe deux approches pour la fusion des scores obtenus par différents systèmes :

- i) La première approche est de traiter le sujet comme un problème de combinaison,
- ii) tandis que l'autre approche est de voir cela comme un problème de classification.

Pour la première approche nous avons utilisé la **logique floue** à base de l'intégrale de **Sugeno** et de **Choquet**. Pour la deuxième approche nous avons utilisé une machine à vecteurs de support (**SVM**). Nous utilisons une machine à vecteurs de support (**SVM**) avec un noyau **RBF** (Radial Basis Function ou fonction à base radiale). Le **SVM** a été implémenté en utilisant la librairie **LibSVM** disponible gratuitement au site (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>). Le **SVM** est un séparateur à **deux classes** mais dans notre problème on a **123 classes**. Nous avons donc utilisé **123 SVMs**.

Le tableau **6.13** présente les performances de combinaison et classification des images de profondeur et de couleur. Afin de s'assurer que la fusion de scores provenant de différents systèmes soit cohérente, les scores doivent d'abord être transformés dans un domaine commun : on parle alors de normalisation de scores, la méthode de normalisation utilisée dans notre cas est la méthode **min-max** donnée par :

$$s'_{ik} = \frac{s_{ik} - \min(\{s_i\})}{\max(\{s_i\}) - \min(\{s_i\})} \quad (6.3)$$

Avec:

s_{ik} le $k^{\text{ième}}$ score de sortie du $i^{\text{ème}}$ SVM, où $i = k=1, 2, \dots, N$.

(N est le nombre de sujets et égale 123).

Tableau 6.13 Taux de reconnaissance pour la combinaison et la classification des scores de couleur HSV et profondeur.

Sous-ensembles de test	Méthodes de fusion		
	Intégrale de Sugeno	Intégrale de Choquet	SVMs
IV	98.98	98.98	100
EV	95.93	95.77	99.72
EVI	93.33	94.1463	99.18

Le **tableau 6.13** montre que la fusion par combinaisons des scores pour les deux intégrales floue de **Sugeno** et **Choquet** donne des résultats équivalents à ceux de la fusion au niveau des caractéristiques.

La fusion par classification de scores en utilisant les **SVMs** associé à l'**ACP+EFM** donne des résultats presque parfaits appliqué à la **BDD Casia 3D** (version 4). Nous pouvons dire que notre système multimodal d'identification des visages 3D (IV, EV, EVI) utilisant la composante **V** du système **HVS** de couleur conjointement à la profondeur est pratiquement insensible à la variation d'éclairage et aux expressions faciales. Une fois de plus nous confirmons que le **SVM** est un outil efficace dans toutes les applications : en détection (dans nos travaux réalisés en 2011, classification pour l'identification monomodale (sections **5.5** (2^{ème} expérience) et **5.7 chapitre 5**, fusion multisystèmes et dans la fusion des scores par classification. Alors que dans le cas de l'identification des visages 2D neutres en monomodale (visage) et en multibiométries (voix, visage) (voir **tableau 6.4** section **6.4.1.1 chapitre 6**) l'intégrale de **Sugeno** nous donne les meilleurs résultats classification.

6.7.2.5 Tableau comparatif et discussion

Tableau comparatif des fusions des caractéristiques et des scores de la couleur HSV et profondeur.

Tableau 6.14 Tableau comparatif des TR des différentes approches et méthodes de fusion HSV et profondeur.

Approche	Combinaison			Classification		
Types de fusion	Fusion des caractéristiques	Fusion des scores		Fusion des scores		
Méthodes de fusion	Concaténation	Choquet	Sugeno	SVM		
Sous ensemble de test	IV	IV	IV	IV	EV	EVI
TR(%)	98.9837	98.98	98.98	100	99.72	99.18

Dans la **fusion** couleur **HSV** et **profondeur** l'approche par classification **SVM** est meilleure que celle combinaison ce qui n'était pas le cas dans les fusions multisystèmes et multibiométries (fusion par combinaison : intégrale floue **de Sugeno** et classification par SVM). Dans toutes les applications étudiées la fusion des scores donne les meilleurs résultats de performance du **SRV**. Le système **HSV** est plus informant que le système **RVB** avec un **TR=79.50%** (pour la fusion **mean**, distance Euclidienne norme **L₁** et **N_p=60**) contre **TR=100%** en système HSV composante V (pour la fusion **SVM**, distance **corrélacion normalisée** et **N_p<70**).

La **fusion des scores** couleur **HSV** et **profondeur** par **classification SVM** est parfaite avec un taux de réussite à 100% dans des visages exposés à la variation d'illumination. Mais elle reste sensible aux changements d'expressions faciales et à la combinaison changement d'expression faciales et lumière.

Nous pouvons dire que :

- l'algorithme **SVM** est très efficace quel que soit l'application envisagée.
- l'algorithme **SVM** est aussi bon en phase détection, qu'en phase classification aussi bien qu'en phase en fusion.
- La meilleure méthode de réduction reste l'ACP+EFM.
- Le filtre de **Gabor** s'il est bien utilisé nous permet d'extraire l'essentiel des caractéristiques globales du visage.
- Les **phases du filtre** sont mieux caractéristiques du visage que les amplitudes.
- Le système **HSV** est plus informant que le système **RVB**
- La fusion des scores est plus efficace que celle des caractéristiques.
- La norme **L₁** et la **corrélacion normalisée** sont les meilleures dans la classification par mesure de similarité.
- La classification par **SVM** est meilleure que celle par mesure de distances c'est plus souple, rapide et nous évite le calcul de seuil.

6.8 Synthèse sur les SRVs multimodaux

La synthèse des résultats obtenus dans cette partie du travail est regroupée dans le tableau

6.15.

Tableau 6.15 Synthèse des résultats des SRVs multimodaux étudiés.

Types SRV	Approches fusion	Types de fusion	Méthodes de fusion	Normalisation	Classification	N _p	BDD	TR(%)
Multisystèmes	Combinaison	scores	Intégrale de Sugeno	Double sigmoïde	Corrélation normalisée	80	XM2VTS	98.36 ±0.75
	Classification	scores	MLP SVM	QLQ Znrm	Corrélation normalisée	<100	XM2VTS	97.84±0.77 97.44±0.77
Multibiométries		Voix , Visage (LFCC,GMM), (DCTs, GMM)	Intégrale de Sugeno	Double sigmoïde	---	---	XM2VTS des scores	99.99±0.04
Multialgorithmes		Caractéristiques +scores	Min	minmax	L ₂	100	XM2VTS	76.75
3D RVB+profondeur	Combinaison	Caractéristiques	Max	minmax	L ₁	40	CASIA3D	75.75
	Combinaison	Scores	Mean	minmax	L ₁	60	CASIA3D	79.50
3D HSV+profondeur	Combinaison	Caractéristiques	Concaténation	minmax	Corrélation normalisée	60	CASIA3D IV	98.9837
		Scores	I.Choquet I.Sugeno	minmax	Corrélation normalisée		CASIA3D IV	98.98 98.98
	Combinaison	Scores	SVM	minmax	Corrélation normalisée		CASIA3D IV	100

Les meilleurs SRV sont à base :

- d'une extraction de caractéristiques par filtre de **Gabor** (pour les SRV multisystèmes , multibiométries et multi algorithmes conçus),
- d'une réduction d'espace **ACP+EFM**,
- la classification par **corrélation normalisée** et **SVM**,
- la fusion par combinaison des scores (intégrale floue de **Sugeno** associée à une normalisation **double sigmoïde**) dans le cas du SRV multibiométries et multisystèmes,
- la fusion par *classification des scores* **SVM** en SRV multisystèmes et multimodales (3D couleur HSV+profondeur).

6.9 Conclusion

D'après l'étude faite et toutes les expériences réalisées nous pouvons dire que la fusion des scores 2D (couleur RVB) et 3D (profondeur) avec les métriques améliore certes les performances du système SRV monomodale. La norme L_1 est la meilleure pour le calcul de scores par rapport à la norme L_2 et la covariance L_3 . La fusion des scores par la **moyenne** est la méthode qui a donné les meilleurs résultats dans ce cas. Ceci dit une étude plus poussée sur les distances reste envisageable. Des améliorations au niveau du prétraitement sont aussi à tenir en compte. Ainsi qu'une meilleure méthode de réduction. D'après les résultats des expériences réalisées nous pouvons dire que l'ACP+EFM reste une méthode de discrimination performante d'une façon générale. La **fusion des scores** reste une méthode très efficace car sans la fusion nous avons avec un $N_p(\text{EFM}) = 100$; un $\text{TR} = 92.25\%$ pour les visages pour une fusion par combinaison **min** et la distance L_2 . Nous pouvons dire que cette dernière appliquée à des images de visages **2D neutres** de la base de données **XM2VTS** en *multi algorithmes* et en *multi biométries* (voix, visage) et affirmons que cette méthode est performante avec un $\text{TR}=99.99\%$ en **2D** et *multi biométrie*. Nous confirmons aussi que cette méthode reste une bonne candidate pour aborder les visages à 3D. Des améliorations restent à réaliser surtout au niveau de la phase prétraitement et aussi et surtout dans la phase de classification car notre utilisation à la mesure métrique reste un choix guidé par sa simplicité. Ceci dit d'autres normes comme Manhattan ou mesures géodésiques restent envisageable, et toujours dans l'objectif de rehausser ces paramètres de performances du système biométriques. Les résultats sont dans l'ensemble satisfaisants, mais le taux de faux rejet reste moyen. Ceci est sûrement dû à l'étape de prétraitement qui reste à améliorer ainsi que le choix d'un meilleur classifieur. Dans l'expérience de fusion 2D et 3D basée sur SVM nous utilisons la profondeur et la couleur des images pour construire un classificateur robuste pour l'identification du visage. Puisque la dimensionnalité des caractéristiques d'images de profondeur et de couleur est très élevée. Nous avons étudié trois méthodes de réduction d'espace de données. Puis nous avons fait des études comparatives sur la fusion de l'information de profondeur et de couleur pour les deux niveaux : caractéristiques et scores pour la construction d'un classifieur efficace. En analysant nos résultats expérimentaux dans la base de données du visage **CASIA 3D** avec des variations complexes, nous illustrons la performance prometteuse du schéma proposé et d'en tirer l'importante conclusion suivante:

- La meilleure méthode de réduction d'espace est **PCA+EFM** pour toutes nos applications.

- La fusion par intégrale floue de **Sugeno** est très efficace dans l'identification des visages neutres 2D (*monomodale* et *multibiométries*(visage,voix))
- La fusion des caractéristiques et des scores de l'information profondeur et image couleur RVB est insuffisante à l'identification des visages 3D.
- L'information de couleur est plus robuste que l'information de profondeur sous des variations d'expression,
- l'information de profondeur est plus robuste que l'information d'intensité de couleur sous des variations d'éclairage.
- La composante V du système HSV est plus performante que le système RVB dans l'identification de visages 3D.
- La performance de classification des scores par les **SVMs** surpasse la fusion au niveau des caractéristiques et la combinaison des scores par les deux intégrales floue de **Sugeno** et **Choquet**.

Nous proposons d'utiliser les filtres de Gabor sur l'information de profondeur et de couleur et étudier d'autres stratégies de fusion.

Conclusion et perspectives

Dans cette thèse, nous avons décrit le problème de la reconnaissance de visages 2D, 3D et 3D expressions en présence de variation d'illumination et poses. Les principales méthodes de la littérature ont été étudiées, et nous nous sommes plus particulièrement concentrés sur le choix du meilleur filtre et de la meilleure méthode de réduction dans notre première approche monomodale. L'extraction de la région du visage à partir de l'image filtrée est une étape fondamentale dans le processus de reconnaissance faciale. L'algorithme d'extraction doit être efficace et doit fournir, sans aucune ambiguïté, la région du visage dans l'image. Notre intérêt est porté sur les **phases** du **filtre de Gabor** sous l'étude d'un **seuil** et le choix du filtre optimal est réalisé ensuite nous procédons à la fusion des meilleurs filtres dans différentes orientations et fréquences afin de récupérer le maximum d'information pertinentes de nos images.

Nous proposons une méthode originale basée sur la réduction d'espace en étudiant plusieurs méthodes linéaires basées sur l'**ACP** (Analyse en Composante Principale), **PCA2D**, **DPCA** (Direct PCA), **ICA** (Independent Component Analysis) et l'**EFM**. La meilleure au sens de performance du système est retenue. Dans notre travail l'**EFM** (Enhanced Fisher Method) est la meilleure candidate. Pour la classification nous nous sommes intéressés à la mesure de similarité pour sa simplicité et une classification par **SVM** pour son efficacité.

Une méthode fondée sur la fusion des scores d'un multi systèmes est étudiée ainsi que la fusion multi biométries et multi algorithmes. Nous avons validé nos résultats sur une **BDD 2D** la **XM2VTS** dans un premier temps puis sur la **XM2VTS** des **scores**. Nous avons mis en revue les deux approches de **fusion** : l'approche fusion par classification et l'approche par combinaison et nous avons introduit la *normalisation dans l'approche de fusion par classification*. Pour contourner les problèmes posés par l'illumination, nous nous sommes intéressés au visage **3D**.

Une autre contribution de cette thèse a consisté à développer une technique **2D** de reconnaissance du visage basée sur une méthode de réduction non linéaire la **KPCA** (Kernel PCA) qui est une extension de la PCA avec l'introduction du noyau (Kernel).

L'architecture du SRV peut se décomposer en trois parties :

- La **première partie**, composée de la partie prétraitement qui est primordiale et particulièrement dans le cas **3D**. Cette phase est basée sur l'extraction de la couleur de l'image 3D et aussi la génération des cartes de profondeur. Elle est suivie d'une phase de réduction d'espace. Qui nous permet d'extraire l'information de l'image d'entrée, de

Conclusion et perspectives

fusionner ces caractéristiques afin d'obtenir une représentation de plus haut niveau. Cette première partie peut être vue comme un moyen de projeter linéairement ou non linéairement les images d'entrée de la base de données **CASIA 3D** (**XM2VTS** pour le cas 2D) sur un espace de plus faible dimension.

- La **deuxième partie** essaie de représenter un visage de référence à partir de la représentation compacte issue de la projection. Lors de l'apprentissage, on construit une image référence de visage par personne, rendant ainsi la projection quasi invariante aux transformations présentées en entrée. L'intérêt des méthodes de réduction c'est la conservation de l'information avec un minimum de paramètres caractéristiques et la garantit d'une bonne discrimination. D'après les expériences menées l'EFM associée à l'ACP est bonne candidate et pour remédier au problème de non linéarité de l'ACP. Nous avons étendue nos recherches sur la KPCA qui donne des résultats satisfaisants. Les scénarios d'identification considérés dans cette thèse n'utilisent qu'un nombre limité d'images d'apprentissage ainsi que peu d'images/personne pour l'enrôlement. Malgré cette limitation, les taux de reconnaissance atteints sont comparables à ceux de l'état de l'art sur la base de données XM2VTS, CASIA 3D et FRGC. En revanche, les taux de reconnaissance sont moyens pour les visages présentant des expressions faciales. Nous pensons que le nombre d'images utilisées pour l'enrôlement influe sur les taux de reconnaissance. Une des limitations majeures des approches fondées sur un apprentissage en général est qu'elles nécessitent un grand nombre d'échantillons d'apprentissage afin d'obtenir une extraction de caractéristiques pertinentes ainsi qu'une bonne généralisation. Une manière de surmonter cette limitation consiste à faire un bon choix des méthodes de réduction d'espace. Pour cela une étude comparative de ces méthodes sur différentes BDD est menée et nous l'avons affirmé plus haut que l'EFM et la KPCA sont les plus performantes. Des tests ont de plus montrés sa bonne robustesse à des dégradations des images tests. Le développement d'une variante de cette méthode a permis d'accroître encore les taux de reconnaissance quand on lui ajoute des modalités c'est le cas du système bimodale étudié. Cette méthode souffre malgré tout de quelques défauts, remédiables cependant.
- Dans un **troisième temps**, la fusion de modalités a été considérée. La fusion de modalités offre une alternative aux systèmes biométriques unimodaux. Partant de l'hypothèse que les modalités offrent des informations complémentaires (ce qui est souvent le cas), la fusion de celles-ci permet globalement d'améliorer la fiabilité d'un système. La fusion de modalités permet en outre de s'affranchir de certaines problématiques inhérentes aux

Conclusion et perspectives

systèmes unimodaux (comme les variations de luminosité pour la modalité visible par exemple). Dans le cadre d'une fusion multi capteurs, elle nécessite cependant un investissement supplémentaire en capteurs, ce qui peut être coûteux.

Différents niveaux de fusion ont été étudiés. La fusion des images (niveau caractéristiques) ainsi que la fusion au niveau des scores ont été considérées via nos approches. La fusion la plus étudiée est la fusion au niveau des scores. Une méthode permettant de pondérer les scores issus des classifieurs est ainsi développée. Celle-ci permet de prendre en compte la pertinence des scores à l'aide d'une fonction de pondération dans le cas de la fusion par combinaison. La somme pondérée s'avère la plus performante dans notre application et particulièrement dans le cas de notre multi algorithmes utilisant la composante couleur **S** et les trois systèmes issus d'orientations différentes retenues du filtre de **Gabor**. Les approches ont également été utilisées comme modules de mise en correspondance avant une fusion des scores des modalités. Les méthodes utilisées comme l'**EFM**, **KPCA** et **SVM** ou mesure de métriques (pour la classification) ont permis l'utilisation de règles simples de fusion (somme des scores, min, max ou moyenne) pour obtenir des taux de reconnaissance élevés. Notons que l'intégrale de **Suegeno** donne aussi de bons résultats. Pour la classification la norme Euclidienne et la corrélation normalisée sont les meilleures pour la mesure de similarité dans notre cas. Le classifieur **SVM** est plus efficace et donne les meilleurs résultats dans toutes nos applications **2D**, **3D** ou **3D expression**.

Nous proposons aussi de développer une technique bimodale :

Dans **un premier temps** nous étudions toutes les caractéristiques issues d'une face.

- i) **2D-3D** de reconnaissance faciale qui combine des mesures anthropologiques 3D du visage avec une technique **2D** basée sur l'Analyse en Composantes Principales ou **ACP**. L'intérêt d'une telle approche est d'exploiter les avantages des deux techniques (**2D** et **3D**) afin d'améliorer le taux de reconnaissance. D'autre part, nous pensons que l'anthropométrie par sa maîtrise de l'anatomie du visage humain peut apporter des informations complémentaires pour améliorer la reconnaissance faciale. En effet, les anthropologues ont été les premiers à étudier de façon scientifique la morphologie humaine en général et celle de la tête et du visage en particulier et cela en utilisant des critères descriptifs et des caractères métriques. Nous proposons de définir des mesures anthropométriques faciales à partir de points caractéristiques **3D** du visage. Ces mesures correspondent à des valeurs indiciaires et angulaires caractérisant les différentes régions du visage.

Conclusion et perspectives

- ii) **2D-3D Expression** qui combine la reconnaissance des expressions faciales et le visage **2D**. L'intérêt dans ce cas est d'extraire les informations fournies par les expressions afin de mieux identifier les personnes. Une technique **3D** expression qui prend en entrée, non pas l'image entière du visage, mais les « imageries » correspondant aux régions caractéristiques du visage (les yeux, le nez et la bouche) et à démontrer que cette méthode donne des taux de reconnaissance aussi bon que l'image complète.
- iii) Enfin, nous proposons de fusionner les résultats de reconnaissance obtenus par les différentes approches (**2D** et **3D**) ainsi que différentes modalités avec le visage afin d'améliorer le taux de reconnaissance. Pour cela, nous avons exploré plusieurs critères de fusion. Un protocole expérimental a été mis en œuvre pour tester cette méthode et analyser leurs performances.

Nous détaillons les principales expériences que nous avons menées durant la thèse ainsi que les résultats monomodaux (sur une seule modalité à la fois) obtenus via les trois principales approches testées :

- l'approche 2D pour laquelle des investigations sur la meilleure orientation du filtre de Gabor et une comparaison des méthodes de réduction sont réalisées. Dans cette approche nous étudions les méthodes de réduction linéaires et non linéaires,
- l'approche 3D dans laquelle la génération de cartes de profondeur est réalisée,
- l'approche 3D expression basée sur la combinaison par concaténation des régions d'intérêts et du visage,
- et enfin l'approche 3D expression basée sur la fusion des caractéristiques du visage neutre et des six expressions (dégout, joie, peur, surprise, colère).

Dans **un deuxième temps** notre étude est orientée vers l'identification des personnes en se basant sur la fusion multi-systèmes, multi-biométries, multi-algorithmes et multimodales et pour tester notre système et évaluer ses performances.

- i) **Fusion multi-systèmes** : combinant la couleur et plusieurs systèmes utilisant la phase de Gabor. La fusion MCS est basée sur quatre systèmes issus des meilleurs filtres de Gabor retenus par nos diverses expériences. Les deux approches par combinaison et par classification de la fusion sont étudiées en association à différentes méthodes de normalisation sur la BDD XM2VTS.
- ii) **Fusion multi-biométries** : combinant deux biométries dépendantes (voix, visage). La fusion multi biométrique est assurée par le filtre de Gabor optimal retenu à

Conclusion et perspectives

travers nos expériences. Une réduction d'espace est choisie selon les meilleures performances de notre système de reconnaissance. La phase de classification est réalisée par calcul de distances ou par SVM. Notre SRV est appliqué à la base des scores XM2VTS.

- iii) **Fusion multi-algorithmes** : C'est une fusion des caractéristiques codées d'une part par un codage entropique et d'autre part par le calcul de la moyenne. Ensuite nous réalisons la réduction des caractéristiques par l'EFM et la KPCA. Les scores obtenus par mesure de similarité par des métriques sont combinés et une décision est prise en fonction d'un seuil prédéfini.
- iv) **Fusion multimodales** : combinant la couleur et la profondeur. Nous avons essayé de contourner le problème de la première expérimentation, à savoir la limite des images 2D face aux changements d'illumination et/ou expressions faciales. Les tests sur la base CASIA 3D sont réalisés :
 - a) Première expérimentation, dans un premier temps, nous avons mis en œuvre un système de reconnaissance multimodale sur le visage 3D couleur. Il s'agit d'une fusion entre les caractéristiques ou scores de l'image couleur 2D (projection de l'image couleur RVB 3D sur le plan 2D) et une image profondeur nommée 2.5D (cartes de profondeur) issue de l'image couleur RVB 3D. Dans ce premier cas nous utilisons un classifieur basé sur la mesure de similarité pour sa simplicité.
 - b) Dans la deuxième expérimentation nous nous sommes intéressés à la composante V du système HSV et nous faisons la fusion des caractéristiques ou scores des images 2D (composante V) et 3D (profondeur). Dans ce deuxième cas le classifieur SVM est choisi pour sa rapidité et son efficacité. Nous tentons de faire l'identification du visage sous diverses variantes (variation de lumière(IV), visage avec expressions faciales (EV) et visage avec expression faciale et variation de lumière(EVI)).

D'après les résultats des expériences réalisées nous pouvons dire que l'ACP+EFM reste une méthode discrimination performante d'une façon générale. La **fusion des scores** reste une méthode très efficace car sans la fusion nous avons avec un $N_p(\text{EFM}) = 100$; un **TR** = **92.25%** pour les visages pour une fusion par combinaison **min** et la distance L_2 .

D'après l'étude faite et toutes les expériences réalisées nous appliquons cette dernière à des images de visages **2D** neutres de la base de données **XM2VTS** en multi algorithmes et en multi biométries (voix et visage). Cette méthode reste performante avec un **TR=99.99%** en **2D** et multi biométrie. Nous confirmons que cette méthode reste une bonne candidate pour aborder les visages à **3D**. Des améliorations reste à réalisées surtout au niveau de la phase

Conclusion et perspectives

prétraitement et aussi et surtout la phase de classification. Car notre utilisation à la mesure métrique reste un choix guidé par sa simplicité. Ceci dit d'autres normes comme **Manhattan** ou mesures **géodésiques** restent envisageable, et toujours dans l'objectif de rehausser ces paramètres de performances du système biométriques. Les résultats sont dans l'ensemble satisfaisants, mais le taux de faux rejet reste moyen. Ceci est sûrement dû à l'étape de prétraitement qui reste à améliorer ainsi que le choix d'un meilleur classifieur.

Dans l'expérience de fusion **2D** et **3D** basée sur **SVM** nous utilisons la **profondeur** et la **couleur** des images pour construire un classificateur robuste pour l'identification du visage. Puisque la dimensionnalité des caractéristiques d'images de profondeur et de couleur est très élevée, nous avons étudié trois méthodes de réduction d'espace de données. Puis nous avons fait des études comparatives sur la fusion de l'information de profondeur et de couleur pour les deux niveaux : caractéristiques et scores pour la construction d'un classifieur efficace. En analysant nos résultats expérimentaux dans la base de données du visage **CASIA 3D** avec des variations complexes, nous illustrons la performance prometteuse du schéma proposé et d'en tirer l'importante conclusion suivante:

- La meilleure méthode de réduction d'espace est **PCA+EFM** pour toutes nos applications.
- La fusion par intégrale floue de **Sugeno** est très efficace dans l'identification des visages neutres **2D** (monomodale et multibiométries (visage,voix)).
- La fusion des caractéristiques et des scores de l'information profondeur et image couleur **RVB** est insuffisante à l'identification des visages **3D**.
- L'information de couleur est plus robuste que l'information de profondeur sous des variations d'expression,
- l'information de profondeur est plus robuste que l'information d'intensité de couleur sous des variations d'éclairage.
- La composante **V** du système **HSV** est plus performante que le système **RVB** dans l'identification de visages **3D**.
- La performance de classification des scores par les **SVMs** surpasse la fusion au niveau des caractéristiques et la combinaison des scores par les deux intégrales floue de **Sugeno** et **Choquet**.

D'après l'étude faite et toutes les expériences réalisées nous pouvons dire que la fusion des scores **2D** (couleur **RVB**) et **3D** (**profondeur**) avec les métriques améliore certes les performances du système **SRV** monomodale.

Conclusion et perspectives

Dans la **fusion** couleur **HSV** et **profondeur** l'approche par classification **SVM** est meilleure que celle par combinaison ce qui n'était pas le cas dans les fusions multi systèmes et multi biométries (fusion par combinaison : intégrale floue de **Sugeno** et fusion par classification : **SVM**).

Dans toutes les applications étudiées la fusion des scores donne les meilleurs résultats de performance du SRV.

Le système **HSV** est plus informant que le système RVB avec un TR=79.50% (pour la fusion mean, distance Euclidienne norme 1 et $N_p=60$) contre TR=100% (pour la fusion SVM, distance corrélation normalisée et $N_p<70$). Ainsi la **fusion des scores** couleur **HSV** et **profondeur** par **classification SVM** est parfaite avec un taux de réussite à 100% dans des visages exposés à la variation d'illumination. Mais elle reste sensible aux changements d'expressions faciales et à la combinaison changement d'expression faciales et lumière.

Nous affirmons à ce stade du travail que :

- l'algorithme **SVM** est très efficace quelque soit l'application envisagée.
- l'algorithme **SVM** est aussi bon en phase détection, qu'en phase classification et en fusion.
- La meilleure méthode de réduction reste l'**EFM**.
- Le filtre de **Gabor** s'il est bien utilisé nous permet d'extraire l'essentiel des caractéristiques globales du visage.
- Les **phases du filtre** sont mieux caractéristiques du visage que les amplitudes.
- Le système **HSV** est plus informant que le RVB.
- La fusion des scores est plus efficace que celles des caractéristiques.
- La norme **L₁** et la **corrélation normalisée** sont les meilleures métriques dans la classification par mesure de similarité.
- La classification par **SVM** est meilleure que celle par mesure de distances. Elle est plus souple, rapide et nous évite le calcul de seuil.

Perspectives

Pour la multimodalité :

- La construction d'un modèle mathématique tel les modèles génétiques ou les PSO permettant d'optimiser notre approche.
- Utiliser les curvelets pour extraire les paramètres caractéristiques.
- Etudier d'autres extensions de l'ACP.

Conclusion et perspectives

- Approfondir les recherches dans les méthodes à noyaux.
- Utiliser la fusion par classification SVM et MLP dans les applications multi algorithmes et multi modales (couleur RVB et profondeur).
- Utiliser des techniques d'optimisation comme les algorithmes génétiques

Pour le visage 3D :

- Penser à utiliser l'information profil.
- Bien détecter les points et régions d'intérêts dans le visage.
- Etudier les techniques de correction de rotation de la tête pour récupérer le maximum d'information.
- S'intéresser par des techniques appropriées à la rotation de la tête, afin de récolter le maximum de caractéristiques utiles pour l'identification.
- S'intéresser à la symétrie du visage pour améliorer le temps de calcul.
- Appliquer notre système à des BDD 3D de jumeaux pour tester son efficacité.
- Mieux étudier les techniques de localisation du nez.

Pour la Classification :

Le problème de la robustesse d'un système biométrique à diverses altérations (luminosité, rotations, mises à l'échelle . . .) peut être attaqué en trois points de la chaîne. La plupart des systèmes biométriques n'en considèrent cependant que deux : améliorer les images avant d'effectuer une extraction de caractéristiques, et/ou extraire des caractéristiques invariantes aux transformations.

- Le premier niveau peut consister par exemple en un recadrage géométrique pour éliminer les transformations géométriques subies par l'image, ou en une modification de la dynamique des niveaux de gris pour pallier le problème de la luminosité.
- Le deuxième niveau généralement considéré consiste à produire des algorithmes capables d'extraire des caractéristiques invariantes aux dites transformations. La multiplication des transformations à considérer peut cependant rendre les algorithmes très complexes, et donc instables à de nouvelles transformations.
- Un troisième niveau pourrait consister à créer de nouvelles images à partir des images d'enrôlement en les modifiant via les transformations auxquelles on souhaite être invariant. Par exemple, à partir d'une image d'enrôlement, créer des images artificielles par rotation, puis les intégrer à la galerie. Une telle approche, combinée à l'algorithme de classification, permettrait à une image test ayant subi une rotation d'être mise en correspondance facilement avec son équivalent artificiel, alors qu'elle

Conclusion et perspectives

ne le serait pas forcément avec l'image originale présente dans la galerie. Une telle approche pour l'invariance par transformations affines serait non seulement simple à réaliser (il suffit de créer autant d'images artificielles que de transformations auxquelles le système doit être robuste), et devrait pouvoir fonctionner avec des algorithmes d'extraction de caractéristiques simples. Un problème se poserait néanmoins avec cette approche : la multiplication des images de la galerie, et donc le temps de calcul nécessaire à l'extraction d'un vecteur caractéristique sur cette galerie.

- Utiliser une classification par PSO

Résumé

Bien qu'ayant suscité des recherches depuis trente ans, le problème de la reconnaissance de visages, sachant que peu d'images par personne sont disponibles pour l'enrôlement, n'est pas encore résolu. Dans ce contexte, les défis les plus difficiles à relever consistent à développer des algorithmes robustes aux variations d'illumination, variations de pose et expressions faciales. De plus, il y'a aussi une contrainte forte sur la complexité en temps et en occupation mémoire des algorithmes à mettre en œuvre dans de tels systèmes. Le travail développé dans cette thèse apporte plusieurs avancées innovantes dans ce contexte. Tout d'abord, une méthode d'extraction des paramètres par les meilleurs filtres de **Gabor** visant à simuler les performances de la rétine est proposée. Ensuite, nous étudions un modèle biométrique basé sur les approches classiques de réduction d'espace à savoir : les méthodes linéaires **PCA, LDA, DLDA, PCA2DOC, PCA2DOL, EFM** et les méthodes non linéaires à **noyaux**, nous nous sommes limités à la **KPCA** la plus répandue en reconnaissance de visages. La méthode de réduction d'espace **EFM** est retenue. La classification est assurée par **mesure de similarité** pour sa simplicité et par **SVM** pour son efficacité. Pour valider ce travail nous avons testé cette approche tout d'abord à la mono modalité sur des images frontales de la base de données **XM2VTS** selon son protocole associé (Protocole de Lausanne). Afin d'améliorer la classification de notre approche la **KPCA** est étudiée. Dans le souci d'amélioration des performances de notre système de reconnaissance nous nous sommes intéressés aux multi systèmes et à la multi biométrie ((visage, voix) réalisés sur la base des scores **MXM2VTS** et (visage, empreintes) sur la **BDD FRVC**. Seulement les images de visages **2D** ont des limites quant à la variation d'illumination. Pour cette raison nous avons étendu nos recherches vers les visages **3D** et aussi à la **fusion** par **ondelettes** associée aux opérateurs de combinaison simples (minmax, moyenne et max) d'information issues du **2D (couleur)** et **3D (profondeur)**. Deux types de fusion sont étudiés : fusion des caractéristiques et fusion des scores. Les résultats obtenus sont encourageants mais pas suffisant car le système de reconnaissance **3D** présente des faiblesses pour les **expressions faciales**. Pour contourner ce problème nous avons étendu nos recherches à l'identification en présence d'expressions faciales sous différentes illumination et différentes poses. Pour cela nous effectuons dans un premier temps la fusion par concaténation des régions d'intérêts (yeux, bouche, nez et visage). Puis nous réalisons une deuxième expérience la fusion des six expressions et du visage neutre. Tous nos travaux **3D** et **3D** expressions sont réalisés sur la **BDD CASIA 3D**. Des résultats satisfaisants et prometteurs sont obtenus et peuvent être davantage améliorés par de meilleurs détecteurs et prétraitements.

Mots clés :

Identification et Authentification, Extraction de caractéristiques, Réduction d'Espace, Scores, Classification, Fusion.

ملخص

هدفنا هو استكشاف تقنيات للتعرف على الوجوه 3D,2D و في جود تعبير في البداية. ثم تنفيذ نظام التعرف من خلال اختيار واحد وأفضل في إطار التحقق من الهوية. ولهذا الصدد يتم استخدام المناهج الحديثة في المجال. تسترشد ما لنا في اختيارنا من التقنيات المستخدمة في تصميم نظامنا من التعرف على تعبير الوجه **3D**. ويستند نهجنا على المعالجة التي تضمن التحضير لتقديم صورة للتحليل. ويستند هذا الأخير على خطوة من استخراج المعلومات في المناطق المهمة من الوجه التي تم دمجها. من المتوقع بعد ذلك على مساحة صغيرة أنشأتها نموذج التمايز الخطي. وفي الأخير حاولنا بالتسلسل المنطقي للوصول إلى **SVM** في التحسين. ثم تقدم هذه المواصفات إلى المصنف المعروف بفعاليتها و هو **EFM** أفضل المرشحات لتحسين معدل النجاح. النتائج التي تم الحصول عليها مشجعة وواحدة. العمل الثاني في هذه المذكرة يندرج ضمن سياق الاستطلاع الآلي للوجوه من صور ذات ثلاث **3D** أبعاد في نظام الاستطلاع الآلي على الوجه نعد ثلاث مراحل أساسية: • المرحلة الأولى المعالجة الأولية المكرسة لتوليد خرائط العمق.

• المرحلة الثانية: تسمى مرحلة التدريب وهي جمع المعلومات الحيوية على الأفراد المراد التعرف عليها. لهذا نحن نطبق خوارزمية تحليل العنصر الرئيسي وتحليل التمايز الخطي لإخراج فضاء الإسقاط وبيانات التمييز على النحو الأمثل. ثم، فإن مرحلة التعرف تستند بدورها على ثلاث مراحل ومن خلالها إسقاط صورة الاختبار في الفضاء ومقارنتها مع صور قاعدة التدريب. وأخيرا، نقوم بإجراء تصنيف على أساس التشابه.

EFM- SVM - تعرف على الوجوه 3D,2D - البيومتري - خرائط العمق - الهوية - صور ذات ثلاث 3D أبعاد.