

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.

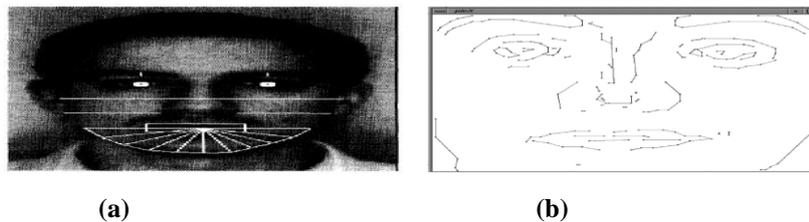


Figure 2.1 (a) Localisation des caractéristiques géométriques utilisées dans [47].
(b) Cartes de contours utilisées dans [68].

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.

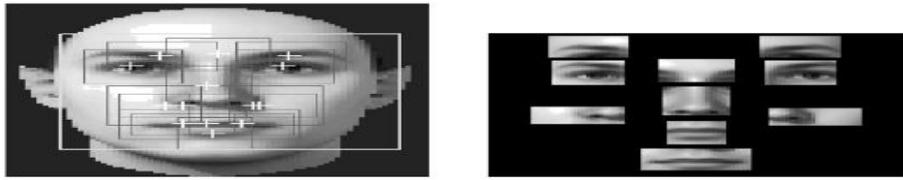


Figure 2.2 Approche proposée dans [70].

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.

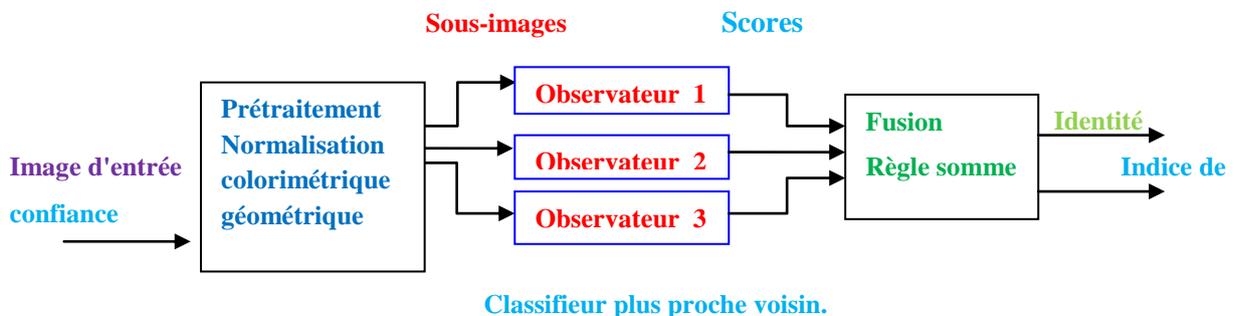


Figure 2.3 Approche de Price et Gee [77].

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**.

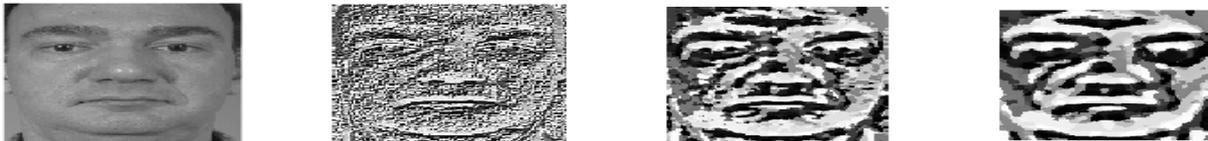


Figure 2.5 Caractéristiques MB-LBP pour un visage, (pour un **masque** de taille $(3, 3)$; $(9,9)$ et $(15,15)$).

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].

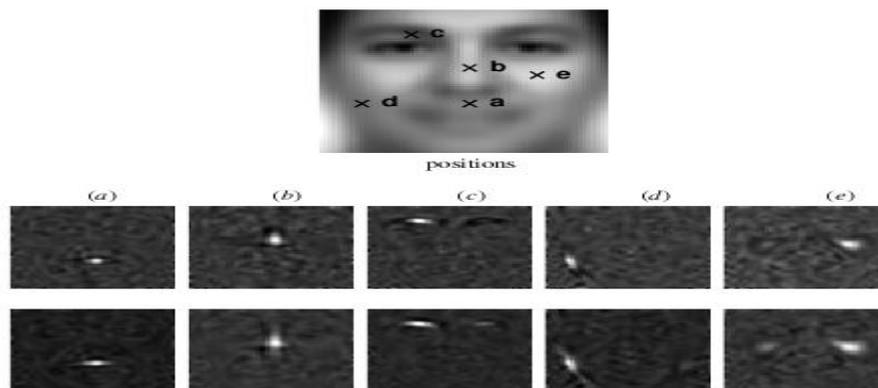


Figure 2.7 Approche Local Component Analysis. [118]

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.

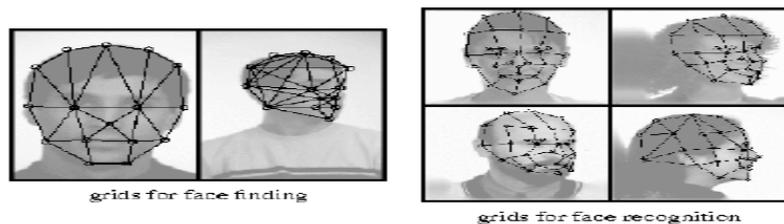


Figure 2.8 Graphe appliqué aux visages pour l'approche EBM. [119]

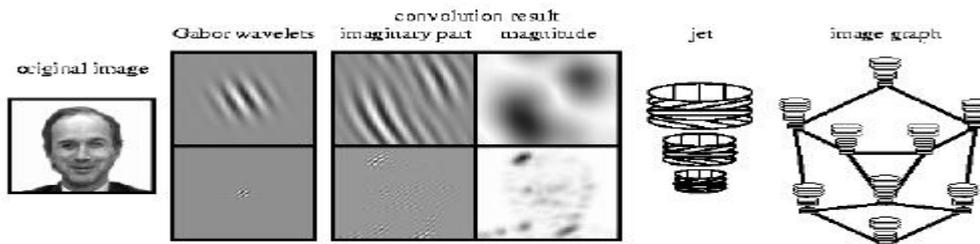


Figure 2.9 Création des jets pour l'approche EBM. [120]

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 appariant 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).

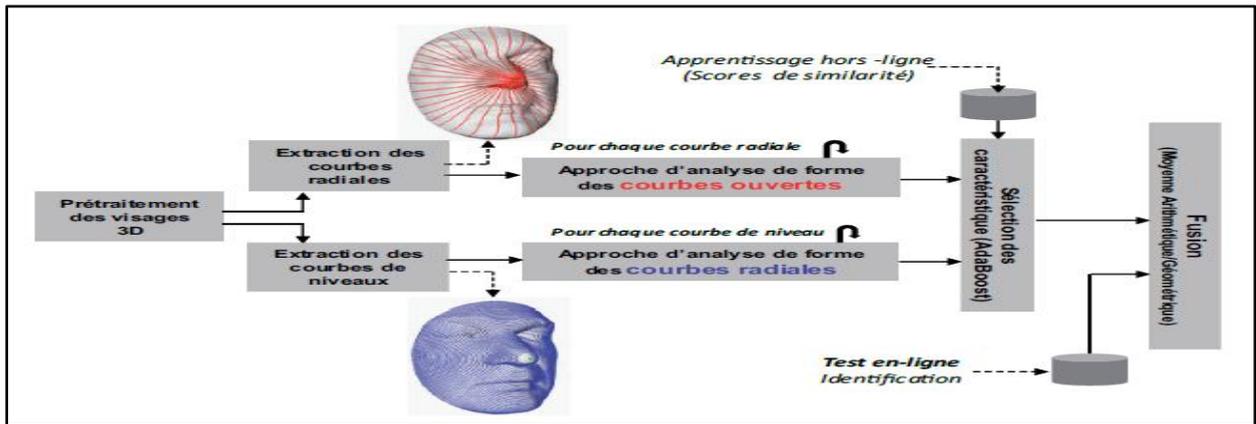


Figure 2.11 Aperçu de l'approche de reconnaissance du visage 3D. [144]

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].

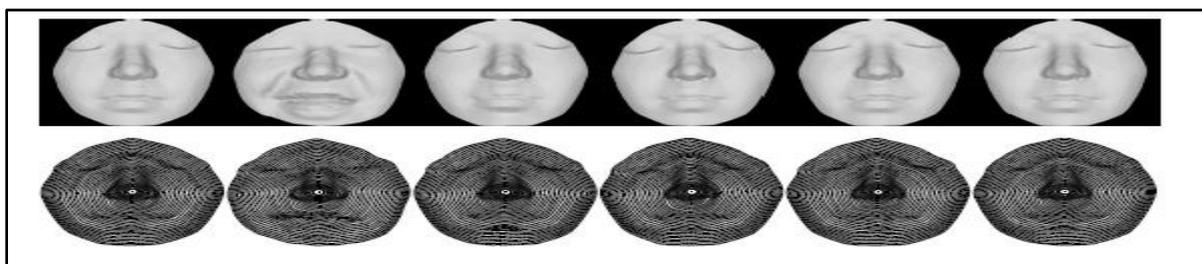


Figure 2.12 Les surfaces faciales de la même personne sous différentes expressions faciales, et leurs représentations dans l'espace H. [149]

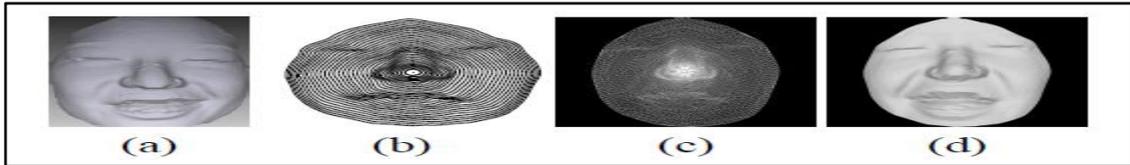


Figure 2.13 Exemple de reconstruction : (a) la surface faciale originale, (b) Sa représentation comme une collection de courbes dans l'espace H, (c) la triangulation, et (d) la surface reconstruite. [149]

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**.

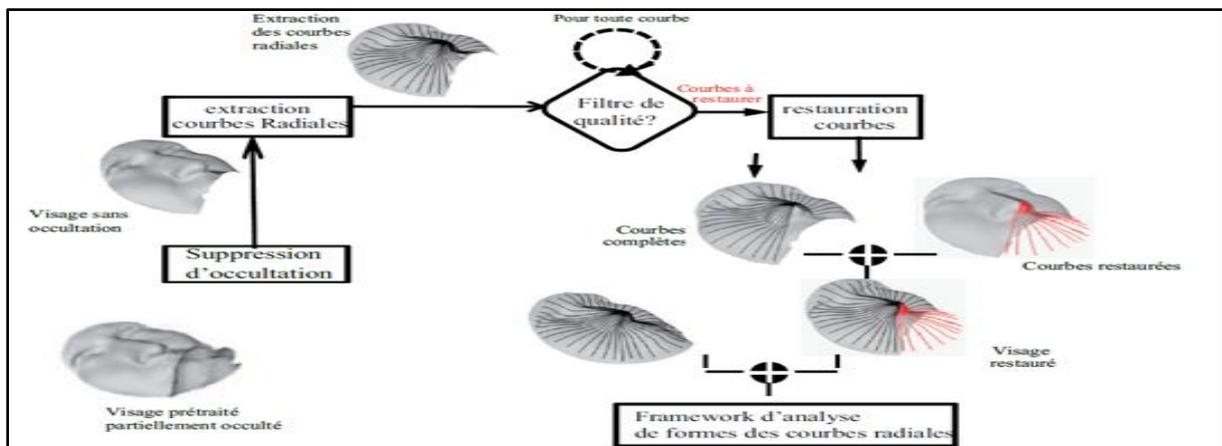


Figure 2.14 Différents étapes de l'approche : de la suppression d'occultation à la reconnaissance passant par la restauration. [153]

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.

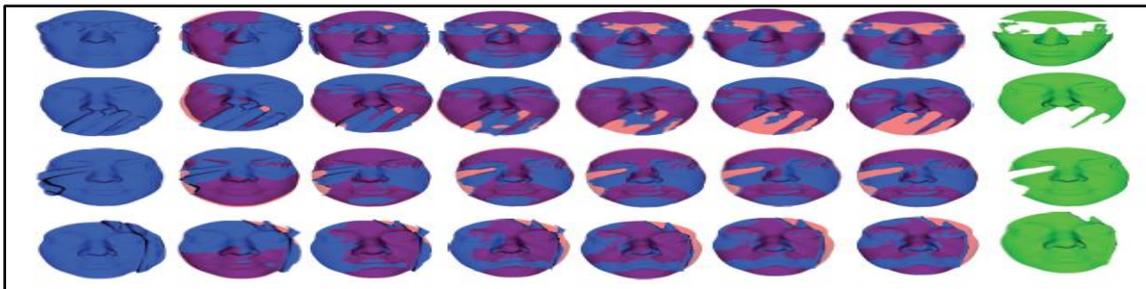


Figure 2.15 Suppression graduelle des parties occultant le scan facial 3D. [153]

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**.

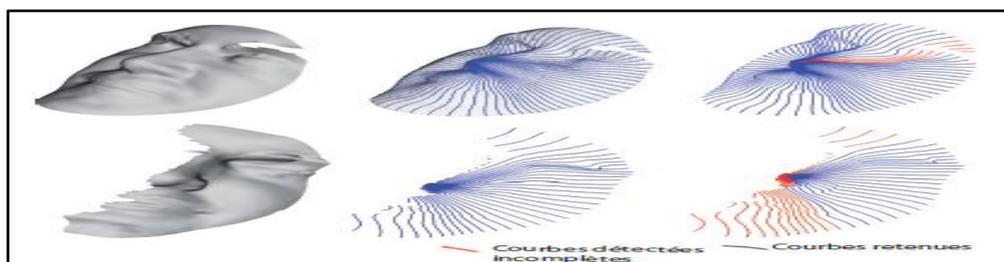


Figure 2.16 Module de qualité [153]

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**.

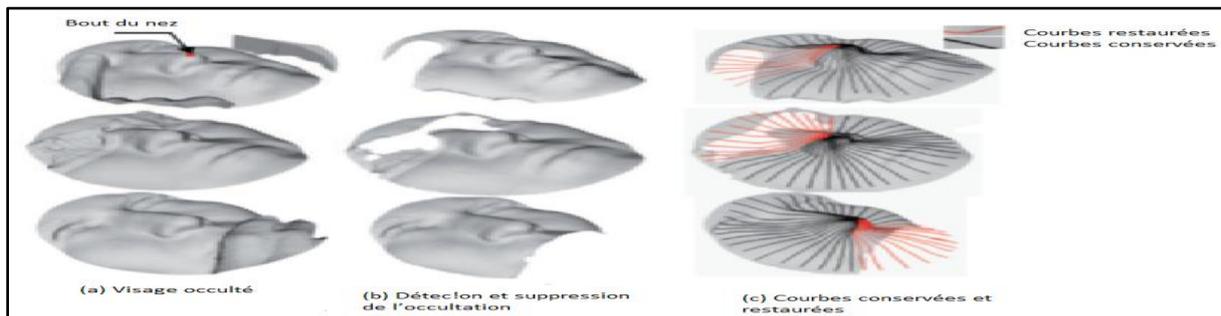


Figure 2.17 (a) Visages avec occultations externes, (b) visages après détection et suppression des parties occultantes, (c) estimation des parties occultées par modèle statistique sur l'espace de forme des courbes. [153]

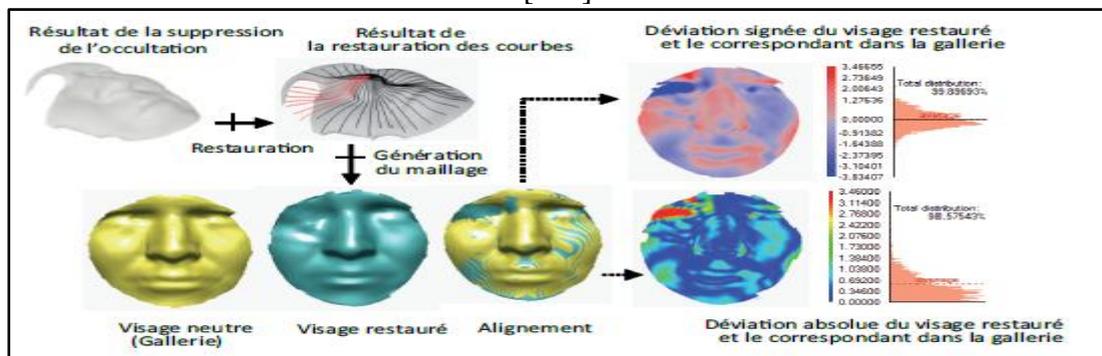


Figure 2.18 Déviation spatiale (absolue et signée) entre un visage restauré et le visage correspondant dans la galerie [153].

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].

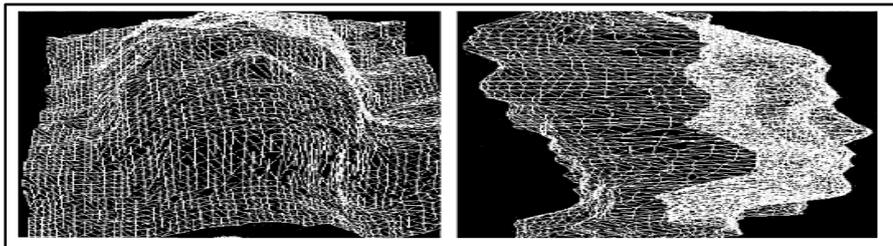


Figure 2.19 Modèle 3D de visage humain constitué de triangles. [161]

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.

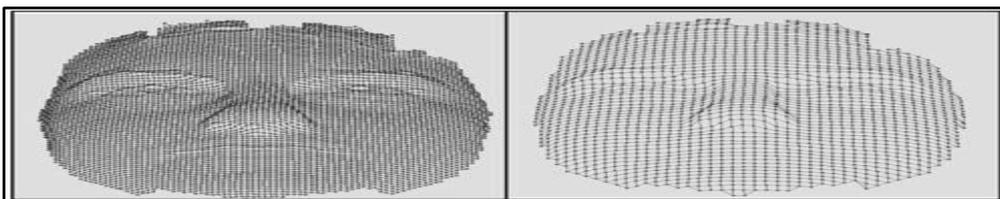


Figure 2.20 (À gauche) en maille originale produite par le scanner **3D** et (à droite) de la même maille à 1-4 la résolution originale [162].

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 de 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].

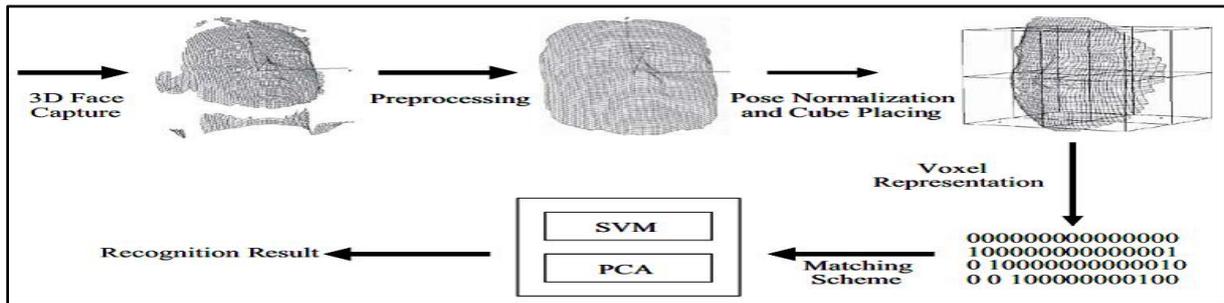


Figure 2.21 Architecture du SRV 3D avec deux génériques schémas de correspondance.

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.

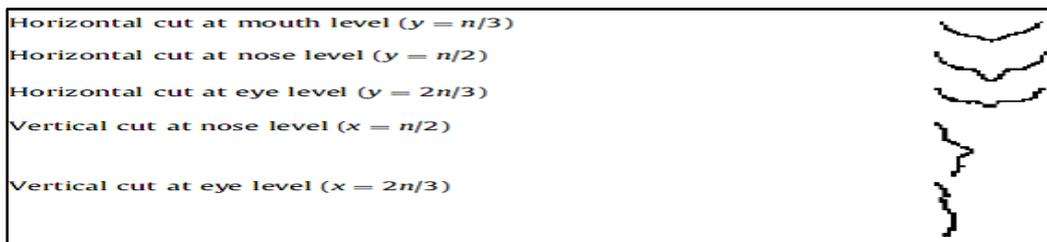


Figure 2.23 Exemples de profils définis simples.

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].

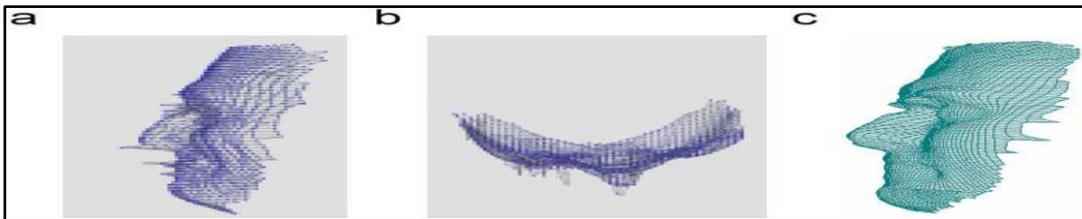


Figure 2.24 (a) et (b) maillage bruité à partir de deux points de vue différents
(c) maillage après le bruit application de filtrage et de lissage [162].

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.

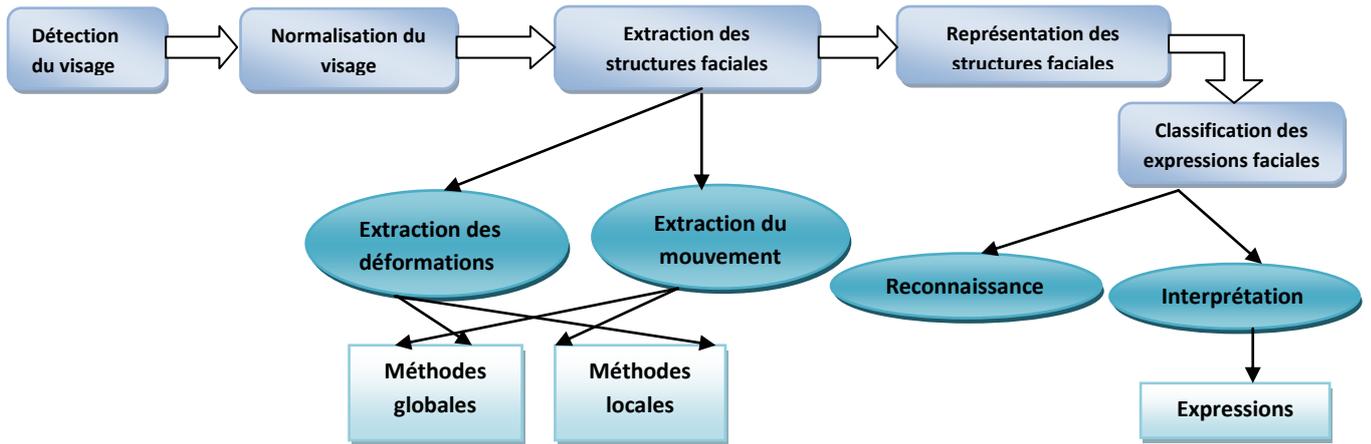


Figure 2.26 Méthodes d'analyse automatique d'un SR d'expressions faciales.

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.

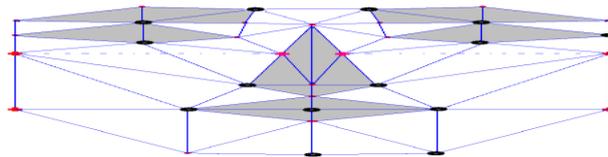


Figure 2.28 Masque pour le visage. [165]

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].

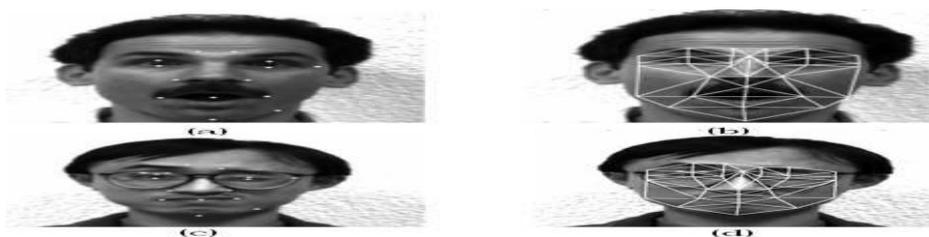


Figure 2.29 Masques de montage. (a), (c): manuellement marqueurs placés (points blancs), (b), (d): équipés masques. [166]

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 information 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**.

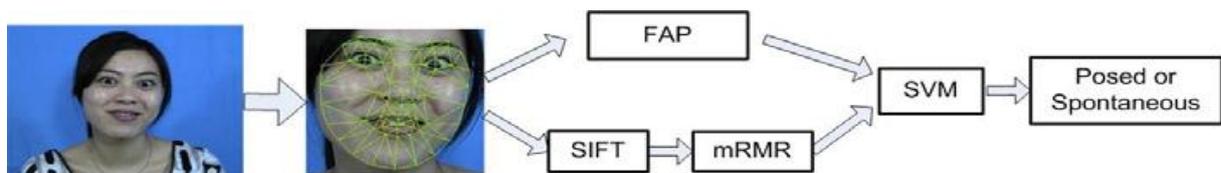


Figure 2.30 Cadre du système d'évaluation. [199]

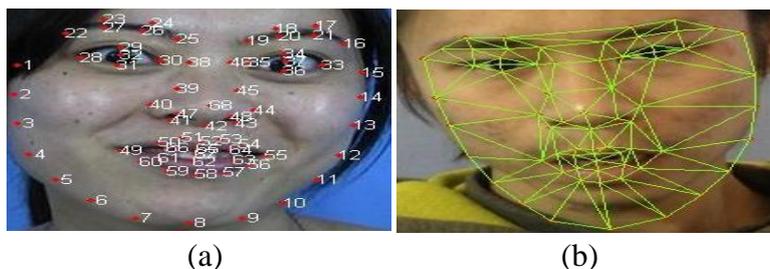


Figure 2.31 (a) 68 points de référence pour la formation, (b) les résultats de détection ASM. [199]

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**.