

Chapitre 6

Conception de l'Approche RV Multi Modales.

6.1 Introduction

Plusieurs expériences sont menées en multi algorithmes et multimodales où les deux approches par combinaison et par classification de la fusion sont étudiées en association à différentes méthodes de normalisation sur un système multi algorithmes appliqué aux images de la **XM2VTS**. Les techniques retenues sont appliquées aux **BDD** Bimodales dépendantes des scores comme (Voix, Visage) et aussi à une application bimodales indépendantes (Empreintes, Visage). Nos travaux se sont étendus au domaine **3D** pour contourner le problème d'illumination et rotation de visages. Seulement dans ce dernier cas le système présente des limites ce qui nous a mené à orienter nos recherches sur le **3D** expressions. Toutes nos expériences en **3D** sont réalisées à base de la **BDD CASIA 3D** et dans le cas des empreintes nous utilisons la **BDD FRVC**. Toutes ces bases de données sont universelles et agréées.

Notre objectif c'est l'identification des personnes en se basant sur une fusion multi algorithmique et multi biométrique.

- La fusion multi algorithmique est basée dans un premier temps sur quatre systèmes issus des meilleurs filtres de Gabor retenus par nos diverses expériences. Puis dans un deuxième temps une fusion des caractéristiques codées d'une part par un codage entropique et d'autre part par le calcul de la moyenne. Ensuite nous réalisons la réduction des caractéristiques par l'**EFM** et la **KPCA**. Les scores obtenus par mesure de similarité par des métriques sont combinés et une décision est prise en fonction d'un seuil prédéfini.
- La fusion multi biométrique est assurée par le filtre de Gabor optimal retenu à travers nos expériences sur la base de données **XM2VTS**. Une réduction d'espace est choisie selon les meilleures performances de notre système de reconnaissance. *Dans le cas de nos travaux l'EFM est bonne candidate*. La phase de classification est réalisée par calcul de distances ou par **SVM**. Les différentes modalités étudiées sont les expressions faciales, les empreintes, la voix et le visage **2D** et **3D**.

6.2 Bases de données

Les expériences ont été menées sur certaines bases spécifiques (**chapitre 5**) :

- La base de données **XM2VTS** [331] est composée de 395 personnes, organisée selon le protocole de Lausanne. Les images présentent des variations d'illumination, de pose et des artefacts (lunettes, barbes, . . .).
- La base de données **FRVC** [331-1] pour les empreintes.
- La base de données **CASIA 3D**[338] est composée de près de $n= 123$ personnes en 3D pour un total de plus de $m= 37 \times 123$ images. Celles-ci présentent de nombreuses variations de luminosité, de pose, rotation de la tête ou d'expressions faciales.
- La base principalement utilisée pour les tests multimodale est la base de données des **scoresXM2VTS**. Elle dispose d'un protocole bien défini permettant une comparaison précise avec des résultats publiés précédemment sur cette base.

Des expérimentations ont été conçues pour tester respectivement les problèmes de luminosité avec expressions ou visages neutres et la reconnaissance à travers le temps. Pour chacune de ces expérimentations, des protocoles sont effectués.

Plusieurs variantes sont prises en compte. Celles-ci dénotent les particularités des images associées et donc les caractéristiques que l'on souhaite tester. Elles sont organisées selon les cas suivants :

- les visages ont une expression faciale neutre,
- les visages présentant des expressions faciales,
- les visages sont capturés avec l'illumination,
- les visages sont capturés avec l'illumination et expressions faciales.

Ces jeux de tests permettent donc de tester différentes combinaisons, comme par exemple :

- un enrôlement où les visages ont une expression faciale neutre avec deux types d'expression, et les visages d'évaluation et de tests présentant des expressions faciales différentes.
- les visages d'enrôlement et de test ont la même expression faciale, mais la luminosité est changée.

La base de données XM2VTS a été largement utilisée car elle possède un protocole bien défini, ce qui est un atout certain pour la comparaison à de précédents travaux, pour peu que ceux-ci suivent le protocole de test. Certains travaux [33] [260] [261] utilisent cette base de données et proposent des algorithmes pour la reconnaissance ainsi que pour la fusion des modalités voix et visage. Dans [33], les fusions au niveau pixels et au niveau caractéristiques

issues d'une ACP sont considérées via des algorithmes génétiques. Les auteurs de [260] et [261] considèrent l'utilisation de multiples Machines à Support de Vecteurs (SVM) [260], ou de transformées en ondelettes [261] pour réaliser la fusion des images visible et infrarouge.

6.2.1 Résultats sur la XM2VTS

Les tests sur la base XM2VTS ont été réalisés en trois temps, ce qui conduit aux deux approches détaillées ci-après :

- a) Première expérimentation, dans un premier temps, nous avons mis en œuvre un système de reconnaissance multi algorithmes sur la modalité visage. Nous avons étudié les cas suivants :
 - la fusion des trois systèmes conçus par les meilleures orientations du filtre de Gabor et la composante **S** du système de couleurs **HSV**.
 - la fusion des caractéristiques issues des algorithmes entropie et moyenne.
- b) Deuxième expérimentation, nous avons essayé de tester notre approche sur deux modalités (bimodales) dépendantes (visage, voix) et sur deux modalités indépendantes (visage, empreintes). Les tests sont réalisés sur la BDD des **scores** XM2VTS pour les deux modalités (visage, voix), sur la BDD XM2VTS (visages) et la BDD FRVC pour les empreintes.

6.2.2 Résultats sur la CASIA 3D

Dans ce qui suit nous avons essayé de contourner le problème de la première expérimentation, à savoir la limite des images 2D face aux changements d'illumination.

Les tests sur la base CASIA 3D ont été réalisés en trois temps, ce qui conduit aux trois approches détaillées ci-après. Nous avons étudié les cas suivants :

- la fusion des caractéristiques 2D et 3D,
 - la fusion des scores 2D et 3D,
 - la fusion de la couleur et de la profondeur des images 3D.
- a) Première expérimentation, dans un premier temps, nous avons mis en œuvre un système de reconnaissance multimodale sur le visage 3D couleur. Il s'agit d'une fusion entre les caractéristiques ou scores de l'image couleur 2D (projection de l'image couleur RVB 3D sur le plan 2D) et une image profondeur nommée 2.5D (cartes de profondeur) issue de l'image couleur RVB 3D. Dans ce premier cas nous utilisons un classifieur basé sur la mesure de similarité pour sa simplicité.
 - b) Dans la deuxième expérimentation nous nous intéressons à la composante **V** du système HSV et nous faisons la fusion des caractéristiques ou scores des images 2D (composante

V) et 3D (profondeur). Dans ce deuxième cas le classifieur SVM est choisi pour sa rapidité et son efficacité. Nous tentons de faire l'identification du visage sous diverses variantes (variation de lumière(IV), visage avec expressions faciales (EV) et visage avec expression faciale et variation de lumière(EVI)).

6.2 Les systèmes d'authentification de visage retenus

Le principe de fonctionnement de base de notre système faciale (**figure 6.1**) est toujours le même que dans les applications monomodales. Il peut être résumé en **quatre étapes**.

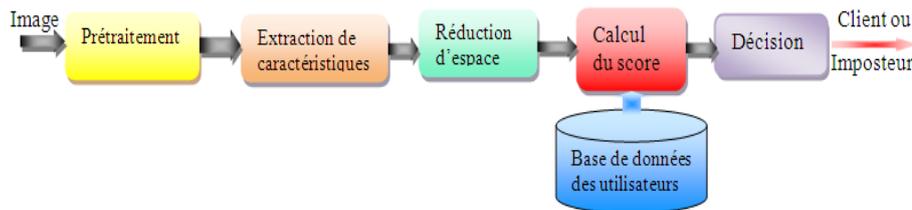


Figure 6.1 Principe de fonctionnement d'un système de reconnaissance de visage.

L'image d'entrée subit le même processus que dans les travaux réalisés au **chapitre 4**.

A.Fusion Multi Systèmes

Au cours des dix dernières années les systèmes **multi-systèmes** ou **multi classifieurs** (**Multiple Classifier Systems MCS**) sont devenus une approche établie pour concevoir des systèmes de classification. Un grand nombre de travaux à la fois théoriques et expérimentaux montre que les systèmes multi-classifieurs peuvent surpasser un classifieur unique dans plusieurs applications réelles, en termes de précision de la classification [339,296]). En particulier, plusieurs auteurs ont montré que les **MCS** peuvent permettre d'améliorer l'authentification biométrique de visages [339]. Toutefois, il n'est pas évident de savoir comment on fusionne les classifieurs. Le mécanisme de **fusion** peut se faire à différents niveaux de la classification [58] : au niveau des **données**, au niveau des **caractéristiques extraites**, au niveau des **scores** et au niveau des **décisions**. Notre **travail** dans ce cas se concentre sur la **fusion des scores** car c'est le type de fusion le plus utilisé. Elle peut être appliquée à tous les types de systèmes (contrairement à la fusion au niveau des données et au niveau des caractéristiques extraites). La fusion des scores est réalisée dans un espace de dimension limité représenté par un vecteur de scores dont la dimension est égale au nombre de sous systèmes, avec des méthodes relativement simples et efficaces mais traitant plus d'information que la fusion de décisions.

Les meilleurs systèmes de vérification de visages retenus (section 4.3.2.1) sont utilisés finalement pour étudier les méthodes de **fusion de scores** associés aux méthodes de **normalisation des scores**.

6.3 Principe de la fusion de scores MCS (Multi Systèmes)

Nous rappelons qu'un système de fusion de scores est constitué de deux modules, un module de fusion et un module de décision (**figure 6.2**).

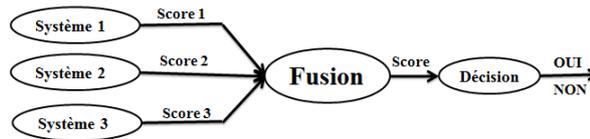


Figure 6.2 Schéma de la fusion de scores.

La fusion de scores consiste donc à la classification : **OUI** ou **NON** pour la décision finale, d'un vecteur de nombres réels. Il existe deux approches pour fusionner les scores obtenus par différents classifieurs :

- (i) Approche par combinaison de scores
- (ii) Approche par classification de scores.

Plus récemment **Morizet** et **Allano** ont essayé d'utiliser les deux approches précédentes et arrivent à des taux de réussite très intéressants.

Allano[58] utilise les deux approches basées sur le **SVM** pour l'approche par classification de scores et les méthodes simples de combinaison citées plus haut et les statistiques pour l'approche par combinaison de scores.

Morizet [3] a travaillé seulement avec l'approche basée sur la combinaison de scores en utilisant les mêmes méthodes que Allano et une nouvelle technique de fusion adaptative appelée "*WaveletScore DenoisingStatisticalFusion*" (**WSDSF**),

Nous nous sommes inspirés des travaux de Morizet et Allano et nous utilisons les deux approches citées précédemment pour l'étude des stratégies de fusion en ajoutant plus de méthodes comme :

La combinaison de scores par **logique floue** à base de l'intégrale de **Sugeno** ou de **Choquet**. Et les statistiques de **Fisher** et les réseaux de neurones artificiels (**MLP**) dans le cas de la première approche. Nous proposons aussi de faire la normalisation des scores dans ce cas bien précis de la fusion par classification de scores et c'est ce qui fait l'originalité de notre travail. Notre objectif est de réduire le coût et le temps d'utilisation des systèmes multimodaux

et améliorer les performances du système biométrique. Puis nous faisons une étude comparative de ces différentes méthodes.

Dans la **première approche** (fusion par combinaison des scores) nous utilisons donc des méthodes simples comme la **moyenne**, le **produit**, la **médiane** et **somme pondérée**[300] et en plus la combinaison de scores par **logique floue** à base de l'intégrale de **Sugeno** ou de **Choquet** [298,301]. Une autre question importante à laquelle nous apportons notre réponse est la normalisation des scores qui est une phase nécessaire avant de faire la combinaison.

Dans le cas de la **deuxième approche** (fusion par classification des scores) en plus de la méthode des machines à vecteurs de support **SVM** utilisée par Allano [58] et dans [287] nous utilisons les statistiques de **Fisher** [180] et les réseaux de neurones artificiels[317,318] (**MLP**). En ajoutant en plus des travaux antérieurs la phase de normalisation dans ce cas bien précis et c'est ce qui fait l'originalité de notre travail.

Nous faisons tout d'abord le choix de la meilleure méthode de réduction d'espace au sens des performances du système biométrique en se basant sur l'image brute (comme au **chapitre 5**). Une fois la méthode de réduction sélectionnée, nous étudions les amplitudes et les phases des filtres de **Gabor** et nous retenons le meilleur filtre présentant les meilleures performances biométriques (**TFA, TFR, TEE**) (voir paragraphe précédent). Ensuite les scores pour chaque système sont calculés puis soumis à la normalisation des scores. Cette dernière représente une étape très importante. Pour cela nous essayons d'étudier à ce stade plusieurs méthodes de normalisation. Nous les associons aux méthodes de fusion dans le cas des approches combinaison des scores et classification des scores et nous retenons la meilleure normalisation.

Finalement la meilleure approche de fusion est appliquée sur une base de données de scores **XM2VTS** utilisant deux modalités visage et parole. Et nous testons ainsi la validité de notre système de fusion retenu.

La solution pour les méthodes de réduction dans notre cas c'est le modèle discriminant Linéaire amélioré de Fisher (Enhanced Fisher Model **EFM**) [336]. Le modèle discriminant Linéaire amélioré de Fisher améliore la capacité de généralisation de la FLD en décomposant la procédure FLD.

La mesure de similitude entre les visages est assurée par un calcul de métriques classiques. Bien que la distance Euclidienne soit optimale dans la théorie. Les diverses expériences ont trouvé que la distance Euclidienne est surpassée par d'autres.

Une d'entre elles est la corrélation normalisée(normalizedcorrelation) [3,65] que nous rappelons est définie par :

$$S(A,B) = \frac{A^T B}{\|A\| \|B\|} \quad (6.1)$$

Cette fonction calcule simplement le cosinus de l'angle entre les deux vecteurs caractéristiques.

Nous étudions un multi systèmes de vérification faciale par deux approches de fusion de scores:

- **combinaison de scores** par des **méthodes simples** (somme, produit, min-max et médiane) et la **logique floue**(**intégrales** de **Sugeno** et **Choquet**),
- **classification de scores** par les méthodes statistiques de **Fisher**, la méthode des machines à vecteurs de support **SVM** et les réseaux de neurones artificiels (**MLP**).

Une normalisation est réalisée au préalable. Différentes méthodes de normalisation sont étudiées et la **Znorm**, **QLQ** (Quadratique Linéaire Quadratique) et la **fonction doublesigmoïde** sont retenues. L'étude de la fusion pour les deux approches combinaison et classification des scores des quatre systèmes d'authentification de visages est présentée par la **figure 6.3**. Nous commençons par comparer les méthodes de normalisation de scores associées aux méthodes de combinaison. Ensuite nous comparons les méthodes de classification de scores sans et avec la phase de normalisation. Toutes les méthodes sont testées sur la base de données de visages **XM2VTS** et selon le protocole de **Lausanne I**.

6.3.1 Conception de la fusion de scores Multi Systèmes

Après une étude des performances biométriques du système de reconnaissance de visages. Les systèmes qui présentent les meilleurs **TEE** (calculé sur l'ensemble d'évaluation) et **TR** (calculé sur l'ensemble de test) sont retenus.

Nous proposons un système multi-classifieurs de vérification faciale par l'étude de fusion de scores des quatre systèmes d'authentification de visages. Qui sont construits par l'extraction des caractéristiques du visage utilisant les phases filtrées des ondelettes de **Gabor** et l'Analyse en Composantes Principales (**ACP**) plus le modèle discriminant linéaire amélioré de Fisher (**EFM**) comme méthode de réduction d'espace de données.

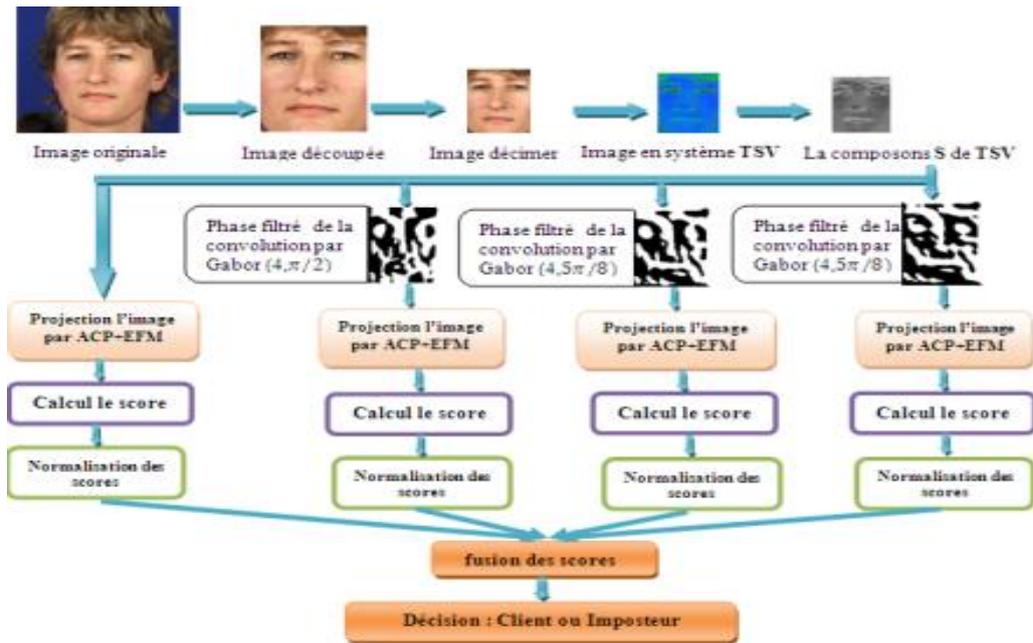


Figure 6.3 Schéma du multi systèmes par fusion de scores.

La représentation spatio-fréquentielle du visage a largement été utilisée et étudiée dans la littérature. Dans la majorité de ces études, seule l'amplitude de la réponse des filtres de Gabor a été employée alors que la phase est omise. Nous avons expliqué les raisons de la limitation de l'utilisation de cette phase et nous avons fourni une solution simple pour pallier cette limitation par une seuillage de la phase qui donne les meilleurs résultats par rapport aux amplitudes (section 4.3.2.1 chapitre 5). La phase de prétraitement, de réduction d'espace et de mesure de score sont commune aux quatre systèmes.

Les meilleurs résultats de **TEE** et **TR** obtenus sont présentés dans le **tableau 6.1**.

Tableau 6.1 Les quatre meilleurs systèmes retenus dans l'ensemble d'évaluation (TEE) et de test. (Intervalle de confiance paramétrique à 95%)

Performance	Système 1	Système 2	Système 3	Système 4
TEE %	2.66 ± 0.72	2.69 ± 0.72	3.3 ± 0.8	3.79 ± 0.85
TR %	94.33 ± 1.48	95.43 ± 0.81	93.59 ± 1.49	95.03 ± 0.82

Pour l'étude de fusion de scores nous utilisons deux approches :

- i) la **classification de scores** par : *méthode statistique de Fisher*, *Support Vector Machine (SVM)* et les *réseaux de neurones artificiels (MLP)*
- ii) les **combinaisons de scores** par : la *somme pondérée* et la *logique floue*.

6.3.2 Choix de la meilleure méthode de normalisation

6.3.2.1 Les distributions et normalisation de scores

Les distributions de scores pour les quatre systèmes d'authentification de visages sont représentées sur la **figure 6.4**. Nous remarquons que les quatre systèmes donnent des distributions Client et Imposteur différentes. Les distributions sont différentes en termes de plage de variation, ce qui rendra nécessaire l'étape de normalisation des scores. Elles sont également différentes dans leurs formes et leurs recouvrements entre les deux classes même si toutes les distributions ont un seul mode (mais pas forcément symétrique).

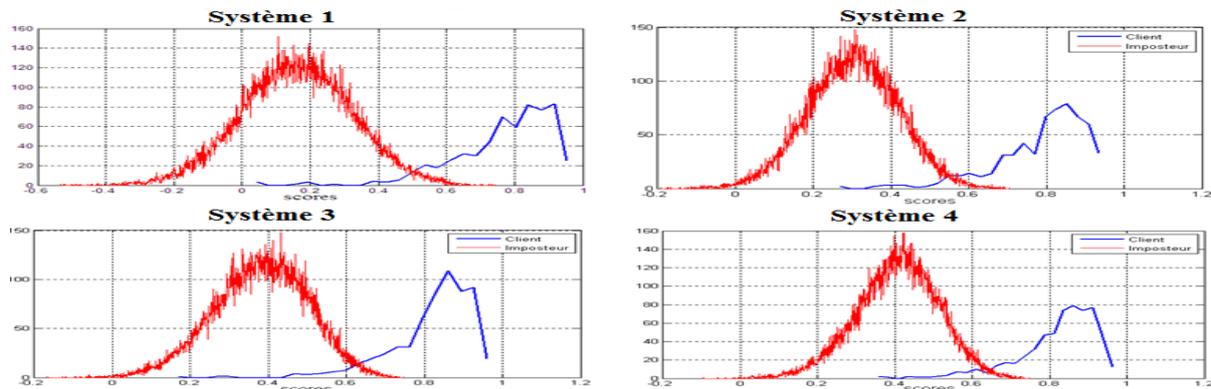


Figure 6.4 Les distributions de scores des quatre systèmes d'authentification de visages

6.3.2.2 Méthodes de normalisation de scores [295]

Les méthodes de normalisation de scores ont pour objectif de transformer individuellement chacun des scores issus des systèmes pour les rendre homogènes avant de les combiner. Les méthodes de normalisation de scores utilisés sont: **Znorm**, la **double sigmoïde** et la fonction quadratique-linéaire-quadratique (**QLQ**). Dans la **figure 6.5** la transformation des scores du premier système est présentée pour les deux normalisations (**Znorm** et la **fonction double sigmoïde**).

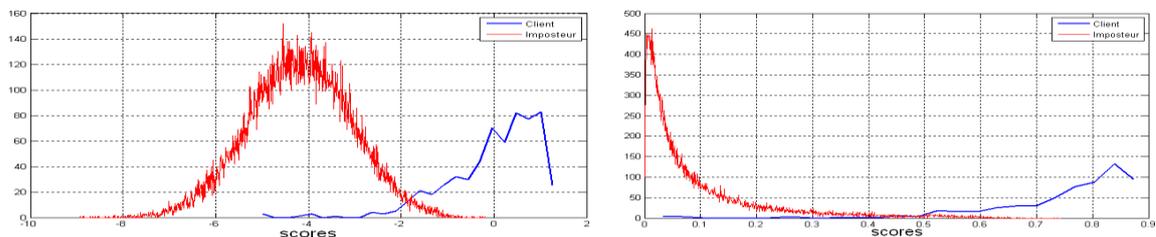


Figure 6.5 Normalisation de scores

Nous remarquons que la méthode de normalisation **Znorm**, ne modifie pas la forme des distributions mais leur différence se situe dans la façon dont chaque distribution sera répartie dans l'intervalle et donc dans la façon dont les systèmes à combiner se juxtaposeront dans

l'intervalle défini (avec Z_{norm} la moyenne de distribution des Clients de chaque système sera en 0). De plus, la méthode de normalisation **fonction double sigmoïde** modifie la forme des distributions.

Choix de la méthode de fusion

Comme dans la section **6.A** nous nous intéressons aux deux approches de fusion de scores la *combinaison* et la *classification*.

6.3.2.3 Fusion par combinaison des scores [340] :Pour ce type de fusion nous utilisons des méthodes simples [341] (moyenne, produit, minimum, maximum, médiane et somme pondérée) et deux intégrales floues [310,311] (**Sugeno** et **Choquet**).La somme pondérée est la meilleure méthode de combinaisons simples. Elle consiste en la définition de poids associés à chaque système, poids calculés de façon à minimiser le taux d'erreur sur la base de développement. On constate que la somme pondérée améliore les performances comparée à la deuxième meilleure méthode de combinaison simples qui est la somme simple (avec des poids identiques pour toutes les systèmes, ici $1/4$). Pour cette raison nous nous limiterons à cette combinaison simple.

Dans ce paragraphe, nous sélectionnons trois méthodes de normalisation de scores associées aux méthodes de combinaisons par logique floue à base de l'intégrale floue de **Sugeno** et de **Choquet**, ainsi que la combinaison simple par somme pondérée:

- une méthode qui *ne modifie pas la forme des distributions* est la **Znorm** puisqu'elle donne un meilleur résultat que la **minmax**, la **médiane** et l'**écart absolu médian** qui est plus simple que la **tangente hyperbolique**,
- deux méthodes qui *modifient la forme des distributions* les fonctions quadratique-linéaire-quadratique (**QLQ**) et **fonction double sigmoïde**.

Les différents taux d'erreur et de succès dans l'ensemble d'évaluation et de test en utilisant la fusion par combinaison de scores se trouvent dans le **tableau 6.2**.

Tableau 6.2 Performances des méthodes de normalisation associées aux méthodes de combinaisons pour MCS (intervalle de confiance paramétrique à 95%).

Méthodes de combinaison	Taux de Performance	Méthode de normalisation		
		Znorm	QLQ	Double sigmoïde
Intégrale floue de Sugeno	TEE	2.08 ± 0.64	1.22 ± 0.49	1.14 ± 0.47
	TR	96.28 ± 0.8	98.05 ± 0.76	98.36 ± 0.75
Intégrale floue de Choquet	TEE	2.16 ± 0.65	1.89 ± 0.61	2.4 ± 0.68
	TR	96.31 ± 0.79	97.08 ± 0.78	95.68 ± 0.81
Somme pondérée	TEE	1.95 ± 0.62	2.13 ± 0.64	2.17 ± 0.65
	TR	96.73 ± 0.79	96.6 ± 0.79	96.36 ± 0.79

Nous pouvons dire que dans le cas des méthodes de normalisation de scores qui modifient la forme des distributions la *fonction double sigmoïde* est la mieux performante associée à différentes méthodes de fusion par combinaison.

La méthode de fusion par combinaison de scores l'intégrale floue de *Sugeno* donne le meilleur résultat avec un **TEE=1.14±0.47%** et **TR=98.36 ± 0.75%** avec un temps de calcul très réduit **t=0.94 seconde**.

6.3.2.4 Fusion par classification des scores [345]: Dans cette section nous étudions la fusion de scores par l'approche de classification de scores (voir paragraphe **chapitre 4**). Normalement les méthodes de fusion par classification de scores cherchent à prendre une décision à partir du vecteur de scores de dimension N sans normalisation de scores. Mais dans notre travail nous proposons d'utiliser les méthodes de normalisation de scores pour transformer individuellement chacun des scores issus des systèmes afin de les rendre homogènes avant la classification, et ceci dans le souci d'augmenter la performance d'authentification. Nous utilisons trois méthodes de normalisation: **Znorm**, fonction quadratique-linéaire-quadratique (**QLQ**) et normalisation par une **fonction double sigmoïde**. Les différents taux d'erreur et de succès dans l'ensemble d'évaluation et de test en utilisant la fusion par classification de scores sans et avec les méthodes de normalisation se trouvent dans le tableau **6.3**. Nous utilisons la classification de Fisher (en utilisant la fusion par fonction linéaire discriminante (**LDF**) et fonction quadratique discriminante (**QDF**), une machine à vecteurs de support (**SVM**) avec un noyau **RBF** (Radial Basis Function) et une classification par MLP. Le **SVM** a été implémenté en utilisant la librairie **LibSVM** disponible gratuitement au site (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>).

Le noyau **RBF** utilisé est de la forme :

$$K_{RBF}(u, v) = e^{-\gamma \|u-v\|^2} \quad (6.2)$$

Où γ est un paramètre qui règle la largeur de la marge.

Les paramètres choisis par notre expérience pour la classification par **MLP** sont :

- Deux entrées sont les scores pour chaque système.
- Une couche cachée avec dix neurones et la fonction d'activation sigmoïde.
- Deux neurones dans la couche de sortie et la fonction d'activation tangente hyperbolique.

Tableau 6.3 Performances des méthodes de normalisation associées aux méthodes de classification pour MCS (intervalle de confiance paramétrique à 95%).

Méthodes de classification	Taux de performance	Méthode de normalisation			
		Sans normalisation	Znorm	QLQ	Double sigmoïde
SVM	TEE	2±0.62	1.5±0.54	1.96±0.62	1.66±0.62
	TR	96.53±0.79	97.44±0.77	97.36±0.78	97.11±0.78
MLP	TEE	1.83±0.6	1.69±0.57	1.66±0.57	1.33±0.51
	TR	96.25±1.07	96.69±1.27	97.84±0.77	97.49±0.77
LDF	TEE	2.14 ± 0.64	1.98 ± 0.62	1.97 ± 0.62	2.17 ± 0.65
	TR	96.32 ± 0.79	96.66 ± 0.79	96.82 ± 0.79	96.35 ± 0.79

Nous pouvons dire que la méthode proposée qui utilise la normalisation de scores avant la fusion par classification (et dans la majorité des cas les méthodes de normalisation qui modifient la forme des distributions) améliore la performance d'authentification de visages comparativement à celle obtenue sans normalisation. La meilleure performance est obtenue par la normalisation QLQ associée au MLP avec des taux **TEE=1.66±0.57%** et **TR=97.84±0.77%**. Seulement nous notons au passage que le SVM est plus sollicité dans les SRV pour sa rapidité, son efficacité et sa facilité d'implémentation malgré les taux obtenus dans notre cas **TEE=1.5±0.54 %** et un **TR=97.44± 0.77%** (pour la normalisation double sigmoïde).

6.3.2.5 Discussion

Nous avons montré comment l'utilisation d'un **système multi classifieurs (MCS)** permet d'améliorer notablement les performances d'un système de vérification d'identité monomodale de visage et nous affirmons que les méthodes de normalisation de scores améliorent la performance en générale de l'authentification de visages pour toutes les méthodes de classification de scores utilisées. Le meilleur résultat est obtenu dans le cas de la fusion par combinaison (intégrale floue de **Sugeno**) avec un **TEE=1.14±0.47%** et **TR=98.36 ± 0.75%** pour la méthode de normalisation de scores la fonction double sigmoïde (**tableau 6.2**). Les deux méthodes de classification de scores **SVM** (TEE=1.5±0.54 % et un TR=97.44± 0.77%) et **MLP** (TEE=1.66±0.57% et TR=97.84±0.77%.) nous donnent satisfaction pour les normalisations de scores respectives **Znorm** et **QLQ** (**tableau 6.3**). Rappelons que dans le cas de *l'approche monomodale* appliquée à la même BDD XM2VTS, pour une classification **SVM** nous sommes arrivés au meilleur des cas à des taux de performance **TEE=0±0%** et **TR=97.72±1.46%** dans l'ensemble de test avec un temps de calcul T_c satisfaisant $t=0.32s$ dans le cas de méthode de réduction linéaire choisie

(ACP+EFM+200SVMs). Alors que pour le cas non linéaire (KPCA) nous obtenons $TR_{\text{test}} = 97.5 \%$ pour un noyau **Gausséen** avec la résolution $\sigma = 0.0004$ et $N_p = 600$.

Le meilleur système de fusion est donc construit par la normalisation de scores :

- 1) Normalisation **double sigmoïde** pour la combinaison par l'intégrale floue de Sugeno [342].
- 2) Normalisation **Znorm** pour la fusion par classification SVM [345].
- 3) Normalisation **QLQ** pour la fusion par classification MLP [345].

B. Fusion Multi Biométries [344]

6.4 Application du meilleur système de fusion sur la base de données de scores XM2VTS

6.4.1 Base des scores XM2VTS

Pour les deux protocoles de Lausanne (LP1 et LP2), Norman Poh a mis à disposition un ensemble de scores issus de **5 classifieurs de visage** et **3 classifieurs de voix** présentés dans [39]. Ces **scores** donnent 21 configurations différentes pour 21 couples de classifieurs **voix** et **visage** (*15 couples pour le protocole I et 6 couples pour le protocole II*). Chaque configuration correspond à un couple de systèmes: un système de voix et un système de visage. Chaque système est composé de :

- un module d'extraction de caractéristiques,
- un classifieur.

Un seul classifieur est utilisé pour la voix : le **GMM** (mélange de gaussiennes). Deux classifieurs sont utilisés pour le visage : le **GMM** (mélange de gaussiennes) et le **MLP** (réseaux de neurones appelé "Multi Layer Perceptron"). Trois modules d'extraction de caractéristiques sont disponibles pour le visage (**FH** : Face Histogram, **DCTs** : Discrete Cosinus Transform Small Size et **DCTb** : Discrete Cosinus Transform big size) et trois modules d'extractions de caractéristiques pour la voix (**LFCC** : Linear Filter-bank Cepstral Coefficient, **PAC-MFCC** : Phase Auto-Correlation Mel Filter-bank Cepstral Coefficient et **SSC** : Spectral Subband Centroid). Tous les détails des classifieurs et des modules d'extraction de caractéristiques sont donnés dans [39]. Rappelons qu'un système de fusion de scores est constitué de deux modules, un module de fusion et un module de décision. Le problème devient donc un problème de classification à 2 classes (OUI ou NON, Client ou Imposteur) à partir d'un vecteur de nombre réels dont la dimension est égale au nombre des sous-systèmes. Afin de s'assurer que la combinaison de scores provenant de différents

si les systèmes sont cohérents, les scores doivent d'abord être transformés dans un domaine commun : on parle alors de *normalisation de score*.

6.4.1.1 Application à la base de données des scores XM2VTS

C'est une BDD multimodale prenant en compte trois systèmes de voix et les cinq systèmes de visage et délivrant les scores obtenus par des systèmes d'authentification biométriques particuliers. Nous nous intéressons uniquement au Protocole I (LP_1) pour lequel nous allons tester notre méthode de fusion. Le **tableau 6.4** présente la performance en terme de TEE dans l'ensemble d'évaluation et en terme TR dans l'ensemble de test pour les quinze couples de systèmes voix et visage (voix,visage). La méthode de normalisation de scores utilisée dans ce cas est la fonction **double sigmoïde**.

Tableau 6.4 Fusion de scores des visages et voix.

Méthodes de fusion	Systèmes de Voix ,Visage	Performance	
		TEE%	TR %
Intégrale floue de Sugeno	(LFCC,GMM),(DCTs, GMM)	0.001±0.01	99.99±0.04
SVM	(LFCC, GMM), (DCTs, MLP)	0.03±0.07	99.30±0.41
MLP	(LFCC, GMM),(FH, MLP)	0.23±0.21	99.57±0.32

La méthode de normalisation et la méthode de fusion de scores sont : la fonction **double sigmoïde** et l'intégrale de floue de **Sugeno** respectivement. Le tableau montre clairement que la fusion de scores appliquée à la multimodalité est toujours mieux que le meilleur système monomodal pour tous les couple des systèmes (voix ,visage). Le meilleur résultat est obtenu avec la fusion par l'intégrale de **Sugeno** dans le cas de la première approche pour le couple voix-visage (**LFCC,GMM),(DCTs,GMM)**. Une performance d'authentification presque parfaite est réalisée avec un **TEE=0.001±0.01%** et un **TR=99.99±0.04%**. Pour la deuxième approche la meilleure performance est atteinte dans le cas de la fusion par **MLP** pour le couple voix-visage ((LFCC, GMM),(FH,MLP)) avec TEE=0.23±0.21 et un TR=99.57±0.32. Nous nous intéressons uniquement au Protocole I (LP_1) pour lequel nous testons notre méthode de fusion. Le **tableau 6.5** présente la performance en terme de TEE dans l'ensemble d'évaluation et en terme TR dans l'ensemble de test pour les 15 couples des systèmes voix et visage (voix,visage). La méthode de normalisation et la méthode de fusion de scores sont : la **fonction double sigmoïde** et l'intégrale floue de **Sugeno** respectivement.

Tableau 6.5 Performances de fusion de scores pour les 15 couples des systèmes (voix, visage) dans d'ensemble d'évaluation et test (intervalle de confiance paramétrique à 95%).

Méthodes Voix ,Visage	Intégrale floue de Sugeno		Le meilleur système	
	TEE%	TR %	TEE%	TR %
(LFCC, GMM),(FH, MLP)	0.002±0.02	99.99±0.04	0.84±0.4	97.7± 1.15
(LFCC, GMM),(DCTs, GMM)	0.001±0.01	99.99±0.04	0.84±0.4	97.7± 1.15
(LFCC, GMM),(DCTb, GMM)	0.007±0.03	99.93±0.12	0.84±0.4	97.7± 1.15
(LFCC, GMM), (DCTs, MLP)	0.03±0.07	99.91±0.14	0.84±0.4	97.7± 1.15
(LFCC, GMM), (DCTb, MLP)	0.04±0.08	99.88±0.17	0.84±0.4	97.7± 1.15
(PAC, GMM),(FH, MLP)	0.17±0.18	99.49±0.35	1.66±0.57	96.25±1.45
(PAC, GMM),(DCTs, GMM)	0.24±0.21	98.69±0.56	4.37±0.91	91.55± 2.04
(PAC, GMM),(DCTb, GMM)	0.03±0.07	99.39±0.28	1.53±0.55	96.62±1.27
(PAC, GMM), (DCTs, MLP)	0.04±0.08	99.66±0.28	2.96±0.76	93.32±1.91
(PAC, GMM), (DCTb, MLP)	0.63±0.53	99.21±0.43	4.36±0.19	91.07±2.15
(SSC, GMM),(FH, MLP)	0.09±0.13	99.72±0.25	1.66±0.57	96.25±1.45
(SSC, GMM),(DCTs, GMM)	0.28±0.23	99.48±0.35	4.36±0.19	91.07±2.15
(SSC, GMM),(DCTb, GMM)	0.03±0.07	99.64±0.29	1.53±0.55	96.62±1.27
(SSC, GMM), (DCTs, MLP)	0.2±0.2	99.40±0.37	2.96±0.76	93.32±1.91
(SSC, GMM), (DCTb, MLP)	0.45±0.3	97.65±0.74	4.36±0.19	91.07±2.15

Le **tableau 6.5** montre clairement que la fusion de scores est toujours mieux que le meilleur système pour tous les couples des systèmes (voix,visage), et le meilleur résultat est obtenu avec le **couple voix visage (LFCC,GMM),(DCTs,GMM)** qui nous donne une performance d'authentification presque parfaite avec un **TEE=0.001±0.01%** et un **TR=99.99±0.04%**.

6.4.1.2 Travaux récents sur la multimodalité appliquée à la BDD XM2VTS

La série de résultats obtenue, est résumée par un tableau de synthèse (**tableau 6.6**) qui nous permet de comparer notre systèmes d'authentification multi-algorithmes et multi-biométries (visage,voix) à d'autres travaux connus en fusion multimodale, et aussi de faire la comparaison entre les systèmes multi-algorithmes et multi-biométries sur la base de données XM2VTS et la base de scores **XM2VTS** selon le protocole de Lausanne I.

Tableau 6.6 Tableau de synthèse des résultats obtenus en fusion multimodale

Date	Auteurs	Multimodalités	Niveau de fusion	Méthodes de fusion	BDD	Meilleurs résultats
2007	Sadeghi et al [85]	Multi algorithmes visage	Scores	Confidence Based Gating of Scores	XM2VTS	TRA=0.48%, TFR= 0.5% TR=98.45%
2009	Villegas et al [346]	Multi biométries visage,voix	Scores	Maximisation de AUC	Base de scores XM2VTS	TEE= 0.28%
2009	Li et al [42]	Multi biométries visage,voix	Scores	Nouvelle méthode basée sur TFR et TFA	Base de scores XM2VTS	TEE=0.27%
2009	E.Jarque et al [314]	Multi biométries visage ,voix	Caractéristiques	SVM	XM2VTS	TEE=0.008%
2009	Chia et al [299]	Multi biométries visage,voix	Scores	Hybride par min ou max ou somme	Base de scores XM2VTS	TEE=0.5%
2010	Lobrano et al [347]	Multi biométries visage,voix	Scores	DynamicLinearCombination(DLC)	Base de scores XM2VTS	TEE= 0.0045%
2012	Notre système [341,342]	Multi systèmes visages	Scores	Intégrales de floue de Sugeno	XM2VTS	TEE=1.14% TR=98.36%
2010	Notre système [340]	Multi biométries visage,voix	Scores	Intégrales de floue de Sugeno	Base des scores XM2VTS	TEE=0.001% TR=99.99%

Nous pouvons remarquer que la performance de notre système multi-algorithmes d'authentification de visages et presque équivalent à celui obtenu en 2007 par **Sadeghi et Khoshrou** avec un **TEE =1.14%** et un **TR=98.36%**.

Notre système multi biométries d'authentification de visage et voix (appliqué à la base des **scores XM2VTS**) est le meilleur comparé aux travaux 2009 et 2010 avec un **TEE=0.001%** et un **TR=99.99%** qui est justifié par l'efficacité de la méthode de normalisation de scores la fonction double sigmoïde et la fusion de scores par l'intégrale floue de **Sugeno**.

Pour la comparaison entre les deux approches multi systèmes et multi biométries, nous affirmons que le système multi-biométries a donné une performance presque parfaite avec un **TR=99.99%** et **TEE=0.001%**, mais avec beaucoup d'inconvénients : temps de capture de visage et voix très lent [5], coûteux, confort d'utilisateurs etc....

Par contre nous avons obtenu une performance satisfaisante par un système multi-algorithmes d'authentification de visages avec un **TR = 98,36%** et **TEE=1.14%** avec beaucoup d'avantages : temps de calcul réduit puisque le temps de capture est très supérieur au temps de traitement [5], confort des utilisateurs etc

C. Fusion Multi Algorithmes

6.5 Fusion basée sur l'entropie, moyenne, EFM et KPCA

Dans cette partie du travail nous essayons de faire la fusion des caractéristiques à la source afin de tirer le maximum d'informations. Et par souci d'encombrements de données nous codons nos images par deux algorithmes statistiques très simples qui sont la moyenne et l'entropie. Ces deux derniers sont ensuite fusionner, une réduction d'espace réalisée par l'EFM ou KPCA est assurée puis une classification par calcul de similarités nous permet d'obtenir nos scores qui à leurs tour sont fusionnés pour faciliter la décision de notre système. Notre approche est validée et testée sur des images frontales de la base de données XM2VTS et nous discutons les résultats en fonction de différents paramètres.

6.5.1 Architecture structurelle du système

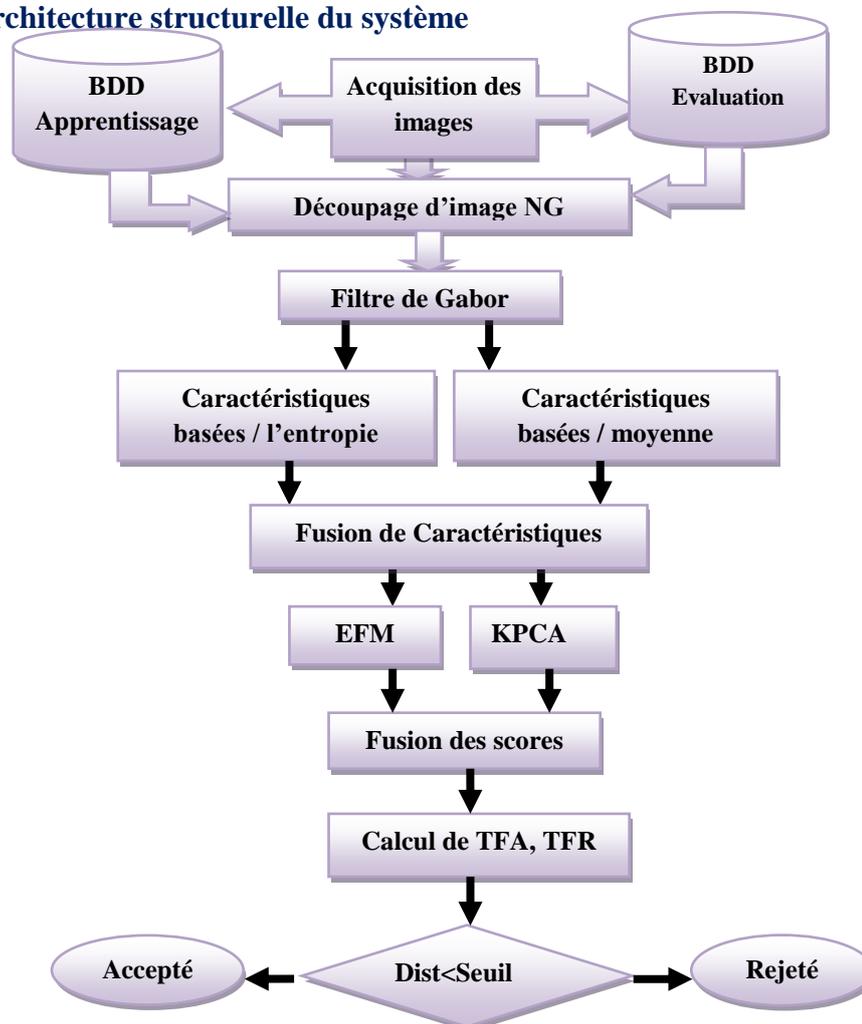


Figure 6.6 Architecture structurelle du système.

Comme le montre si bien la figure 6.6 notre système est structuré par plusieurs composantes fonctionnelles. Le prétraitement représente une fonction fondamentale du processus SRV. Nous préparons notre image pour être traitée dans cette phase. Ensuite, nous codons l'image pour en extraire des paramètres caractéristiques réduits. Cette caractérisation est réalisée par

deux algorithmes en parallèle : l'algorithme *moyenne* et l'algorithme *entropie*. Les matrices ainsi obtenues sont fusionnées par des méthodes de combinaison simples et le résultat est ainsi soumis à l'analyse. Il s'agit là d'une fusion multi algorithmes des caractéristiques (niveau caractéristiques). Cette fusion a l'avantage de traiter les caractéristiques brutes de l'image, donc nous agissons sur l'information à l'entrée. Et nous codons l'image pour contourner le traitement de taille de données énorme dû au nombre d'images issues de la base de données. Maintenant pour l'analyse nous procédons tout d'abord à la réduction de la taille des paramètres caractéristiques. Les méthodes que nous utilisons dans notre approche sont finement sélectionnées suite à notre étude bibliographique et l'état de l'art des travaux récents présenté au **chapitre 2**. Donc nous choisissons de faire une réduction d'espace par méthode le modèle de Fisher Amélioré **EFM** (Enhanced Fischer Model) et par Analyse en Composantes Principales à Kernel (Noyau) **KPCA** (Kernel Principle Component Analysis). Ces deux transformations sont réalisées en parallèle. L'**EFM** nous renforce la discrimination des classes et leur généralisation et la **KPCA** est là pour nous faciliter la tâche dans la classification non linéaire. Ces deux algorithmes sont une fois encore fusionnés et cette fois ci il s'agit d'une fusion des scores de même nature que la première c'est-à-dire une fusion multi algorithmes. Ensuite les paramètres de performances du **SRV** sont calculés. Finalement la classification est réalisée par mesure de similarité et une décision est prise se basant sur un test sur un seuil.

Organigramme de test

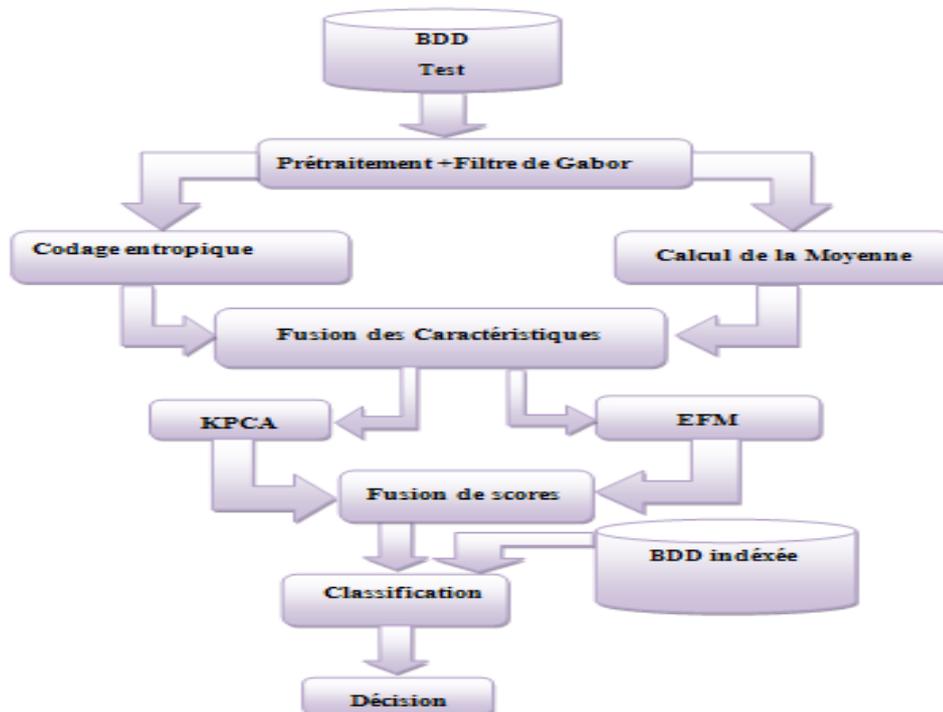


Figure 6.7 Organigramme de test

6.5.2 Implémentation du modèle

La mise en œuvre et essais de notre approche se présente en deux variantes :

- 1) La première basée sur les différents noyaux de la **KPCA**.
 - i) La prise en compte de plusieurs fonctions de noyaux (noyaux Gauss, Linear et Poly existant dans le toolbox Matlab 2009).
 - ii) La prise en compte des fonctions réalisant la fusion par combinaison. Dans notre travail nous nous limitons aux algorithmes de fusion par min, max et mean pour leurs simplicités. (Ces opérateurs sont utilisés dans la fonction **wfusmat** du Matlab 2009).
- 2) La deuxième sur le nombre de paramètres de l'**EFM**.
 - i) La variation du nombre de paramètres caractéristiques,
 - ii) Les différentes méthodes de fusion par combinaison citées précédemment.

6.5.3 Comparaison des performances

Nous effectuons des mesures pour diverses fonctions de noyau (Gauss, Linear, Poly) afin de retenir la meilleure au sens performances en changeant aussi les méthodes de fusion. Nous menons nos expériences pour $N_p(\text{PCA}) = 100$ avec la variation de $N_p(\text{EFM})$ et différentes fusions par combinaison (**mean, min, max**) dans le cas des trois noyaux du **Noyau Gausséen**, **linéaire** et **polynomial** sur la **BDD** évaluation et test. Les meilleurs résultats sont répertoriés dans le tableau suivant :

Tableau 6.7 Valeurs **TEE** et **TR** pour différents noyaux de la **KPCA**(fusion multialgorithmes).

Kernels	Gausséen	Linéaire	Polynomial
TR	Min=76.75%	Min =76.20%	Min=76.20%
TEE	Min=23.24%	Mean= 23.79%	Mean=25.78%

D'après ces résultats nous pouvons dire que le **Kernel Gausséen** reste celui qui présente le meilleur taux de réussite pour la fusion **min**. Dans le cas du **Kernel Gausséen**, nous constatons que la fusion par **min** donne le meilleur **TR = 76.75%** dans l'ensemble d'évaluation. Les valeurs de **TR** obtenus dans le cas de la fusion par **max** et **mean** sont semblables et stables autour de **TR = 64.75%**. Ce qui confirme l'importance et l'apport de la fusion basée sur **min**. Nous constatons une stationnarité dans les valeurs de **TR** pour la fusion min. Ce qui prouve la **stabilité** de ce type de fusion pour **divers Kernels**.

D. Fusion des informations profondeur et couleur

6.6 Fusion de la couleur RVB et profondeur

6.6.1 Fusion des caractéristiques par combinaison (couleur RVB + profondeur)

La reconnaissance faciale **3D**, combinant la forme et l'intensité / couleur **2D** dans la formation, est un sujet en développement. La combinaison d'information **2D** et **3D** fournit une occasion d'améliorer les performances de reconnaissance des visages. Wang et al. [158] décrivent les points caractéristiques du visage à l'aide de réponses des filtres de Gabor dans un domaine **2D** et les signatures de points dans un domaine **3D**. La classification a été faite ensuite par des machines à vecteurs de support à une décision "graphe orienté acyclique". Tsalakanidou et al. [161] ont construit un classifieur "embarqué modèles de Markov cachés (EHMM)" basé sur la profondeur et l'intensité, puis utilisé des règles de fusion pour combiner les scores. Dans [35], Chang et al. ont évalué le système de reconnaissance avec différentes combinaisons de l'information **2D** et **3D** et ont montré que la combinaison des informations **2D** et **3D** était plus efficace dans la caractérisation d'un individu. Dans les méthodes existantes susmentionnées, certains problèmes demeurent distincts et en suspens:

- Certains travaux [348,349,350] ont été effectués pour comparer la reconnaissance faciale **2D** et **3D**. Cependant, certains détails sont ignorés sur la façon dont la profondeur et l'intensité d'information contribue à la reconnaissance des visages avec une expression des variations d'éclairage.
- Seule la fusion au niveau de décision est considérée.

Nous consacrons cette partie à l'implémentation du modèle d'identification basé sur la fusion des caractéristiques et des scores *profondeur* et *couleur RVB* basée sur une classification métrique. Les différentes étapes de la conception sont :

- Projection de l'image **3D** en **2D** : Pour des facilités d'implémentation nous projetons nos images **3D** sur un espace **2D**. Nous réalisons deux projections en parallèle :
 - Génération des cartes de profondeur **2.5**.
 - Projection de l'image **3D** dans l'espace des couleurs **2D** couleur (**RGB**).
- Fusion de données (image **2D RGB** et **2.5D** profondeur).
- Extraction de paramètres (Filtrage **GABOR**).
- Réduction d'espace (**LDA** et **EFM**).
- Classification : Les descripteurs extraits par le filtre de **Gabor** sont projetés d'une façon parallèle sur deux espaces. La première est calculée par **LDA** et la deuxième obtenue par l'**EFM**. Les vecteurs projetés dans chacune des espaces séparément sont

comparés aux vecteurs de la base d'apprentissage indexée et un calcul de distances intra et extra est effectué. Ce calcul est réalisé par des normes Euclidiennes (Euclidienne norme 1 (L_1)).

- Décision

6.6.1.1 Architecture fonctionnelle du SRV à base de fusion des caractéristiques

Cesystème de reconnaissance de visages 3D se présente comme suit :

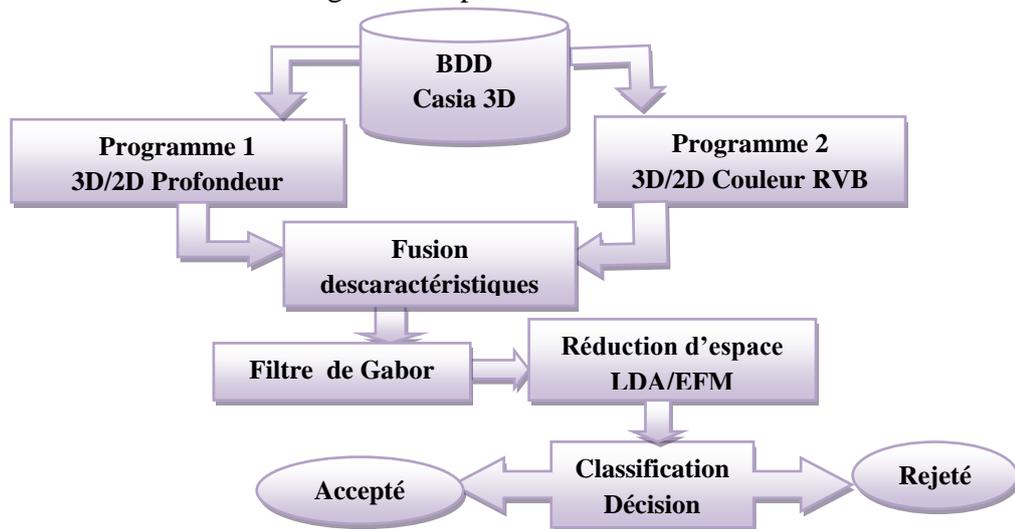


Figure 6.8 SRV 3D à base de fusion des caractéristiques : *profondeur* et *couleur RVB*

Nous utilisons une fusion des caractéristiques. Cette fusion présente l'avantage de prendre en compte l'information contenue dans l'image. Mais elle a l'inconvénient de mettre en jeu une taille importante de données. Cet inconvénient est remédié par le filtre de **Gabor** qui nous permet de retirer que les détails dans nos images. Ces détails représentent des contours dans l'image qui sont très représentatifs puisqu'il s'agit de points et régions d'intérêt dans le visage (exemple : contour du nez (narines), des yeux, la bouche, sourcils et tout autre détails dans le visage). Cette étape de filtrage est suivie d'une méthode de réduction d'espace à pouvoir discriminatoire c'est l'**EFM** qui nous prépare à la phase classification. Le choix de l'**EFM** est justifié par les résultats obtenus dans les travaux de M.Belahcène et A.M.Ouamane en 2011[55].

6.6.1.2 Implémentation de l'Approche SRV 3D (couleur RVB+profondeur)

Pour augmenter la lisibilité de notre programme et faciliter les différents tests nécessaires, nous avons essayé de le partitionner en un ensemble demodules.

1 Prétraitement de visage

En projetant l'image **3D** couleur sur un espace plan nous obtenons notre image couleur **2D**. Par un programme de prétraitement nous obtenons la carte profondeur appelée

aussi l'image **2.5D**(figure 6.9). Ces deux images après avoir subies des traitements séparés seront fusionnées.

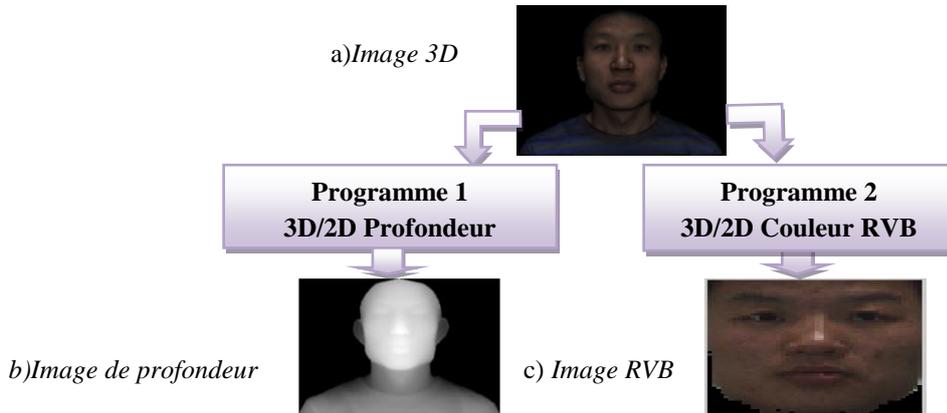


Figure 6.9 Phase de prétraitement de l'image **3D** couleur

2 Fusion des caractéristiques

Nous fusionnons deux images originales : l'image **3D** couleur projetée sur un espace **2D** et l'image **2.5D** qui représente les cartes de profondeur. Pour cela nous utilisons des méthodes de fusions appliquées aux coefficients d'approximations et les coefficients de détails de l'ondelette. Les deux images doivent être de la même taille et sont censées être associées à des images indexées sur une palette de couleurs commune. Les méthodes de fusion par combinaison des coefficients de l'ondelette sont ainsi calculées. Nous changeons le mode d'extension de rembourrage symétrique, en utilisant la commande: **dwtmode** ("**symétrique**"), et la fonction (**wfusing**) pour la fusion d'images. L'algorithme suivant nous présente une méthode de fusion de données réalisée sur les images profondeur et couleur transformées par ondelettes. Cette dernière transformation nous permet de réduire les données considérablement à traiter en tenant compte des détails contenus dans l'image. En effet, il existe différentes méthodes d'extraction de paramètres, et plusieurs types de fusion de données. Nous effectuons nos expériences sur plusieurs types de fusion et nous gardons le type qui convient le mieux à notre étude.

3 Extraction des paramètres et classification basée sur LDA/EFM

Il existe des différents filtres de **Gabor** pour diverses résolutions et orientations. Celui qui nous donne le meilleur score est maintenu. La classification est réalisée par la norme Euclidienne L_1 . Nous calculons les paramètres de performances du **SRV** pour différentes paramètres caractéristiques (N_p variable).

4 Résultats

Tableau 6.8 Tableau des résultats de fusion couleur RVB+profondeur

Approche	Combinaison		
Types de fusion	Fusion des caractéristiques		
Méthodes de fusion	Mean	Max	Min
Méthodes de réduction	LDA	EFM	EFM
N_p	10	40	70
Distance	L1		
TR(%)	54.76	75.75	75.25

Les résultats obtenus avec la classification par **LDA** ne sont pas satisfaisants. Car le meilleur **TEE** est de l'ordre de **45.24%** avec un taux de réussite **TR=54.76%** en utilisant seulement **10 paramètres** de **LDA**. Après le choix du filtre optimal ainsi que la mesure de distance, nous calculons les taux d'erreur et les taux de réussite dans le cas où nous faisons une projection sur l'espace **LDA/EFM** du vecteur caractéristique. Nous appliquons la fusion de données par trois méthodes (**max**, **min**, **mean**). La méthode de fusion par la moyenne (**mean**) donne un bon résultat par rapport aux méthodes de fusion par **max** et **min**. Les meilleures performances du système se trouvent dans l'intervalle de N_p entre **[70,100]**. La méthode **EFM** est une fois de plus la meilleure réduction dans le SRV. Pour N_p (**EFM**) entre **[10,70]** nous remarquons que les différences entre les erreurs dans les trois ensembles (**max**, **min**, **mean**) sont faibles. Ceci est une caractéristique très importante, car on peut dire que le système de reconnaissance de visages est stable. D'après ces résultats nous pouvons dire que la méthode de fusion par la moyenne (**mean**) donne le meilleur taux de réussite **75.75%** avec le nombre de paramètres égal à $N_p = 40$. Et la fusion par **min** donne un taux égal à **75.25%** avec un nombre de paramètres $N_p = 70$. Malgré que le taux d'erreur reste élevé pour dire que notre système est stable, mais nous pouvons affirmer que c'est un résultat qui nous encourage à faire davantage d'investigations.

6.6.2 Fusion des scores par combinaison (RVB + profondeur)

Notre but est d'étudier la fusion des scores entre les images 3D couleur et 3D profondeur à partir de la base de données **CASIA 3D**, et de classifier des personnes particulières pour les identifier. Nous présentons les différentes étapes de la conception de notre système. Nous effectuons nos expériences sur plusieurs types de fusion et nous gardons le type qui convient à notre étude.

Notre programme principal se présente comme suit :

- 1) Prétraitement de l'image
 - Extraction de la couleur du visage 3D et sa projection sur 2D.
 - Génération des cartes de profondeur du visage 3D.
- 2) Projection des images sur l'espace **PCA** (visages 2D et visages 2D de profondeur).
- 3) Mesure des distances (pour visages 2D et visages 2.5D de profondeur): Euclidienne norme 1 (L_1), Euclidienne Norme 2 (L_2), **covariance**.
- 4) Fusion des scores.
- 5) Classification et décision.

6.6.2.1 Architecture du système de conception

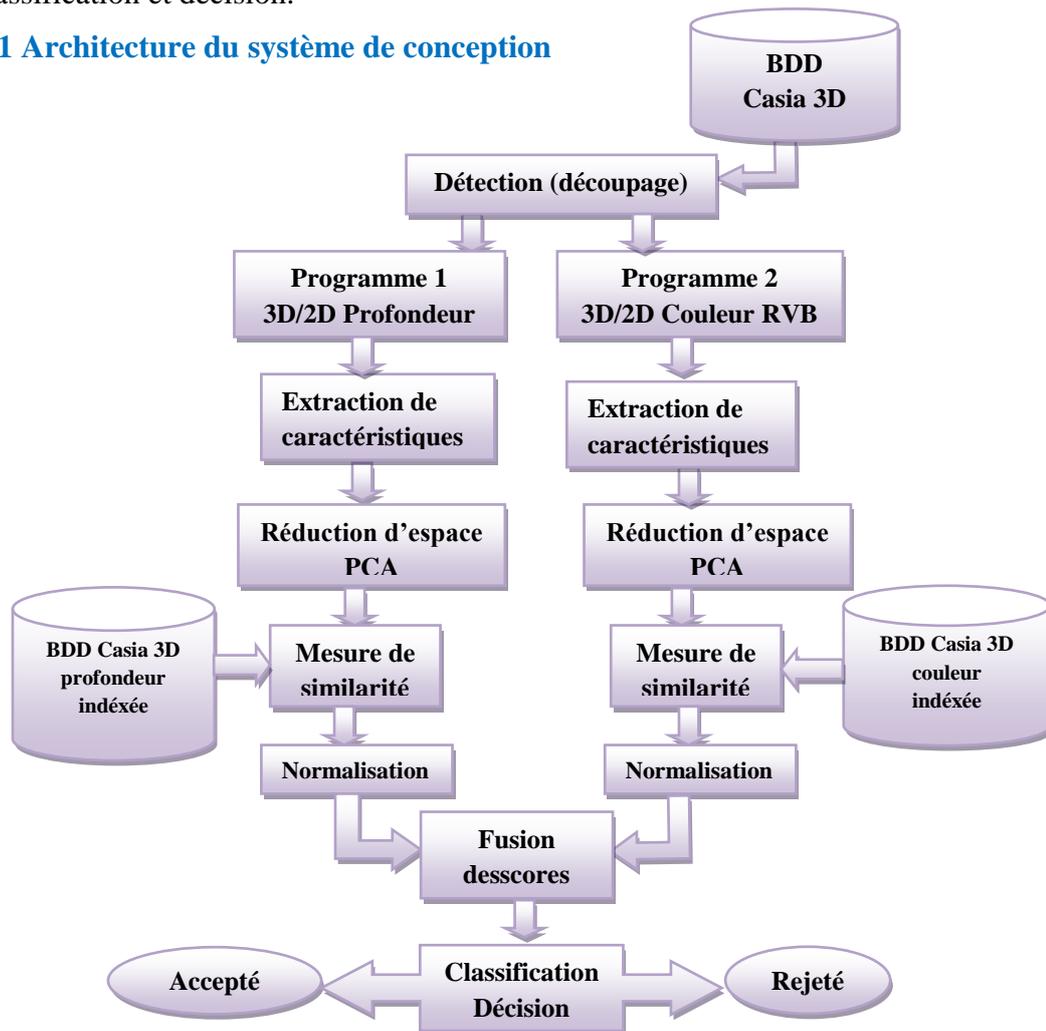


Figure 6.10 Architecture structurelle du système de conception.

Dans l'étape de fusion nous réalisons la fusion des scores, ces derniers sont représentés par les distances calculées précédemment, la méthode de fusion que nous utilisons est la fusion par combinaisons simples des scores basés sur les calculs de moyenne, max et min. Nous utilisons des instructions disponibles dans Matlab, cette instruction fusionne deux matrices ou deux vecteurs.

Dans un processus d'authentification, la stratégie de décision nous permet de choisir entre les deux alternatives suivantes: l'identité de l'utilisateur correspond bien à l'identité de la personne proclamée (ou recherchée) ou elle ne correspond pas. En comparant le modèle de l'utilisateur à celui qui existe dans la base de données. Cette phase est basée sur le seuil prédéfini. L'algorithme suivant explique le processus de la phase de décision :

Algorithme de processus de la phase de décision

Lire V_{test}
 $Y_1 = \text{prétraitement 1}(V_{test})$
 $Y_2 = \text{prétraitement 2}(V_{test})$
Extraction des paramètres ($Y_1|Y_2$)
 $DIS_V = \text{mesurer distance entre } XV, Y_1$
 $DIS_P = \text{mesurer distance entre } XP, Y_2$
 $DIS = \text{Fusion entre } DIS_V, DIS_P$
Seuillage
Si $DIS < \text{seuil}$
 Personne connue (identité parfaite)
 Sinon
 Personne inconnue (identité non parfaite)
Fin

V_{test} : image de visage test.(image de l'utilisateur)

$XV \setminus XP$: modèle (couleur,profondeur) de la personne proclamée stocké dans la base des données

DIS_V2D couleur : distance entre XV et Y_1 (Projetées dans le même sous espace)

DIS_V2D profondeur : distance entre XP et Y_2 (Projetées dans le même sous espace)

6.6.2.2 Résultats

Nous réalisons plusieurs types de fusion par combinaison des scores (mean, max, min).Nous étudions d'abord le cas où le nombre de paramètres $N_p(\text{PCA})$ varie de 20 à 200. Nous présentons la comparaison de TR entre les normes : L_1 , L_2 et covariance L_3 :

Tableau 6.9 TR de la fusion des scores (couleur RVB et profondeur) pour L_1 , L_2 et L_3

	Mean	Min	Max
L_1	79.50	78.85	74.74
L_2	65.79	78.75	63.33
L_3	76.45	76.75	67.48

D'après ces résultats nous pouvons dire que dans le cas de la norme L_1 la méthode de fusion des scores par moyenne pour les images de visages 2D couleur RVB et 2.5D profondeur donne le meilleur taux de précision 79.50% avec un nombre de paramètre égal à $N_p= 60$. Par contre la fusion par min donne un taux égal à 78.85% avec $N_p= 20$.

Malgré que le taux d'égale erreur reste élevé pour dire que notre système est stable. Mais nous pouvons affirmer que c'est un bon résultat dans le cas de la reconnaissance de visage en 3D.

Tableau 6.10 Tableau comparaison des résultats de la **fusion des caractéristiques** et des **scores** de la couleur RVB et profondeur.

Approche	Combinaison					
	Fusion des caractéristiques			Fusion des scores		
Types de fusion	Fusion des caractéristiques			Fusion des scores		
Méthodes de fusion	Mean	Max	Min	Mean	Max	Min
Méthodes de réduction	LDA	EFM	EFM	ACP		
Np	10	40	70	60	20	
Distance	L1			L1		
TR(%)	54.76	75.75	75.25	79.50	78.85	74.74

Ce tableau nous montre que la meilleure performance **SRV** est le cas de fusion des scores malgré une réduction d'espace primaire. Ceci est dû au fait que l'information traitée est prise sur l'image de visage directement sans aucun filtrage. En plus le traitement des scores est plus rapide que celui des caractéristiques vu la taille de données mises en jeu.

6.7 Fusion 2D couleur HSV et 3D profondeur

Dans cette partie du travail, nous tentons de résoudre ces problèmes restants dans la reconnaissance du visage et nous examinons comment l'information de profondeur et de couleur HSV contribue à la reconnaissance faciale lorsque la variation d'éclairage et l'expression sont prises en compte. La classification utilisée est basée sur **SVM** et ceci d'après les travaux de M.Belahcène en 2011[351].

Les principales contributions dans cette partie du travail sont les suivantes:

- Faire l'étude comparative des méthodes de réduction d'espace de données suivantes : L'Analyse Discriminante Linéaire (ADL), le Modèle Discriminant Linéaire Amélioré de Fisher (EFM) et la ADL Directe (ADLD) pour choisir le meilleur espace de projection des données de profondeur et de couleur HSV.
- Etudier la fusion au niveau des caractéristiques.
- Etudier la fusion au niveau de scores par: l'intégrale floue de **Sugeno**, de **Choquet** et les machines à vecteurs de support (**SVMs**).

6.7.1 Identification de visages à base de fusion par combinaison

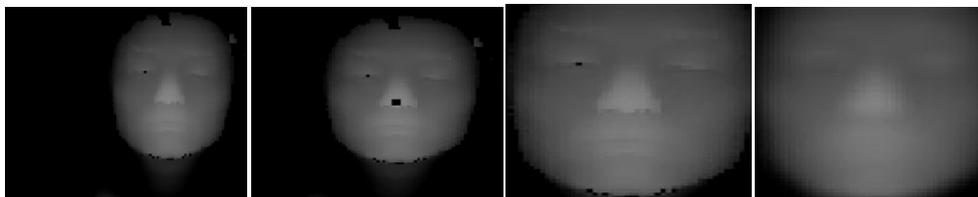
(couleur HSV+ profondeur)

Idéalement, un système de reconnaissance faciale doit pouvoir identifier des visages grâce à l'information profondeur et/ou couleur. Il est supposé dans ce travail que l'une des faces est

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

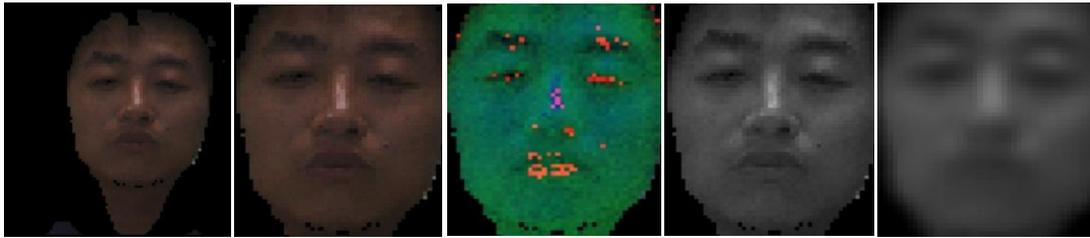
Les données **3D** sont converties en une image en profondeur (voir **Figure 6.11(a)**) et une image de couleur (voir **Figure 6.12(a)**). Dans la plupart des images, le nez est la plus proche partie du visage au scanner **3D**, c'est à dire, il a la valeur la plus élevée en profondeur entre tous les points du visage. En utilisant une fenêtre de taille **3x3** qui calcule la somme des valeurs de profondeur de ses pixels correspondants, le **nez** est détecté comme le point de coordonnées du pixel central de la fenêtre qui renvoie la valeur maximale (voir **6.11(b)**). Après avoir détecté le nez, toutes les images dans la base de données sont découpées par une fenêtre rectangulaire de taille **57x47** centrée autour du centre du nez (voir **Figure 6.11(c)**, **6.12(b)**). Pour l'image de couleur **RGB**, nous avons utilisé l'espace de couleur **HSV** (Hue, Saturation, Value) (voir **Figure 6.12(c)**) et nous considérons la composante **V** " Value " parce qu'elle est moins sensible à la variation d'éclairage (voir **Figure 6.12(d)**) [20].

Toutefois, en raison de la qualité des données originales en **3D**, les images de profondeur et de couleur que nous utilisons contiennent généralement beaucoup de bruit, tels que les trous et les valeurs aberrantes. Nous pouvons obtenir des images claires par les procédés suivants. Le prétraitement des images de profondeur et l'image de la composante **V** de l'espace de couleur **HSV** comprend la suppression du bruit et le trou de remplissage. Pour chaque pixel, la moyenne est calculée par la sous-fenêtre de taille **5 × 5** autour de lui. Le résultat est montré dans les **Figures 6.11 (d)**, **6.11(e)**.



(a) Image de profondeur (b) Détection du bout de nez (c) Image découpée (d) Image après suppression de bruit et remplissage de trous.

Figure 6.11 Prétraitement d'image de profondeur



(a) Image de couleur (b) Image découpée (c) Image en système HSV (d) Composante V de HSV (e) Image après suppression de bruit et remplissage de trous.

Figure 6.12 Prétraitement de l'image couleur 3D.

Bien que la distance euclidienne soit optimale dans la théorie, les diverses expériences ont montré que la distance euclidienne est surpassée par d'autres distances. Une d'elles est la corrélation normalisée (normalized correlation) équation 6.1.

6.7.2 Résultats

6.7.2.1 La base de données CASIA 3D

Nous utilisons la base de données des visages CASIA 3D [338] pour tester notre système d'identification proposé. La base est construite par un numériseur 3D sans contact **Minolta VIVID 910** travaillant sur le mode rapide. Cette base de données contient 123 sujets, chaque sujet ayant 37 ou 38 images avec les variations individuelles de poses, d'expression, d'illumination, de variations combinées des expressions sous illumination et de pose sous des expressions. Cette base de données contient des variations complexes qui sont difficiles à n'importe quel algorithme.

Dans notre travail, nous avons étudié les variations d'illuminations (**Figure 6.13**), expressions (**Figure 6.14**) et les variations combinées des expressions sous illumination. Nous avons donc utilisé 15 images pour chaque sujet. La base de données de **1845 images** est divisée en deux sous-ensembles, l'ensemble de **galerie** et l'ensemble de **test**. L'ensemble de galerie contient une image pour chaque sujet (sous la condition de **vue de face, l'éclairage des bureaux et l'expression neutre**).

L'ensemble de **test** est en outre divisé en **trois sous-ensembles**:

- **IV** (400 images): variations d'éclairage, y compris l'éclairage en haut, bas, gauche et à droite.
- **EV** (500 images): variations d'expression, y compris sourire, rire, la colère, la surprise et les yeux fermés.
- **EVI** (500 images): variations d'expression en vertu de l'éclairage du côté droit.

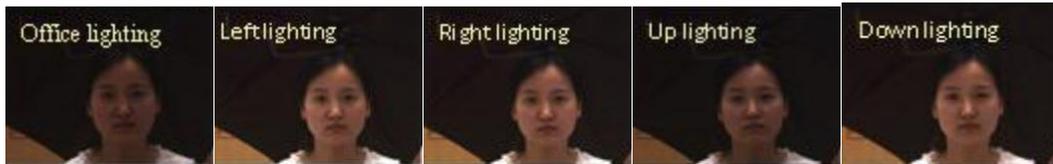


Figure 6.13 Illumination variations of the CASIA 3D Face Database.



(a) sourire (b) rire (c) colère (d) surprise (e) yeux fermés

Figure 6.14 Variations d'expression de la base de données des visages 3D CASIA

6.7.2.2 Comparaison des méthodes de réduction sur les images de profondeur et couleur HSV

Dans notre expérience, des images de profondeur et des images de la composante **V** de l'espace de couleur **HSV** sont utilisées pour caractériser les sujets. Les figures 6.15 et 6.16 présentent les performances des méthodes de réduction pour d'identification des images de profondeur et de couleur.

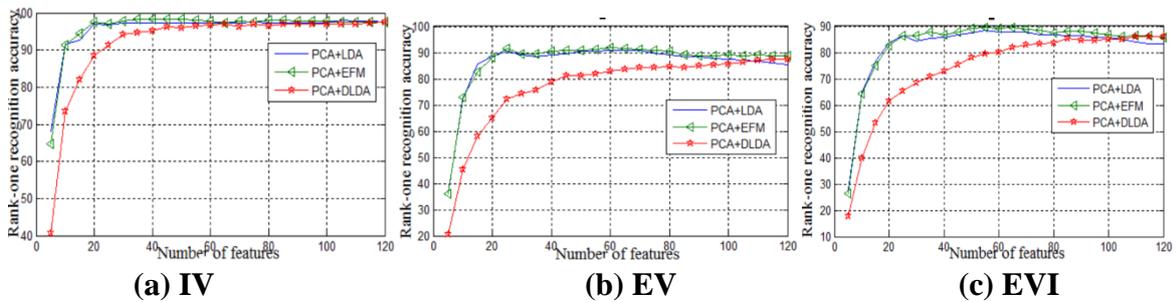


Figure 6.15 Performances des méthodes de réduction pour d'identification des images de profondeur.

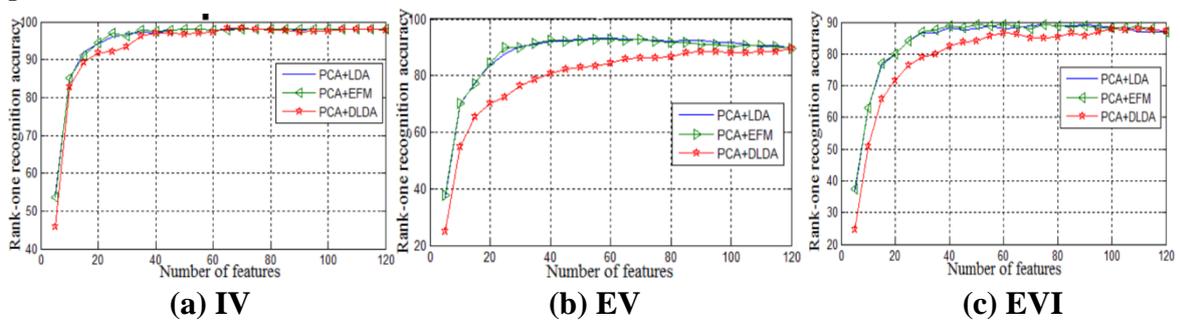


Figure 6.16 Performances des méthodes de réduction pour d'identification des images de couleur.

Nous pouvons constater dans les **figures 6.15** et **6.16** que la méthode **PCA+EFM** est mieux performante que la **PCA+LDA** suivie de la **PCA+DLDA** en employant un nombre restreint de caractéristiques dans notre base de données **CASIA 3D** pour les images de profondeur et de couleur. Le tableau 1 donne les meilleurs résultats pour la méthode de réduction d'espace **PCA+EFM**.

Tableau 6.11 Les meilleurs résultats pour la méthode de réduction d'espace PCA+EFM.

Sous-ensembles de test	Images de profondeur		Images en couleur	
	TR(%)	Nombre de caractéristiques	TR(%)	Nombre de caractéristiques
IV	98.3740	35	98.1707	70
EV	92.0325	60	93.0081	55
EVI	89.5935	55	89.1057	60

Le tableau montre que l'information de profondeur est presque équivalent à l'information de couleur, mais l'information de profondeur est moins sensible à la variation d'éclairage et l'information de couleur est plus performante pour des variations d'expression.

6.7.2.3 Fusion par combinaison au niveau caractéristiques (FeatureLevel)

Pour la **fusion** au niveau **caractéristiques** nous avons utilisé deux types :

- La fusion par la **concaténation** pour les deux vecteurs caractéristiques de profondeur et de couleur.
- La fusion des pixels de l'image de profondeur et de couleur par: la **moyenne**, le **produit**, le **minimum** et le **maximum**.

La **figure 6.17** présente les performances des méthodes de fusion au niveau caractéristiques pour l'identification multimodale des visages.

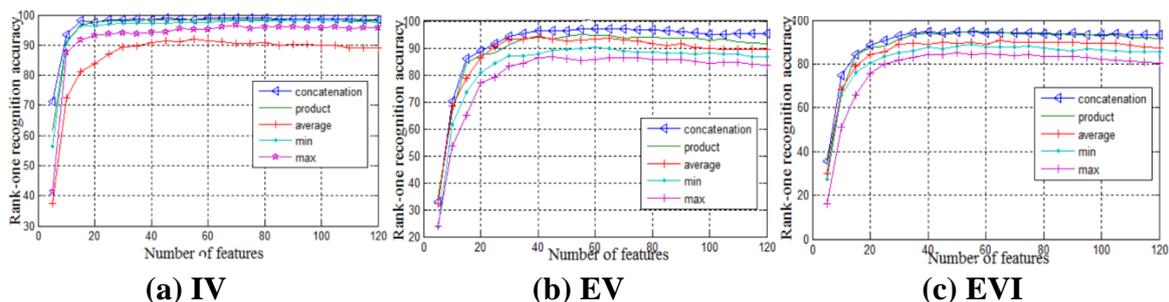


Figure 6.17 Performances des méthodes de fusion au niveau caractéristique pour l'identification multimodale des visages.

Nous avons remarqué dans la **figure 6.17** que la fusion par la concaténation et le produit ont donnés presque la même performance pour les trois sous-ensembles de test. Le **tableau 6.12** présente les meilleurs résultats.

Tableau 6.12 Les meilleurs résultats pour la fusion des caractéristiques de couleur **HSV** et **profondeur**

Sous-ensembles de test	Méthodes de fusion	TR(%)	Nombre de caractéristiques
IV	Concaténation	98.9837	60
EV	Concaténation	97.0732	60
EVI	Produit	94.9593	55

6.7.2.4 Fusion par combinaison et classification au niveau scores (Score Level)

La fusion au niveau des scores est le type de fusion le plus utilisé car elle peut être appliquée à tous les types de systèmes (contrairement à la fusion des caractéristiques), dans un espace de dimension, avec des méthodes relativement simples et efficaces mais traitant plus d'information que la fusion de décisions. La fusion de scores consiste donc à la classification pour la décision finale.

Il existe deux approches pour la fusion des scores obtenus par différents systèmes :

- i) La première approche est de traiter le sujet comme un problème de combinaison,
- ii) tandis que l'autre approche est de voir cela comme un problème de classification.

Pour la première approche nous avons utilisé la **logique floue** à base de l'intégrale de **Sugeno** et de **Choquet**. Pour la deuxième approche nous avons utilisé une machine à vecteurs de support (**SVM**). Nous utilisons une machine à vecteurs de support (**SVM**) avec un noyau **RBF** (Radial Basis Function ou fonction à base radiale). Le **SVM** a été implémenté en utilisant la librairie **LibSVM** disponible gratuitement au site (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>). Le **SVM** est un séparateur à **deux classes** mais dans notre problème on a **123 classes**. Nous avons donc utilisé **123 SVMs**.

Le tableau **6.13** présente les performances de combinaison et classification des images de profondeur et de couleur. Afin de s'assurer que la fusion de scores provenant de différents systèmes soit cohérente, les scores doivent d'abord être transformés dans un domaine commun : on parle alors de normalisation de scores, la méthode de normalisation utilisée dans notre cas est la méthode **min-max** donnée par :

$$s'_{ik} = \frac{s_{ik} - \min(\{s_i\})}{\max(\{s_i\}) - \min(\{s_i\})} \tag{6.3}$$

Avec:

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

(N est le nombre de sujets et égale 123).

Tableau 6.13 Taux de reconnaissance pour la combinaison et la classification des scores de couleur HSV et profondeur.

Sous-ensembles de test	Méthodes de fusion		
	Intégrale de Sugeno	Intégrale de Choquet	SVMs
IV	98.98	98.98	100
EV	95.93	95.77	99.72
EVI	93.33	94.1463	99.18

Le **tableau 6.13** montre que la fusion par combinaisons des scores pour les deux intégrales floue de **Sugeno** et **Choquet** donne des résultats équivalents à ceux de la fusion au niveau des caractéristiques.

La fusion par classification de scores en utilisant les **SVMs** associé à l'**ACP+EFM** donne des résultats presque parfaits appliqué à la **BDD Casia 3D** (version 4). Nous pouvons dire que notre système multimodal d'identification des visages 3D (IV, EV, EVI) utilisant la composante **V** du système **HVS** de couleur conjointement à la profondeur est pratiquement insensible à la variation d'éclairage et aux expressions faciales. Une fois de plus nous confirmons que le **SVM** est un outil efficace dans toutes les applications : en détection (dans nos travaux réalisés en 2011, classification pour l'identification monomodale (sections **5.5** (2^{ème} expérience) et **5.7 chapitre 5**, fusion multisystèmes et dans la fusion des scores par classification. Alors que dans le cas de l'identification des visages 2D neutres en monomodale (visage) et en multibiométries (voix, visage) (voir **tableau 6.4** section **6.4.1.1 chapitre 6**) l'intégrale de **Sugeno** nous donne les meilleurs résultats classification.

6.7.2.5 Tableau comparatif et discussion

Tableau comparatif des fusions des caractéristiques et des scores de la couleur HSV et profondeur.

Tableau 6.14 Tableau comparatif des TR des différentes approches et méthodes de fusion HSV et profondeur.

Approche	Combinaison			Classification		
	Fusion des caractéristiques	Fusion des scores		Fusion des scores		
Types de fusion		Choquet	Sugeno	SVM		
Méthodes de fusion	Concaténation					
Sous ensemble de test	IV	IV	IV	IV	EV	EVI
TR(%)	98.9837	98.98	98.98	100	99.72	99.18

Dans la **fusion** couleur **HSV** et **profondeur** l'approche par classification **SVM** est meilleure que celle combinaison ce qui n'était pas le cas dans les fusions multisystèmes et multibiométries (fusion par combinaison : intégrale floue **de Sugeno** et classification par SVM). Dans toutes les applications étudiées la fusion des scores donne les meilleurs résultats de performance du **SRV**. Le système **HSV** est plus informant que le système **RVB** avec un **TR=79.50%** (pour la fusion **mean**, distance Euclidienne norme **L₁** et **N_p=60**) contre **TR=100%** en système HSV composante V (pour la fusion **SVM**, distance **corrélacion normalisée** et **N_p<70**).

La **fusion des scores** couleur **HSV** et **profondeur** par **classification SVM** est parfaite avec un taux de réussite à 100% dans des visages exposés à la variation d'illumination. Mais elle reste sensible aux changements d'expressions faciales et à la combinaison changement d'expression faciales et lumière.

Nous pouvons dire que :

- l'algorithme **SVM** est très efficace quel que soit l'application envisagée.
- l'algorithme **SVM** est aussi bon en phase détection, qu'en phase classification aussi bien qu'en phase en fusion.
- La meilleure méthode de réduction reste l'ACP+EFM.
- Le filtre de **Gabor** s'il est bien utilisé nous permet d'extraire l'essentiel des caractéristiques globales du visage.
- Les **phases du filtre** sont mieux caractéristiques du visage que les amplitudes.
- Le système **HSV** est plus informant que le système **RVB**
- La fusion des scores est plus efficace que celle des caractéristiques.
- La norme **L₁** et la **corrélacion normalisée** sont les meilleures dans la classification par mesure de similarité.
- La classification par **SVM** est meilleure que celle par mesure de distances c'est plus souple, rapide et nous évite le calcul de seuil.

6.8 Synthèse sur les SRVs multimodaux

La synthèse des résultats obtenus dans cette partie du travail est regroupée dans le tableau

6.15.

Tableau 6.15 Synthèse des résultats des SRVs multimodaux étudiés.

Types SRV	Approches fusion	Types de fusion	Méthodes de fusion	Normalisation	Classification	N _p	BDD	TR(%)
Multisystèmes	Combinaison	scores	Intégrale de Sugeno	Double sigmoïde	Corrélation normalisée	80	XM2VTS	98.36 ±0.75
	Classification	scores	MLP SVM	QLQ Znrm	Corrélation normalisée	<100	XM2VTS	97.84±0.77 97.44±0.77
Multibiométries		Voix , Visage (LFCC,GMM), (DCTs, GMM)	Intégrale de Sugeno	Double sigmoïde	---	---	XM2VTS des scores	99.99±0.04
Multialgorithmes		Caractéristiques +scores	Min	minmax	L ₂	100	XM2VTS	76.75
3D RVB+profondeur	Combinaison	Caractéristiques	Max	minmax	L ₁	40	CASIA3D	75.75
	Combinaison	Scores	Mean	minmax	L ₁	60	CASIA3D	79.50
3D HSV+profondeur	Combinaison	Caractéristiques	Concaténation	minmax	Corrélation normalisée	60	CASIA3D IV	98.9837
		Scores	I.Choquet I.Sugeno	minmax	Corrélation normalisée		CASIA3D IV	98.98 98.98
	Combinaison	Scores	SVM	minmax	Corrélation normalisée		CASIA3D IV	100

Les meilleurs SRV sont à base :

- d'une extraction de caractéristiques par filtre de **Gabor** (pour les SRV multisystèmes , multibiométries et multi algorithmes conçus),
- d'une réduction d'espace **ACP+EFM**,
- la classification par **corrélation normalisée** et **SVM**,
- la fusion par combinaison des scores (intégrale floue de **Sugeno** associée à une normalisation **double sigmoïde**) dans le cas du SRV multibiométries et multisystèmes,
- la fusion par *classification des scores* **SVM** en SRV multisystèmes et multimodales (3D couleur HSV+profondeur).

6.9 Conclusion

D'après l'étude faite et toutes les expériences réalisées nous pouvons dire que la fusion des scores 2D (couleur RVB) et 3D (profondeur) avec les métriques améliore certes les performances du système SRV monomodale. La norme L_1 est la meilleure pour le calcul de scores par rapport à la norme L_2 et la covariance L_3 . La fusion des scores par la **moyenne** est la méthode qui a donné les meilleurs résultats dans ce cas. Ceci dit une étude plus poussée sur les distances reste envisageable. Des améliorations au niveau du prétraitement sont aussi à tenir en compte. Ainsi qu'une meilleure méthode de réduction. D'après les résultats des expériences réalisées nous pouvons dire que l'ACP+EFM reste une méthode de discrimination performante d'une façon générale. La **fusion des scores** reste une méthode très efficace car sans la fusion nous avons avec un $N_p(\text{EFM}) = 100$; un $\text{TR} = 92.25\%$ pour les visages pour une fusion par combinaison **min** et la distance L_2 . Nous pouvons dire que cette dernière appliquée à des images de visages **2D neutres** de la base de données **XM2VTS** en *multi algorithmes* et en *multi biométries* (voix, visage) et affirmons que cette méthode est performante avec un $\text{TR}=99.99\%$ en **2D** et *multi biométrie*. Nous confirmons aussi que cette méthode reste une bonne candidate pour aborder les visages à 3D. Des améliorations restent à réaliser surtout au niveau de la phase prétraitement et aussi et surtout dans la phase de classification car notre utilisation à la mesure métrique reste un choix guidé par sa simplicité. Ceci dit d'autres normes comme Manhattan ou mesures géodésiques restent envisageable, et toujours dans l'objectif de rehausser ces paramètres de performances du système biométriques. Les résultats sont dans l'ensemble satisfaisants, mais le taux de faux rejet reste moyen. Ceci est sûrement dû à l'étape de prétraitement qui reste à améliorer ainsi que le choix d'un meilleur classifieur. Dans l'expérience de fusion 2D et 3D basée sur SVM nous utilisons la profondeur et la couleur des images pour construire un classificateur robuste pour l'identification du visage. Puisque la dimensionnalité des caractéristiques d'images de profondeur et de couleur est très élevée. Nous avons étudié trois méthodes de réduction d'espace de données. Puis nous avons fait des études comparatives sur la fusion de l'information de profondeur et de couleur pour les deux niveaux : caractéristiques et scores pour la construction d'un classifieur efficace. En analysant nos résultats expérimentaux dans la base de données du visage **CASIA 3D** avec des variations complexes, nous illustrons la performance prometteuse du schéma proposé et d'en tirer l'importante conclusion suivante:

- La meilleure méthode de réduction d'espace est **PCA+EFM** pour toutes nos applications.

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

Nous proposons d'utiliser les filtres de Gabor sur l'information de profondeur et de couleur et étudier d'autres stratégies de fusion.