

Conclusion et perspectives

Dans cette thèse, nous avons décrit le problème de la reconnaissance de visages 2D, 3D et 3D expressions en présence de variation d'illumination et poses. Les principales méthodes de la littérature ont été étudiées, et nous nous sommes plus particulièrement concentrés sur le choix du meilleur filtre et de la meilleure méthode de réduction dans notre première approche monomodale. L'extraction de la région du visage à partir de l'image filtrée est une étape fondamentale dans le processus de reconnaissance faciale. L'algorithme d'extraction doit être efficace et doit fournir, sans aucune ambiguïté, la région du visage dans l'image. Notre intérêt est porté sur les **phases** du **filtre de Gabor** sous l'étude d'un **seuil** et le choix du filtre optimal est réalisé ensuite nous procédons à la fusion des meilleurs filtres dans différentes orientations et fréquences afin de récupérer le maximum d'information pertinentes de nos images.

Nous proposons une méthode originale basée sur la réduction d'espace en étudiant plusieurs méthodes linéaires basées sur l'**ACP** (Analyse en Composante Principale), **PCA2D**, **DPCA** (Direct **PCA**), **ICA** (Independent Component Analysis) et l'**EFM**. La meilleure au sens de performance du système est retenue. Dans notre travail l'**EFM** (Enhanced Fisher Method) est la meilleure candidate. Pour la classification nous nous sommes intéressés à la mesure de similarité pour sa simplicité et une classification par **SVM** pour son efficacité.

Une méthode fondée sur la fusion des scores d'un multi systèmes est étudiée ainsi que la fusion multi biométries et multi algorithmes. Nous avons validé nos résultats sur une **BDD 2D** la **XM2VTS** dans un premier temps puis sur la **XM2VTS** des **scores**. Nous avons mis en revue les deux approches de **fusion** : l'approche fusion par classification et l'approche par combinaison et nous avons introduit la *normalisation dans l'approche de fusion par classification*. Pour contourner les problèmes posés par l'illumination, nous nous sommes intéressés au visage **3D**.

Une autre contribution de cette thèse a consisté à développer une technique **2D** de reconnaissance du visage basée sur une méthode de réduction non linéaire la **KPCA** (Kernel **PCA**) qui est une extension de la PCA avec l'introduction du noyau (Kernel).

L'architecture du SRV peut se décomposer en trois parties :

- La **première partie**, composée de la partie prétraitement qui est primordiale et particulièrement dans le cas **3D**. Cette phase est basée sur l'extraction de la couleur de l'image 3D et aussi la génération des cartes de profondeur. Elle est suivie d'une phase de réduction d'espace. Qui nous permet d'extraire l'information de l'image d'entrée, de

Conclusion et perspectives

fusionner ces caractéristiques afin d'obtenir une représentation de plus haut niveau. Cette première partie peut être vue comme un moyen de projeter linéairement ou non linéairement les images d'entrée de la base de données **CASIA 3D** (**XM2VTS** pour le cas 2D) sur un espace de plus faible dimension.

- La **deuxième partie** essaie de représenter un visage de référence à partir de la représentation compacte issue de la projection. Lors de l'apprentissage, on construit une image référence de visage par personne, rendant ainsi la projection quasi invariante aux transformations présentées en entrée. L'intérêt des méthodes de réduction c'est la conservation de l'information avec un minimum de paramètres caractéristiques et la garantit d'une bonne discrimination. D'après les expériences menées l'EFM associée à l'ACP est bonne candidate et pour remédier au problème de non linéarité de l'ACP. Nous avons étendue nos recherches sur la KPCA qui donne des résultats satisfaisants. Les scénarios d'identification considérés dans cette thèse n'utilisent qu'un nombre limité d'images d'apprentissage ainsi que peu d'images/personne pour l'enrôlement. Malgré cette limitation, les taux de reconnaissance atteints sont comparables à ceux de l'état de l'art sur la base de données XM2VTS, CASIA 3D et FRGC. En revanche, les taux de reconnaissance sont moyens pour les visages présentant des expressions faciales. Nous pensons que le nombre d'images utilisées pour l'enrôlement influe sur les taux de reconnaissance. Une des limitations majeures des approches fondées sur un apprentissage en général est qu'elles nécessitent un grand nombre d'échantillons d'apprentissage afin d'obtenir une extraction de caractéristiques pertinentes ainsi qu'une bonne généralisation. Une manière de surmonter cette limitation consiste à faire un bon choix des méthodes de réduction d'espace. Pour cela une étude comparative de ces méthodes sur différentes BDD est menée et nous l'avons affirmé plus haut que l'EFM et la KPCA sont les plus performantes. Des tests ont de plus montrés sa bonne robustesse à des dégradations des images tests. Le développement d'une variante de cette méthode a permis d'accroître encore les taux de reconnaissance quand on lui ajoute des modalités c'est le cas du système bimodale étudié. Cette méthode souffre malgré tout de quelques défauts, remédiables cependant.
- Dans un **troisième temps**, la fusion de modalités a été considérée. La fusion de modalités offre une alternative aux systèmes biométriques unimodaux. Partant de l'hypothèse que les modalités offrent des informations complémentaires (ce qui est souvent le cas), la fusion de celles-ci permet globalement d'améliorer la fiabilité d'un système. La fusion de modalités permet en outre de s'affranchir de certaines problématiques inhérentes aux

Conclusion et perspectives

systèmes unimodaux (comme les variations de luminosité pour la modalité visible par exemple). Dans le cadre d'une fusion multi capteurs, elle nécessite cependant un investissement supplémentaire en capteurs, ce qui peut être coûteux.

Différents niveaux de fusion ont été étudiés. La fusion des images (niveau caractéristiques) ainsi que la fusion au niveau des scores ont été considérées via nos approches. La fusion la plus étudiée est la fusion au niveau des scores. Une méthode permettant de pondérer les scores issus des classifieurs est ainsi développée. Celle-ci permet de prendre en compte la pertinence des scores à l'aide d'une fonction de pondération dans le cas de la fusion par combinaison. La somme pondérée s'avère la plus performante dans notre application et particulièrement dans le cas de notre multi algorithmes utilisant la composante couleur **S** et les trois systèmes issus d'orientations différentes retenues du filtre de **Gabor**. Les approches ont également été utilisées comme modules de mise en correspondance avant une fusion des scores des modalités. Les méthodes utilisées comme l'**EFM**, **KPCA** et **SVM** ou mesure de métriques (pour la classification) ont permis l'utilisation de règles simples de fusion (somme des scores, min, max ou moyenne) pour obtenir des taux de reconnaissance élevés. Notons que l'intégrale de **Suegeno** donne aussi de bons résultats. Pour la classification la norme Euclidienne et la corrélation normalisée sont les meilleures pour la mesure de similarité dans notre cas. Le classifieur **SVM** est plus efficace et donne les meilleurs résultats dans toutes nos applications **2D**, **3D** ou **3D expression**.

Nous proposons aussi de développer une technique bimodale :

Dans **un premier temps** nous étudions toutes les caractéristiques issues d'une face.

- i) **2D-3D** de reconnaissance faciale qui combine des mesures anthropologiques 3D du visage avec une technique **2D** basée sur l'Analyse en Composantes Principales ou **ACP**. L'intérêt d'une telle approche est d'exploiter les avantages des deux techniques (**2D** et **3D**) afin d'améliorer le taux de reconnaissance. D'autre part, nous pensons que l'anthropométrie par sa maîtrise de l'anatomie du visage humain peut apporter des informations complémentaires pour améliorer la reconnaissance faciale. En effet, les anthropologues ont été les premiers à étudier de façon scientifique la morphologie humaine en général et celle de la tête et du visage en particulier et cela en utilisant des critères descriptifs et des caractères métriques. Nous proposons de définir des mesures anthropométriques faciales à partir de points caractéristiques **3D** du visage. Ces mesures correspondent à des valeurs indiciaires et angulaires caractérisant les différentes régions du visage.

Conclusion et perspectives

- ii) **2D-3D Expression** qui combine la reconnaissance des expressions faciales et le visage **2D**. L'intérêt dans ce cas est d'extraire les informations fournies par les expressions afin de mieux identifier les personnes. Une technique **3D** expression qui prend en entrée, non pas l'image entière du visage, mais les « imageries » correspondant aux régions caractéristiques du visage (les yeux, le nez et la bouche) et à démontrer que cette méthode donne des taux de reconnaissance aussi bon que l'image complète.
- iii) Enfin, nous proposons de fusionner les résultats de reconnaissance obtenus par les différentes approches (**2D** et **3D**) ainsi que différentes modalités avec le visage afin d'améliorer le taux de reconnaissance. Pour cela, nous avons exploré plusieurs critères de fusion. Un protocole expérimental a été mis en œuvre pour tester cette méthode et analyser leurs performances.

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

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

Dans **un deuxième temps** notre étude est orientée vers l'identification des personnes en se basant sur la fusion multi-systèmes, multi-biométries, multi-algorithmes et multimodales et pour tester notre système et évaluer ses performances.

- i) **Fusion multi-systèmes** : combinant la couleur et plusieurs systèmes utilisant la phase de Gabor. La fusion MCS est basée sur quatre systèmes issus des meilleurs filtres de Gabor retenus par nos diverses expériences. Les deux approches par combinaison et par classification de la fusion sont étudiées en association à différentes méthodes de normalisation sur la BDD XM2VTS.
- ii) **Fusion multi-biométries** : combinant deux biométries dépendantes (voix, visage). La fusion multi biométrique est assurée par le filtre de Gabor optimal retenu à

Conclusion et perspectives

travers nos expériences. Une réduction d'espace est choisie selon les meilleures performances de notre système de reconnaissance. La phase de classification est réalisée par calcul de distances ou par SVM. Notre SRV est appliqué à la base des scores XM2VTS.

- iii) Fusion multi-algorithmes :** C'est une fusion des caractéristiques codées d'une part par un codage entropique et d'autre part par le calcul de la moyenne. Ensuite nous réalisons la réduction des caractéristiques par l'EFM et la KPCA. Les scores obtenus par mesure de similarité par des métriques sont combinés et une décision est prise en fonction d'un seuil prédéfini.
- iv) Fusion multimodales :** combinant la couleur et la profondeur. Nous avons essayé de contourner le problème de la première expérimentation, à savoir la limite des images 2D face aux changements d'illumination et/ou expressions faciales. Les tests sur la base CASIA 3D sont réalisés :
 - a)** Première expérimentation, dans un premier temps, nous avons mis en œuvre un système de reconnaissance multimodale sur le visage 3D couleur. Il s'agit d'une fusion entre les caractéristiques ou scores de l'image couleur 2D (projection de l'image couleur RVB 3D sur le plan 2D) et une image profondeur nommée 2.5D (cartes de profondeur) issue de l'image couleur RVB 3D. Dans ce premier cas nous utilisons un classifieur basé sur la mesure de similarité pour sa simplicité.
 - b)** Dans la deuxième expérimentation nous nous sommes intéressés à la composante V du système HSV et nous faisons la fusion des caractéristiques ou scores des images 2D (composante V) et 3D (profondeur). Dans ce deuxième cas le classifieur SVM est choisi pour sa rapidité et son efficacité. Nous tentons de faire l'identification du visage sous diverses variantes (variation de lumière(IV), visage avec expressions faciales (EV) et visage avec expression faciale et variation de lumière(EVI)).

D'après les résultats des expériences réalisées nous pouvons dire que l'ACP+EFM reste une méthode discrimination performante d'une façon générale. La **fusion des scores** reste une méthode très efficace car sans la fusion nous avons avec un $N_p(\text{EFM}) = 100$; un **TR** = **92.25%** pour les visages pour une fusion par combinaison **min** et la distance L_2 .

D'après l'étude faite et toutes les expériences réalisées nous appliquons cette dernière à des images de visages **2D** neutres de la base de données **XM2VTS** en multi algorithmes et en multi biométries (voix et visage). Cette méthode reste performante avec un **TR=99.99%** en **2D** et multi biométrie. Nous confirmons que cette méthode reste une bonne candidate pour aborder les visages à **3D**. Des améliorations reste à réalisées surtout au niveau de la phase

Conclusion et perspectives

prétraitement et aussi et surtout la phase de classification. Car notre utilisation à la mesure métrique reste un choix guidé par sa simplicité. Ceci dit d'autres normes comme **Manhattan** ou mesures **géodésiques** restent envisageable, et toujours dans l'objectif de rehausser ces paramètres de performances du système biométriques. Les résultats sont dans l'ensemble satisfaisants, mais le taux de faux rejet reste moyen. Ceci est sûrement dû à l'étape de prétraitement qui reste à améliorer ainsi que le choix d'un meilleur classifieur.

Dans l'expérience de fusion **2D** et **3D** basée sur **SVM** nous utilisons la **profondeur** et la **couleur** des images pour construire un classificateur robuste pour l'identification du visage. Puisque la dimensionnalité des caractéristiques d'images de profondeur et de couleur est très élevée, nous avons étudié trois méthodes de réduction d'espace de données. Puis nous avons fait des études comparatives sur la fusion de l'information de profondeur et de couleur pour les deux niveaux : caractéristiques et scores pour la construction d'un classifieur efficace. En analysant nos résultats expérimentaux dans la base de données du visage **CASIA 3D** avec des variations complexes, nous illustrons la performance prometteuse du schéma proposé et d'en tirer l'importante conclusion suivante:

- La meilleure méthode de réduction d'espace est **PCA+EFM** pour toutes nos applications.
- La fusion par intégrale floue de **Sugeno** est très efficace dans l'identification des visages neutres **2D** (monomodale et multibiométries (visage,voix)).
- La fusion des caractéristiques et des scores de l'information profondeur et image couleur **RVB** est insuffisante à l'identification des visages **3D**.
- L'information de couleur est plus robuste que l'information de profondeur sous des variations d'expression,
- l'information de profondeur est plus robuste que l'information d'intensité de couleur sous des variations d'éclairage.
- La composante **V** du système **HSV** est plus performante que le système **RVB** dans l'identification de visages **3D**.
- La performance de classification des scores par les **SVMs** surpasse la fusion au niveau des caractéristiques et la combinaison des scores par les deux intégrales floue de **Sugeno** et **Choquet**.

D'après l'étude faite et toutes les expériences réalisées nous pouvons dire que la fusion des scores **2D** (couleur **RVB**) et **3D** (**profondeur**) avec les métriques améliore certes les performances du système **SRV** monomodale.

Conclusion et perspectives

Dans la **fusion** couleur **HSV** et **profondeur** l'approche par classification **SVM** est meilleure que celle par combinaison ce qui n'était pas le cas dans les fusions multi systèmes et multi biométries (fusion par combinaison : intégrale floue de **Sugeno** et fusion par classification : **SVM**).

Dans toutes les applications étudiées la fusion des scores donne les meilleurs résultats de performance du SRV.

Le système **HSV** est plus informant que le système RVB avec un TR=79.50% (pour la fusion mean, distance Euclidienne norme 1 et $N_p=60$) contre TR=100% (pour la fusion SVM, distance corrélation normalisée et $N_p<70$). Ainsi la **fusion des scores** couleur **HSV** et **profondeur** par **classification SVM** est parfaite avec un taux de réussite à 100% dans des visages exposés à la variation d'illumination. Mais elle reste sensible aux changements d'expressions faciales et à la combinaison changement d'expression faciales et lumière.

Nous affirmons à ce stade du travail que :

- l'algorithme **SVM** est très efficace quelque soit l'application envisagée.
- l'algorithme **SVM** est aussi bon en phase détection, qu'en phase classification et en fusion.
- La meilleure méthode de réduction reste l'**EFM**.
- Le filtre de **Gabor** s'il est bien utilisé nous permet d'extraire l'essentiel des caractéristiques globales du visage.
- Les **phases du filtre** sont mieux caractéristiques du visage que les amplitudes.
- Le système **HSV** est plus informant que le RVB.
- La fusion des scores est plus efficace que celles des caractéristiques.
- La norme L_1 et la **corrélation normalisée** sont les meilleures métriques dans la classification par mesure de similarité.
- La classification par **SVM** est meilleure que celle par mesure de distances. Elle est plus souple, rapide et nous évite le calcul de seuil.

Perspectives

Pour la multimodalité :

- La construction d'un modèle mathématique tel les modèles génétiques ou les PSO permettant d'optimiser notre approche.
- Utiliser les curvelets pour extraire les paramètres caractéristiques.
- Etudier d'autres extensions de l'ACP.

Conclusion et perspectives

- Approfondir les recherches dans les méthodes à noyaux.
- Utiliser la fusion par classification SVM et MLP dans les applications multi algorithmes et multi modales (couleur RVB et profondeur).
- Utiliser des techniques d'optimisation comme les algorithmes génétiques

Pour le visage 3D :

- Penser à utiliser l'information profil.
- Bien détecter les points et régions d'intérêts dans le visage.
- Etudier les techniques de correction de rotation de la tête pour récupérer le maximum d'information.
- S'intéresser par des techniques appropriées à la rotation de la tête, afin de récolter le maximum de caractéristiques utiles pour l'identification.
- S'intéresser à la symétrie du visage pour améliorer le temps de calcul.
- Appliquer notre système à des BDD 3D de jumeaux pour tester son efficacité.
- Mieux étudier les techniques de localisation du nez.

Pour la Classification :

Le problème de la robustesse d'un système biométrique à diverses altérations (luminosité, rotations, mises à l'échelle . . .) peut être attaqué en trois points de la chaîne. La plupart des systèmes biométriques n'en considèrent cependant que deux : améliorer les images avant d'effectuer une extraction de caractéristiques, et/ou extraire des caractéristiques invariantes aux transformations.

- Le premier niveau peut consister par exemple en un recadrage géométrique pour éliminer les transformations géométriques subies par l'image, ou en une modification de la dynamique des niveaux de gris pour pallier le problème de la luminosité.
- Le deuxième niveau généralement considéré consiste à produire des algorithmes capables d'extraire des caractéristiques invariantes aux dites transformations. La multiplication des transformations à considérer peut cependant rendre les algorithmes très complexes, et donc instables à de nouvelles transformations.
- Un troisième niveau pourrait consister à créer de nouvelles images à partir des images d'enrôlement en les modifiant via les transformations auxquelles on souhaite être invariant. Par exemple, à partir d'une image d'enrôlement, créer des images artificielles par rotation, puis les intégrer à la galerie. Une telle approche, combinée à l'algorithme de classification, permettrait à une image test ayant subi une rotation d'être mise en correspondance facilement avec son équivalent artificiel, alors qu'elle

Conclusion et perspectives

ne le serait pas forcément avec l'image originale présente dans la galerie. Une telle approche pour l'invariance par transformations affines serait non seulement simple à réaliser (il suffit de créer autant d'images artificielles que de transformations auxquelles le système doit être robuste), et devrait pouvoir fonctionner avec des algorithmes d'extraction de caractéristiques simples. Un problème se poserait néanmoins avec cette approche : la multiplication des images de la galerie, et donc le temps de calcul nécessaire à l'extraction d'un vecteur caractéristique sur cette galerie.

- Utiliser une classification par PSO