



N° d'ordre : 2012-61

Thèse

Présentée par

Wael Ben Soltana

Pour obtenir le grade de Docteur de l'Ecole Centrale de Lyon

Spécialité : Informatique

11 Décembre 2012

Optimisation de stratégies de fusion pour la reconnaissance de visages 3D

COMPOSITION DU JURY

Pr. Bernadette DORIZZI	Institut Télécom Sud-Paris	Rapporteur
Pr. Michèle ROMBAUT	Université Joseph Fourier Grenoble	Rapporteur
Pr. Christophe GARCIA	Institut National des Sciences Appliquées de Lyon	Examineur
Dr. Boulbaba BEN AMOR	Institut Mines-Télécom/TELECOM Lille1	Examineur
Dr. Sami ROMDHANI	Safran Morpho	Examineur
Pr. Liming CHEN	Ecole Centrale de Lyon	Directeur de thèse
Pr. Chokri BEN AMAR	Ecole Nationale d'Ingénieurs de Sfax	Directeur de thèse
Dr. Mohsen ARDABILIAN	Ecole Centrale de Lyon	Co-encadreur de thèse

Remerciements

Je tiens à remercier Liming Chen, professeur à l'école centrale de Lyon, Chokri Ben Amar, professeur à l'école nationale d'ingénieurs de Sfax, pour avoir dirigé mes travaux et permis la réalisation de cette thèse. J'adresse également mes remerciements à Monsieur Mohsen Ardabilian, maître de conférences à l'école centrale de Lyon, pour ses précieux conseils qui m'ont grandement aidé dans mes travaux, son aide, sa disponibilité et sa gentillesse en toute circonstance.

Je veux aussi témoigner de ma gratitude à Mme Colette Vial et Mme Isabelle Dominique qui font partie du personnel du laboratoire LIRIS à Ecully. Je remercie également toute l'équipe d'enseignement en informatique du département Math-Info de l'école centrale de Lyon avec qui j'ai eu plaisir de travailler.

J'adresse évidemment un grand merci aux doctorants et aux stagiaires du laboratoire que j'ai côtoyés et qui se sont tous révélés être des collègues très sympathiques.

J'adresse également des remerciements tout particuliers aux rapporteurs Bernadette Dorizzi, professeur à l'Institut Télécom Sud-Paris, et, Michèle Rombaut, professeur à Gipsa-lab, qui ont pris le temps de lire et évaluer mes travaux ainsi que pour leurs remarques judicieuses qui m'ont permis d'améliorer ce manuscrit. Je remercie enfin l'ensemble des membres du jury pour l'intérêt qu'ils ont porté à mes travaux.

Je terminerai en remerciant ma famille plus particulièrement mes parents Jamel et Chadia, ma chère épouse Yosra, mon frère Majdi, ma sœur Fatma et mes amis tunisiens qui m'ont toujours soutenu.

Table des Matières

1. Introduction	20
1.1 Contexte et Motivation	20
1.2 Problématique et approche proposée	21
1.3 Contributions	23
1.4 Organisation de la thèse.....	24
2. Etat de l'art.....	28
2.1 Scénarios applicatifs et mesures de performance	28
2.2 Représentation Faciale.....	30
2.2.1 Représentation 2D de l'image de visage	31
2.2.1.1 Image d'intensité et les méthodes sous-espace	32
2.2.1.2 Les filtres de Gabor	32
2.2.1.3 Local Binary Pattern (LBP).....	34
2.2.2 Représentation de la forme géométrique 3D.....	35
2.2.2.1 Nuage de Points 3D.....	36
2.2.2.2 Image de profondeur	37
2.2.2.3 Les descripteurs de la Courbure	38
2.2.2.4 Vecteurs normaux.....	38
2.2.3 Conclusion	39
2.3 Stratégies de fusion.....	39
2.3.1 Fusion Avant Appariement	42
2.3.1.1 Fusion au niveau des données des capteurs	42
2.3.1.2 Fusion au niveau des caractéristiques	43
2.3.1.3 Conclusion.....	45
2.3.2 Fusion Après Appariement	45
2.3.2.1 Fusion parallèle.....	45
2.3.2.1.1 Fusion au niveau des scores	46
2.3.2.1.2 Fusion au niveau des rangs.....	50

2.3.2.1.3	Fusion au niveau des décisions.....	51
2.3.2.2	Fusion en cascade	51
2.3.2.3	Fusion multi-niveaux.....	52
2.3.2.4	Conclusion.....	54
2.4	Conclusion générale	54
3.	<i>Reconnaissance de visages par une méthode de fusion adaptative au niveau des scores.....</i>	58
3.1	Vue d'ensemble de l'approche proposée	59
3.2	Prétraitement des visages	60
3.3	Les descripteurs du visage	62
3.3.1	Les descripteurs de texture 2D	62
3.3.2	Les descripteurs de forme 3D.....	62
3.4	Classifieurs	64
3.5	La fusion adaptative	65
3.5.1	Normalisation des scores.....	66
3.5.2	Méthode de fusion proposée	66
3.5.3	Variantes de la Méthode de fusion adaptative	67
3.6	Les résultats expérimentaux	68
3.6.1	Protocole d'expérimentation	68
3.6.2	Analyse des performances individuelles des descripteurs	70
3.6.3	Résultats expérimentaux de la méthode de fusion des scores	72
3.7	Conclusion.....	75
4.	<i>Reconnaissance de visages par une méthode heuristique de fusion au niveau des scores.....</i>	78
4.1	Mélange d'experts gardés (Mixture of Gated Experts)	79
4.2	Optimisation du mélange d'experts gardés par l'algorithme du recuit simulé.....	79
4.2.1	Sélection dynamique des experts.....	80

4.2.2	Application de l'algorithme du recuit simulé à l'optimisation du mélange d'experts gardés dans un système de reconnaissance de visages 3D.....	82
4.3	Les mesures de similarité (les experts)	82
4.4	Protocole Expérimental.....	83
4.4.1	Base de visages	83
4.4.2	Les scénarios d'expérimentation.....	84
4.4.3	Détermination des paramètres de l'algorithme RS (température initiale (T) et la fonction de refroidissement (α)).....	84
4.5	Résultats Expérimentaux.....	85
4.5.1	Expérience 1: la règle de somme simple (SOM)	86
4.5.2	Expérience 2: la règle de somme pondérée utilisant le poids de la décision de l'expert (SPA_PDE)	87
4.5.3	Expérience 3 : la règle de somme pondérée utilisant le poids de la qualité de l'expert (SPA_PQE)	89
4.5.4	Expérience 4: Schéma de pondération adaptative (SPA)	91
4.6	Analyse des performances	92
4.7	Conclusion.....	93
5.	<i>Reconnaissance de visages par une méthode heuristique de fusion au niveau des caractéristiques et des scores</i>	97
5.1	La fusion multi-niveaux	98
5.1.1	Principe de la fusion multi-niveaux	98
5.1.2	Espace des stratégies de fusion.....	99
5.2	Recherche heuristique d'une stratégie de fusion par l'algorithme génétique.....	100
5.2.1	Principe général de l'algorithme génétique	100
5.2.2	Application à la reconnaissance de visages 3D.....	101
5.2.2.1	Encodage d'une stratégie de fusion.....	101
5.2.2.2	Sélection, Croisement et Mutation.....	103
5.2.2.2.1	Sélection des chromosomes.....	103
5.2.2.2.2	Croisement uniforme	104
5.2.2.2.3	Mutation.....	104

5.2.2.3	Critère de qualité (fitness)	105
5.3	Descripteurs et classifieurs	105
5.3.1	Les descripteurs	105
5.3.2	Les classifieurs	106
5.4	Résultats expérimentaux.....	106
5.4.1	Protocole expérimental	106
5.4.2	Analyse des résultats expérimentaux.....	107
5.5	Conclusion.....	110
6.	<i>Conclusion générale</i>	113
6.1	Contribution.....	113
6.2	Perspectives	115
7.	<i>Bibliographie</i>.....	118

Table des figures

Figure 1. Des exemples de modalités utilisées dans la biométrie : (a) l’empreinte digitale, (b) l’iris, (c) l’oreille, (d) la main, (e) les veines de la main	20
Figure 2. Modalité Faciale : (a) Image 3D de Visage, (b) Images avec différentes intensités de luminosité (c) Images avec différentes expressions	21
Figure 3. Modalité Faciale : (a) Image 3D de Visage, (b) Image de couleur	21
Figure 4. Vue d'ensemble d'un système de reconnaissance de visages (Identification et vérification) avec les deux phases de fonctionnement et d’enrôlement	28
Figure 5. Exemple de taux d’égale erreur pour trois systèmes de reconnaissance de visages S1, S2 et S3	30
Figure 6. Les différents formats représentatifs du modèle du visage en 3D (de gauche à droite : l’image de texture superposée sur la forme du visage, l’image de profondeur, le modèle ombré, le modèle filaire et le modèle de nuage de points)	31
Figure 7. Filtre de Gabor : (a) Représentation de Gabor (b) Exemple de traitement	34
Figure 8. Exemple de traitement de l’opérateur LBP [Ahonen et al. 2004].....	35
Figure 9. Représentation nuage de points pour un visage 3D	36
Figure 10. Image de profondeur d'un visage 3D	37
Figure 11. Images LBP générées à partir d’une image de profondeur	37
Figure 12. Surfaces Normales d'un Visage 3D	39
Figure 13. Schémas de fusion précoce (a) et tardive (b).....	41
Figure 14. Les niveaux de fusion	42
Figure 15. Acquisition des visages.....	43
Figure 16. Un exemple de visages avec les points d’intérêts détectés.....	43
Figure 17. Un exemple de profil et sa représentation avec les points 3D	44
Figure 18. Vue d’ensemble de l’approche proposée	59
Figure 19. Un exemple de visage prétraité avec ses deux images de profondeur (Gauche) et de texture (Droite) en présence de 15 points annotés manuellement.....	60
Figure 20. L’alignement, R-ICP (a) les régions rigides de deux échantillons de visage, (b) Superposition des deux modèles de textures 3D (c) une étape d'alignement globale dont chaque modèle 3D de visage est représenté par un couleur différent (d) une étape d'alignement fin	61
Figure 21. Les points caractéristiques du Visage 3D	63
Figure 22. Exemple d’images de texture : la colonne (a) les images avec expression neutre, les colonnes (b), (c) les images en présence d’expressions	69

Figure 23. Algorithme du recuit simulé	81
Figure 24. Les différents TERs des deux bases (a) B1 et (b) B2	89
Figure 25. Comparaison de performances de différentes méthodes de fusion	93
Figure 26. Utilisation de l'algorithme génétique pour la sélection des descripteurs (F_1, F_2, \dots, F_N), des classifieurs (D_1, D_2, \dots, D_r) et de leur stratégie de fusion	98
Figure 27. Aperçu général de l'approche proposée avec les différentes étapes de prétraitement des visages, de l'extraction des descripteurs et du codage de la stratégie de fusion. Le schéma de fusion final est généré par l'algorithme génétique.	100
Figure 28. Exemple de stratégie de codage d'un chromosome avec deux parties : Partie A et Partie B. Partie A représente les descripteurs sélectionnés (1 au niveau de caractéristique, 0 au niveau de score et -1 inactif). Partie B modélise les nombres des descripteurs à concaténer parmi l'ensemble sélectionné.	102
Figure 29. Exemples de combinaisons représentées dans la partie P1 pour la concaténation des descripteurs sélectionnés dans la Partie A.....	102
Figure 30. Exemple de sélection SUS pour choisir 4 chromosomes parents	103
Figure 31. Exemple de croisement entre deux chromosomes parents pour la génération de nouveaux chromosomes fils	104
Figure 32. Exemple de mutation pour la génération d'un nouveau chromosome fils.....	105

Liste des tableaux

Tableau 1. Organisation de la base FRGC v.2.0	69
Tableau 2. Taux d'identification individuel avec la base FRGC v2.0 - Neutre vs Tous.....	70
Tableau 3. Taux d'identification individuel avec la base FRGC v2.0 - Neutre vs Neutre.....	71
Tableau 4. Taux d'identification individuel avec la base FRGC v2.0 - Neutre vs non Neutre...	72
Tableau 5. Taux d'identification dans le cas « Neutre vs Tous » avec trois méthodes de normalisation et différents schémas de fusion	73
Tableau 6. Taux d'identification des modalités 2D, 3D et leur fusion adaptative	74
Tableau 7. Comparaison avec d'autres méthodes de fusion dans le cas « Neutre vs Tous ».....	74
Tableau 8. Organisation de la base FRGC v.2.0	84
Tableau 9. Taux d'identification des différents experts avec les deux bases B1 et B2.....	85
Tableau 10. Les taux d'identification de l'expérience SOM (B1 : base d'apprentissage, B2 : base de test)	86
Tableau 11. Les taux d'identification de l'expérience SOM (B2 : base d'apprentissage, B1 : base de test)	87
Tableau 12. Les taux d'identification de l'expérience SPA_PDE (B1 : base d'apprentissage, B2 : base de test)	88
Tableau 13. Les taux d'identification de l'expérience SPA_PDE (B2 : base d'apprentissage, B1 : base de test)	88
Tableau 14. Les taux d'identification de l'expérience SPA_PQE (B1 : base d'apprentissage, B2 : base de test)	90
Tableau 15. Les taux d'identification de l'expérience SPA_PQE (B2 : base d'apprentissage, B1 : base de test)	90
Tableau 16. Les taux d'identification de l'expérience SPA (B1 : base d'apprentissage, B2 : base de test)	91
Tableau 17. Les taux d'identification de l'expérience SPA (B2 : base d'apprentissage, B1 : base de test)	92
Tableau 18. Organisation de la base FRGC v.2.0	107
Tableau 19. Les taux d'identification des classifieurs sélectionnés par l'algorithme génétique de la meilleure stratégie de fusion.....	108
Tableau 20. Taux d'identification avec l'expérience « Neutre vs. Tous ».....	108
Tableau 21. Taux d'identification avec les expériences « Neutre vs. Neutre (NN) », « Neutre vs. Non-neutres (N-nonN) ».....	109

Résumé :

La reconnaissance faciale (RF) est un domaine de recherche très actif en raison de ses nombreuses applications dans le domaine de la vision par ordinateur en général et en biométrie en particulier. Cet intérêt est motivé par plusieurs raisons. D'abord, le visage est universel. Ensuite, il est le moyen le plus naturel par les êtres humains de s'identifier les uns des autres. Enfin, le visage en tant que modalité biométrique est présente un caractère non intrusif, ce qui le distingue d'autres modalités biométriques comme l'iris ou l'emprunte digitale. La RF représente aussi des défis scientifiques importants. D'abord parce que tous les visages humains ont des configurations similaires. Ensuite, avec les images faciales 2D que l'on peut acquérir facilement, la variation intra-classe, due à des facteurs comme le changement de poses et de conditions d'éclairage, les variations d'expressions faciales, le vieillissement, est bien plus importante que la variation inter-classe.

Avec l'arrivée des systèmes d'acquisition 3D capables de capturer la profondeur d'objets, la reconnaissance faciale 3D (RF 3D) a émergé comme une voie prometteuse pour traiter les deux problèmes non résolus en 2D, à savoir les variations de pose et d'éclairage. En effet, les caméras 3D délivrent généralement les scans 3D de visages avec leurs images de texture alignées. Une solution en RF 3D peut donc tirer parti d'une fusion avisée d'informations de forme en 3D et celles de texture en 2D. En effet, étant donné que les scans 3D de visage offrent à la fois les surfaces faciales pour la modalité 3D pure et les images de texture 2D alignées, le nombre de possibilités de fusion pour optimiser le taux de reconnaissance est donc considérable. L'optimisation de stratégies de fusion pour une meilleure RF 3D est l'objectif principal de nos travaux de recherche menés dans cette thèse.

Dans l'état d'art, diverses stratégies de fusion ont été proposées pour la reconnaissance de visages 3D, allant de la fusion précoce "early fusion" opérant au niveau de caractéristiques à la fusion tardive "late fusion" sur les sorties de classifieurs, en passant par de nombreuses stratégies intermédiaires. Pour les stratégies de fusion tardive, nous distinguons encore des combinaisons en parallèle, en cascade ou multi-niveaux. Une exploration exhaustive d'un tel espace étant impossible, il faut donc recourir à des solutions heuristiques qui constituent nos démarches de base dans le cadre des travaux de cette thèse.

En plus, en s'inscrivant dans un cadre de systèmes biométriques, les critères d'optimalité des stratégies de fusion restent des questions primordiales. En effet, une stratégie de fusion est dite optimisée si elle est capable d'intégrer et de tirer parti des différentes modalités et, plus largement, des différentes informations extraites lors du processus de reconnaissance quelque soit leur niveau d'abstraction et, par conséquent, de difficulté.

Pour surmonter toutes ces difficultés et proposer une solution optimisée, notre démarche s'appuie d'une part sur l'apprentissage qui permet de qualifier sur des données d'entraînement les experts 2D ou 3D, selon des critères de performance comme ERR, et d'autre part l'utilisation de stratégie d'optimisation heuristique comme le recuit simulé qui permet d'optimiser les mélanges des experts à fusionner.

Dans un premier temps, nous avons étudié et comparé les caractéristiques 2D et 3D et avons constaté qu'aucune caractéristique 2D ou 3D est suffisamment discriminante pour l'application de l'identification de visages. En conséquence, nous avons proposé une méthode de fusion qui tient en compte des propriétés intra-classifieur et inter-classifieur. Cette méthode utilise les scores de similarité et se base sur une règle de somme pondérée avec deux types de poids. Un premier poids hors ligne, calculé lors d'une phase d'apprentissage, évalue la performance de chaque classifieur et définit la relation inter-classifieur. Un deuxième poids en ligne évalue la qualité de décision de chaque classifieur et définit la relation intra-classifieur. Il est calculé lors de la phase de reconnaissance en utilisant les différents scores de similarité. Des expériences menées sur la base FRGC v.2.0 ont montré l'efficacité de la méthode proposée.

Dans une deuxième étape, nous nous sommes intéressés à la présence d'un nombre important de classifieurs. Dans ce cas, la cardinalité de l'espace de recherche de la stratégie de fusion optimale est importante. Une recherche exhaustive de la meilleure stratégie de fusion dans cet espace est quasi-impossible. Ceci nous a conduit à recourir à des techniques d'optimisation heuristique, notamment au couplage de l'algorithme du recuit simulé avec la méthode de fusion précédemment décrite. Elle a été proposée afin de chercher une stratégie de fusion presque-optimale des classifieurs.

Une étude a été élaborée aussi concernant la stabilité de la génération des schémas de fusion dans le cas d'augmentation de la base d'apprentissage. Les résultats trouvés ont montré que l'utilisation de 60% de la base d'apprentissage permet de générer une stratégie de fusion stable et (presque)-optimisé indépendamment de la taille des données d'apprentissage.

Dans une troisième étape, nous avons analysé un cas plus général qui est la combinaison en présence de deux niveaux de fusion: fusion au niveau des scores et fusion au niveau des caractéristiques. Une nouvelle stratégie basée sur une telle combinaison a été développée. Elle utilise comme technique l'algorithme génétique avec un nouveau codage afin d'assurer deux aspects : la sélection des meilleurs caractéristiques et la recherche d'une stratégie de fusion presque-optimale. Notre nouvelle approche proposée ne contient pas seulement la stratégie de fusion pour combiner les caractéristiques les plus pertinentes, mais elle possède aussi la capacité étendue pour sélectionner les caractéristiques les plus discriminantes et leurs classifieurs appropriés. Des expériences sont menées sur la base de données FRGC v.2.0 et ont montré la performance de la méthode proposée. Notre méthode de fusion proposée a été intégrée, dans le

cadre des travaux de recherche menés au sein de notre laboratoire, dans un système de reconnaissance de visages utilisant de nouveaux descripteurs. Les performances de ce système ont été améliorées grâce à ce schéma de fusion.

Mots Clés: fusion des données, fusion précoce, fusion tardive, fusion hybride, qualité des experts, descripteurs, classifieurs, expression faciale, reconnaissance faciale 3D, optimisation, stratégies de fusion, méthode heuristique, recuit simulé, algorithme génétique, sélection des meilleurs caractéristiques, sélection des classifieurs, experts.

Abstract :

Face recognition (FR) was one of the motivations of computer vision for a long time, but only in recent years reliable automatic face recognition has become a realistic target of biometrics research. This interest is motivated by several reasons. First, the face is one of the most preferable biometrics for person identification and verification related applications, because it is natural, non-intrusive, and socially well accepted. The second reason relates to the challenges encountered in the FR domain, in which all human faces are similar to each other and hence offer low distinctiveness as compared with other biometrics, e.g., fingerprints and irises. Furthermore, when employing facial texture images, intra-class variations due to various factors as illumination and pose changes are usually greater than inter-class ones, preventing 2D face recognition systems from being completely reliable in real conditions.

Recent, 3D acquisition systems are capable to capture the shape information of objects. Thus, 3D face recognition (3D FR) has been extensively investigated by the research community to deal with the unsolved issues in 2D face recognition, i.e., illumination and pose changes. Indeed, 3D cameras generally deliver the 3D scans of faces with their aligned texture images. 3D FR can benefit from the fusion of 2D texture and 3D shape information.

This Ph.D thesis is dedicated to the optimization of fusion strategies based on three dimensional data. However, there are some problems. Indeed, since the 3D face scans provide both the facial surfaces for the 3D model and 2D texture images, the number of fusion method is high.

In the literature, many fusion strategies exist that have been proposed for 3D face recognition. We can roughly classify the fusion strategies into two categories: early fusion and late fusion. Some intermediate strategies such as serial fusion and multi-level fusion have been proposed as well. Meanwhile, the search for an optimal fusion scheme remains extraordinarily complex because the cardinality of the space of possible fusion strategies. It is exponentially proportional to the number of competing features and classifiers. Thus, we require fusion technique to efficiently manage all these features and classifiers that constitute our contribution in this work. In addition, the optimality criteria of fusion strategies remain critical issues. By definition, an optimal fusion strategy is able to integrate and take advantage from different data.

To overcome all these difficulties and propose an optimized solution, we adopted the following reflection. Firstly, we studied and compared the distinctiveness of popular features extracted from both the registered 2D face images and 3D face models. We show that none of these 2D texture and 3D shape related features is distinctive enough for reliable face recognition. We prove that existing 3D shape based features perform better in the case of

Neutral vs. Neutral than their counterpart 2D texture related ones, while they prove to be much more sensitive to facial expression variations. We also proposed an adaptive score level fusion based on weighted sum rule. The weight associated to each feature is a combination of an offline weight using the EER value on a learning dataset and an online one from the matching scores. We demonstrate that 2D and 3D features are complementary and their competitive fusion ameliorates overall performance. The new adaptive fusion strategy shows its effectiveness when fusing non discriminative 2D and 3D features. Experimental results based the FRGC v.2.0 database showed the effectiveness of the proposed method.

Secondly, we investigated the presence of a large number of classifiers. In this case, the cardinality of the search space of optimal fusion strategy is important. A comprehensive search of the best fusion strategy in this space is almost impossible. We present a mixture of gated experts, which is a generalized weighted sum rule, to fuse different classifiers for 3D face recognition. The novel fusion strategy consists of two weight-learning processes, both of which automatically select the most relevant weights of all the scores for each probe face in each modality. The first weight, expert quality weight, has the ability to handle biometric performance of each expert based on Equal Error Rates. The second expert decision weight is dynamically obtained according to the matching scores. To search the optimal fusing strategy in an exponential space of possible fusion schemes, we further propose a simulated annealing (SA) based stochastic search approach. We modify also the SA algorithm to compute automatically the expert quality weight of each expert. Moreover, we study the stability of the searched optimal mixture of gated experts as learning data is increased. The experimental results on the FRGC v2.0 database show the effectiveness of the proposed approach both in terms of performance and stability.

Finally, we consider the problem of feature selection and classifier fusion and discuss how they should be reflected in the fusion system architecture. We employed the genetic algorithm with a novel coding to search the worst performing fusion strategy. The proposed algorithm tunes itself between feature and matching score levels, and improves the final performance over the original on two levels, and as a fusion method, it not only contains a fusion strategy to combine the most relevant features so as to achieve adequate and optimized results, but also has the extensive ability to select the most discriminative features and their appropriate classifiers. Experiments are provided on the FRGC database and show that the proposed method produces significantly better results than the baseline fusion methods. Our proposed fusion method has been integrated in a face recognition system using new 3D face descriptors. The performances of this system have been improved by our fusion strategy.

Keywords: Data fusion, early fusion, late fusion, multi level fusion, fusion des données, fusion précoce, fusion tardive, fusion hybride, features, classifiers, expression, 3D face recognition, expert quality, optimization, fusion scheme, heuristic method, simulated annealing, genetic algorithm, feature selection, classifier selection, expert selection.

CHAPITRE I : INTRODUCTION

1. Introduction

1.1 Contexte et Motivation

La biométrie faciale a été l'objet d'une recherche intensive des communautés de recherche depuis maintenant plus de 40 ans. Cet intérêt est justifié par des multiples raisons. D'abord, le visage est universel et il est non intrusif car il est le moyen naturel pour les être humains de s'identifier les uns des autres, ce qui le distingue des autres modalités biométriques comme l'iris ou l'empreinte digitale (voir figure 1). Ensuite, les applications potentielles de la reconnaissance de visages (RF) par la machine sont nombreuses et diversifiées. Elles se définissent par la reconnaissance d'une personne particulière parmi un ensemble de sujets. Nous citons par exemple : le contrôle d'accès, la vidéo surveillance, l'indexation d'images et de vidéos, le commerce électronique, etc.

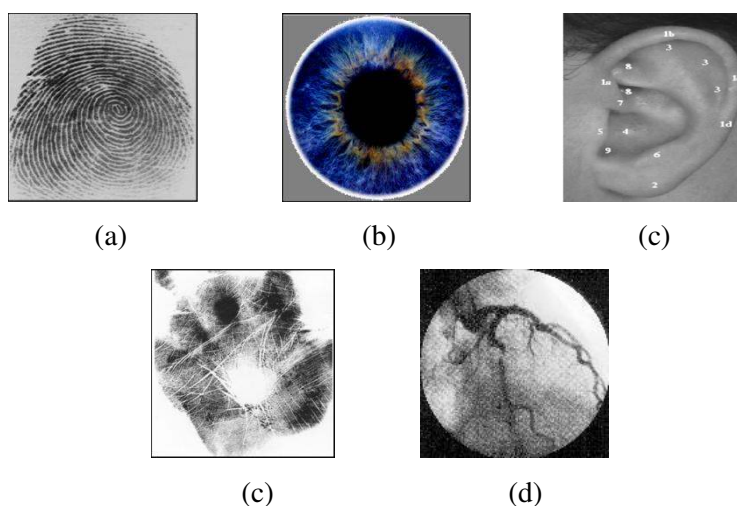


Figure 1. Des exemples de modalités utilisées dans la biométrie : (a) l'empreinte digitale, (b) l'iris, (c) l'oreille, (d) la main, (e) les veines de la main

Néanmoins, la RF représente aussi des défis scientifiques importants parce que les visages humains ont des caractéristiques similaires et l'effet de changement de poses, des conditions d'éclairage (voir figure 2), des variations des expressions faciales (voir figure 2) et de l'aspect vieillissement, etc. rendent la variation intra-classe souvent bien plus importante que la variation inter-classe [W. Zhao et al., 2003].

Avec l'arrivée de systèmes d'acquisition en 3D capables de capturer la profondeur d'objets, la reconnaissance faciale 3D (RF 3D) a émergé comme une voie prometteuse pour traiter deux problèmes non résolus en 2D, à savoir les variations de pose et d'éclairage [Bowyer et al., 2006]. En effet, la surface faciale capturée par une caméra 3D est théoriquement

insensible aux conditions d'éclairage alors que la pose peut être plus facilement normalisée grâce à la disponibilité d'informations géométriques 3D. Néanmoins, en capturant précisément la surface faciale en 3D, les acquisitions 3D se révèlent aussi plus sensibles aux expressions faciales que leur équivalent en 2D [Bowyer et al., 2006]. En conséquence, le développement d'une solution robuste aux variations des expressions faciales a été la principale cible de la majorité des travaux de la littérature.

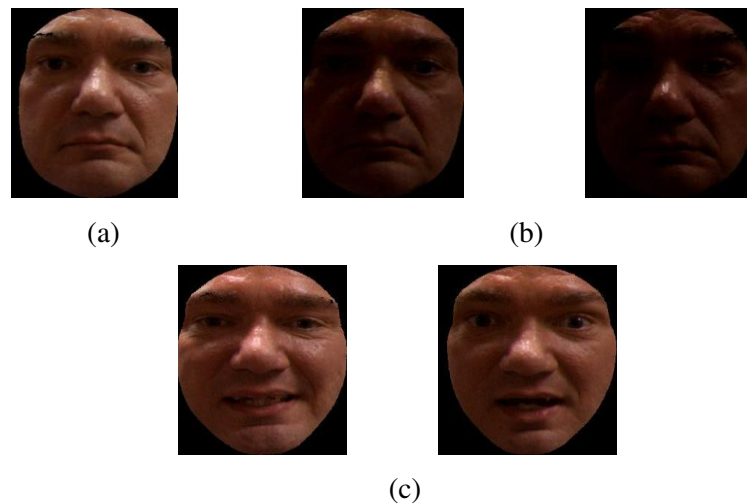


Figure 2. Modalité Faciale : (a) Image 3D de Visage, (b) Images avec différentes intensités de luminosité
(c) Images avec différentes expressions

Les travaux de cette thèse ont été déroulés dans un tel contexte. L'objectif ici est l'optimisation de stratégies de fusion pour une meilleure RF 3D dès lors que les caméras 3D délivrent généralement les modèles 3D de visages avec leur image de texture alignée (voir figure 3). Une solution en RF 3D peut et doit donc tirer parti d'une fusion avisée d'informations 3D et celles de texture 2D.



Figure 3. Modalité Faciale : (a) Image 3D de Visage, (b) Image de couleur

1.2 Problématique et approche proposée

Il existe dans la littérature un nombre important d'algorithmes en RF 3D tirant parti de différentes propriétés qui peuvent être extraites de la forme géométrique d'un visage en 3D. Ces

algorithmes s'appuient sur des représentations 3D de surfaces, e.g. les nuages de points, les images de profondeur, les images géométriques [Kakadiaris et al., 2007], les cartes LBP multi-échelles [Huang et al., 2011], les courbes faciales radiales [Samir et al., 2006][Drira et al.,2009], etc., et utilisent soit des approches de mise en correspondance holistiques (e.g. l'ICP), soit des approches locales basées sur des caractéristiques locales [Huang et al., 2010], ou encore des approches hybrides [Faltemier et al., 2008].

Les efforts de recherche intensifs de ces 30 dernières années portant sur la RF2D ont accumulé un nombre impressionnant de méthodes et techniques [Zhao et al.,2003] [Li & Jain 2005], que ce soit en représentations faciales comme les "Eigen face" [Turk & Pentland, 1997], ou les "Fischer face" [Belhumeur et al. 2000] ou encore en algorithmes de mise en correspondance [Brunelli & Poggio, 1993].

Etant donné que les acquisitions 3D de visage offrent à la fois les surfaces faciales pour la modalité 3D pure et les images de texture 2D alignées, le nombre de possibilités de fusion pour optimiser le taux de reconnaissance est donc considérable.

Dans l'état d'art, diverses stratégies de fusion ont été proposées [Lao et al., 2000], [Wang et al., 2002], [Papatheodorou & Rueckert, 2004], [Tsalakanidou et al., 2003], [Chang et al., 2003], [Hüsken et al., 2005], [Lu & Jain, 2005], [Maurer et al., 2005], [Gokberk et al.,2008] pour la reconnaissance de visages en 3D, allant de la fusion précoce "early fusion" opérant au niveau de caractéristiques à la fusion tardive "late fusion" opérant sur les sorties de classifieurs, en passant par de nombreuses stratégies intermédiaires. Pour les stratégies de fusion tardive, nous distinguons encore des combinaisons en parallèle, en cascade ou multi-niveaux. Le défi est de trouver la meilleure stratégie de fusion dans un espace de stratégies de fusion qui est exponentiel. En effet, si nous possédons N classifieurs qui utilisent par exemple chacun une représentation faciale spécifique, alors il existe au moins 2^N stratégies de fusion possibles simplement pour les stratégies de fusion tardive. Si par exemple, $N=12$, un cas étudié dans les travaux de cette thèse, le nombre de stratégies de fusion possible est alors $2^{12} = 4\ 096$. A cela s'ajoutent des stratégies précoces ou encore des stratégies de fusion intermédiaires incluant la présence simultanée des caractéristiques de bas niveau et des scores de similarité (ou distance) considérés comme des informations de haut niveau. Une exploration exhaustive d'un tel espace étant impossible, il faut donc recourir à des solutions heuristiques qui constituent nos démarches de base dans le cadre des travaux de cette thèse.

L'objectif des travaux menés dans le cadre de cette thèse étant l'étude de stratégies optimisées de fusion dans un contexte de reconnaissance de visages 3D, la première question est naturellement le critère d'optimalité. S'inscrivant dans un cadre de systèmes biométriques, les critères d'optimalité qui ont été utilisés sont naturellement les indicateurs de performance, à savoir le taux de reconnaissance, le taux de vérification, le taux de fausse acceptation, le taux de

faux rejet et le taux d'égalité des erreurs (Equal Error Rate ou ERR). Une stratégie de fusion est dite *optimisée* si elle est capable d'intégrer et de tirer parti des différentes modalités et, plus largement, des différentes informations extraites lors du processus de reconnaissance quelque soit leur niveau d'abstraction et de difficulté, et de produire des performances meilleures à d'autres stratégies de fusion dans un espace de stratégies de fusion à définir. La mise en œuvre de telles stratégies de fusion optimisées n'aboutira pas sans trouver des réponses à un certain nombre de questions : Comment modélise-t-on la combinaison des informations produites par des descripteurs et des classifieurs ? Comment exploite-t-on cette complémentarité pour notre problème de RF 3D ? En fonction de quels critères évalue-t-on cette complémentarité ? Comment procède-t-on pour trouver une stratégie de fusion quasi-optimale ? Pour répondre à tous ces problèmes, nous avons proposé trois stratégies de fusion.

1.3 Contributions

Concrètement, nous nous sommes intéressés à toutes les phases du processus de fusion : analyse de données pour la sélection de caractéristiques, apprentissage pour la classification et la fusion d'informations. Nous avons ainsi adopté le chemin de réflexion suivant : dans une première étape, et avant la phase de fusion, une étude a été élaborée sur l'étude des performances des caractéristiques extraites de l'information 2D et 3D du visage [Ben Soltana et al., 2010a], ainsi que leurs robustesses par rapport aux changements d'expressions. Nous avons remarqué qu'aucune caractéristique 2D ou 3D est suffisamment discriminante pour l'application de l'identification de visages. En conséquence, nous avons proposé une méthode de fusion qui tient en compte des propriétés intra-classifieur et inter-classifieur [Ben Soltana et al., 2010a], [Ben Amor et al., 2012]. Cette méthode utilise les scores de similarité et se base sur une règle de somme pondérée avec deux types de poids. Un premier poids hors ligne évalue la performance de chaque classifieur et définit la relation inter-classifieur. Il est calculé lors de la phase d'apprentissage. Un deuxième poids en ligne évalue la qualité de décision de chaque classifieur et définit la relation intra-classifieur. Il est calculé lors de la phase de reconnaissance utilisant les différents scores de similarité. Des expériences réalisées avec la base FRGC v.2.0 ont montré l'efficacité de la méthode proposée ce qui représente notre première contribution. Dans un deuxième temps, un multi-matcher a été développé en se basant sur la méthode de fusion précédemment décrite [Ben Soltana et al., 2012]. Il a été notamment testé en présence de trois niveaux de dégradations (le bruit, la décimation, les lacunes) des scans 3D de visage chacune à trois niveaux. Ici nous avons cherché à tester la robustesse du multi-matcher par rapport aux dégradations. Dans ce cas de figure, un indice de performance a été calculé suivant trois différents paramètres : le taux de reconnaissance, le taux de vérification et le taux d'égalité

d'erreur. Les résultats expérimentaux ont démontré que le multi-matcher est plus performant que chacune des méthodes prise individuellement dans les différents scénarios de dégradation. En plus, la baisse des performances du multi-matcher due aux diverses dégradations est moins importante que celles de chaque méthode prise séparément.

Dans une deuxième étape, nous nous sommes intéressés, en premier lieu, à la présence d'un nombre important de classifieurs. Dans ce cas, la cardinalité de l'espace des stratégies de fusion possible est exponentielle en regard au nombre de classifieurs en compétition. Une recherche exhaustive de la meilleure stratégie de fusion dans un tel espace est impossible. Ceci nous a incité à recourir à des techniques d'optimisation heuristique. Une approche, qui se base sur le couplage de l'algorithme du recuit simulé et de la méthode de fusion précédemment décrite, a été proposée afin de chercher une stratégie de fusion presque-optimale des classifieurs [Ben Soltana et al., 2011a] [Ben Soltana et al., 2011b]. Une étude a été élaborée aussi concernant la stabilité de la génération des schémas de fusion dans le cas d'augmentation de la base d'apprentissage. Les résultats trouvés ont montré que l'utilisation d'une partie de la base d'apprentissage permet de générer une stratégie de fusion stable. En plus, le fait d'ajouter des personnes n'affecte pas la performance totale de l'approche proposée. Cette nouvelle approche de fusion constitue notre deuxième contribution.

En troisième lieu, nous avons analysé un cas plus général qui est la combinaison de classifieurs en présence de deux niveaux de fusion: niveau des scores et niveau des caractéristiques. Une nouvelle stratégie [Ben Soltana et al., 2010b] [Ben Soltana et al., 2010c] envisageant cette présence a été développée. Elle utilise comme technique l'algorithme génétique avec un nouveau codage afin d'assurer deux aspects : la sélection des meilleurs caractéristiques et la recherche d'une stratégie de fusion presque-optimale. Notre approche proposée ne contient pas seulement la stratégie de fusion pour combiner les caractéristiques les plus pertinentes, mais a aussi la capacité étendue pour sélectionner les caractéristiques les plus discriminantes et leurs classifieurs appropriés. Des expériences ont été réalisées sur la base de données FRGC v.2.0 et ont montré la performance de la méthode proposée ce qui confirme l'apport de notre troisième contribution.

Notre méthode de fusion proposée a été intégrée, dans le cadre des travaux de recherche menés au sein de notre laboratoire, dans un système de reconnaissance de visages utilisant de nouveaux descripteurs de visages [Huang et al., 2011]. Les performances en terme de taux de reconnaissance et de taux de vérification de ce système ont été améliorées grâce à notre schéma de fusion.

1.4 Organisation de la thèse

Le manuscrit de thèse s'articule autour de cinq chapitres. Le chapitre 2 présente un état de l'art sur les différents descripteurs qui peuvent caractériser un visage 3D ainsi qu'une analyse des approches et techniques de fusion existantes.

Le chapitre 3 détaille le fruit de notre première contribution basée sur l'utilisation d'une nouvelle stratégie de fusion adaptative au niveau des scores. Dans le chapitre 4, nous proposons une nouvelle stratégie de fusion basée sur une nouvelle méthode heuristique de fusion au niveau des scores. Dans le chapitre 5, nous avons abordé le problème de sélection des caractéristiques et des classifieurs adéquats en proposant un schéma de combinaison en présence des niveaux de fusion au niveau des scores et des caractéristiques. Un algorithme génétique avec un nouveau schéma de codage a été proposé pour une meilleure sélection. Enfin, le chapitre 6 conclut ce travail en mettant en avant nos contributions et en présentant les différentes perspectives possibles.

CHAPITRE II : ETAT DE L'ART

2. Etat de l'art

Depuis maintenant plus de 40 ans, la reconnaissance faciale (RF) par la machine est un sujet de recherche très actif en vision par ordinateur en général et en biométrie en particulier. Cet intérêt toujours renouvelé par les communautés de recherche a été motivé par plusieurs raisons. D'abord parce que le visage en tant que biométrie est naturel, non-intrusif et sans contact, ceci le distingue d'autres biométries (e.g., l'iris, l'emprunte digitale, etc.). Ensuite, parce que les besoins applicatifs en matière de sécurité ne cessent de croître dans le monde d'aujourd'hui ; Enfin, parce que les défis scientifiques que représente la RF, e.g., le changement de poses et variations de conditions d'éclairage, les variations d'expressions faciales, le vieillissement, etc., ont été des sources de motivation pour introduire de nouveaux paradigmes, techniques et outils en vision par ordinateur et reconnaissance de forme.

2.1 Scénarios applicatifs et mesures de performance

Dans un système biométrique, e.g. visages, on distingue généralement deux principaux scénarios ou deux formes de fonctionnement: l'identification ou la vérification (ou l'authentification). La figure 4 montre une vue d'ensemble d'un système de reconnaissance de visages avec ces deux formes.

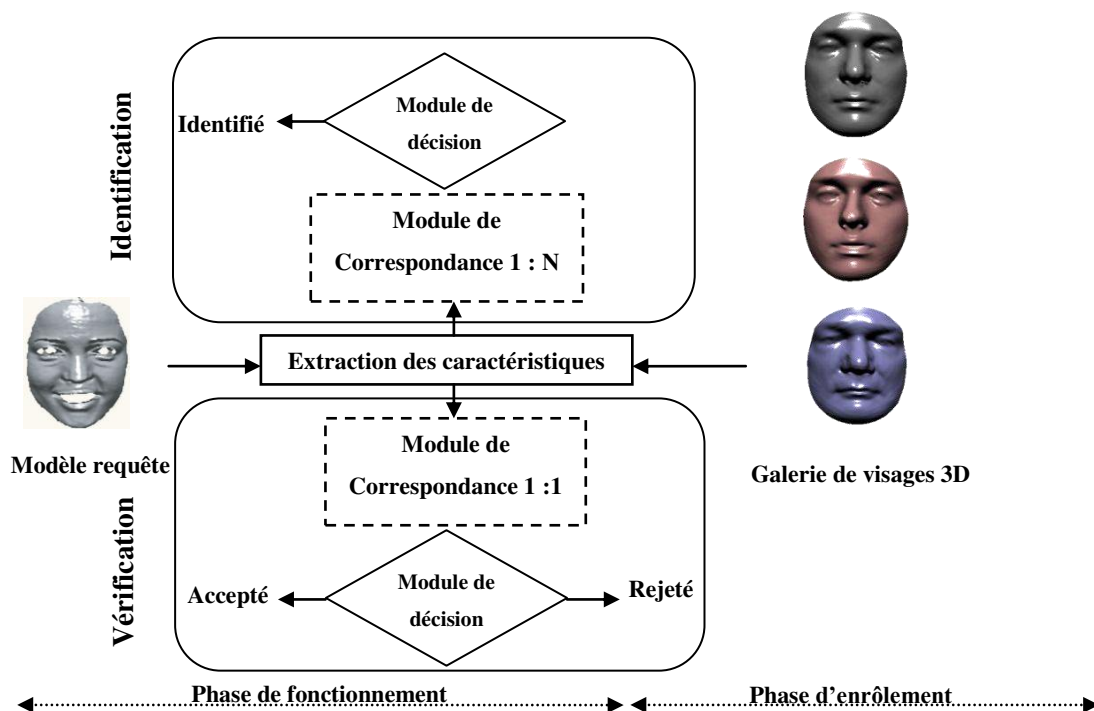


Figure 4. Vue d'ensemble d'un système de reconnaissance de visages (Identification et vérification) avec les deux phases de fonctionnement et d'enrôlement

Pour ces deux formes, une phase d'enrôlement est nécessaire avant la phase de fonctionnement. Il s'agit essentiellement de représenter un utilisateur dans le système tout en capturant et enregistrant ses informations biométriques dans une base de données. Pour la phase de fonctionnement, elle est répartie suivant trois modules : un premier module pour l'extraction des caractéristiques, un deuxième module pour la mise en correspondance et un module final pour la prise de décision. Lorsque le système biométrique opère suivant une application d'authentification ou vérification, l'utilisateur affirme son identité et le système vérifie si cette affirmation est valide ou non. On est donc dans le cas de la mise en correspondance 1:1 dans lequel une mesure de similitude est effectuée entre l'entrée biométrique de l'utilisateur et le modèle enregistré dans la base qui est réclamé. Ainsi, si le degré de similitude, entre l'identité affirmée et le modèle enregistré est élevé, l'affirmation est validée et l'utilisateur est considéré comme étant un authentique. Dans le cas contraire, l'affirmation de la validation de l'identité est rejetée et l'utilisateur est considéré comme étant un imposteur.

Dans le cas d'une application d'identification, l'utilisateur ne dévoile pas explicitement son identité. Cependant, l'affirmation implicite faite par l'utilisateur est qu'il s'agit d'une personne déjà représentée dans le système. Ainsi, l'échantillon biométrique de l'individu est comparé avec les modèles de toutes les personnes de la base de données. On parle alors de la mise en correspondance 1:N. La sortie du système biométrique est constituée par l'identité de la personne dont le modèle possède le degré de similitude le plus élevé avec l'échantillon biométrique présenté en entrée. Typiquement, si la plus grande similarité entre l'échantillon et tous les modèles est inférieure à un seuil de sécurité minimum fixé, la personne est rejetée, ce qui implique que l'utilisateur n'était pas une des personnes représentées dans le système. Dans le cas contraire, la personne est acceptée.

Pour ces deux scénarios applicatifs, la performance des systèmes biométriques est traditionnellement évaluée selon quatre critères de performance suivants. Le taux d'identification représente le premier critère. Il définit le pourcentage de visages correctement identifiés parmi les visages présentés au système.

$$\text{Taux d'identification} = \frac{\text{Nombre de visages correctement identifiés}}{\text{Nombre total de visages}} * 100 \quad (\text{eq.1})$$

Le deuxième critère s'appelle le taux de faux rejet ("False Reject Rate" ou FRR). Ce taux représente le pourcentage de personnes censées être reconnues mais qui sont rejetées par le système. Le troisième critère est le taux de fausse acceptation ("False Accept Rate" ou FAR).

Ce taux représente le pourcentage de personnes censées ne pas être reconnues mais qui sont tout de même acceptées par le système. Le critère final est connu sous le nom de taux d'égalé erreur ("Equal Error Rate" ou EER). Ce taux est calculé à partir des deux critères (FRR, FAR) et constitue un point de mesure de performance courant. Ce point correspond à l'endroit où

$$FRR = FAR \quad (\text{eq.2})$$

Ce taux représente le meilleur compromis entre les faux rejets et les fausses acceptations. La figure 5 permet d'illustrer les taux d'égalé erreur pour trois systèmes de reconnaissance de visages S1, S2 et S3.

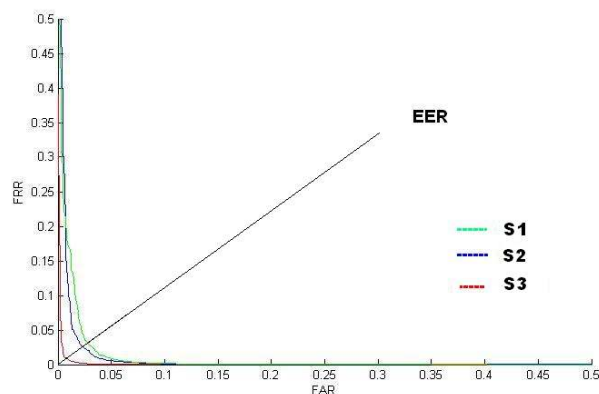


Figure 5. Exemple de taux d'égalé erreur pour trois systèmes de reconnaissance de visages S1, S2 et S3

Dans un scénario de vérification, on présente la performance d'un système biométrique sous forme d'une courbe ROC "Receiver Operating Characteristic". Cette courbe trace les deux premiers critères à savoir le taux de faux rejet en fonction du taux de fausse acceptation. Dans de nombreux travaux de l'état de l'art, le taux de vérification est représenté par le taux de bonne acceptation ("True Accept Rate" ou TAR) ($TAR = 1 - FRR$) où la valeur du taux de fausse acceptation (FAR) est fixée à 0.1%. Ce taux est un point spécifique de la courbe ROC. En revanche, dans un scénario d'identification, nous utilisons ce que l'on appelle une courbe CMC "Cumulative Match Characteristic". Cette courbe donne le taux d'identification en fonction d'une variable que l'on appelle le rang R . On dit qu'un système identifie au rang R lorsqu'il choisit l'image la plus proche, parmi les R premières images, comme résultat de la reconnaissance. Nous pouvons donc dire que plus le rang augmente, plus le taux de reconnaissance correspondant est lié à un niveau de sécurité faible.

2.2 Représentation Faciale

A l'heure actuelle, la majorité des travaux en FR était basée sur les images de visage 2D. Néanmoins, depuis quelques années, la forme 3D du visage a émergé comme une nouvelle

modalité faciale. Dès lors que les visages 3D sont théoriquement insensibles aux conditions d'éclairage et à la rotation, cette nouvelle modalité faciale a été intensivement étudiée par la communauté des chercheurs [Scheenstra et al., 2005] [Bowyer et al., 2006] pour développer des solutions en FR robustes aux variations d'expressions faciales. L'émergence de cette modalité a été accompagnée par un développement considérable de caméras 3D. Ainsi, les modèles 3D de visages retenus par plusieurs bases de test (FRGC [Phillips et al., 2005], Bosphorus [Savran et al. 2008], etc.) ont la particularité de contenir les deux informations de forme et de texture. Profitant de ces avantages, plusieurs représentations ont été proposées afin de décrire les modèles 3D. La figure 5 présente les différents formats représentatifs du modèle du visage en 3D : l'image de texture (superposée sur la forme du visage), l'image de profondeur, le modèle ombré, le modèle filaire, et le modèle de nuage de points.

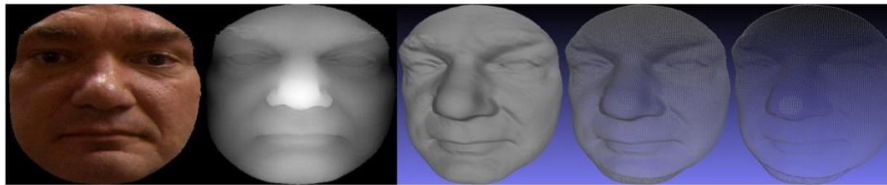


Figure 6. Les différents formats représentatifs du modèle du visage en 3D (de gauche à droite : l'image de texture superposée sur la forme du visage, l'image de profondeur, le modèle ombré, le modèle filaire et le modèle de nuage de points)

Avec ces deux modalités faciales, différents schémas FR peuvent être mis en œuvre suivant les modalités utilisées au niveau de la galerie (gallery dataset) et de test (probe dataset). Le premier schéma est du type *symétrique* étant donné que le même type d'information est exploité pour la galerie et le test : l'information 3D de forme de visage, ou les informations de texture (2D) et de forme 3D du visage. Le deuxième schéma est considéré comme un nouveau thème de recherche dans le domaine de la reconnaissance de visages : la reconnaissance asymétrique. Il convient à utiliser différents types de données pour la galerie et le test. Par exemple, la forme de visage 3D est enregistrée dans la galerie tandis que les images d'intensité 2D sont utilisées comme image de test. Dans nos travaux, nous avons adopté un schéma symétrique.

2.2.1 Représentation 2D de l'image de visage

Dans la mesure où les visages humains sont similaires en configuration, il convient donc de mettre en évidence, à travers des représentations spécifiques, la différence faciale entre deux visages. Différentes techniques d'analyses sont exploitées afin de décrire convenablement l'information 2D du visage.

2.2.1.1 *Image d'intensité et les méthodes sous-espace*

Le visage 2D peut être simplement représenté sous forme d'une image de taille $N \times M$ en niveau d'intensité (niveaux de gris). Plusieurs techniques sont proposées afin d'analyser ces images et de générer une représentation plus compacte. Parmi ces techniques, deux approches sont très connues dans la littérature: l'approche des visages propres ("eigenfaces") [Turk and Pentland, 1991] et celle des visages de Fisher ("Fisherfaces") [Belhumeur et al., 1997]. La méthode de visages propres (eigenfaces) consiste à appliquer l'analyse en composante principale (ACP) sur les images faciales 2D décrites comme des vecteurs de dimension $N \times M$ [Turk and Pentland, 1991]. Il s'agit de trouver un sous-espace avec les modes principaux de dimension largement inférieure à $N \times M$ qui capture les variations majeures des visages. Une partie de la base de visages, appelée base d'apprentissage, est utilisée pour produire ce nouveau sous-espace de projection. Toutes les images faciales de galerie ou de test sont alors projetées et représentées dans ce nouveau sous-espace. En reconnaissance, une image du visage sous forme de vecteur est projetée dans ce nouvel-espace de représentation en vue d'extraire ses nouvelles coordonnées. Elles sont comparées aux coordonnées des visages de la galerie en utilisant par exemple une distance euclidienne. Cette méthode est caractérisée par la réduction de la taille des données en les projetant dans un sous espace de taille plus réduit en optimisant la conservation des variations des visages. Cependant, cette représentation n'est pas toujours la meilleure car elle ne permet pas d'accroître la séparabilité entre les classes. La méthode de visages de Fisher ("Fisherfaces") est alors proposée pour tenir compte de cette spécificité. Elle est fondée sur le principe de l'analyse linéaire discriminante (LDA) en maximisant la dispersion inter-classes tout en minimisant celle de l'intra-classes. La méthode FR utilisant des visages de Fisher s'avère plus efficace que celle des visages propres dans les expériences [Belhumeur et al., 1997]. D'autres méthodes ont été également proposées dans cette catégorie des méthodes dites de sous-espace, comme par exemple la méthode ICA (Independent Component Analysis) [Bartlett et al. 2002] pour une meilleure représentation faciale.

2.2.1.2 *Les filtres de Gabor*

La représentation des images faciales par les filtres de Gabor est largement utilisée en traitement d'images. Ceci est motivé par le fait que les filtres de Gabor présentent deux propriétés intéressantes : la localisation fréquentielle et la sélectivité en orientation. Elles s'apparentent aussi à celles du système visuel humain [Hubel and Wiesel, 1977]. Les filtres de Gabor en deux dimensions ont été introduits en biométrie par [Daugman, 1993] pour la reconnaissance d'iris. Elles ont été utilisées au début par [Lades et al., 1993] en reconnaissance faciale.

Un noyau de filtre de Gabor est le produit d'une sinusoïde complexe avec une enveloppe gaussienne. Une ondelette de Gabor est définie comme suit :

$$\psi_{u,v}(z) = \frac{\|k_{u,v}\|^2}{\sigma^2} e^{-\left(\frac{\|k_{u,v}\|^2 \|z\|^2}{2\sigma^2}\right)} \left[e^{ik_{u,v}z} - e^{-\frac{\sigma^2}{2}} \right] \quad (\text{eq.3})$$

où $z = (x, y)$ est le point de coordonnées (x, y) dans le plan image. Les paramètres u et v définissent l'orientation et la fréquence des noyaux de Gabor. $\|\cdot\|$ est l'opérateur norme, et σ l'écart-type de l'enveloppe gaussienne. k est le vecteur d'onde dont $k_{u,v} = k_v \cdot e^{i\phi_u}$, $k_v = k_{max}/f^v$ et $\phi_u = \pi u/8$. k_{max} et f représentent respectivement la fréquence maximale et le facteur d'espace dans le domaine fréquentiel entre les noyaux de Gabor.

La représentation en filtre de Gabor d'une image résulte du produit de convolution de l'image avec une famille de noyaux de Gabor de fréquence et d'orientation différente comme définis par l'équation 4. La convolution de l'image I et d'un noyau de Gabor $\psi_{u,v}(z)$ est définie par :

$$G_{u,v}(z) = I(z) * \psi_{u,v}(z) \quad (\text{eq.4})$$

L'intérêt d'utiliser les filtres de Gabor pour extraire des caractéristiques faciales est qu'ils permettent de capturer les informations de visage dans des orientations et des résolutions différentes. De plus, ils sont robustes aux changements d'illumination, aux distorsions et aux variations d'échelle. En effet, la convolution d'une image avec une banque de 40 noyaux de Gabor (5 échelles et 8 orientations) conduit à 40 cartes d'amplitude et 40 cartes de phase qui sont de même taille que l'image originale, comme illustre la figure 7.

Par conséquent, si on ne considère que la réponse en amplitude, chaque pixel est décrit par un vecteur de dimension 40. Ce vecteur de dimension 40 est également appelé "Jet" et il a été utilisé dans les méthodes de reconnaissance de visages 2D, tels que l'approche "Elastic Bunch Graph Matching" (EBGM) [Wiskott et al., 1997] basées sur ces points caractéristiques. En effet, à partir d'une image de visage, des points caractéristiques sont localisés comme les coins des yeux, la bouche, le nez. Un treillis élastique virtuel est ensuite appliqué sur l'image de visage à partir de ces points. Chaque point représente un nœud labélisé auquel on associe un "Jet". Pour effectuer une reconnaissance avec une image test, une mesure de similarité est calculée entre les différents Jets et les longueurs des segments du treillis des deux images.

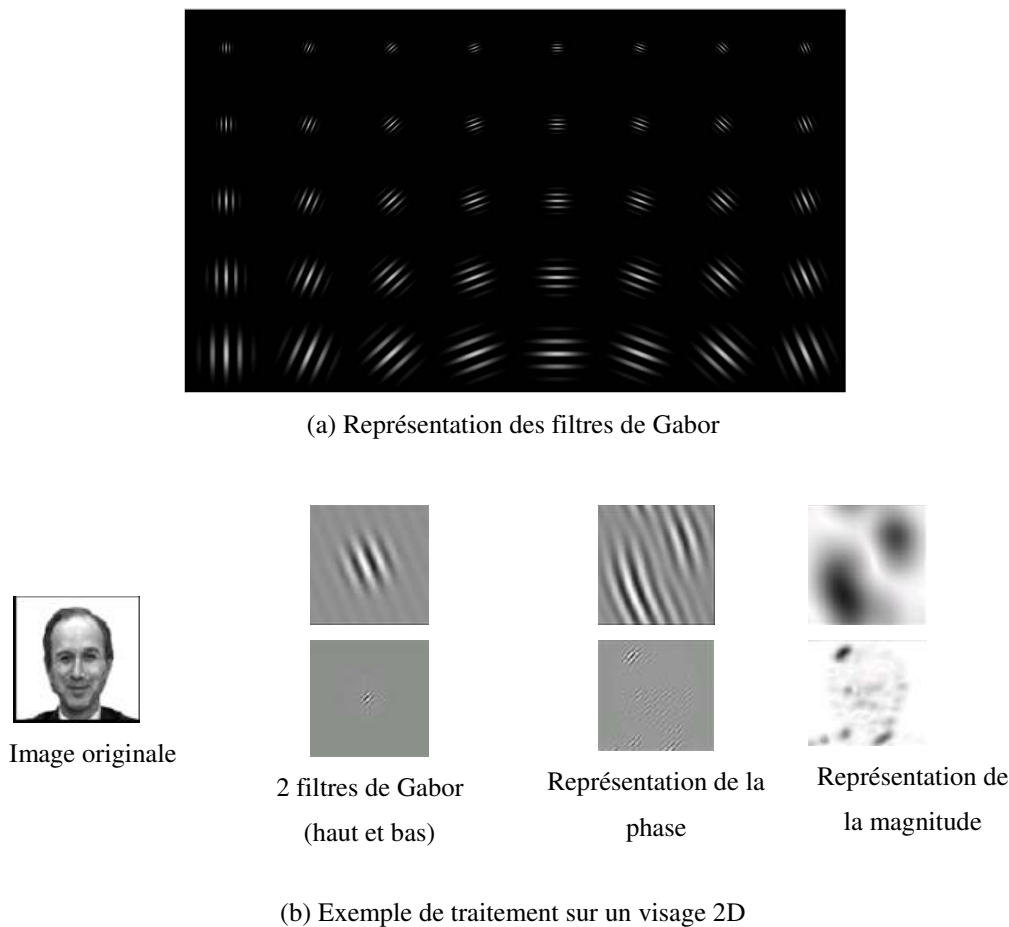


Figure 7. Filtre de Gabor : (a) Représentation de Gabor (b) Exemple de traitement

2.2.1.3 Local Binary Pattern (LBP)

L'opérateur LBP a été proposé à la fin des années 90 par [Ojala et al., 1996]. L'idée de cet opérateur de texture est de donner à chaque pixel un code dépendant des niveaux de gris de son voisinage. Le niveau de gris du pixel central (i_c) est comparé à ceux de ses voisins (i_n) suivant la formule suivante :

$$LBP(x_c, y_c) = \sum_{n=0}^P s(i_n - i_c) 2^n \quad (\text{eq.5})$$

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Le code LBP du pixel courant est alors produit en concaténant ces 8 valeurs pour former un code binaire. La figure. 8 donne un exemple de traitement de l'opérateur LBP. On obtient donc, comme pour une image en niveaux de gris, une matrice des valeurs LBP contenant des valeurs d'intensité comprises entre 0 et 255.

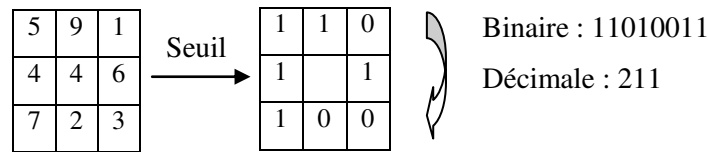


Figure 8. Exemple de traitement de l'opérateur LBP [Ahonen et al. 2004]

Le LBP a été étendu ultérieurement en utilisant des voisinages de taille différente [Ojala et al. 2002]. Dans ce cas, un cercle de rayon R autour du pixel central est considéré. Les valeurs des P points échantillonnés sur le bord de ce cercle sont prises et comparées avec la valeur du pixel central. Pour obtenir les valeurs des P points échantillonnés dans le voisinage pour tout rayon R , une interpolation est nécessaire. La notation (P, R) est adoptée pour définir le voisinage de P points de rayon R d'un pixel. $LBP_{P,R}$ est le code LBP pour le rayon R et le nombre de voisins P . L'opérateur LBP obtenu avec $P = 8$ et $R = 1$ est très proche de l'opérateur LBP d'origine. La principale différence est que les pixels doivent d'abord être interpolés pour obtenir les valeurs des points sur le cercle. La propriété importante du code LBP est que ce code est invariant aux changements uniformes d'illumination parce que le LBP d'un pixel ne dépend que des différences entre son niveau de gris et celui de ses voisins.

Dans la reconnaissance faciale, une fois le code LBP calculé pour tous les pixels de l'image du visage, l'histogramme de cette image LBP est calculé pour former un vecteur de caractéristiques représentant l'image faciale. Dans plusieurs travaux [Shan et al. 2009] [Ahonen et al. 2004] et afin d'incorporer plus d'informations spatiales au vecteur représentant le visage, l'image codée par l'opérateur LBP est divisée en petites régions et l'histogramme est construit pour chaque région. Finalement, tous les histogrammes des régions sont concaténés afin de former un grand histogramme représentant l'image des caractéristiques faciales. Etant donné deux histogrammes de LBP de deux visages, l'idée consiste à utiliser une métrique pour calculer la similarité entre ces deux histogrammes. Dans [Ahonen et al., 2004], trois métriques χ^2 , l'intersection des histogrammes et le maximum de vraisemblance sont testés. Les meilleurs résultats sont observés avec la première métrique.

2.2.2 Représentation de la forme géométrique 3D

Un modèle 3D de visage capture l'information de forme d'une surface faciale. Ceci le rend invariant à la variation d'illumination et de la pose. Aussi, diverses caractéristiques de forme peuvent être extraites à partir de ce modèle, allant de la représentation de base telle que les nuages de points 3D aux représentations les plus élaborées telles que les images de profondeur, les courbures, les vecteurs normaux, etc.

La plupart de ces caractéristiques 3D ont été plus ou moins utilisées dans la littérature. Par exemple, nous pouvons voir une étude réalisée par [Gökberk et al., 2008] regroupant différentes caractéristiques pour la reconnaissance de visages.

2.2.2.1 Nuage de Points 3D

Le nuage de points 3D est l'ensemble 3D des coordonnées (x, y, z) des points du visage. Un visage avec N échantillons peut donc être représenté par trois vecteurs, X , Y et Z de longueur N . La figure 9 montre un exemple de nuage de points d'un visage 3D.

Avec cette représentation, ICP (Iterative Closest Point) [Besl and Mckay, 1992] peut être utilisé pour le recalage de deux surfaces faciales 3D par itération successive en minimisant la somme des distances entre les paires de points entre un modèle référence et un modèle test. L'appariement ou la mise en correspondance vise à lier chaque point du modèle de test avec le point le plus proche du modèle de référence en employant l'heuristique du plus proche voisin au sens euclidien. Par la suite, un recalage rigide ou transformation est calculée aux sens des moindres carrés entre les points liés afin d'avoir le modèle test le plus proche du modèle de référence. Ceci permet de calculer une erreur qui servira comme critère d'arrêt de l'itération. En effet, l'algorithme s'arrête dès que la valeur de cette erreur chute au dessous d'une valeur seuil. [Medioni and Waupotitsch, 2003] ont exploité l'algorithme ICP pour le recalage et le calcul de similarité entre les surfaces faciales 3D. Ils reportent un taux de reconnaissance de 98% avec une base de 700 modèles 3D (100 sujets). Plus récemment, Boulbaba et al. ont utilisé le même principe pour recalibrer deux surfaces faciales sur la base de leurs régions statiques pour réduire l'effet d'expressions faciales [Ben Amor et al., 2007]. Ils proposent que le recalage des deux surfaces faciales se fasse sur les parties ou régions peu déformables d'un visage via ICP. Pour la reconnaissance, ils ont employé une métrique de similarité qui calcule une distance globale sous la forme d'une somme pondérée de distances élémentaires des différentes régions. Un taux de reconnaissance de 97.56% a été obtenu avec la base IV¹.

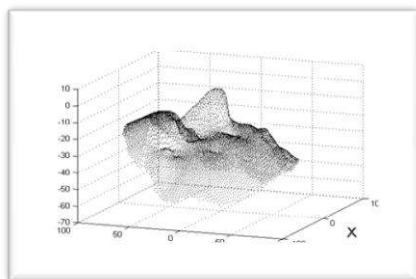


Figure 9. Représentation nuage de points pour un visage 3D

¹ <http://lsc.univ-evry.fr/techno/iv2/PageWeb-IV2.html>

2.2.2.2 Image de profondeur

Dans cette représentation, les coordonnées Z des points 3D de visage sont tracées sur une grille régulière X-Y en appliquant une interpolation linéaire. L'image de profondeur (Figure 10) a la forme d'une fonction $I(x, y) = Z(x, y)$ semblable à une image en niveau de gris 2D.

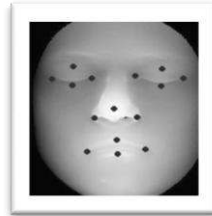


Figure 10. Image de profondeur d'un visage 3D

[Achermann et al. 1997] appliquent l'approche des visages propres sur les images de profondeur. Ils reportent un taux de reconnaissance de 100% avec une base de visages de 24 sujets dont chacun possède 10 images. Dans leur expérience, la base d'apprentissage est construite avec 5 images pour chaque sujet. Dans [Hesher et al., 2003], les auteurs utilisent la même technique avec une base composée de 37 sujets dont chacun possède 6 images de visages avec différentes expressions. Un taux de reconnaissance de 83% est obtenu. Dans une application de vérification de visage, [Heseltine et al., 2004] évaluent cette même approche avec trois distances différentes: mahalanobis, euclidienne et cosinus. Ils obtiennent le taux minimal d'égale erreur (17.8%) avec la distance mahalanobis pour une base de 100 sujets et 230 images de profondeur. Récemment, [Huang et al., 2011b] combinent une variante de l'algorithme LBP pour la description de la surface faciale à partir des images de profondeurs (voir Figure 11) et l'algorithme SIFT pour extraire de cette représentation des points d'intérêt. Ces derniers sont par la suite appariés entre un modèle test et un modèle de la galerie.

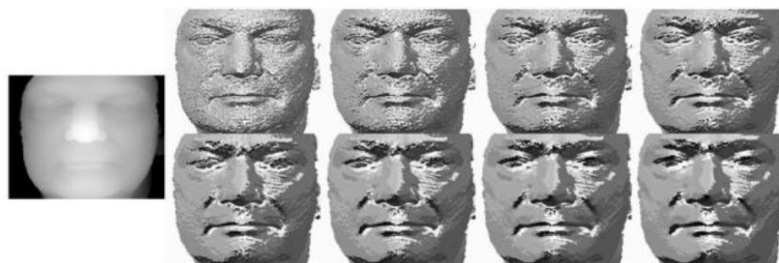


Figure 11. Images LBP générées à partir d'une image de profondeur

La mesure de similarité appliquée tient compte des appariements, des contraintes spatiales et de configuration. Les résultats expérimentaux ont montré l'efficacité de cette approche par rapport aux variations d'expressions et aux occultations partielles.

2.2.2.3 *Les descripteurs de la Courbure*

Plusieurs approches utilisent les propriétés de la géométrie différentielle des surfaces, e.g. les courbures principales maximales, minimales, etc., qui permettent de caractériser une surface faciale en des localités selon leur concavité, convexité et points de selles. Ces caractéristiques locales de la surface peuvent offrir des outils intéressants pour la reconnaissance d'un visage 3D. Au début, [Tanaka et al., 1998] calculent des cartes de courbures principales minimales et maximales à partir des cartes de profondeur des visages. Ensuite, des lignes de sommets et de vallées sont extraites de ces cartes afin de construire des images gaussiennes étendues (EGI) pour chaque visage. La mise en correspondance entre l'image de profondeur référence et celle de test s'effectue moyennant la corrélation sphérique de Fisher entre les EGI de sommets et de vallées respectifs. Un taux de reconnaissance de 100% a été obtenu avec 37 images de sujets différents. [Lee et Milios, 1990] représentent aussi les différentes régions faciales du visage par une EGI. Pour établir la correspondance entre deux visages, ils procèdent par une mise en correspondance de graphe en calculant la corrélation entre les images gaussiennes étendues des régions convexes. [Moreno et al., 2003] proposent de segmenter le visage en plusieurs régions en utilisant la courbure gaussienne. Ainsi, chaque modèle facial 3D est subdivisé en 86 régions différentes grâce à un algorithme de segmentation qui exploite les signes de la médiane et des courbures gaussiennes pour isoler les régions ayant des courbures significatives. Cet espace de caractéristiques est ensuite réduit pour augmenter l'efficacité de la méthode. Les expérimentations ont été menées sur un ensemble de 420 modèles 3D de visages appartenant à 60 sujets, comprenant des images avec des variations de lumière, de rotation et d'expressions de visages. Le taux de reconnaissance a atteint 78 %.

Dans [Samir et al., 2005], les auteurs ont enrichi leur approche topologique en utilisant les graphes de Reeb par la courbure moyenne comme information géométrique. Ultérieurement, les auteurs caractérisent les formes des surfaces des visages modulo des déformations qui correspondent aux déformations faciales [Samir et al., 2006]. Ils présentent une surface de visage en utilisant une famille de courbes fermées [Bronstein et al., 2005]. Ces courbes seront calculées comme les courbes de niveaux d'une fonction continue F de la surface du visage, et les indexes correspondent aux valeurs de cette fonction. Ils utilisent la fonction de profondeur (valeur des coordonnées z) comme fonction où les courbes à niveau fournissent les courbes des visages désirés. Les formes de deux surfaces de visages seront comparées à travers les formes de courbes faciales des visages.

2.2.2.4 *Vecteurs normaux*

La surface d'un visage peut être considérée comme un exemple d'un objet de forme libre à partir duquel nous pouvons extraire des descripteurs locaux pour représenter l'information de

visage. Dans la représentation avec les vecteurs normaux extérieurs (figure 12), chaque point de la surface faciale est décrit par un vecteur à trois dimensions (N_x, N_y, N_z).

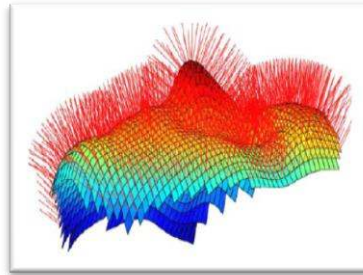


Figure 12. Surfaces Normales d'un Visage 3D

[Abate et al. 2006] projettent les points du visage 3D sur une sphère 3D pour obtenir leurs coordonnées sphériques. Les angles entre deux vecteurs normaux des facettes adjacentes sont calculés et représentés sous forme d'image 2D. Les opérateurs de Fourier 2D sont appliqués sur ces images et par la suite sont exploités pour la classification. Un taux de reconnaissance de l'ordre de 96 % est obtenu avec une base de test contenant 120 sujets dont chacun possède 9 modèles. Dans [Gökberk et al., 2008], les auteurs mettent en œuvre un algorithme de sélection de meilleurs descripteurs. Dans leurs expérimentations, les vecteurs normaux à la surface du visage 3D qui possèdent un taux de reconnaissance individuelle de l'ordre de 83.79 %, sont sélectionnés parmi les descripteurs les plus performants. Ceci a permis d'aboutir à une performance finale de 95.45 % avec la base FRGC v.2.0 comme base de test et la base FRGC v.1.0 comme base d'apprentissage.

2.2.3 Conclusion

Dans cette section, nous avons présenté un ensemble de représentations permettant de décrire les informations de texture 2D ou la forme géométrique 3D du visage. S'il est possible d'exploiter chacune de ces différentes représentations faciales d'une façon indépendante pour la reconnaissance faciale, il serait plus intéressant de les fusionner, e.g., la prise en compte simultanée des deux modalités 2D et 3D, ainsi que des différents descripteurs et algorithmes, en vue d'un système de reconnaissance plus robuste et plus performant. Dans la section suivante, nous dressons un court état de l'art sur les différentes techniques de fusion.

2.3 Stratégies de fusion

L'importance de la fusion d'information dans le contexte biométrique est soulignée dans plusieurs travaux [Dorizzi and García-Mateo, 2007] [Raghavendra et al., 2011]. La fusion pour une meilleure prise de décision est une nécessité pour beaucoup d'applications car les sources

d'informations sont généralement multiples mais elles sont aussi très souvent imparfaites. Dans le cas de la reconnaissance faciale 3D, nous avons vu dans la section précédente que les sources d'informations peuvent être d'une part la forme géométrique 3D d'un visage et d'autre part la texture associée à celui-ci. Etant donnée une source d'information, par exemple géométrique, nous pouvons encore en extraire des différents descripteurs, tels que les courbures et les vecteurs normaux conduisant à différentes représentations faciales qui mettent en évidence des propriétés différentes.

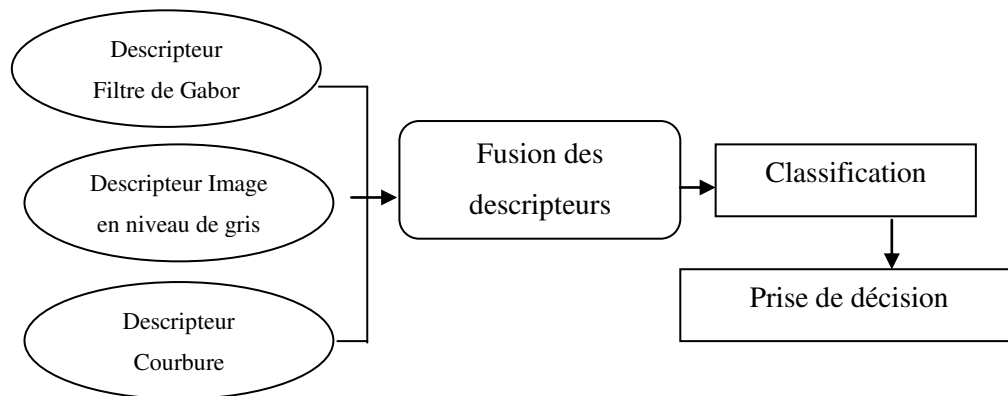
La fusion peut aussi être nécessaire car les méthodes de classification basées sur différentes théories, e.g., SVM, réseaux neuronaux, classifieurs Bayésiens, peuvent aussi faire apparaître des différences de comportement face aux données. Enfin, la fusion peut aussi être souhaitable pour des experts du même type, e.g. SVM, mais entraînés sur des données différentes, donc pouvant conduire à des expertises complémentaires. Par conséquent, la prise en compte des deux modalités 2D et 3D, ainsi que des différentes caractéristiques et algorithmes, nécessite des techniques de traitements adaptés et éventuellement de fusion 3D et 2D préalable à une étape de reconnaissance. Ces stratégies de fusion doivent être capables d'intégrer et de tirer parti des différentes modalités et, plus largement, des différentes informations extraites lors du processus de reconnaissance.

Dans l'état d'art, diverses stratégies de fusion, e.g., [Lao et al., 2000], [Wang et al., 2002], [Papatheodorou & Rueckert, 2004], [Tsalakanidou et al., 2003], [Chang et al., 2003], [Hüsken et al., 2005], [Lu & Jain, 2005], [Maurer et al., 2005], ont été proposées pour la reconnaissance de visages, allant de la fusion précoce "early fusion" opérant au niveau de caractéristiques à la fusion tardive "late fusion" opérant sur les sorties de classifieurs, en passant par de nombreuses stratégies intermédiaires. En raffinant cette catégorisation, la fusion des informations des deux modalités 2D et 3D peut encore s'opérer à plusieurs niveaux différents en regard de l'opération de l'appariement qui a pour but de générer une mesure de similitude ou un score d'appariement entre le vecteur caractéristique du modèle de la base et l'entrée du problème (visage 2D, visage 3D,...). On distingue ainsi la fusion avant l'appariement (pré-classification) de la fusion après l'appariement (post-classification). Par exemple, si nous considérons trois types de descripteurs : les filtres de Gabor, image en niveau de gris, courbure gaussienne, on peut procéder soit à une fusion précoce ou une fusion tardive comme l'illustre la figure 13.

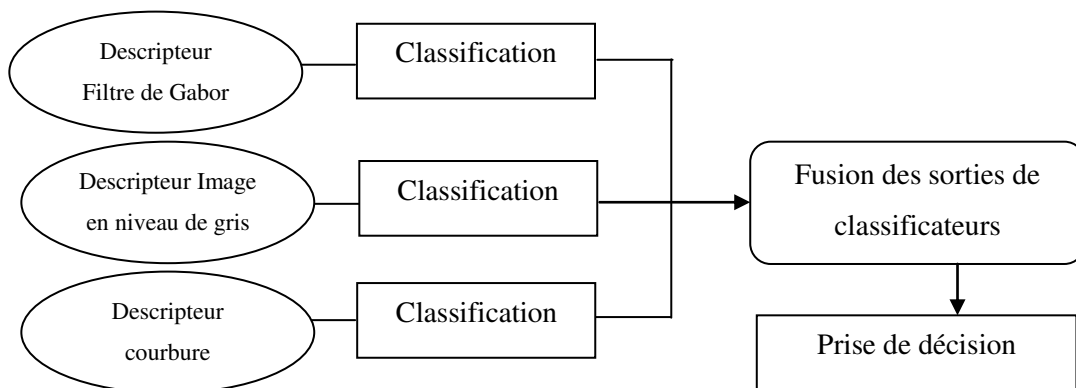
D'une façon générale, la combinaison des informations de visage peut intervenir aux différents niveaux :

- *Niveau des échantillons*: Chaque modalité biométrique produit une collection d'échantillons à partir des différents capteurs. Le processus de fusion combine ces données dans un simple échantillon.

- *Niveau des caractéristiques*: chaque modalité biométrique produit un ensemble de caractéristiques. Le processus de fusion combine ces collections de caractéristiques en un unique ensemble ou vecteur de caractéristiques.
- *Niveau du score*: chaque classifieur produit typiquement un score d'appariement ou plusieurs scores. Le processus de fusion combine ces derniers dans un résultat unique représentant le score final.
- *Niveau de rang*: chaque classifieur produit un rang suivant le score d'appariement. Le processus de fusion assemble ces rangs pour générer le rang final.
- *Niveau de décision*: chaque classifieur produit un résultat unique qui peut être un résultat booléen ou binaire. Le processus de fusion les assemble pour avoir la décision finale.



(a)



(b)

Figure 13. Schémas de fusion précoce (a) et tardive (b)

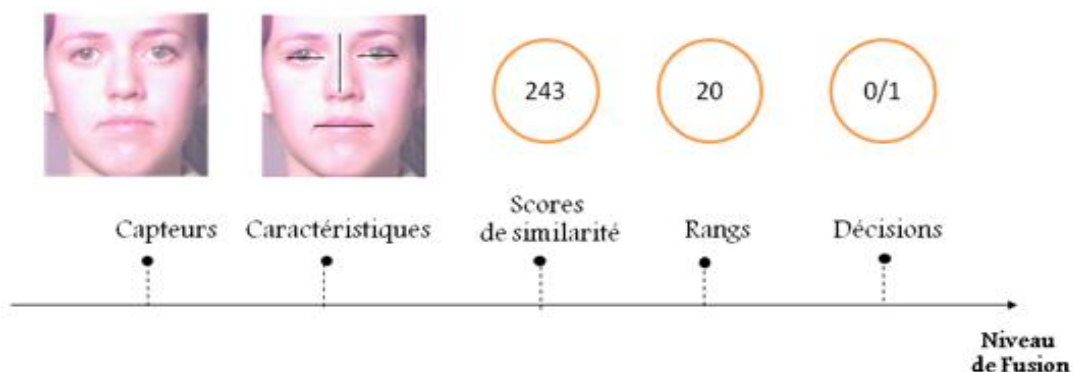


Figure 14. Les niveaux de fusion

La figure 14 illustre la relation entre les données manipulées et le niveau de fusion correspondant. Néanmoins, nous pouvons aussi trouver d'autres stratégies de fusion telles que : la fusion en cascade [Mian et al., 2007] ou la fusion multi-niveau [Li et al., 2007].

2.3.1 Fusion Avant Appariement

Il s'agit principalement d'une fusion précoce. Elle correspond à la fusion des informations issues de plusieurs données biométriques au niveau du capteur (images brutes) ou au niveau des caractéristiques extraites par le module d'extraction de caractéristiques.

2.3.1.1 Fusion au niveau des données des capteurs

Dans le cas de la reconnaissance de visage, la fusion des données peut être réalisée au niveau du capteur. Dans ce cas, le capteur peut être, par exemple, un scanner 3D² ou un système stéréoscopique avec deux ou trois caméras [Ouji et al., 2011] (voir figure 15). Dans [Papatheodorou & Rueckert, 2004], les informations de texture 2D (T) et de profondeur 3D (Sx, Sy, Sz) sont enregistrées sous forme de points 4D (Sx, Sy, Sz, p*T) où p est un facteur de pondération déterminant l'importance de la texture. Pour le calcul de la similarité, un algorithme modifié d'ICP ("Iterative Closest Point") est appliqué en se basant sur le calcul de la distance euclidienne entre deux points (S_q, t_q) dans cet espace 4D :

$$4D_{eval}(s, T) = \frac{\sum_{q=1}^N \sqrt{(s_{qx} - t_{qx})^2 + (s_{qy} - t_{qy})^2 + (s_{qz} - t_{qz})^2 + \alpha (s_{qi} - t_{qi})^2}}{N} \quad (\text{eq.6})$$

Une base de galerie composée de 62 visages neutres et frontaux et cinq bases de test contenant chacune 62 visages avec différentes expressions (sourire, colère) et différentes poses (frontal, 45°, 20°) ont permis d'évaluer cette approche. Les expériences ont montré une amélioration en

² Minolta 3D Digitizers, Non-Contact 3D Laser Scanner, <http://www.minolta3d.com>, 2006.

termes de taux de reconnaissance dans le cas de variation de l'expression du visage par rapport à l'utilisation seule de l'information de forme. Par exemple, le taux de reconnaissance passe de 98.4 % à 100% pour toutes les valeurs fixées de p avec des images frontales neutres. Une diminution de performance est aperçue dans le cas de changement de pose.



Figure 15. Acquisition des visages

2.3.1.2 Fusion au niveau des caractéristiques

La fusion au niveau des caractéristiques appelée aussi fusion bas-niveau est considérée comme une fusion complémentaire. Elle consiste à combiner dans un seul vecteur caractéristique les différents ensembles de descripteurs qui sont extraits des sources biométriques (2D, 3D). Le vecteur résultant devrait être la synthèse des informations les plus pertinentes pour guider au mieux la prise de décision finale. Dans [Wang et al., 2002] les auteurs ont développé une approche 2D+3D combinant deux méthodes. Une première méthode, qui utilise un filtre de Gabor, localise les points d'intérêt 2D. Une deuxième méthode, qui utilise l'algorithme de [Chua et al., 1997], pour extraire les points de signature 3D (voir figure 16).



Figure 16. Un exemple de visages avec les points d'intérêts détectés

Les points 3D et les points de texture 2D ainsi obtenus sont projetés chacun dans leur propre sous-espace, en utilisant l'ACP, pour générer respectivement des vecteurs de forme et de texture. Ces vecteurs sont regroupés en un seul vecteur qui est normalisé afin de représenter l'image faciale. Le processus d'identification est basé sur une fonction de similitude utilisant la méthode SVM. Les expérimentations, faisant participer 50 sujets ayant des expressions faciales différentes pour des points de vue différents, ont démontré l'efficacité de l'algorithme avec un taux de reconnaissance de l'ordre de 90 %.

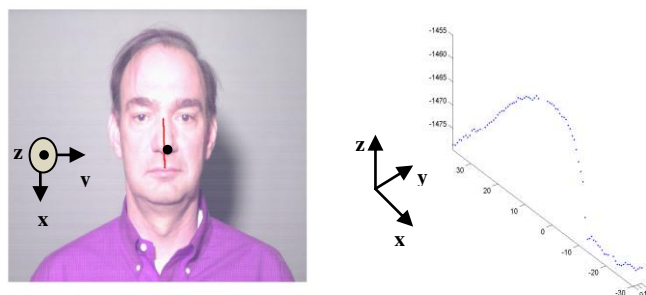


Figure 17. Un exemple de profil et sa représentation avec les points 3D

[Arca et al., 2007] extraient automatiquement 27 points clés de chaque visage dans le domaine 2D. Au niveau de chaque point, un ensemble de filtres de Gabor est appliqué avec 5 échelles et 8 orientations. Un vecteur 2D de dimension (27×40) est obtenu. Dans le domaine 3D, un ensemble de 15 profils locaux (3D) sont extraits (voir figure 17) avec leurs points 3D correspondants. Chaque profil 3D est projeté sur les deux principales directions en appliquant une ACP. Un vecteur 3D est formé avec les 15 profils projetés. Le vecteur 2D-3D est formé avec la concaténation des 42 caractéristiques (27 caractéristiques 2D + 15 caractéristiques 3D). Le calcul de la similarité, lors de la présence d'une image de test dans un scénario d'identification, s'effectue avec toutes les composantes du vecteur, accompagné du calcul de leur poids respectif. La fusion des scores de similarité applique une somme pondérée en fonction du poids précédemment calculé. Les 22 premiers scores ayant les meilleurs poids sont sélectionnés. Un taux d'identification de l'ordre de 98.1 % est obtenu avec une partie de la base de visages FRGCv.2.

[Mian et al., 2008] concatènent aussi deux descripteurs locaux en se basant sur des points 3D clés détectés au niveau des régions de visages 3D. Dans le domaine 3D, une base de coordonnées 3D unique est définie localement à chaque point clé pour faciliter l'extraction de descripteur invariant au changement de pose. Dans le domaine 2D, les descripteurs SIFT sont extraits des différents visages 2D au niveau des points 3D détectés précédemment. Les différents descripteurs des sujets présents dans la base de galerie et de test sont projetés en utilisant l'ACP. Dans cette opération, chaque descripteur est normalisé dans son sous-espace tout en prenant en compte la quantité d'information conservée. Le vecteur multimodal est formé par la concaténation des deux descripteurs et a été normalisé. Pour la prise de décision finale soit pour l'identification ou la vérification, un vecteur final est formé par quatre scores générés à partir du vecteur multimodal, des points clés 3D, du nombre de points 3D clés détectés, et de la longueur du contour formé par les points clés 3D. L'algorithme proposé atteint un taux d'identification de 93.1 % et 97.2% pour le taux de vérification avec tous les sujets de la base de visages FRGC v2.0. Ces performances sont légèrement plus faibles en comparaison avec

l'utilisation seule des caractéristiques 3D (93.5 % identification et 97.4 % vérification). Un module de fusion au niveau de scores a été développé en utilisant les quatre scores décrits précédemment. Une nette amélioration des performances a été observée.

2.3.1.3 Conclusion

D'après ces travaux, nous constatons que la combinaison au niveau des caractéristiques doit prendre en considération plusieurs contraintes telles que : la taille des descripteurs, le domaine de définition des descripteurs et la taille des bases de données. Si la démarche est simple et facile, son principal inconvénient est la malédiction de dimension (curse of dimensionality) ainsi que la non-homogénéité de données fusionnées. Une alternative à cette fusion précoce est la fusion au niveau des scores ou la fusion après l'appariement [Mian et al., 2008] [Godil et al., 2005] qui est largement étudiée dans la littérature.

2.3.2 Fusion Après Appariement

Il s'agit bien d'une fusion tardive. Elle est connue aussi comme une fusion de haut niveau, vu qu'elle opère sur les sorties des différents classifieurs. Comme sortie générée par un classifieur, on peut avoir : les scores de similarité, une liste de candidats avec les rangs respectifs ou une décision unique. Elle est caractérisée par sa capacité à prendre en compte un grand nombre de classifieurs différents en exploitant le comportement de chacun d'eux [Gökberk et al., 2008] [Gökberk et al., 2005] [Cook et al., 2007] [Faltelier et al., 2006] [Mian et al., 2006]. Avec une telle procédure de fusion, nous pouvons exploiter la compétitivité de ces classifieurs. Sur ce point, la littérature a fourni beaucoup de travaux présentant des méthodes de combinaison qui se différencient aussi bien par le type d'informations apportées par chaque classifieur que par leurs capacités d'apprentissage et d'adaptation [Hüsken et al., 2005]. Cette section a pour objectif de présenter les recherches actuelles sur la combinaison parallèle, en cascade et multi-niveau de classifieurs.

2.3.2.1 Fusion parallèle

La fusion parallèle laisse tous les classifieurs du système opérer indépendamment les uns des autres ; Ensuite elle fusionne leurs réponses respectives. Dès lors se posent les problèmes de précision des informations fournies par les classifieurs et de la confiance qu'on peut accorder à chacun d'eux. Ainsi, le problème de la combinaison parallèle de classifieurs peut se poser de la façon suivante : étant donné un ensemble de K classifieurs, comment élaborer une réponse finale unique à partir des K résultats fournis. Rappelons qu'un classifieur reçoit en entrée un visage x et génère par la suite des informations à propos de l'appartenance aux modèles de la

galerie. Cet outil décide d'attribuer au visage x la classe C_i parmi un nombre fini de visages possibles dans la galerie, $i = 1, \dots, M$.

$$X \rightarrow E(X) = C_i \quad (\text{eq.7})$$

Les réponses fournies par un classifieur peuvent être divisées en trois catégories suivant le niveau d'information apporté par celui-ci :

- décision unique « la classe d'appartenance »: $e_j(x) = C_i$ ($i \in \{1, \dots, M\}$) ; indique que le classifieur j a attribué la classe C_i à x .
- liste de candidats avec les rangs respectifs: $e_j(x) = [r_1^j, r_2^j, \dots, r_M^j]$ où r_i^j est le rang attribué à la classe i par le classifieur j .
- score de similarité: $e_j(x) = [m_1^j, m_2^j, \dots, m_M^j]$ où m_i^j est le score de similarité attribué x pour la classe i par le classifieur j .

Chaque type de sortie correspond à un niveau d'information différent fourni par le classifieur. La sortie « décision unique » est a priori l'information la plus naturelle que l'on attend d'un classifieur mais elle fournit aussi le moins d'informations. La sortie « liste de candidats » reflète l'ordre de préférence des propositions fournies par le classifieur. La sortie « score de similarité » est la plus riche en information puisqu'elle reflète le niveau de confiance du classifieur dans ses propositions. Toutefois, comme ces mesures ne sont pas toujours comparables (une distance, une probabilité a posteriori, une valeur de confiance ou une fonction de croyance), une normalisation s'avère nécessaire [Xu et al. 1992] [Huang et al. 1994].

2.3.2.1.1 Fusion au niveau des scores

Rappelons que l'opération de l'appariement pour un classifieur donné permet de générer une mesure de similitude entre un modèle de galerie et un visage d'entrée en fonction d'une caractéristique utilisée par le classifieur. Cette mesure reflète le degré de confiance d'un classifieur sur l'appartenance du visage d'entrée à reconnaître à chacune des classes de la galerie. Le fait de fusionner plusieurs mesures de similitude pour pouvoir produire une décision finale s'appelle *fusion au niveau des scores*. Par exemple, nous pouvons avoir des mesures de distance ou de dissimilarité pour lesquelles une plus petite distance indique une meilleure correspondance. Néanmoins, les sorties des classifieurs doivent être sur la même échelle de valeurs. D'où la nécessité d'une étape de normalisation. Soit S_i le score de similarité cumulé pour la i -ème classe de la galerie par les différents classifieurs à travers une règle de fusion. La règle de décision $E(x)$ pour un visage d'entrée x consiste alors à choisir la classe C_i pour laquelle le score S_i est la plus élevée :

$$E(x) = \begin{cases} C_i = \operatorname{argmax}_{m \in 1..M} S_m & \text{si } C_i \geq \delta \\ \text{rejet} & \text{sinon} \end{cases} \quad (\text{eq.8})$$

On décide donc la classe C_i si le score de similarité cumulé des différents classifieurs pour la classe C_i est supérieur à un seuil δ . En choisissant $\delta=0$, la règle de décision précédente ne génère jamais de rejets.

Il existe différentes variantes de règles pour cumuler les scores produits par les classifieurs en fusion. Rappelons que m_i^j désigne le score de similarité produit par le j -ème classifieur entre l' i -ème classe de la galerie et le visage x de l'individu d'entrée à reconnaître tandis que K le nombre de classifieurs. Ces différentes variantes sont les suivantes :

- la règle de maximum [Snelick et al., 2005] [Godil et al., 2005] [Gökberk et al., 2008],

$$S_i = \max_{j=1}^K m_i^j \quad (\text{eq.9})$$

- la règle de minimum [Gökberk et al., 2008] [Godil et al., 2005] [Snelick et al., 2005] [Chang et al., 2005],

$$S_i = \min_{j=1}^K m_i^j \quad (\text{eq.10})$$

- la règle de moyenne [Godil et al., 2005], [Gökberk et al., 2008],

$$S_i = \frac{1}{K} \sum_{j=1}^K m_i^j \quad (\text{eq.11})$$

- la règle de produit [Tsalakanidou et al., 2003] [Gökberk et al., 2008] [Mian et al., 2006],

$$S_i = \prod_{j=1}^K m_i^j \quad (\text{eq.12})$$

- la règle de somme simple [Gökberk et al., 2008] [Snelick et al., 2005] [Tsalakanidou et al., 2007], [Lin et al., 2007], [Lin et al., 2006]

$$S_i = \alpha \sum_{j=1}^K m_i^j \quad (\text{eq.13})$$

La règle maximum s'apparente à un opérateur logique OU. Une classe est reconnue dès lors qu'elle est reconnue par le classifieur ayant le plus de confiance (score de similarité). La règle minimum s'apparente à l'opérateur logique ET. En effet, une classe de visages serait reconnue si et seulement si tous les classifieurs l'avaient reconnue. Enfin, la règle de moyenne est une sorte de vote majoritaire où chaque classifieur vote avec son score de similarité. [Kittler et al., 1998] ont fait une étude systématique d'un ensemble d'opérateurs de fusion, notamment les règles de maximum, minimum, moyenne, etc. et ont montré clairement que la règle de moyenne

est plus robuste que les deux règles maximum et minimum. Ils ont aussi prouvé que la règle de produit donne de bonnes performances si les classifieurs sont indépendants. Cependant, cette règle reste sensible aux erreurs estimées par les classifieurs (dues à la présence de bruit dans les données ou à l'utilisation de base d'apprentissage de petite taille). [Godil et al., 2005] comparent les performances respectives des quatre premières règles en utilisant la méthode de normalisation z-score et le plus proche voisin avec les deux distances L1 et Mahalanobis comme classifieur. Ils utilisent les descripteurs 2D de couleur et les points 3D de la surface du visage et appliquent une étape de réduction de dimensionnalité avec l'ACP sur tous les sujets de la base d'apprentissage (200 sujets) et la base de test (200 sujets). Le meilleur taux de reconnaissance, de l'ordre de 82 %, a été réalisé avec la distance L1 par la règle de produit.

Dans la dernière règle de somme simple (eq.13), α est une constante. Ainsi, pour $\alpha = 1$, nous obtenons la règle de somme. Si $\alpha = 1/K$ alors la règle est la moyenne simple. La somme fonctionne bien dans le cas d'utilisation de caractéristiques différentes par des classifieurs de même structure [Ho et al., 1994] ou de structures différentes [Kittler et al., 1998]. Dans [Kittler et al., 1998] par exemple, il est montré théoriquement que la règle de somme est la plus résistante à la mauvaise estimation des erreurs des classifieurs indépendants. Une comparaison expérimentale entre les différentes règles (somme, produit, maximum, minimum, moyenne) a été aussi menée sur une base de chiffres manuscrits pour confirmer le résultat théorique obtenu. En effet, avec 4 classifieurs indépendants (utilisant des caractéristiques distinctes), il est constaté la supériorité de la somme sur les autres règles. La moyenne et le vote à la majorité donnent des performances proches de la somme. Alors que le produit et le minimum donnent de mauvaises performances (inférieures même aux performances de chacun des classifieurs). En effet, ces deux règles sont sensibles aux erreurs produites par les classifieurs. Il suffit que l'un des classifieurs propose des probabilités faibles (proches de 0) pour que le résultat de la combinaison soit proche de 0. Une autre étude comparative sur les règles de somme, minimum et maximum a été réalisée dans [Snelick et al., 2005]. Pour cela, les scores générés par plusieurs systèmes d'authentification "Commercially available Off-The-Shelf" ont été combinés. Le meilleur taux d'EER (eq. 2) est généré par la règle de somme avec trois méthodes de normalisation différentes.

Généralisant la règle de moyenne ou somme pour fusion, nous trouvons aussi dans la littérature des méthodes de fusion utilisant la règle de somme pondérée qui consiste tout simplement à appliquer des pondérations aux sorties des classifieurs. Chaque score S_i d'une classe de visage C_i peut être obtenue par la règle suivante:

$$S_i = \sum_{j=1}^K \beta_j m_i^j \quad (\text{eq.14})$$

β_j est le coefficient qui détermine l'importance attribuée au $k^{\text{ème}}$ classifieur dans la combinaison alors que m_i^j est sa réponse pour la classe i .

[Lu & Jain, 2005] [Lu et al., 2006] utilisent cette règle de fusion avec deux scores. Le premier score est généré par un module d'appariement de surface utilisant l'ICP. Le deuxième score est produit par un deuxième module d'appariement d'apparence (texture) utilisant l'analyse discriminante linéaire LDA. Ce dernier score correspond à la similarité cosinus entre les projections respectives du visage test et celui de la base d'apprentissage sur l'espace LDA. Dans la phase d'évaluation, deux bases ont été conçues : une base de galerie avec 200 sujets et une base de test avec 598 sujets. Les deux scores sont normalisés suivant la méthode z-score. En fixant les deux poids à la valeur '1', l'opération de la fusion a permis d'obtenir un taux de reconnaissance de l'ordre de 90 % avec des améliorations respectifs de 4 % et 13 % par rapport au premier et deuxième module. En modifiant la valeur du poids du deuxième module entre 0 et 7, tout en gardant le premier poids à 1, le taux de reconnaissance a été légèrement amélioré pour une valeur = 2.

[Chang et al., 2003] [Chang et al., 2005] préfèrent calculer le poids β_j en utilisant les trois premiers distances générées par chaque classifieur. L'analyse en composantes principales a permis de transformer et classifier les images de texture et de profondeur. Le poids p relatif à chaque classifieur est donné par la formule suivante:

$$p = \frac{2^{\text{ème}} \text{ distance} - 1^{\text{er}} \text{ distance}}{3^{\text{ème}} \text{ distance} - 1^{\text{er}} \text{ distance}} \quad (\text{eq.15})$$

Un taux de reconnaissance finale est obtenu par [Chang et al., 2003] de l'ordre de 98.5 % (2D : 89 %, 3D : 94.5 %) avec une base totale de 400 sujets divisées équitablement entre les bases de galerie et de test. [Mian et al., 2008] génèrent quatre types de scores de similarité entre un probe et l'ensemble de visages de la galerie. Pour calculer le score final, un poids est assigné à chaque type de score par l'équation suivante :

$$\mathcal{K}_q = \frac{\bar{S}_q - \min(S_q)}{\bar{S}_q - \min_2(S_q)} \quad (\text{eq.16})$$

Avec \bar{S}_q représente la moyenne des score de similarité de type q , $\min(S_q)$ et $\min_2(S_q)$ sont respectivement le premier et le deuxième score parmi la liste des scores générés par le type de similarité q . L'algorithme proposé atteint un taux d'identification de 96.1 % et 98.6 % le taux de vérification sur la base de visages FRGC v2.0.

2.3.2.1.2 Fusion au niveau des rangs

Dans cette catégorie, les différentes méthodes combinent les listes ordonnées de propositions de classe produites par les différents classifieurs ordonnées de manière décroissante. [Gökberk et al., 2006] utilisent le Borda Count et le meilleur rang pour la fusion des décisions par quatre classifieurs spécifiques aux descripteurs : le nuage de points 3D, l'image de profondeur, l'image de profil et l'index de forme (*shape index*). Borda count consiste à calculer, pour une classe particulière C_i de galerie, la somme des rangs proposés par les classifieurs pour cette classe.

$$Borda\ Count(C_i) = \sum_{j=1}^K r_i^j \quad (\text{eq.17})$$

où K désigne le nombre de classifieurs tandis que r_i^j le rang du visage x d'entrée pour la classe i par le classifieur j . Concrètement, étant donné un visage x d'entrée à reconnaître, le classifieur j va le comparer avec chacune des classes de la galerie et produit un score de similarité. Ces scores de similarité pourront alors être triés dans un ordre décroissant, avec r_i^j désignant le rang du score de la classe i par le classifieur j . Nous attribuons $r_i^j = N$ si le score de la classe C_i est placé au début de la liste, $r_i^j = N-1$ si le score de la classe C_i est placée à la place suivante, etc. On peut donc calculer pour chaque classe de galerie le Borda Count qui cumule les rangs produits par tous les classifieurs selon l'eq.17. Ces Borda counts déterminent donc un nouvel ordre des classes, les classes ayant un grand Borda count étant celles qui sont reconnues par les classifieurs avec plus de confiance.

La méthode du meilleur rang pour la fusion des décisions consiste à attribuer à chacune des classes le rang le plus élevé parmi les rangs proposés par les classifieurs et d'ordonner la liste selon ces rangs :

$$Meilleur\ Rang(C_i) = \max_{j=1}^K r_i^j \quad (\text{eq.18})$$

Pour une classe donnée, ceci revient à faire confiance uniquement au classifieur ayant la plus de certitude pour une classe donnée tout en ignorant informations apportées par les autres classifieurs.

Dans leur étude [Gökberk et al., 2006], ces deux méthodes de fusion se révèlent moins performantes que les méthodes de fusion des scores : la règle de somme simple et la règle de produit.

Les méthodes précédentes traitent les classifieurs de manière égale et ne prennent pas en compte de leur capacité individuelle. Il est proposé dans la littérature des méthodes de fusion

avec pondération des experts. Nous trouvons ainsi dans la littérature des variantes de la méthode de Borda Count, par exemple la somme non linéaire des rangs [Gökberk et al., 2005], la somme des rangs hiérarchiques [Datta et al., 2011], etc. [Monwar & Gavrilova, 2008] utilisent la méthode de régression logistique pour le calcul d'une somme pondérée des rangs individuels. Les poids sont déterminés par régression logistique [Agresti, 1990]. Ils comparent cette méthode de régression linéaire avec la méthode de fusion par simple Borda count à travers un système multimodal qui utilise un classifieur simple du type le plus proche voisin avec la distance euclidienne. La méthode régression logistique a permis d'avoir les meilleures performances avec un taux d'égalité d'erreur (EER) de l'ordre de 9.76 %.

2.3.2.1.3 Fusion au niveau des décisions

Dans cette catégorie, chaque classifieur génère une étiquette de classe unique de galerie. La fusion s'opère alors à travers la règle de vote majoritaire selon l'équation suivante :

$$E(x) = \begin{cases} c_i = \operatorname{argmax}_{c \in \{1, \dots, M\}} \#\{i \in \{1, \dots, K\} : e(i) = c\} & \text{si } c_i \geq \alpha K \\ \text{rejet} & \text{sinon} \end{cases} \quad (\text{eq.19})$$

où K est le nombre de classifieurs à combiner tandis que l'opérateur $\#$ donne la cardinalité d'un ensemble.

Ainsi pour $\alpha = 1$, la classe finale est choisie si tous les classifieurs proposent cette réponse sinon la réponse finale serait le rejet. Cette méthode restrictive qui accepte le moins de risque possible est appelée *majorité unanime*. Pour $\alpha = 0$, il s'agit de la *majorité simple* appelée aussi vote majoritaire où le résultat de la combinaison est la classe de la galerie la plus proposée parmi les K classifieurs. [Gökberk et al., 2005] [Gökberk et al., 2006] utilisent cette dernière méthode dans une étude comparative de plusieurs méthodes de fusion pour la reconnaissance de visages. Une version pondérée de cette méthode, appelée vote majoritaire amélioré, a été proposée dans [Gökberk et al., 2006] [Gökberk et al., 2008]. Ici, la réponse de chaque classifieur est pondérée par un coefficient indiquant son importance dans la combinaison (le poids de confiance). Ce coefficient est calculé en se basant sur les deux premiers rangs générés par chaque classifieur. En effet, une valeur de la pente de la droite formée par les premiers scores relatifs aux premiers rangs est calculée pour chaque classifieur. En cas de conflit, une décision est prise en choisissant la classe qui possède le maximum de confiance. Mais ces méthodes se sont révélées moins performantes que les méthodes de fusion des scores à travers la règle de somme simple et la règle de produit.

2.3.2.2 Fusion en cascade

La fusion en cascade, appelée également, la *fusion séquentielle* ou *en série*, est organisée en niveaux successifs de décision permettant de réduire progressivement le nombre de classes

de la galerie possibles. Dans chaque niveau, il existe un classifieur qui prend en entrée la décision ou la réponse du classifieur placée en amont afin de la confirmer ou de la rejeter. La motivation principale de la fusion en cascade réside principalement dans le gain de temps de recherche en filtrant les décisions des classifieurs. Toutefois, ce type de fusion demeure sensible à l'ordre dans lequel les classifieurs sont placés. En effet, les premiers classifieurs utilisés doivent identifier la solution réelle du visage en entrée, sinon, l'erreur va se propager jusqu'au dernier classifieur.

Dans [Gökberk et al., 2006], une fusion hiérarchique de deux phases a été développée. La première phase consiste à sélectionner les 20 premiers candidats générés par le classifieur propre au descripteur nuage de points 3D en utilisant un algorithme modifié d'ICP. Ces visages candidats sont transférés au deuxième classifieur qui exploite l'information de profondeur projetée dans un sous espace construit par l'analyse discriminante linéaire. Le système génère un taux de reconnaissance moyen de l'ordre de 97.93 % en expérimentant quatre sous bases de la base de visage 3D_RMA³ dont chacune comporte 318 visages dans la base galerie et 193 visages dans la base de test. Cette méthode génère la meilleure performance par rapport à plusieurs autres méthodes, comprenant notamment: la somme simple, la somme pondérée, le produit, la vote majoritaire, la vote majoritaire pondérée. [Mian et al., 2007] proposent également un système de fusion en cascade utilisant deux phases. La première phase consiste à filtrer les classes de la galerie en rejetant celles qui ne seront pas les visages de la galerie dont on est certain de ne pas contenir de réponse. Pour cela, ils utilisent une somme pondérée des scores générés par deux classifieurs différents. Le premier classifieur est basé sur la mise en correspondance de points d'intérêt SIFT sur les images de texture 2D tandis que le deuxième s'appuie sur l'image 3D à travers la présentation sphérique du visage (SFR). La deuxième phase consiste à classifier les visages candidats en s'appuyant sur la somme pondérée d'un nombre défini de scores. Ces derniers sont générés par un algorithme d'appariement de différentes parties de visages en utilisant un algorithme modifié d'ICP. Cette approche multimodale 3D+2D a permis d'atteindre un taux de reconnaissance de 99.3 % qui se révèle légèrement meilleur que l'utilisation seule des informations 3D (98.5 %).

2.3.2.3 Fusion multi-niveaux

La fusion multi-niveaux consiste à combiner des schémas de fusion d'architecture différente, par exemple la combinaison d'un schéma de fusion séquentiel avec un autre qui est plutôt parallèle. Ceci permet d'engendrer un schéma de fusion permettant la réduction de l'ensemble de classes à travers un schéma de fusion séquentiel et la recherche d'un consensus entre les classifieurs à travers un schéma parallèle. Nous trouvons aussi des méthodes de fusion

³ [Beumier and Acheroy 2001]

combinant des méthodes complémentaires (*fusion avant appariement*) avec des méthodes compétitives (*fusion après appariement*). Ce type d'approche de fusion permet de générer de nombreux schémas de combinaison. L'objectif principal de la fusion multi-niveaux est de combiner les descripteurs extraits de l'information suivant différents niveaux.

[Kusuma & Chua, 2008], concatènent l'information 2D issue d'une image en niveau de gris avec l'information de profondeur d'un visage 3D. Ensuite, une ACP est appliquée sur l'image résultante pour construire l'information fusionnée (non corrélée). Deux images résultent de cette opération : une image 2D contenant des informations additionnelles 3D et une image 3D qui contient des informations additionnelles 2D. Chaque nouvelle image est projetée dans le sous espace Fisherface. Pour la classification, deux distances sont générées, en utilisant chaque nouvelle image, entre un visage test et un visage de la galerie à travers la distance euclidienne. La méthode z-score est ensuite appliquée pour normaliser ces deux distances. Celles-ci génèrent finalement une somme pondérée qui est utilisée comme le score final. Finalement, le visage test est assigné à l'une des classes (client ou imposteur) suivant que son score final soit le plus proche de la moyenne des scores de chacune des classes. Cette moyenne a été calculée dans une phase d'apprentissage. Les expériences sont effectuées avec une base de visages composée de 80 sujets dont chacun possède 16 images avec des expressions et des poses différentes. Parmi ces 80 sujets, 60 sujets sont des classes connues dont chaque classe est répartie en 6 images pour l'apprentissage, 5 images pour la validation et 5 images pour le test. Les 20 classes restantes sont considérées comme des classes inconnues avec 8 images pour la validation et 8 images pour le test. Les résultats expérimentaux ont permis d'obtenir un taux d'égale erreur (EER) de l'ordre de 0.85 % dans une application de vérification de visage.

[Li et al., 2007] utilisent les descripteurs 2D issus du filtre de Gabor et l'image en niveau de gris. L'information 3D de visage a été utilisée pour estimer la pose du visage. Ceci permet d'aligner les visages de galerie sur la même pose que le visage de test. Deux approches de fusion ont été évaluées. La première approche concatène les deux vecteurs de descripteurs après une réduction de dimension par ACP. LDA est ensuite appliqué au vecteur résultant pour la mise en correspondance finale. La deuxième approche fusionne à travers une somme pondérée les scores générés par chaque classifieur propre à chaque descripteur. Les poids sont fixés expérimentalement. En s'appuyant sur les résultats expérimentaux des deux approches, les auteurs proposent une troisième approche de fusion qui consiste dans un premier niveau à fusionner les données de la phase de Gabor et de l'amplitude de Gabor dans un vecteur V . Dans un deuxième niveau de fusion, le score généré par la classification de V est fusionné avec celui produit par la classification en se basant sur l'information de l'image en niveau de gris. Un taux de reconnaissance de l'ordre de 81.9 % est obtenu avec une base contenant 160 sujets dans la

base de galerie et 320 images dans la base de test. Ce taux est supérieur au taux de reconnaissance des deux approches précédemment décrites.

2.3.2.4 Conclusion

Dans cette section, nous avons étudié les différentes stratégies de fusion après l'appariement dans le domaine de la reconnaissance de visage 3D. Ces stratégies sont variées, allant de la fusion parallèle à la fusion multi-niveaux en passant par la fusion en cascade. La fusion parallèle peut encore s'opérer au niveau des scores, au niveau des rangs ou encore au niveau des décisions. Le nombre de choix est donc considérable. Si plusieurs études semblent suggérer que la fusion tardive au niveau des scores est plus performante en terme de taux de reconnaissance par rapport à des fusions au niveau des rangs ou de décision, cette conclusion est-elle générale par rapport d'autres possibilités encore, e.g. fusion hybride au niveau des scores et des rangs, ou encore la fusion multi-niveaux ? Nous tenterons d'apporter des réponses à ces questions dans le cadre de cette thèse.

2.4 Conclusion générale

Nous venons de passer en revue les différentes stratégies de fusion qui ont été proposées dans la littérature, notamment dans le domaine de la reconnaissance faciale 3D. Comme nous l'avons présenté dans les sections précédentes, les scans 3D de visage contiennent généralement à la fois les surfaces faciales pour la modalité 3D et les images de texture 2D alignées, et offrent donc un nombre de possibilités de fusion considérable pour optimiser le taux de reconnaissance. Le défi consiste alors à trouver la meilleure stratégie de fusion en terme de taux de reconnaissance dans un espace de stratégies de fusion qui est exponentiel. En effet, si on possède N classifieurs qui utilisent par exemple chacun une représentation faciale spécifique, alors il existe 2^N stratégies de fusion possibles simplement pour les stratégies de fusion tardive si l'on se permet de choisir un sous-ensemble de classifieurs pour fusion. A cela s'ajoute des stratégies précoces ou encore des stratégies de fusion intermédiaires comme la fusion en cascade ou encore la fusion multi-niveaux.

Il est donc impossible d'effectuer une recherche exhaustive dans un tel espace de stratégies de fusion et on est amené à recourir à des méthodes d'optimisation heuristiques pour trouver des stratégies de fusion presque-optimale en terme de taux de reconnaissance. Ceci constitue le cadre général de cette thèse. Dans un premier temps, nous allons étudier et comparer des descripteurs 2D et 3D en vue de la reconnaissance faciale 3D. Nous proposons ensuite une stratégie de fusion adaptative permettant de prendre en compte non seulement la qualité intrinsèque des experts (classifieurs) mais aussi la confiance que chaque expert a dans le

score produit. Enfin, deux méthodes d'optimisation heuristiques, i.e. recuit simulé et génétique, ont été explorées pour la recherche de stratégies de fusion presque-optimale.

CHAPITRE III : RECONNAISSANCE DE VISAGES
PAR UNE METHODE DE FUSION ADAPTATIVE
AU NIVEAU DES SCORES

3. Reconnaissance de visages par une méthode de fusion adaptative au niveau des scores

Les modèles 3D de visage offrent généralement la forme de la surface faciale 3D ainsi que l'image de texture alignée. Ils permettent donc un nombre considérable de fusions d'informations, à travers les modalités 3D et 2D, et pour chacune de ces modalités les différentes représentations faciales ainsi que les différents descripteurs variés qui peuvent en être extraits.

Dans ce chapitre, nous étudions d'abord un ensemble de représentations faciales et descripteurs 3D comparé à leurs équivalents en 2D en vue de la RF 3D plus précisément pour un scénario d'identification 1:N. Cette première étude révèle qu'aucun descripteur 3D ou 2D n'est suffisamment discriminant pour offrir une reconnaissance fiable de visages. Aussi, nous proposons dans un deuxième temps un nouveau schéma de fusion au niveau des scores, appelé *fusion adaptative*, pour explorer la complémentarité d'informations issues de ces descripteurs 3D et 2D. Ce nouveau schéma de fusion s'inscrit dans la lignée des stratégies de fusion au niveau des scores qui, comme nous l'avons vu au chapitre précédent sur l'état de l'art, constitue non seulement un compromis raisonnable entre la préservation d'information pour une prise de décision et la complexité de calcul mais aussi s'avèrent très performantes dans plusieurs études comparatives en RF. L'originalité de notre schéma de fusion adaptative réside dans le fait que la pondération des scores des experts⁴ tienne en compte d'une part la qualité intrinsèque des experts, *e.g.* EER dans ce travail, et d'autre part la confiance que chacun place en ses scores.

Les contributions de nos travaux exposés dans ce chapitre peuvent donc être résumées comme suit :

- Etude et comparaison de descripteurs 3D et 2D en vue de la reconnaissance faciale ainsi que leur sensibilité aux expressions faciales ;
- Proposition d'un nouveau schéma de fusion au niveau des scores, *fusion adaptative*, qui prenne en compte, à travers la règle de somme pondérée, la qualité intrinsèque des experts mais aussi leur propre confiance sur les scores qu'ils ont produits.

Le reste du chapitre est organisé comme suit. Dans la première partie, nous exposons toutes les étapes de la reconnaissance faciale allant de la représentation de l'information faciale à la classification et l'obtention de scores. Ensuite, nous étudions les performances individuelles des descripteurs extraits de l'information 2D et 3D des visages de la base FRGC v2.0, ainsi que leur comportement face aux expressions faciales. Dans la deuxième partie, nous détaillons notre méthode de fusion proposée en exploitant l'analyse effectuée sur les descripteurs. L'évaluation

⁴ Nous appelons un expert l'association d'un descripteur avec un classifieur, *e.g.* SRC, dans ce travail.

de la performance de notre méthode de fusion est effectuée sur la même base de visages. L'étude comparative entre les meilleures performances individuelles et celle obtenue par la fusion permet de les confronter et d'apprécier l'avantage de la fusion.

3.1 Vue d'ensemble de l'approche proposée

Une vue générale de notre schéma de fusion adaptatif est présentée dans la figure 18. Nous y distinguons deux phases : une phase d'apprentissage qui utilise une base d'apprentissage pour évaluer les qualités intrinsèques des experts à travers une métrique appropriée, ici *EER*, (eq.2) puis une phase de test qui utilise notre schéma de fusion adaptatif pour combiner les scores des experts individuels et produire une décision finale dans un scénario d'identification. Les deux phases comprennent les trois étapes suivantes en commun. La première étape est le prétraitement qui permet de filtrer et/ou de corriger les sources d'informations qui présentent différentes formes de dégradation. Par exemple, les modèles de visage 3D générés par les capteurs actuels sont généralement entachés par du bruit et comportent des aberrations et trous. L'extraction des différents descripteurs de l'information de texture 2D et de forme 3D du visage se fait dans la deuxième étape. La troisième étape est la mise en correspondance et la classification de données.

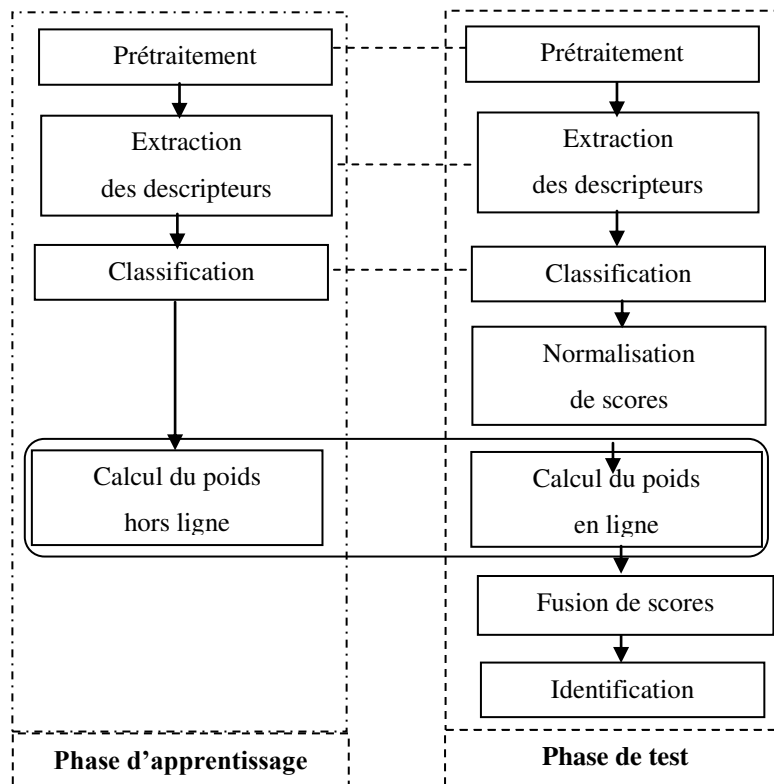


Figure 18. Vue d'ensemble de l'approche proposée

Pour la phase d'apprentissage, chaque expert peut donc être évalué sur une base d'apprentissage, à travers une métrique appropriée, pour apprécier la fiabilité de ses décisions et donner naissance à un poids hors ligne qui le caractérise. Pour la phase de test, après l'étape de la classification par les experts, on normalise leurs scores et calcule aussi un poids en ligne qui caractérise la confiance que chaque expert place en ses propres scores. Les scores des experts sont ensuite fusionnés à travers la règle de la somme pondérée qui tient en compte les poids calculés hors ligne pour chaque expert lors de la phase de l'apprentissage et les poids en ligne. Il s'en suit une décision finale d'identification.

Les résultats des trois premières étapes exposées dans ce chapitre seront réutilisés ultérieurement dans les chapitres suivants.

3.2 Prétraitement des visages

Le prétraitement joue un rôle crucial en reconnaissance faciale car il permet de corriger la représentation brute de visage de ses défauts et d'en normaliser la pose par exemple. Cette normalisation pourrait à son tour, plus ou moins, impacter le résultat final en fonction de la robustesse ou l'invariance des caractéristiques extraites, à la pose.

Dans ce travail, notre objectif est de comparer le caractère distinctif de descripteurs 2D et 3D et d'étudier le bénéfice éventuel de leur fusion, nous avons cherché à réduire au maximum l'impact lié à l'erreur d'alignement. Pour cela, nous avons utilisé 15 points anthropométriques qui ont été manuellement annotés sur tous les visages de la base. La figure 19 présente un exemple de visage avec une image de texture et une image de profondeur annotées manuellement avec ces 15 points. Comme nous pouvons voir, ces différents points localisent les yeux, le nez et la bouche.

Nous disposons donc d'une image de texture 2D recalée avec le modèle 3D (voir figure 19) à la sortie du scanner. Ces modèles 3D présentent d'importants bruits liés aux conditions ou au procédé d'acquisition. Les principales formes de bruits, traités ici, sont les points aberrants et les trous. Aussi, la partie informative de visage est découpée.

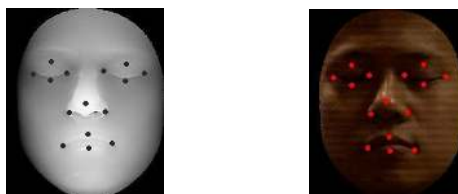


Figure 19. Un exemple de visage prétraité avec ses deux images de profondeur (Gauche) et de texture (Droite) en présence de 15 points annotés manuellement

Les algorithmes utilisés ici sont ceux développés par notre équipe [Szeptycki et al., 2009]. Ils visent à extraire la partie informative du visage, supprimer les points aberrants et combler les trous. La partie informative du visage est celle se plaçant à l'intérieur d'une sphère centrée sur la pointe du nez et d'un rayon de 80 mm. Une technique décisionnelle utilisant le filtre médian et une valeur de seuil est appliquée pour localiser et supprimer les points aberrants qui sont supprimés. Les trous sont localisés en identifiant les points de leur contour comme ayant moins de 8 voisins. Ils sont comblés par l'estimation de la surface à partir des frontières des trous. Plus de détails de ces approches sont fournis dans [Szeptycki et al., 2009].

Vient enfin, la normalisation de pose qui utilise l'algorithme de *Region-Iterative Closest Point* (R-ICP) [Ben Amor et al., 2006]. Pour cela, un modèle de visage frontale avec une expression neutre est sélectionné comme un modèle de référence dans la base de visages 3D. Pour chaque modèle, une région rigide formée par le nez et le front, réputé comme le moins sensible aux variations de l'expression faciale, est extraite pour le processus d'alignement (voir figure 20).

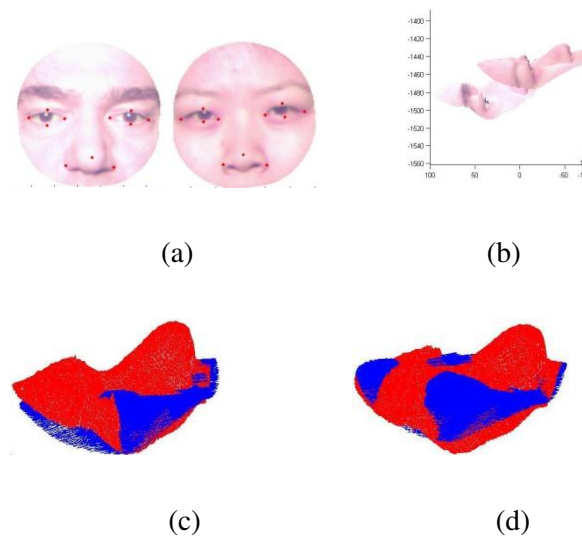


Figure 20. L'alignement, R-ICP (a) les régions rigides de deux échantillons de visage, (b) Superposition des deux modèles de textures 3D (c) une étape d'alignement globale dont chaque modèle 3D de visage est représenté par un couleur différent (d) une étape d'alignement fin

L'alignement s'opère en deux phases, une phase d'alignement grossier suivi d'une phase d'alignement fin. La première phase utilise un ensemble de 11 points, parmi les 15 points annotés manuellement, situé sur la partie supérieure du visage. La méthode SVD est ensuite appliquée pour déduire les paramètres de rotation et de translation 3D pour une transformation rigide. Au niveau de la deuxième phase, « l'*Iterative Closest Point* » (ICP) est exploitée pour la

mise en correspondance des surfaces en plus de l'amélioration de l'estimation des paramètres de rotation et de translation.

3.3 Les descripteurs du visage

Une fois les visages 3D texturés prétraités et normalisés, nous procédons à l'extraction de plusieurs descripteurs 3D et 2D pour étudier leur puissance discriminative ainsi que leur complémentarité dans le cadre d'une fusion. Ces descripteurs sont décrits dans les sections suivantes.

3.3.1 Les descripteurs de texture 2D

Trois types de descripteurs sont extraits des images de texture 2D.

- **Image en niveau de gris** : les images de textures sont converties en niveau de gris, redimensionnées à 80×92 pixels, puis leur histogramme est égalisé afin de normaliser la variation d'éclairage.

- **Banc de filtres de Gabor** : Les filtres de Gabor sont appliqués sur les images en niveaux de gris [Wiskott et al., 1997]. Dans nos travaux, cinq fréquences et huit orientations sont utilisées pour générer le banc de filtres de Gabor. Ainsi, la dimension du vecteur caractéristique est 40 (5×8) fois la dimension des images originales.

- **Image Local Binary Pattern (LBP)** : L'opérateur non-paramétrique LBP est utilisé ici [Ahonen et al., 2004]. Le traitement consiste à diviser d'abord le visage normalisé en pose en plusieurs régions locales. Un histogramme LBP est extrait de chaque région. Par la suite, ces histogrammes sont concaténés, décrivant ainsi des informations à la fois locale et globale de chaque visage.

3.3.2 Les descripteurs de forme 3D

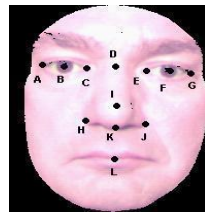
Un modèle 3D de visage décrit l'information de forme d'une surface faciale. Diverses descripteurs de forme peuvent être extraites d'un modèle 3D du visage, telles que les courbures, les vecteurs normaux, les tangents, etc. A ces descripteurs s'ajoutent les représentations brutes comme image de profondeur ou encore le nuage de points.

- **Nuage de points 3D** : Un visage est représenté comme un nuage de points 3D dont les coordonnées sont décrites grâce à trois vecteurs : X, Y, et Z.

- **Vecteurs Normal, Tangent, Binormal** : La surface d'un visage humain peut être considérée comme un exemple d'objet de forme libre représenté par des descripteurs locaux. En utilisant la géométrie différentielle des surfaces [Zucker, 2005], chaque point de la surface faciale est décrit par trois vecteurs à trois dimensions qui sont le vecteur normal, vecteur tangent et vecteur binormal.

- **Courbures Minimum, Maximum, Moyenne, Gaussienne** : Les deux premières courbures utilisées s'appuient sur les directions principales correspondant aux courbures minimales et maximales [Tanaka et al., 1998]. Les deux courbures suivantes, moyenne et gaussienne, sont des courbures dérivées calculées à partir des deux premières courbures.

- **Mesures Anthropométriques** [Ben Soltana et al., 2009]: la méthodologie anthropométrique préconise souvent l'extraction d'une signature à partir d'un certain nombre de points anthropométriques jugés les plus pertinents en terme de discrimination et de stabilité.



A.C.E.G. : Quatre coins des yeux; B.F.: Centres pupilles;
D.: Nasion, K.: subnasal ;
H. J. : Extrémités du Nez ; I. : Pointe du Nez ;
L. : Centre de la bouche

Figure 21. Les points caractéristiques du Visage 3D

Mian et al., montrent que la partie supérieure est plus stable car moins affectée par les expressions faciales [Mian et al., 2006]. Faltemier et al. établissent que la partie du visage qui entoure le nez est très stable [Faltemier et al., 2006] [Faltemier et al., 2008]. Arca et al. utilisent des segments 2D du visage autour des yeux et du nez pour en extraire les profils 3D [Arca et al., 2007]. En se basant sur ces travaux, mais aussi sur le travail de Perrot sur les mesures anthropométriques du visage [Perrot, 1997], nous avons sélectionné un ensemble de points caractéristiques illustrés par la figure 21. Cet ensemble est utilisé pour calculer une signature 3D qui correspond à un vecteur de distances, d'indices et d'angles. Les distances générées représentent les distances entre tous les points caractéristiques 3D du visage. Les angles calculés à partir de l'information 3D du visage correspondent aux angles entre les points *A, D, G, I, L*. L'indice nasale est quant à lui obtenu par l'équation (eq.20). Il représente le rapport entre la largeur horizontale du nez et sa hauteur verticale.

$$Indice\ Nasal = 100 * \left(\frac{dist(H,J)}{dist(D,K)} \right) \quad (eq.20)$$

avec $dist(H,J)$ et $dist(D,K)$, respectivement, les distances euclidiennes entre $[H, J]$ et $[D, K]$.

Les caractéristiques ainsi extraits sont corrélés, l'analyse en composantes principales (ACP) permet de ne retenir que l'information principale.

3.4 Classifieurs

Pour l'étape de classification, deux classifieurs sont mis en œuvre.

Le premier est basé sur l'algorithme du plus proche voisin utilisant la distance euclidienne. Dans ce cas, pour identifier un visage test a , ici un vecteur de caractéristiques d'un visage 3D, la méthode du plus proche voisin consiste à prendre en compte les visages de la base de galerie. Ces visages sont représentés aussi sous forme de vecteur de caractéristiques d'un visage 3D, par exemple $b = [b_1, b_2, \dots, b_m]$, avec m représente le nombre de visages dans la base de galerie. Le visage test a est alors identifié comme un visage i de la base de galerie selon que sa distance est la plus minimale selon la distance euclidienne :

$$d(a, b_i)^2 = (a - b_i)(a - b_i)' \quad (eq.21)$$

Le deuxième classifieur utilise la représentation parcimonieuse (Sparse Representation Classifier - SRC). L'utilisation de cette représentation pour la classification du signal a été proposée par Huang et Aviyente [Huang & Aviyente, 2006]. Elle intègre les propriétés de reconstruction et de discrimination pour une classification robuste. Wright et al. ont proposé une approche générale de reconnaissance d'objets utilisant une minimisation de la norme L1 [Wright et al., 2009] associée à un dictionnaire de données [Xie & Feng, 2009].

Soient k classes distinctes et n_i vecteurs de caractéristiques, nommés $v_{i,j} \in R_m$, disponibles pour l'apprentissage de la $i^{\text{ème}}$ classe, $i = \{1, 2, \dots, k\}$ et j l'indice de l'échantillon, $j = \{1, 2, \dots, n_i\}$.

Nous plaçons dans une matrice A_i toutes les données d'apprentissage de la $i^{\text{ème}}$ classe, appelées aussi atomes de cette classe: $A_i = [v_{i,1}, v_{i,2}, \dots, v_{i,n_i}] \in R_m \times n_i$. Nous développons une matrice dictionnaire, A , pour toutes les classes k par concaténation A_i , $i = 1, 2, \dots, k$ comme suit:

$$A = [A_1, A_2, \dots, A_k] \in R_{m \times n_{j,k}} \quad (eq.22)$$

Un modèle test y peut alors être représenté comme une combinaison linéaire de tous les échantillons d'apprentissage n ($n = n_i \times k$):

$$y = Ax \quad (eq.23)$$

où x est un vecteur de coefficients inconnus.

Intuitivement, l'équation (eq.23) exprime le fait que l'échantillon de test y , ici un vecteur de caractéristiques d'un visage 3D, doit être une combinaison linéaire des atomes de la classe de y , les valeurs de x correspondant à des atomes d'autres classes du dictionnaire A étant non nuls. Ainsi, la solution à l'équation (eq.23) est une solution parcimonieuse et résoudre l'équation (eq.23) permet de trouver la classe du modèle test y . Si la solution de l'équation (eq.23) est suffisamment parcimonieuse, la résolution de celle-ci peut être obtenue par une minimisation avec la norme $L1$ [Candes et al., 2006] [Donoho, 2006] :

$$(L1): x_1 = \arg \min \|x\|_1 ; Ax = y \quad (\text{eq.24})$$

L'avantage de l'équation (eq.24) est qu'elle peut être résolue par des techniques standards de programmation linéaire. En fait, les méthodes de classification utilisant la norme Euclidienne pour trouver le plus proche voisin pourraient être modélisées comme une minimisation similaire à l'équation (eq.24) à la seule différence que l'objective à minimiser est au sens de la norme $L2$. Ainsi, avec la solution x_i de l'équation (eq.24), une erreur résiduelle entre un visage test et chaque visage de la base d'apprentissage est calculée en se basant sur la norme $L2$:

$$r_i = \left\| y - \sum_{j=1}^k x_{i,j} v_{i,j} \right\|_2 \quad (\text{eq.25})$$

L'identité d'un visage test est déterminée comme celle ayant la plus petite valeur du résiduel r_i .

3.5 La fusion adaptative

Nous venons de voir qu'il est possible d'extraire différents descripteurs, 2D ou 3D, d'un scan de visage 3D. Chacun de ces descripteurs associés à un classifieur définit un *expert* qui explore une propriété différente de visages 3D en vue de leur reconnaissance. On est naturellement amené à se poser la question de savoir si leur fusion permet tirer de parti de cette diversité et/ou complémentarité et d'améliorer en conséquence les performances.

Nous l'avons vu au chapitre sur l'état de l'art que la fusion au niveau des scores est une méthode souple et compétitive qui est largement utilisée en FR 3D. Notre méthode de fusion que l'on propose ici est aussi une fusion au niveau des scores. L'idée de base consiste à pondérer les scores produits par un expert donné selon d'une part sa propre qualité intrinsèque mesurée à travers une métrique, ici *EER* (eq.2), et d'autre part la confiance que l'expert place dans les scores qu'il a produits. Pour cela, une règle de somme pondérée avec deux poids a été proposée. Un premier poids hors-ligne est calculé en fonction du taux d'égalité des erreurs (EER) réalisé par un expert sur une base d'apprentissage. Ce poids évalue la qualité intrinsèque de celui-ci. Un deuxième poids est obtenu en-ligne dynamiquement à partir des scores de similarité générés par l'expert donné. Il permet d'évaluer la qualité de décision de chaque

expert. Néanmoins, avant l'application de notre fusion adaptative, les scores générés par un expert doivent être normalisés.

3.5.1 Normalisation des scores

L'opération de normalisation vise à modifier les scores de similarités générées par les experts pour qu'ils aient la même dynamique. Concrètement, soient s un score à partir d'un ensemble S des scores produits par un expert donné et n le score correspondant normalisé. Trois approches de normalisation sont introduites : Min-Max, Z-Score et Tanh :

- Min-Max (MM) [Godil et al., 2005], [Mian et al., 2006], [Snelick et al., 2005] génère des scores appartenant à l'intervalle $[0, 1]$:

$$n = \frac{s - \min(s)}{\max(s) - \min(s)} \quad (\text{eq.26})$$

$\max(s)$ et $\min(s)$ spécifient respectivement le maximum et le minimum de l'ensemble des scores.

- Z-Score (ZS) [Godil et al., 2005], [Snelick et al., 2005] transforme tous les scores correspondant à des nouveaux scores dont leur distribution est caractérisée par une moyenne nulle et un écart-type égale à 1 :

$$n = \frac{s - \text{mean}(s)}{\text{std}(s)} \quad (\text{eq.27})$$

$\text{mean}(s)$ et $\text{std}(s)$ désignent respectivement la moyenne arithmétique et l'écart type.

- Tanh (TH) méthode tangente hyperbolique [Godil et al., 2005], [Snelick et al., 2005] permet de fournir des scores normalisés dans l'intervalle $[0, 1]$:

$$n = \frac{1}{2} \left[\tanh \left(\frac{s - \text{mean}(S)}{\text{std}(S)} \times 0.01 \right) + 1 \right] \quad (\text{eq.28})$$

3.5.2 Méthode de fusion proposée

Notre méthode de fusion proposée se base sur une règle de somme pondérée opérant avec deux poids. Le premier poids reflète le degré de performance de chaque expert. Dans notre cas, le taux d'égalité d'erreur (EER) [Snelick et al., 2005] est utilisé comme mesure de performance. Le deuxième poids vise à évaluer la confiance qu'un expert place en ses scores. Les scores

élevés produit par un expert traduisent une meilleure confiance de celui-ci dans ses décisions et vont recevoir un poids plus élevé en conséquence.

Soit e^m la valeur EER réalisée par un expert m sur une base d'apprentissage, $m = 1, 2, \dots, M$ où M est le nombre total des experts. Le poids P^m , associée à un expert m , est alors calculé comme suit :

$$P^m = \frac{1}{e^m}, \quad \sum_{k=1}^M P^k = 1, \quad 0 \leq P^m \leq 1 \quad (\text{eq.29})$$

Il est à noter que le poids P^m associé à un expert m est inversement proportionnel à son EER. En conséquence, plus la valeur de l'EER est faible, plus la valeur du poids attribuée est grande.

Le deuxième poids est calculé automatiquement à partir d'un ensemble de score de similarité $S_{g,f}$ produit par un expert f . Ce poids est calculé entre un visage test et chaque visage g de la galerie pour un expert utilisant un descripteur f . Concrètement, les scores de similarité par l'expert f sont d'abord triées dans un ordre décroissant. Chaque score $S_{g,f}$ sera pondéré par un poids $w^{g,f}$ qui est fonction de sa position ordonnée $p^{g,f}$ dans la liste triée. Plus précisément, le poids $w^{g,f}$ est défini comme suit :

$$w^{g,f} = f(p) = \ln(N_g / P^{g,f}) \quad (\text{eq.30})$$

Où N_g est le nombre de sujets dans la galerie.

Eq.30 donne donc des poids plus grands aux scores par l'expert qui sont les mieux placés dans la liste des scores triés d'une façon décroissante. Elle favorise les visages de la galerie qui ont obtenu des scores élevés par l'expert en question, sachant qu'un score de similarité traduit la confiance d'un expert sur l'appartenance d'un visage de test à une classe de galerie.

Le score final de correspondance entre un visage g de la galerie et un visage de test est:

$$S_{Final}(g) = \sum_{m \in Experts} P^m \cdot w^{g,m} \cdot S^{g,m} \quad (\text{eq.31})$$

Le visage test est reconnu comme un visage de la galerie possédant le meilleur score obtenu selon l'équation 31.

3.5.3 Variantes de la Méthode de fusion adaptative

A partir de l'eq.31, il est facile d'en dériver deux variantes de la fusion adaptative. La première, appelée dans la suite *somme simple avec la pondération hors-ligne*, utilisera uniquement le poids P^m qui caractérise la qualité intrinsèque de chaque expert :

$$S_{Final}(g) = \sum_{m \in Experts} P^m \cdot S^{g,m} \quad (\text{eq.32})$$

La deuxième variante de la fusion adaptative, appelée dans la suite *somme simple avec la pondération en ligne*, utilisera seulement le poids, $w^{g,m}$, qui caractérise la confiance d'un expert dans ses décisions :

$$S_{Final}(g) = \sum_{m \in Experts} w^{g,m} \cdot S^{g,m} \quad (\text{eq.33})$$

3.6 Les résultats expérimentaux

Dans cette section, nous présentons et analysons les résultats expérimentaux. Nous introduisons d'abord le protocole d'expérimentation, puis comparons les descripteurs 2D et 3D, et finalement analysons notre fusion adaptative en comparaison d'autres schémas de fusion populaires.

3.6.1 Protocole d'expérimentation

Pour nos expériences, les deux bases de visages FRGC v.1.0 et FRGC v.2.0 [Phillips et al., 2005] ont été utilisées: la base FRGC v1.0 pour l'apprentissage et la base FRGC v.2.0 pour le test. La base FRGC v.1.0 contient 900 modèles de visages, vue de face et sans expression, recueillie en 2003. FRGC v.2.0, quant à elle, contient 4007 modèles 3D ainsi que leur image de texture. Ces données représentent 466 sujets avec une répartition en homme de 57 % et en femme de 43 %. La répartition selon l'âge est, 65 % pour les 18-22 ans, 18 % pour les 23-27 ans et 17 % pour les 28 ans ou plus. L'ensemble des données ont été recueillies au cours des années 2003-2004 avec les variations d'expression "neutre" et "non neutre" annotées.

Dans ces deux bases, chaque visage est représenté par un modèle 3D et son image de texture. Le modèle 3D est disponible sous la forme de quatre matrices, chacune de taille 480× 640. La première matrice est un masque binaire indiquant les pixels du visage dans les trois autres matrices qui contiennent respectivement les coordonnées des sommets x, y et z.

La base FRGC v1.0 a permis de former les sous espaces de projection de l'analyse discriminante linéaire (Linear Discriminant Analysis - LDA) en choisissant 116 sujets. Chaque sujet possède 4 modèles de visage avec leurs images de texture. Une réduction de dimensionnalité est opérée sur les descripteurs suivants pour la base FRGC v2.0 : l'image niveau de gris, les vecteurs normaux, bi-normaux, tangents, les courbures gaussiennes, moyennes, minimums, maximums et les mesures anthropométriques.

Afin de constituer la galerie pour la phase de test, un modèle 3D de visage avec une expression neutre et son image de texture de FRGC v2.0 a été choisi parmi les sujets ayant plus de deux modèles de visage 3D, soit 410 sujets (410 = 466 - 56). Les 56 sujets ne possédant

qu'un seul modèle de visage 3D ont été éliminés. Les modèles restants avec leurs images de texture correspondantes ($4007-466 = 3541$) ont été retenus pour former la base de test.

Une première expérience « Neutre vs Tout » est réalisée afin d'étudier l'efficacité des différents descripteurs dans un scénario d'identification. Ensuite, la base de test est divisée en deux sous ensembles en fonction de l'expression faciale. Le premier sous ensemble contient les modèles de visage caractérisés par une expression neutre avec leurs images de texture (1979 modèles). Le deuxième sous ensemble regroupe le reste des modèles et leurs images de texture présentant différentes expressions faciales (1562 modèles). La figure 22 illustre des exemples de visages avec et sans la présence d'expression faciale. Deux nouvelles expériences sont alors effectuées : « Neutre vs Neutre », « Neutre vs Non-neutre ». Le Tableau.1 présente l'organisation de la base FRGC v.2.0.

Tableau 1. Organisation de la base FRGC v.2.0

	Nombre de sujets	Nombre de modèles 3D avec leurs images de texture	Nombre de modèles avec leurs images de texture sans présence d'expression	Nombre de modèles avec leurs images de texture en présence d'expression
Base de galerie	410	410	410	0
Base de test	410	3541	1979	1562

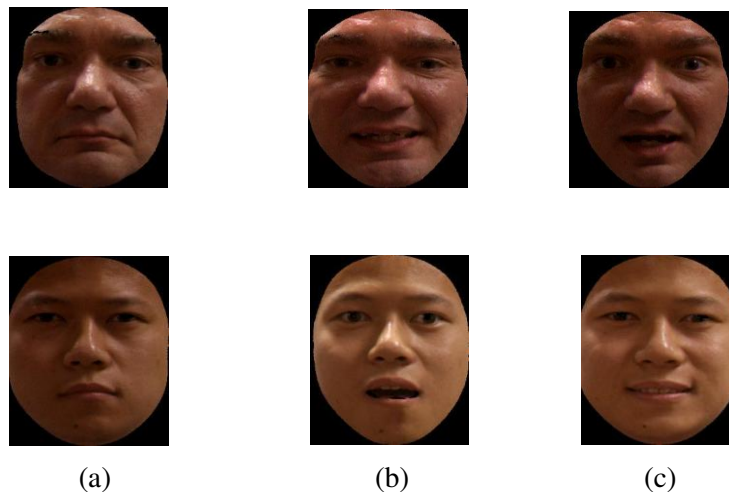


Figure 22. Exemple d'images de texture : la colonne (a) les images avec expression neutre, les colonnes (b), (c) les images en présence d'expressions

Les expériences « Neutre vs Neutre » et « Neutre vs Non-neutre » permettent, en premier lieu, une analyse détaillée des performances des descripteurs extraits face aux variations

de l'expression faciale ; et en deuxième lieu, l'étude du comportement de l'approche de fusion proposée à ces mêmes variations.

3.6.2 Analyse des performances individuelles des descripteurs

Nous analysons le comportement individuel des descripteurs dans un scénario d'identification. Une première expérience « Neutre vs Tout » est mise en œuvre avec les deux classifieurs : le plus proche voisin en utilisant la distance euclidienne (NN) (eq. 21) et le classifieur SRC (eq. 25). Le Tableau.2 expose la performance individuelle de chaque descripteur en terme de taux d'identification (eq.1) sur la base FRGC v.2.0. Pour le descripteur Points 3D, la similarité est calculée directement avec l'algorithme *ICP* en utilisant la distance euclidienne.

Tableau 2. Taux d'identification individuel avec la base FRGC v2.0 - Neutre vs Tous

	Taux d'identification avec SRC	Taux d'identification avec NN
Filtres de Gabor	77.89 %	57.92 %
LBP	71.82 %	47.64 %
Image d'intensité	49.82 %	40.30 %
Mesures Anthropométriques	46.48 %	42.60 %
Les Points 3D	-	58.57 %
Courbure Gaussienne	59.02 %	56.28 %
Courbure Moyenne	71.62 %	70.87 %
Courbure Maximale	67.81 %	66.82 %
Courbure Minimum	66.73 %	66.25 %
Vecteurs Bi-normaux	70.63 %	69.78 %
Vecteurs Normaux	70.01 %	69.26 %
Vecteurs tangents	79.72 %	79.13 %

D'après ce tableau, de meilleures performances sont obtenues par le classifieur SRC pour l'expérience « Neutre vs Tous ». L'écart par rapport au classifieur NN est particulièrement important pour la modalité 2D. Ces résultats montrent qu'aucune des deux modalités n'est suffisamment discriminante pour une reconnaissance fiable de visages. En effet, le meilleur descripteur 3D, vecteurs tangents, permet d'atteindre seulement 79.72 % de taux d'identification, tandis que le meilleur taux d'identification, en 2D, est obtenu par les filtres de

Gabor avec 77.89 %. Aussi, au sein de chaque modalité, la variation des performances est importante.

Par la suite, nous examinons, à travers les deux tableaux suivants, le comportement des différents descripteurs proposés contre les variations de l'expression faciale. Le Tableau 3 présente en détail les performances de chaque descripteur dans le cas du « Neutre vs Neutre » en terme de taux d'identification (eq.1).

Tableau 3. Taux d'identification individuel avec la base FRGC v2.0 - Neutre vs Neutre

	Taux d'identification avec SRC	Taux d'identification avec NN
Filtres de Gabor	84.59 %	62.81 %
LBP	78.83 %	54.53 %
Image D'intensité	56.85 %	45.65 %
Mesures Anthropométriques	55.99 %	51.58 %
Les Points 3D	-	72.26 %
Courbure Gaussienne	73.67 %	70.88 %
Courbure Moyenne	85.25 %	83.74 %
Courbure Maximale	82.57 %	81.44 %
Courbure Minimum	81.05 %	80.85 %
Vecteurs Bi-normaux	84.49 %	83.29 %
Vecteurs Normaux	83.78 %	82.57 %
Vecteurs tangents	89.64 %	88.86 %

Comme nous pouvons voir, tous les descripteurs génèrent une performance nettement supérieure à l'expérience « Neutre vs Tout ». Néanmoins, aucun de ces descripteurs 3D ou 2D n'affiche un résultat suffisamment discriminant pour la reconnaissance faciale. Le meilleur taux d'identification de l'ordre de 89.64 % est obtenu avec les vecteurs tangents.

Avec un gain en moyenne de 10 points, les descripteurs 3D améliorent significativement leur performance, confirmant ainsi notre intuition que les descripteurs de la forme 3D sont beaucoup plus sensibles aux expressions faciales que les descripteurs 2D. Cette intuition est confirmée une fois de plus par le Tableau 4 qui affiche les résultats de reconnaissance dans le cas du « Neutre vs non Neutre ». La dégradation des performances est beaucoup plus importante pour les descripteurs 3D que pour les descripteurs 2D.

Tableau 4. Taux d'identification individuel avec la base FRGC v2.0 - Neutre vs non Neutre

	Taux d'identification avec SRC	Taux d'identification avec NN
Filtres de Gabor	69.4 %	49.49 %
LBP	63.06 %	41.68%
Image D'intensité	40.91 %	30.65 %
Mesures Anthropométriques	34.44 %	31.24 %
Les Points 3D	-	41.23 %
Courbure Gaussienne	40.46 %	37.85 %
Courbure Moyenne	54.35 %	53.63 %
Courbure Maximale	49.1 %	48.23 %
Courbure Minimum	48.59 %	47.96 %
Vecteurs Bi-normaux	53.07 %	52.78 %
Vecteurs Normaux	52.56 %	51.87 %
Vecteurs tangents	67.16 %	66.73 %

Le comportement non discriminant dégagé par les descripteurs ainsi que leur sensibilité aux variations d'expressions, nous a mené à proposer la stratégie de fusion adaptative. Cette stratégie vise à favoriser une complémentarité entre les différents classifieurs propres à chaque descripteur. Ce qui permet l'obtention d'un système plus performant et plus robuste aux variations d'expression. Par la suite, nous examinons les avantages potentiels de notre méthode de fusion des informations 2D et 3D.

3.6.3 Résultats expérimentaux de la méthode de fusion des scores

Comme nous venons de constater, les descripteurs individuels 2D et 3D se révèlent loin d'être discriminatifs pour la reconnaissance faciale, affichant des comportements différents au regard des expressions faciales. Il serait donc intéressant de voir si la combinaison de ces descripteurs à travers une méthode de fusion permet de fournir un système plus robuste et plus efficace. Pour cela, nous avons expérimenté plusieurs méthodes de fusion au niveau des scores, fusion au niveau du rang, règle de produit, règle de somme simple, en comparaison de notre méthode de fusion adaptative (eq.31), ses variantes comprises, i.e. somme simple avec la pondération hors-ligne (eq.32), somme simple avec pondération en ligne (eq.33), qui ont été proposées dans ce chapitre. Le même protocole d'expérimentation a été mis en œuvre avec trois méthodes de normalisation différentes.

Tableau 5. Taux d'identification dans le cas « Neutre vs Tous » avec trois méthodes de normalisation et différents schémas de fusion

Méthodes de Normalisation	Max-Min	Z-score	Tanh
Méthodes de Fusion			
Fusion au niveau du rang	88.87 %		
Règle de Produit	51.60 %	67.64 %	27.40 %
Règle de Somme Simple	87.49 %	85.12 %	86.44 %
Somme Simple avec pondération hors-ligne	93.98 %	91.89 %	89.33 %
Somme Simple avec pondération en-ligne	90.45 %	85.23 %	89.69 %
Fusion Adaptative	95.48 %	92.01 %	94.69 %

Comme nous pouvons le conclure, d'après le Tableau 5, toutes les stratégies de fusion, à l'exception de la règle du produit, permettent d'atteindre des taux d'identifications nettement supérieures aux performances individuelles affichées par les descripteurs 2D (77.89 %) et 3D (79.72 %) en utilisant la normalisation Max-Min. Ces résultats confirment que les descripteurs 3D et 2D sont des modalités complémentaires dont une fusion adéquate et convenable peut améliorer la décision finale.

Le tableau 5 montre aussi l'importance de pondérer les scores des différents experts à travers les trois variantes de notre méthode de fusion proposée. En prenant en compte la confiance de chaque expert dans ses scores, la variante somme simple avec pondération enligne améliore de 3 points le taux de reconnaissance par la règle de la somme simple des scores. Maintenant en prenant en compte la qualité intrinsèque de chaque expert, la variante Somme simple avec pondération hors-ligne améliore plus de 6 points le taux d'identification par la règle de la somme simple. En combinant la pondération enligne et celle hors-ligne, notre méthode de fusion adaptative affiche un taux d'identification de 95,48% et améliore par 8 points le taux d'identification de la règle de somme simple. Remarquons que cette performance est aussi nettement supérieure à celle des méthodes de fusion au niveau du rang ou encore par la règle de produit.

Le tableau 5 permet aussi une autre comparaison intéressante qui consiste à comparer notre méthode de fusion adaptative avec ses deux variantes, à savoir la somme simple utilisant seulement les poids hors-ligne et la somme simple utilisant seulement les poids en-ligne. Comme nous pouvons voir, la méthode de fusion adaptative qui utilise la combinaison des deux poids réagit mieux que ses deux variantes. Nous remarquons une amélioration de 5.03 % de taux d'identification par rapport aux résultats affichés par la règle de somme pondérée utilisant le poids en-ligne, ainsi qu'un gain de 1.5 % est détecté par rapport aux performances de la règle de somme pondérée utilisant le poids hors-ligne.

Comme le montre le Tableau 6, nous avons aussi évalué la robustesse des descripteurs 2D ou 3D fusionnés par rapport aux expressions faciales. Comme nous pouvons y voir, la fusion de descripteurs 3D seule puis 2D seule par notre méthode de fusion adaptative génère une performance de 84.07% et de 85.31% dans le cas de Neutre vs. Tous, alors que la fusion des descripteurs 3D et 2D en même temps permet d'améliorer ces performances de 10 points et affiche un taux d'identification de 95.48%, ce qui démontre une fois de plus la complémentarité des informations véhiculées par les descripteurs 3D et 2D. Dans le cas du scénario du Neutre vs Neutre, nous observons d'abord que l'utilisation d'informations de forme à travers les descripteurs 3D a permis un meilleur taux de reconnaissance (91.61%) par rapport à sa contrepartie 2D (89.54%), mais dans le cas du scénario Neutre vs Non neutre, comme nous pouvons y attendre, les descripteurs 3D et 2D enregistrent une chute de performance très importante mais les descripteurs 3D se révèlent bien plus sensibles aux expressions faciales par rapport à leur contreparties 2D avec un écart de performance de 8 points (67.99% vs 76.63%). Nous observons encore une fois de plus la complémentarité entre les descripteurs 3D et 2D à travers notre fusion adaptative dans ces deux scénarios car leur fusion a permis une amélioration significative de performance, que ce soit avec le scénario Neutre vs. Neutre (7 points) ou encore le scénario Neutre vs Non Neutre (14 points).

Tableau 6. Taux d'identification des modalités 2D, 3D et leur fusion adaptative

	Neutre vs Tous	Neutre vs Neutre	Neutre vs Non-neutre
3D	84.07 %	91.61 %	67.99 %
2D	85.31 %	89.54 %	76.63 %
2D+3D	95.48 %	98.64 %	90.65 %

Tableau 7. Comparaison avec d'autres méthodes de fusion dans le cas « Neutre vs Tous »

	Taux d'identification Neutre vs. Tous
La fusion adaptative	95.48 %
Gökberk et al., 2008	95.45 %
Mian et al., 2008	96.10 %

Dans le tableau 7, nous avons aussi comparé nos résultats avec des méthodes de l'état de l'art utilisant la fusion. On voit que notre fusion adaptative de scores affiche des performances comparables à celles de [Mian et al., 2008] et [Gökberk et al., 2008]. Ces deux algorithmes proposés sont évalués sur la base de visages FRGC v.2.0 et en adoptant le même protocole

expérimental dans le cas « Neutre vs Tous ». [Mian et al., 2008] détectent des points clés dans chaque visage de la base afin d'en extraire des descripteurs locaux (cf. sections 2.3.2.1 et 2.3.2.1.1). [Gökberk et al., 2008] utilisent plusieurs descripteurs 2D et 3D du visage. Ils s'inspirent d'une méthode de sélection de caractéristiques afin de sélectionner les meilleurs descripteurs pour l'identification (cf. section 2.2.2.4). En analysant les résultats de fusion, la méthode de fusion proposée par [Gökberk et al., 2008] affiche un taux d'identification de 95.45% et apporte une amélioration de 3.57 % sur la performance du meilleur classifieur individuel qui est de l'ordre de 91.88 %. Dans [Mian et al., 2008], la méthode de fusion proposée affiche une performance de 96.1% et améliore de 2.6% la performance du meilleur classifieur individuel qui est de 93.5 %. Dans nos résultats, la performance du meilleur classifieur est de 79.72 % et notre méthode de fusion tardive améliore celle-ci de 15% pour atteindre un taux d'identification de 95.48 %. Néanmoins, cette comparaison ne peut être qu'indicatif car le gain de performance dépend beaucoup de classifieurs individuels mis en jeu et de leur complémentarité.

3.7 Conclusion

Dans ce chapitre, nous avons d'abord comparé les descripteurs 2D et 3D pour le problème de la reconnaissance faciale, notamment en terme de robustesse vis-à-vis d'expressions faciales. Nous avons montré sur la base FRGC v2.0 qu'aucun des descripteurs 2D ou 3D n'est suffisamment discriminant pour une reconnaissance fiable du visage, notamment en présence d'expressions faciales. Si les descripteurs 3D, en capturant précisément la forme 3D de visage, permettent une meilleure reconnaissance faciale que leurs homologues 2D dans le cas des expressions neutres, ils s'avèrent bien plus sensibles que les descripteurs 2D en présence d'expressions faciales. Ces avantages et défauts différenciés entre les descripteurs 2D et 3D nous ont conduit à proposer une méthode de fusion adaptative au niveau de scores pour prendre en compte cette complémentarité entre les différents classifieurs propres à chaque descripteur. Les expériences sur la base FRGC v2.0 montrent l'efficacité de notre méthode de fusion en améliorant significativement la performance des experts (classifieurs) pris individuellement.

Néanmoins, faut-il toujours fusionner tous les experts pouvant avoir des comportements très disparates dans la reconnaissance faciale ? Nos expériences montrent que la fusion de tous les experts peut conduire à une chute de performance, raison du bruit ou redondance d'information. Nous proposons d'étudier cette question au chapitre suivant.

CHAPITRE IV : RECONNAISSANCE DE VISAGES
PAR UNE METHODE HEURISTIQUE DE FUSION
AU NIVEAU DES SCORES

4. Reconnaissance de visages par une méthode heuristique de fusion au niveau des scores

La fusion de plusieurs experts pour la prise de décision peut conduire à une amélioration de la performance d'un système, dans notre cas le système de reconnaissance de visages 3D. Néanmoins, la prise en compte de tous les experts ne conduit pas nécessairement à la meilleure stratégie de fusion. En effet, les experts peuvent être redondants ou parfois en conflit entre eux pour une prise de décision. Il peut être parfois intéressant de n'en sélectionner qu'un sous-ensemble.

Dans ce chapitre, nous proposons d'étudier l'apport d'une étape de sélection d'experts à notre stratégie de fusion adaptative étudiée au chapitre précédent dans un scénario d'identification. L'objectif est de définir une stratégie de fusion presque-optimale dans un espace de stratégies de fusion qui est exponentiel. En effet, avec N experts en compétition, nous aurons 2^N stratégies de fusion possibles. Comme une recherche exhaustive de stratégies de fusion optimales dans un tel espace est impraticable, nous nous sommes orientés vers des techniques d'optimisation heuristique, en occurrence l'algorithme du recuit simulé qui a fait ses preuves dans des nombreux domaines applicatifs [Sadegheih, 2011] [Zhao, 2011], en vue d'aboutir à des stratégies de fusion presque-optimales. Pour cela, nous proposons d'étudier la sélection d'experts en combinaison avec les différentes stratégies de fusion, notamment les variantes de notre fusion adaptative proposée au chapitre précédent, et aboutissons à des nouveaux schémas de fusion qui sont aussi comparés avec les stratégies de fusion sans la sélection d'experts. Au final, nous étudions aussi la stabilité de ces schémas de fusion par rapport à l'ajout de sujets dans la base d'apprentissage.

Les contributions de nos travaux exposés dans ce chapitre peuvent donc être résumées comme suit :

- Proposition de nouveaux schémas de fusion en combinant un algorithme de sélection d'experts, le recuit simulé, à des stratégies de fusion tardives au niveau score pour un système de reconnaissance faciale 3D ;
- Etude et comparaison de ces nouveaux schémas de fusion avec leurs équivalents sans la sélection d'experts;
- Etude du comportement de ces schémas par rapport à la modification de la base d'apprentissage.

Le présent chapitre est organisé comme suit. Dans la première partie, nous exposons la problématique sur la sélection de stratégies de fusion optimales dans un espace des stratégies de fusion dont la cardinalité est exponentielle. Ensuite, nous présentons notre solution pour la recherche de la stratégie de fusion presque-optimale utilisant l'algorithme du recuit simulé pour

notre problème de la reconnaissance faciale 3D. Finalement, nous évaluons notre solution sur la base FRGC v2.0 en fonction de deux paramètres essentiels : la règle de fusion et la taille de la base d'apprentissage. L'étude comparative des différentes expériences permet d'apprécier l'avantage de l'utilisation d'un algorithme de sélection d'experts, et la mise en valeur de notre schéma de pondération adaptatif.

4.1 Mélange d'experts gardés (Mixture of Gated Experts)

Le mélange des experts gardés (mixture of gated experts) modélise la densité de probabilité globale (pdf) d'un système comme étant une somme pondérée des pdfs des experts individuels. Dans sa forme la plus générale, ce mélange d'experts gardés est décrit par :

$$p(t/x_1 \dots x_K) = \sum_{k=1}^K \pi_k(x_k) p_k(t/x_k) \quad (\text{eq.34})$$

où $p_k(t|x_k)$ représente la probabilité postérieure du $k^{\text{ème}}$ expert pour une valeur donnée t suivant le descripteur x_k , et les coefficients $\pi_k(x_k)$ sont les fonctions de pondération connus sous « *gating functions* » [Jacobs et al., 1991].

Pour notre problème de reconnaissance de visages 3D, les experts sont les mesures de similarité s_k calculées sur des descripteurs 3D ou 2D. La formule (eq.34) peut alors être exprimée de la manière suivante :

$$s(A, B) = \sum_{k=1}^K \pi_k s_k(x_k^A, x_k^B) \quad (\text{eq.35})$$

où π_k représente le poids associé à chaque s_k . s_k étant la $k^{\text{ème}}$ mesure de similarité suivant le descripteur x_k extrait respectivement des sujets A et B. Lorsque π_k est fixé à la constante 1, nous obtenons la règle bien connue de la somme simple pour la fusion. Lorsque l'on limite π_k dans l'intervalle $[0,1]$ de telle façon que leur addition somme à 1, nous obtenons la règle de la somme pondérée pour la fusion. Enfin, si l'on limite π_k à la fonction indicatrice binaire 0 ou 1, donc permettant à des experts d'être absents de la fusion, nous aboutissons à 2^N combinaisons possibles d'experts. Comme une recherche exhaustive dans un tel espace est impossible, il faut que l'on fasse appel à des solutions heuristiques pour trouver la meilleure combinaison d'experts.

4.2 Optimisation du mélange d'experts gardés par l'algorithme du recuit simulé

Nous avons introduit au chapitre précédent un schéma de fusion adaptative qui se révèle être efficace pour la fusion des experts utilisant des descripteurs 3D et 2D. Ce schéma de fusion est une fusion au niveau de scores à travers la règle de la somme pondérée (cf. eq.31) et prend en compte d'une part la qualité de chaque expert, pqe_{off}^k (eq.29), évalué sur une base d'apprentissage, et la confiance, $pde_{on}^{A,B}$ (eq.30), que chaque expert place dans ses scores. En permettant une sélection d'experts avant une telle fusion adaptative, on aboutit à un schéma de pondération adaptative (SPA) et la règle de la fusion adaptative par (eq.31) se transforme à l'équation suivante :

$$s(A, B) = \sum_{k=1}^N \pi_k pqe_{off}^k \cdot pde_{on}^{A,B} \cdot s_k(x_k^A, x_k^B) \quad (\text{eq.36})$$

Où π_k est la fonction indicative binaire qui prend la valeur 0 si le k-ième expert n'est pas sélectionné et 1 dans le cas contraire. Le problème consiste maintenant à rechercher le schéma optimal de fusion des N experts en se basant sur le schéma de pondération adaptatif (eq.36).

Le schéma de pondération adaptatif défini par la formule (eq.36) permet la combinaison de tous les sous-ensembles d'experts en assignant à chacun un poids adaptatif. Le nombre de stratégies de fusion est donc exponentiellement proportionnel au nombre d'experts N mis en jeu. Comme la recherche exhaustive dans un tel espace est impossible, nous proposons de recourir à des méthodes heuristiques, en occurrence l'algorithme du recuit simulé, afin de trouver une stratégie de fusion presque-optimale.

4.2.1 Sélection dynamique des experts

Le recuit simulé (RS) est un algorithme de recherche stochastique inspiré de la dynamique des systèmes en physique d'abord introduit par [Metropolis et al., 1953] puis démocratisé par [Kirkpatrick et al., 1983]. Il a prouvé son efficacité dans la résolution de nombreux problèmes d'optimisation [Sadegheih, 2011] [Zhao, 2011]. Dans notre contexte de reconnaissance de visages, le RS cherche le meilleur sous-ensemble d'experts parmi les différentes solutions permises au cours d'une phase d'apprentissage. Dans son processus de recherche de solutions, il se permet d'accepter des solutions moins performantes afin de minimiser le risque d'être tenu dans un optimum local.

Cet algorithme est caractérisé par deux types de paramètres : les paramètres généraux et les paramètres spécifiques. La figure 23 présente l'algorithme avec ces différents paramètres ainsi que les étapes de recherche.

Les paramètres généraux fixent la température initiale T , la fonction de refroidissement α et les conditions d'arrêt. Plus précisément, la température T contrôle l'évolution de la recherche de la solution s . Une haute température signifie que l'algorithme est libre de se déplacer dans l'espace des solutions en choisissant des solutions moins performantes. A basse température, les meilleures solutions sont choisies, mais d'autres peuvent être acceptées, empêchant ainsi l'algorithme de tomber dans un optimum local. La fonction de refroidissement α permet la génération de nouvelles températures. Enfin, les conditions d'arrêt stoppent la procédure de recherche. Elles peuvent se résumer en deux conditions suivantes. La première condition est la température finale qui est fixée avant le lancement de l'algorithme. La seconde condition est que l'algorithme ne trouve plus de meilleures solutions.

Les paramètres spécifiques concernent l'espace des solutions S , le voisinage des solutions $V(S)$ et la fonction d'énergie f . Dans ce cas, la fonction d'énergie quantifie la qualité d'une solution s .

```

Sélection de la solution initiale  $s$  ;
Sélection de la température initiale  $T > 0$ ;
Définition de la fonction  $\alpha$  pour la réduction de la température
Répéter
  Répéter
    Sélection  $s' \in V(s)$ ;
     $\Delta = f(s') - f(s)$ ;
    si  $\Delta \geq 0$  alors  $s' = s$  ;
    sinon
      Générer aléatoirement  $R \in (0, 1)$ ;
      si  $R < \exp(\Delta / (k * T))$  alors  $s' = s$ ;
  Jusqu'à compteur_itération =  $n_{rep}$ ;
   $T = \alpha(T)$ ;
Jusqu'à condition d'arrêt.

```

Figure 23. Algorithme du recuit simulé

Dans son fonctionnement, l'algorithme du recuit simulé (RS) recherche une solution optimale par itération. La solution initiale s est prise au hasard dans l'espace des solutions possibles. Sa pertinence est calculée en fonction du critère que l'on cherche à optimiser (fonction d'énergie f). Une température initiale T élevée est également choisie. A chaque itération, une nouvelle solution s' est produite dans le voisinage $V(s)$ de la solution actuelle s . La meilleure solution selon l'objectif Δ ($\Delta \geq 0$, $\Delta = f(s') - f(s)$) est retenue. Le choix d'une mauvaise solution ($\Delta < 0$) est autorisé ou envisageable, avec une fréquence qui est régie par la

fonction de probabilité suivante : $R < \exp(\Delta/T)$. Cette fonction dépend de (T, Δ) avec R qui est une valeur tirée aléatoirement de l'intervalle $[0, 1]$.

Après un nombre d'itération prédéfini, la température est progressivement réduite par la fonction $\alpha(T) = \alpha(T)$. Ceci permet de diminuer progressivement le taux d'acceptation des mauvaises solutions. L'algorithme s'arrête si la température atteint un seuil fixé à l'avance ou lorsque le système devient stable en générant la même solution. La convergence vers une solution globale optimale est en relation avec les deux paramètres : la fonction de refroidissement α et le nombre d'itération pour chaque température [Rayward-Smith et al., 1996]. Dans la suite, nous explicitons l'application du RS à notre problème d'optimisation du mélange d'experts gardés en vue de la reconnaissance de visages 3D.

4.2.2 Application de l'algorithme du recuit simulé à l'optimisation du mélange d'experts gardés dans un système de reconnaissance de visages 3D

La recherche d'un schéma de fusion (presque)-optimal des N experts (eq.34) est assurée par l'algorithme du recuit simulé. Dans ce cas, chaque expert (ou mesure de similarité) est codé suivant une valeur binaire simple.

Un ensemble de N experts en compétition pour un schéma de fusion s est codée suivant un vecteur à valeurs binaires de taille N . La valeur "1" à la position i indique que le $i^{\text{ème}}$ expert (i -ème mesure de similarité) est actif ou encore sélectionné dans la stratégie de fusion, par contre la valeur "0" indique son absence. Dans notre travail, nous avons considérés $N= 12$ experts. Cela signifie que nous possédons $2^{12}= 4096$ stratégies de fusion possibles. Par exemple, un schéma de fusion $s= [1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0]$ indique que les 1^{er}, 3^{ème} et 4^{ème} mesures de similarité sont sélectionnées dans l'équation (eq.34).

L'algorithme de base de recuit simulé maintient à la fois un état représenté par la solution actuelle s et une température T . A chaque itération, une nouvelle solution s' est générée dans le voisinage $V(s)$ de la solution actuelle s en se basant sur de multiples mutations aléatoires. La qualité des deux solutions (s et s') est évaluée en utilisant une fonction d'énergie f . Dans le scénario de reconnaissance de visages 3D, différentes fonctions d'énergie peuvent être envisagées, comme par exemple le taux d'identification, le taux d'égalité des erreurs, le taux de fausse acceptation et le taux de faux rejet, etc. Dans notre travail, le taux d'identification (eq.1) est adopté comme fonction d'énergie. Une fois un schéma de fusion s fixé, il est appliqué sur une base de test.

4.3 Les mesures de similarité (les experts)

Les experts utilisés pour la reconnaissance faciale 3D ont été introduits au chapitre précédent. Nous les rappelons rapidement ici. Une fois les visages 3D texturés prétraités et normalisés, nous procédons à l'extraction de plusieurs descripteurs 3D et 2D. Ces descripteurs sont décrits dans la section 3.3. Etant donné deux visages A et B, leur mesure de similarité $s_k(A, B)$ est calculée suivant un descripteur donné k , appartenant à l'espace des descripteurs. Le classifieur SRC (section 3.4) est alors utilisé pour générer différentes mesures de similarité.

Dans ce travail, nous avons utilisé 12 mesures de similarité calculées à partir de 12 descripteurs. Trois types de descripteurs sont extraits des images de texture 2D : Image en niveau de gris, Banc de filtres de Gabor et Image Local Binary Pattern (LBP). Pour l'information de forme 3D, divers descripteurs peuvent être extraits telles que les courbures minimum, maximum, moyenne et gaussienne; les vecteurs normaux, les vecteurs tangents, les vecteurs bi-normaux et les mesures anthropométriques extraites à partir d'un certain nombre de points anthropométriques. A ces descripteurs s'ajoutent les représentations brutes comme l'image de profondeur ou encore le nuage de points 3D.

Au final, les douze différents experts en compétition pour un schéma de fusion sont organisés comme suit : filtres de Gabor, LBP, Image en niveau de gris, mesures anthropométriques, points 3D, courbure gaussienne, courbure moyenne, courbure maximum, courbure minimum, vecteurs bi-normaux, vecteurs normaux, vecteurs tangents.

4.4 Protocole Expérimental

Afin de mettre en évidence l'intérêt de sélection d'experts à travers le recuit simulé, nous avons conçu plusieurs scénarios expérimentaux utilisant la base de visage FRGC v2.0 dans un scénario d'identification.

4.4.1 Base de visages

Dans nos expériences, nous avons utilisé la base de visages FRGC v2.0 (voir la section 3.6.1) avec une base de galerie (410 personnes) et une base de test (3541 modèles). La stabilité de la solution générée par chaque expérience est étudiée par rapport à l'ajout de nouveaux sujets dans la base d'apprentissage. Pour cela, les 3541 modèles de visage 3D avec leurs images de texture dans la base de test ont été divisés aléatoirement en deux bases (B1) et (B2). Chaque base est composée de 5 ensembles de modèles avec leurs images de texture correspondantes: l'ensemble (P1) avec 60% des sujets de la base, et les ensembles (P2), (P3), (P4) et (P5) avec 10% des sujets restants. Chaque ensemble contient des modèles de visages avec différentes expressions. Cette décomposition est illustrée dans le Tableau 8 avec les différents ensembles.

Tableau 8. Organisation de la base FRGC v.2.0

		Nombre de sujets	Nombre de modèles 3D avec leurs images de texture
Base (B1) 172 sujets	(P1)	103	989
	(P2)	16	179
	(P3)	16	176
	(P4)	16	171
	(P5)	21	184
Base (B2) 238 sujets	(P1)	134	1122
	(P2)	24	181
	(P3)	16	186
	(P4)	31	178
	(P5)	33	174

4.4.2 Les scénarios d'expérimentation

Pour évaluer notre méthode, quatre scénarios d'expériences ont été définis dans un scénario d'identification en utilisant les différentes variantes du schéma de pondération adaptatif (SPA) défini par (eq.36). La première expérience étudie l'intérêt d'avoir une sélection d'experts par rapport à la règle de somme simple (eq.35) où tous les experts ont des poids égaux. La deuxième expérience étudie l'intérêt de sélectionner un sous-ensemble d'experts à travers le schéma de pondération adaptatif utilisant uniquement le poids de la décision de l'expert (eq.30) tout en ignorant le poids de la qualité de l'expert (eq.29). Le troisième scénario étudie cette fois-ci l'intérêt de sélectionner des experts à travers le schéma de pondération adaptatif qui utilise uniquement le poids de la qualité de l'expert (eq.29). Enfin, la quatrième expérience évalue la performance de notre schéma de pondération adaptatif (SPA) proposée en combinaison avec la sélection d'experts. Pour ces différentes expériences, le recuit simulé (RS) est utilisé pour la sélection dynamique des experts comme introduit dans la section 4.2.

Dans chaque scénario d'expériences, la première stratégie de fusion est générée en utilisant l'ensemble P1 (voir Tableau 8). Les autres ensembles sont ajoutés les uns après les autres à l'ensemble P1. Nous obtenons finalement pour chaque expérience une base de test et 5 bases d'apprentissage différentes: (1) P1 ; (2) P1+ P2 ; (3) P1+ P2 + P3 ; (4) P1+ P2 + P3 + P4 ; (5) P1+ P2 + P3 + P4 + P5. Dans un premier temps, la base B1 est utilisée comme base d'apprentissage et la base B2 comme base de test. Dans un deuxième temps, les rôles des deux bases sont inversés.

4.4.3 Détermination des paramètres de l'algorithme RS (température initiale (T) et la fonction de refroidissement (α))

Dès lors que le recuit simulé est utilisé dans la recherche d'une stratégie de fusion optimale, les paramètres : température initiale (T) et fonction de refroidissement (α) doivent être

fixés comme dans la section 4.2. Pour cela, nous avons utilisé les deux premiers ensembles P1 des deux bases B1 et B2 de la première expérience qui met en évidence l'intérêt de sélectionner les experts à travers la règle de somme simple. Différentes valeurs ont été choisies pour $(T) = \{1, 100, 1000\}$ et $(\alpha) = \{0.01, 0.1, 0.6, 0.9\}$. En testant les différents pairs de valeurs $\{T, \alpha\}$, nous avons obtenu deux taux d'identification (eq.1) de l'ordre de 95.15 % et 92.34 % respectivement pour les bases B1 et B2.

Les résultats obtenus montrent qu'avec les valeurs $T = \{1, 100, 1000\}$ et $\alpha = \{0.01, 0.1\}$, la température finale atteint la température limite prédéfinie à $10^{(-8)}$. Tandis qu'avec ces mêmes valeurs T mais d'autres valeurs $\alpha = \{0.6, 0.9\}$, la température finale est proche de $10^{(-4)}$. Avec la valeur $\alpha = 0.9$, nous remarquons que le temps de convergence vers le meilleur sous-ensemble est plus grand que celui avec $\alpha = 0.6$. Ceci s'explique par le fait que la diminution de la température ou la fonction de refroidissement est beaucoup plus lente avec $\alpha = 0.9$. En conséquence, pour les quatre séries d'expérimentation, nous avons considéré les paramètres $T = 1000$, et $\alpha = 0.6$.

4.5 Résultats Expérimentaux

Afin d'analyser l'apport de notre méthode de sélection et de fusion d'experts, la performance de chaque expert est d'abord évaluée sur les deux bases B1 et B2. Le Tableau 9 présente les différents experts (numérotés de 1 à 12) et leur performance individuelle en terme de taux d'identification (eq.1).

Tableau 9. Taux d'identification des différents experts avec les deux bases B1 et B2

	Taux d'identification (B1)	Taux d'identification (B2)
Descripteurs 2D		
1 Filtres de Gabor	82.29	73.82
2 LBP	75.29	68.60
3 Image D'intensité	54.94	45.08
Descripteurs 3D		
4 Mesures anthropométriques	46.82	46.17
5 Points 3D	63.06	54.43
6 Courbure gaussienne	62.71	55.62
7 Courbure moyenne	75.18	68.33
8 Courbure maximum	71.00	64.86
9 Courbure minimum	69.94	63.77
10 Vecteurs bi-normaux	74.88	66.70
11 Vecteurs normaux	74.94	65.45
12 Vecteurs tangents	83.71	76.05

Il est clair d'après ce tableau que les différents experts n'ont pas un comportement suffisamment discriminant pour une reconnaissance fiable du visage. Ceci nous a d'ailleurs conduit à proposer une sélection d'experts à travers le RS en combinaison avec les différentes variantes du schéma de fusion adaptatif.

4.5.1 Expérience 1: la règle de somme simple (SOM)

Dans cette première expérience, nous évaluons la performance de la méthode de fusion de base, à savoir la règle de somme simple. L'algorithme du recuit simulé a été couplé avec cette méthode de fusion dans la phase d'apprentissage pour les différents ensembles des deux bases B1 et B2. Les différents résultats sont donnés par Tableau 10 et Tableau 11.

Tableau 10. Les taux d'identification de l'expérience SOM (B1 : base d'apprentissage, B2 : base de test)

Base 1	Taux d'identification de la phase d'apprentissage (Base B1)	Taux d'identification de la phase de test (Base B2)
Règle de Simple Somme	-	85.55
Règle de Maximum	-	81.15
Règle de Produit	-	55.73
Meilleur classifieur	-	76.05
P1 (sans Recuit Simulé)	88.78	85.55
(1) P1	95.15	91.31
(2) P1+ P2	95.63	91.31
(3) P1+ P2 + P3	95.39	91.74
(4) P1+ P2 + P3 + P4	95.51	91.74
(5) P1+ P2 + P3 + P4 + P5	95.47	91.74

D'après le Tableau 10, en utilisant les données d'apprentissage (1) et (2) de la base B1, RS génère la même stratégie de fusion $S^1_{SOM} = [1, 2, 3, 4, 5, 0, 7, 0, 0, 0, 11, 0]$ c'est à dire les experts $\{1, 2, 3, 4, 5, 7, 11\}$ (voir Tableau 9) ont été sélectionnés dans la stratégie de fusion finale. Cette stratégie a permis d'obtenir un taux d'identification de 91.31 % sur la phase de test. Avec les autres données d'apprentissage de (3), (4) et (5), RS produit une nouvelle stratégie de fusion $S^1_{SOM} = [1, 2, 3, 4, 5, 0, 0, 0, 0, 0, 11, 12]$.

Les mêmes expériences ont été réalisées avec les sous ensembles de la base B2 utilisée cette fois-ci pour l'apprentissage et la base B1 pour le test. Le Tableau 11 liste les différents taux d'identifications obtenus. En employant les données d'apprentissage (1), (2), (3) et (4), on obtient la même stratégie de fusion $S^2_{SOM} = [1, 2, 3, 4, 5, 0, 0, 0, 0, 10, 0, 0]$.

Tableau 11. Les taux d'identification de l'expérience SOM (B2 : base d'apprentissage, B1 : base de test)

Base 2	Taux d'identification de la phase d'apprentissage (Base B2)	Taux d'identification de la phase de test (Base B1)
Règle de Simple Somme	-	89.59
Règle de Maximum	-	86.74
Règle de Produit	-	65.18
Meilleur classifieur	-	83.71
P1 (sans Recuit Simulé)	85.55	89.59
(1) P1	92.34	95.06
(2) P1+ P2	92.56	95.06
(3) P1+ P2 + P3	92.75	95.06
(4) P1+ P2 + P3 + P4	92.50	95.06
(5) P1+ P2 + P3 + P4 + P5	92.56	95.29

D'après ces premières expériences, il est évident que la sélection dynamique des experts utilisant le recuit simulé conduit à une amélioration de la règle de somme simple en terme de taux d'identification. Cette sélection dynamique d'experts a permis d'améliorer le taux d'identification aussi bien en phase d'apprentissage et qu'en phase de test. En effet, sans la sélection d'experts, la méthode de fusion à travers la règle de la somme simple, « P1 sans recuit simulé », enregistre un écart de performance de 6 points par rapport à la même stratégie de fusion où l'algorithme du recuit simulé est utilisé pour sélectionner les experts sur l'ensemble P1. Ces expériences montrent aussi que la méthode proposée génère de meilleures performances en comparaison avec les autres méthodes de fusion (règle de simple somme, règle de maximum, règle de produit et règle du meilleur classifieur). Par exemple, nous obtenons un gain moyen de 5.8 points en comparaison avec la meilleure méthode de base (règle de somme simple). En plus, les experts sélectionnés par le RS sont relativement stables dans le cas d'augmentation de la taille des données d'apprentissage de (1) à (5). Ceci confirme le fait que le RS converge vers une solution stable et optimisée indépendamment de la taille des données d'apprentissage.

4.5.2 Expérience 2: la règle de somme pondérée utilisant le poids de la décision de l'expert (SPA_PDE)

Dans cette deuxième expérience, nous évaluons le schéma de pondération adaptatif (SPA) avec seulement le poids de la décision de l'expert (PDE) (voir eq. 30). Le RS est utilisé de nouveau pour la sélection dynamique des experts participant au schéma de fusion SPA_PDE. Les résultats expérimentaux avec les différents ensembles des deux bases B1 et B2 sont rassemblés dans les Tableaux 12 et 13.

Tableau 12. Les taux d'identification de l'expérience SPA_PDE (B1 : base d'apprentissage, B2 : base de test)

Règle de fusion	Taux d'identification de la phase d'apprentissage (Base B1)	Taux d'identification de la phase de test (Base B2)
Règle de Simple Somme	-	85.55
Règle de Maximum	-	81.15
Règle de Produit	-	55.73
Meilleur classifieur	-	76.05
P1 (sans Recuit Simulé)	90.80	87.89
(1) P1	95.45	92.12
(2) P1+ P2	95.98	92.12
(3) P1+ P2 + P3	95.76	92.83
(4) P1+ P2 + P3 + P4	95.97	92.83
(5) P1+ P2 + P3 + P4 + P5	95.94	92.83

Tableau 13. Les taux d'identification de l'expérience SPA_PDE (B2 : base d'apprentissage, B1 : base de test)

Règle de fusion	Taux d'identification de la phase d'apprentissage (Base B2)	Taux d'identification de la phase de test (Base B1)
Règle de Simple Somme	-	89.59
Règle de Maximum	-	86.74
Règle de Produit	-	65.18
Meilleur classifieur	-	83.71
P1 (sans Recuit Simulé)	87.89	91.35
(1) P1	92.87	95.59
(2) P1+ P2	93.25	95.59
(3) P1+ P2 + P3	93.42	95.59
(4) P1+ P2 + P3 + P4	93.46	95.76
(5) P1+ P2 + P3 + P4 + P5	93.48	95.76

Dans ces deux séries d'expérimentation, deux stratégies de fusion ont été générées. Dans la première série d'expériences (Tableau. 12), $S^1_{pde} = [1, 2, 3, 4, 0, 6, 7, 0, 0, 0, 11, 12]$ est générée avec les données d'apprentissage de (1) et (2). L'utilisation de données d'apprentissage (3), (4) et (5) a permis d'obtenir une nouvelle stratégie $S^1_{pde} = [1, 2, 3, 4, 0, 0, 7, 0, 0, 0, 0, 12]$. Au niveau de la seconde série (Tableau 13), nous obtenons $S^2_{pde} = [1, 2, 3, 4, 0, 6, 0, 0, 0, 10, 0, 12]$ et $S^2_{pde} = [1, 2, 3, 4, 0, 0, 0, 0, 0, 10, 0, 12]$ respectivement à partir des données de (1), (2) et (3) et des données de (4) et (5).

Ces résultats montrent que la sélection dynamique d'experts, pour participer à la fusion à travers SPA_PDE qui utilise uniquement le poids de la décision de l'expert, améliore encore de 4 à 5 points, la performance de SPA_PDE de base sans la sélection d'experts qui, lui-même, enregistre un gain de performance de 2 à 3 points par rapport à la stratégie de fusion à travers la

somme simple. Ceci confirme que, d'une part, l'affectation d'un poids possède un effet positif sur les performances du système de reconnaissance, et que d'autre part, ces performances sont améliorées aussi en utilisant une méthode de sélection des experts se basant sur le recuit simulé.

4.5.3 Expérience 3 : la règle de somme pondérée utilisant le poids de la qualité de l'expert (SPA_PQE)

Dans cette expérience, nous évaluons une autre variante du schéma de pondération adaptatif (SPA). La différence par rapport à l'expérience 2 réside au niveau du poids accordé à l'expert. Ce poids dépend de sa qualité et de son comportement global par rapport à la base d'apprentissage (voir eq. 29). Dans ce travail, la qualité des experts se base sur la métrique du taux d'égalité des erreurs (TER) ou encore Equal Error Rate (EER) (eq.2), propre à chaque expert. La Figure 24 présente les différentes valeurs TERs propres aux experts sur des deux bases B1 et B2 pendant la phase d'apprentissage.

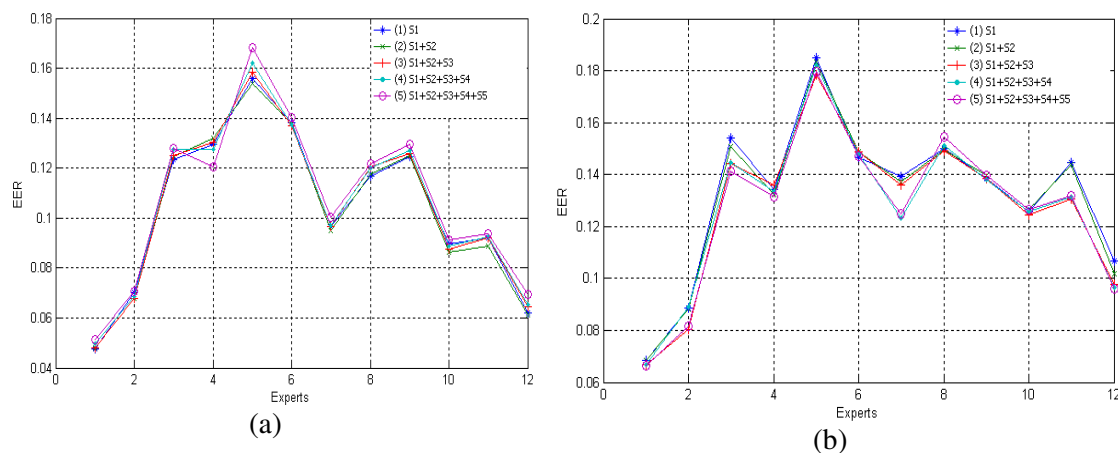


Figure 24. Les différents TERs des deux bases (a) B1 et (b) B2

D'après la figure ci-dessus, les valeurs TER des différents experts de la base B1 sont relativement stables dans le cas de changement de la taille de la base d'apprentissage. Ces valeurs varient entre 0.0478 et 0.1683 avec une moyenne de 0.1048. Ces valeurs sont légèrement différentes pour la base B2 avec une moyenne égale à 0.1289 et deux valeurs limites : 0.0664 et 0.1849.

Après avoir calculé les différents EERs propres aux différents experts, nous procédons à la génération des différentes stratégies de fusion.

Avec la première série d'expérimentation, nous notons qu'une seule stratégie de fusion est générée (Tableau 14) $S_{pqe}^1 = [1, 2, 3, 4, 5, 0, 7, 0, 0, 0, 11, 0]$. Dans ce cas, nous possédons deux

taux de reconnaissance puisque le poids assigné aux experts n'est pas identique pour toutes les données d'apprentissage de (1) à (5).

Tableau 14. Les taux d'identification de l'expérience SPA_PQE (B1 : base d'apprentissage, B2 : base de test)

Règle de fusion	Taux d'identification de la phase d'apprentissage (Base B1)	Taux d'identification de la phase de test (Base B2)
Règle de Simple Somme	-	85.55
Règle de Maximum	-	81.15
Règle de Produit	-	55.73
Meilleur classifieur	-	76.05
P1 (sans Recuit Simulé)	92.21	88.43
(1) P1	96.16	92.49
(2) P1+ P2	96.49	92.88
(3) P1+ P2 + P3	96.35	92.49
(4) P1+ P2 + P3 + P4	96.37	92.88
(5) P1+ P2 + P3 + P4 + P5	96.18	92.88

Tableau 15. Les taux d'identification de l'expérience SPA_PQE (B2 : base d'apprentissage, B1 : base de test)

Règle de fusion	Taux d'identification de la phase d'apprentissage (Base B2)	Taux d'identification de la phase de test (Base B1)
Règle de Simple Somme	-	89.59
Règle de Maximum	-	86.74
Règle de Produit	-	65.18
Meilleur classifieur	-	83.71
P1 (sans Recuit Simulé)	89.24	93.06
(1) P1	93.05	95.88
(2) P1+ P2	93.55	95.88
(3) P1+ P2 + P3	93.62	96.00
(4) P1+ P2 + P3 + P4	93.52	96.18
(5) P1+ P2 + P3 + P4 + P5	93.37	96.18

Pour la deuxième série d'expérimentation (Tableau 15), deux stratégies sont générées $S^2_{pqc} = [1, 2, 3, 4, 5, 0, 0, 8, 0, 10, 0, 0]$ pour les données de (1) à (4), et $S^{2'}_{pqc} = [1, 2, 3, 4, 5, 0, 0, 0, 0, 10, 0, 12]$ pour les données de (5).

D'après ces résultats expérimentaux, on remarque une fois de plus que la fusion SPA_PQE sans la sélection d'experts améliore de 3 à 4 points la performance de la fusion par la somme simple sur les deux bases B1 et B2. Maintenant la sélection d'experts par le RS améliore encore de 3 à 4 points le taux d'identification de la fusion SPA_PQE sans la sélection d'experts. Quand la taille de données d'apprentissage est accrue de (1) vers (5), on constate que la performance du

mélange d'experts sélectionnés reste stable lorsque la base B1 est utilisée comme données d'apprentissage ; elle s'améliore de 1 à 2 points lorsque la base B2 est utilisée comme données d'apprentissage.

4.5.4 Expérience 4: Schéma de pondération adaptative (SPA)

Dans cette expérience, le schéma de pondération adaptatif est utilisé avec les deux poids de décision de l'expert et de sa qualité. Les valeurs de la qualité des différents experts sont celles utilisées dans l'expérience 3, c'est-à-dire les valeurs EER pour les différents experts sur les données d'apprentissage (1) à (5). Les différents résultats sont illustrés dans les Tableaux 16 et 17. Nous remarquons d'abord que le schéma de fusion par la pondération adaptative (SPA) améliore de 6 points la performance de la fusion par la somme simple, et réalise donc un gain de performance significativement plus important que les deux variantes précédentes, i.e. Expérience 2 et 3, par rapport à la fusion des experts à travers la somme simple. Plus précisément, le couplage des deux poids a permis d'améliorer considérablement le taux d'identification des expériences (1) à (5) en comparaison avec les deux expériences précédentes: Expérience 2 et Expérience 3. Maintenant, lorsque le recuit simulé est utilisé pour la sélection d'experts participant au mélange d'experts par la fusion SPA, nous enregistrons encore un gain de performance de 1 à 2 points sur les bases de test.

En analysant la stabilité de la génération des stratégies de fusion par le RS, nous notons le même comportement avec les deux séries d'expérimentation. En effet, pour la première série (Tableau 16), $S^1_{spa} = [1, 2, 3, 4, 0, 6, 7, 0, 0, 10, 0, 12]$ est produite par les données d'apprentissage de (1), (2), (3) et (4), tandis que les données de (5) génère $S^1_{spa} = [1, 2, 3, 4, 0, 0, 7, 0, 0, 10, 0, 12]$. Pour la deuxième série (Tableau 17), nous obtenons respectivement $S^2_{spa} = [1, 2, 3, 4, 0, 0, 0, 8, 0, 0, 11, 12]$ et $S^2_{spa} = [1, 2, 3, 4, 0, 0, 7, 0, 0, 0, 11, 12]$ pour les mêmes données d'apprentissage.

Tableau 16. Les taux d'identification de l'expérience SPA (B1 : base d'apprentissage, B2 : base de test)

Règle de fusion	Taux d'identification de la phase d'apprentissage (Base B1)	Taux d'identification de la phase de test (Base B2)
Règle de Simple Somme	-	85.55
Règle de Maximum	-	81.15
Règle de Produit	-	55.73
Meilleur classifieur	-	76.05
P1 (sans Recuit Simulé)	94.74	91.04
(1) P1	96.26	93.05
(2) P1+ P2	96.58	92.94
(3) P1+ P2 + P3	96.50	93.05
(4) P1+ P2 + P3 + P4	96.50	93.00

(5) P1+ P2 + P3 + P4 + P5	96.35	93.05
---------------------------	-------	-------

Tableau 17. Les taux d'identification de l'expérience SPA (B2 : base d'apprentissage, B1 : base de test)

Règle de fusion	Taux d'identification de la phase d'apprentissage (Base B2)	Taux d'identification de la phase de test (Base B1)
Règle de Simple Somme	-	89.59
Règle de Maximum	-	86.74
Règle de Produit	-	65.18
Meilleur classifieur	-	83.71
P1 (sans Recuit Simulé)	91.53	95.12
(1) P1	92.69	96.29
(2) P1+ P2	93.09	96.35
(3) P1+ P2 + P3	93.55	96.18
(4) P1+ P2 + P3 + P4	93.76	96.35
(5) P1+ P2 + P3 + P4 + P5	93.70	96.35

4.6 Analyse des performances

Les expérimentations présentées dans ce chapitre ont été établies afin d'étudier certains comportements des méthodes de fusion. Nous avons effectués quatre expériences en exploitant l'algorithme du recuit simulé pour chercher la stratégie de fusion presque-optimale dans une application d'identification de visages 3D. Chacune des expériences emploie une méthode de fusion avec des poids de pondération spécifiques. Les résultats de ces expérimentations permettent de tirer les conclusions suivantes :

- La sélection des experts par le recuit simulé (RS) permet de mieux concevoir une stratégie de fusion. En effet, la sélection dynamique d'experts pour participer aux différentes variantes du schéma de pondération adaptatif améliore en performance ces mêmes variantes sans la sélection d'experts. D'après la Figure.25, l'algorithme du recuit simulé est bénéfique pour toutes les méthodes de fusion et surtout pour les règles de fusion pondérées. Par exemple, un gain moyen de 5.8 points est obtenu par rapport à la meilleure méthode de fusion à travers la somme simple (**Expérience 1**). Ceci prouve que nous pouvons construire une stratégie de fusion d'experts sans déployer tous les experts disponibles.
- La procédure d'affectation d'un poids a un impact positif sur la performance du système d'identification. Cette procédure offre de meilleurs résultats comparés aux méthodes n'utilisant pas de poids. En plus, les résultats obtenus dans l'expérience 4 indique que le choix des deux poids couplés (le poids de la qualité de l'expert et le poids de la décision de l'expert) est le meilleur choix par rapport aux cas dans l'absence du poids (Expérience 1), d'utilisation du poids de la décision de l'expert (Expérience 2), et d'utilisation du poids de la qualité de l'expert (Expérience 3).

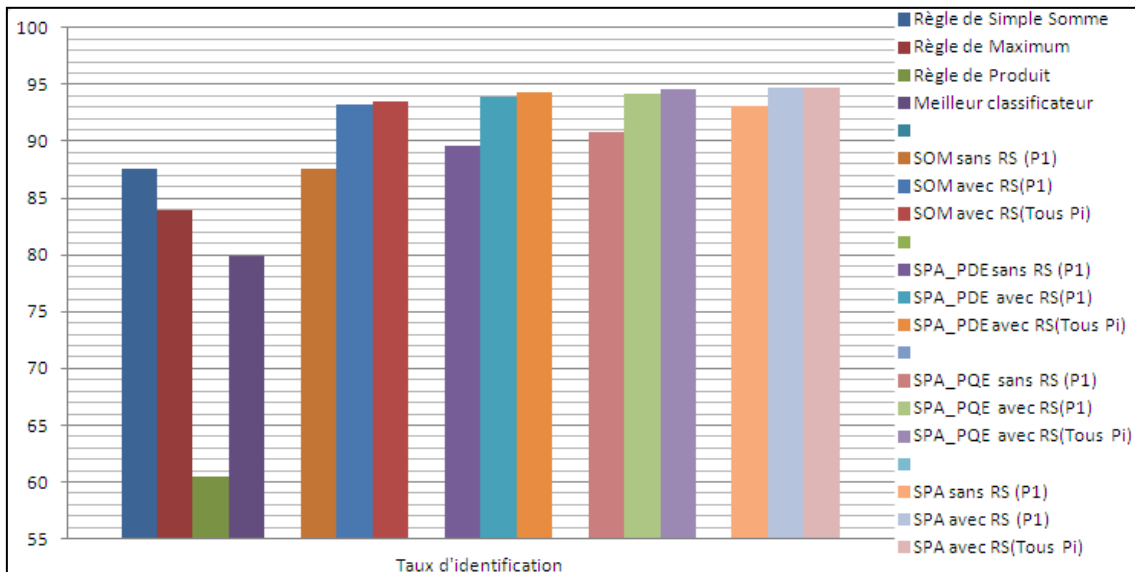


Figure 25. Comparaison de performances de différentes méthodes de fusion

- L'utilisation d'une partie de la base d'apprentissage est considérée suffisante pour prédire la stratégie de fusion convenable. La plupart des expériences utilisent seulement la première base d'apprentissage pour générer une stratégie de fusion stable et (presque)-optimisé indépendamment de la taille des données d'apprentissage.

4.7 Conclusion

Dans ce chapitre, une approche d'optimisation a été proposée afin de chercher une stratégie de fusion presque-optimale d'experts. Cette approche se base sur le couplage de l'algorithme du recuit simulé et d'un schéma de pondération adaptatif (SPA). Quatre séries d'expérimentation ont été mises en œuvre avec différentes variantes du SPA. Ceci a permis de valider certaines propositions. La première proposition confirme que la procédure de pondération a un impact positif sur la performance du système d'identification. En plus, le choix des deux poids couplés (le poids de la qualité de l'expert et le poids de la décision de l'expert) est le meilleur choix par rapport à l'absence du poids, l'utilisation seule du poids de la décision de l'expert, ou le poids de la qualité de l'expert. La deuxième proposition montre que l'utilisation d'une méthode de sélection d'expert permet d'avoir un système d'identification plus performant. La troisième proposition prouve que l'utilisation d'une partie de la base d'apprentissage est considérée suffisante pour prédire une stratégie de fusion stable et (preque)-optimisé. Toutes ces propositions sont évaluées au niveau de fusion de scores avec la base FRGC v2.0.

Néanmoins, pouvons-nous appliquer le même schéma de fusion dans le cas de présence de plusieurs descripteurs et classifieurs ? Ceci requiert certainement une nouvelle architecture qui

tient compte de deux niveaux de fusion : la fusion au niveau des caractéristiques et la fusion au niveau des scores. Nous proposons d'étudier cette nouvelle architecture au chapitre suivant.

CHAPITRE V: RECONNAISSANCE DE VISAGES
PAR UNE METHODE HEURISTIQUE DE FUSION
AU NIVEAU DES CARACTERISTIQUES ET DES
SCORES

5. Reconnaissance de visages par une méthode heuristique de fusion au niveau des caractéristiques et des scores

Nous avons étudié au chapitre 4 l'optimisation d'un mélange d'experts gardés et abouti à des stratégies de fusion tardive au niveau des scores. Il existe aussi dans la littérature des stratégies de fusion dite précoce qui consistent à combiner les divers vecteurs de descripteurs en un seul avant l'étape de classification et qui présentent théoriquement l'avantage de préserver toute information extraite de l'objet avant son classification. Néanmoins, comme ces divers vecteurs de descripteurs sont généralement de nature très différente et de mesures très disparates comme c'est le cas avec les différents descripteurs extraits d'un visage 3D, leur combinaison brute conduit très souvent à une dégradation de performance.

Dans ce chapitre, nous nous proposons d'étudier des stratégies de fusion hybride, mélange de stratégies précoces et tardives, en vue de la reconnaissance faciale 3D. Nous les qualifions de stratégies de fusion multi-niveaux, ou plus exactement fusion à 2-niveaux car l'ensemble de vecteurs de descripteurs seront d'abord partitionnés ; Puis, chaque partie de vecteurs de descripteurs est combinée en un seul vecteur de descripteurs selon une stratégie de fusion précoce ; Enfin, les scores de ces différents vecteurs de descripteurs issues de la stratégie de fusion précoce sont ensuite fusionnés selon une stratégie de fusion tardive. Cela signifie que la cardinalité de l'espace de stratégies de fusion est plus grande que celui où les stratégies de fusion s'opèrent à un seul niveau, précoce ou tardif. Comme une recherche exhaustive de stratégies de fusion optimales dans un tel espace est impossible, nous nous sommes encore une fois de plus orientés vers des techniques d'optimisation heuristique.

Dans ce chapitre, une approche de sélection de la meilleure stratégie de fusion basée sur l'algorithme génétique a été proposée. Pour cela un nouveau codage est adapté afin d'assurer trois critères essentiels : la sélection des meilleurs descripteurs, la sélection de leurs classifieurs adéquats et la recherche d'une stratégie de fusion presque-optimale. Dans ce codage, la règle de la somme simple est utilisée comme règle de fusion au niveau du score. Les expériences menées sur la base FRGC v2.0 montrent l'intérêt de notre méthode.

Les contributions de nos travaux exposés dans ce chapitre peuvent donc être résumées comme suit :

- Proposition de stratégies de fusion hybride, mélange de fusion précoce et tardive, en vue de la reconnaissance faciale 3D ;
- La recherche heuristique d'une stratégie de fusion presque-optimale à travers la mise en œuvre de l'algorithme génétique ;
- Etude des performances de notre méthode de fusion basée sur l'algorithme génétique à travers une série d'expérimentation avec la base de visages FRGC v.2.0.

Dans la suite, nous exposons, d'abord, le principe de la fusion multi-niveau qui est en relation étroite avec la cardinalité de l'espace des stratégies de fusion. Ensuite, nous présentons l'application de l'algorithme génétique comme solution pour sélectionner la meilleure stratégie de fusion dans notre système de reconnaissance de visages 3D. Enfin, une série d'expérimentations est effectuée afin de mettre en valeur notre nouvelle approche.

5.1 La fusion multi-niveaux

5.1.1 Principe de la fusion multi-niveaux

La fusion multi-niveau consiste à prendre en compte deux niveaux de fusion : la fusion au niveau des caractéristiques et la fusion aux niveaux des scores. Dans ce cas de figure, nous avons comme entrée un ensemble de descripteurs de visages 3D représentant l'information bas niveau et un ensemble de classifieurs. Le but est de chercher la stratégie de fusion presque-optimale qui permet de sélectionner le sous ensemble minimal de descripteurs et de leurs classifieurs adéquats. La figure 26 illustre ce concept.

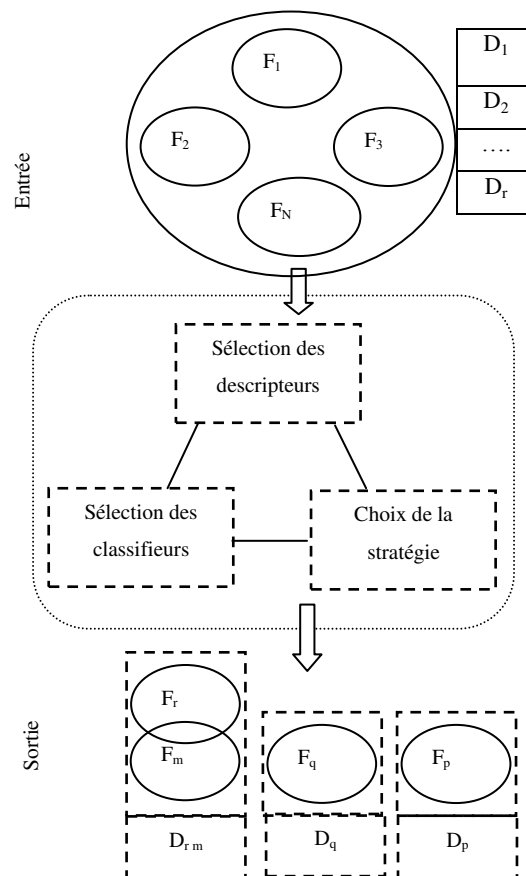


Figure 26. Utilisation de l'algorithme génétique pour la sélection des descripteurs (F_1, F_2, \dots, F_N), des classifieurs (D_1, D_2, \dots, D_r) et de leur stratégie de fusion

Plus précisément, la recherche de cette stratégie s'effectue avec une base de galerie et une base d'apprentissage. Ces deux bases sont formées par les modèles de visages 3D avec leurs images de textures correspondantes. Un ensemble de descripteurs 3D et 2D est extrait des scans 3D de deux bases. En présence de cet ensemble de descripteurs F_i et de différents classifieurs D_i , des opérations de sélection des meilleurs descripteurs et de leurs classifieurs adéquats sont mises en œuvre. La sélection de ces descripteurs peut être une sélection passive ou active. La sélection passive suppose que le descripteur est inhibé contrairement à la sélection active (voir section 5.2.2.1). Cette sélection spécifie le niveau de fusion (niveau caractéristique ou niveau score) des descripteurs. La sélection des classifieurs pour ces descripteurs passe par l'évaluation de tous les classifieurs D_i . La décision finale est prise par la fusion des décisions des classifieurs sélectionnés pour tous les sujets de la base d'apprentissage. Le critère de l'optimalité est obtenu suivant la performance globale affichée par ces sujets.

5.1.2 Espace des stratégies de fusion

Dans le chapitre précédent, l'espace de stratégies de fusion est peuplé de schémas de fusion à un seul niveau, i.e. au niveau des scores. La cardinalité de cet espace est exponentiellement proportionnelle au nombre d'experts N mis en jeu. Dans le cas de stratégies de fusion à deux niveaux, l'espace de stratégies de fusion possibles comprend non seulement les stratégies de fusion tardives pures mais aussi toutes les autres stratégies de fusion intermédiaires où des caractéristiques sont d'abord fusionnées d'une façon précoce puis de nouveau combinées avec d'autres experts d'une manière tardive au niveau des scores, ceci sans encore mentionner les divers classifieurs ou mesures de similarité que l'on peut appliquer. Cet espace de stratégies de fusion est donc encore plus grand que celui des mélanges des experts gardés, déjà exponentiel, que nous avons étudié au chapitre précédent. Encore une fois, la recherche exhaustive d'une stratégie de fusion dans un tel espace est impossible. Nous proposons donc de recourir à des méthodes heuristiques, en occurrence l'algorithme génétique, afin de trouver une stratégie de fusion presque-optimale.

Le choix de l'algorithme génétique a été justifié pour les deux raisons suivantes. D'abord, l'algorithme génétique offre un mécanisme d'encodage de la solution à chercher. Ceci nous a permis de tenir en compte les différents descripteurs et les différents classifieurs mis en jeu. Ensuite, il a fait preuve dans la résolution de nombreux problèmes d'optimisation vu sa puissance exploratoire [Sedaaghi et al., 2007] [Gökberk et al., 2007] [Kanan et al., 2008] [Perez et al., 2011]. Ainsi, cette nouvelle approche proposée ne génère pas seulement la stratégie de fusion pour combiner les descripteurs les plus pertinents, mais elle possède aussi la capacité étendue pour sélectionner les descripteurs les plus discriminants et leurs classifieurs appropriés.

5.2 Recherche heuristique d'une stratégie de fusion par l'algorithme génétique

Rappelons que notre contexte est la reconnaissance faciale 3D qui s'appuie sur un ensemble de descripteurs 3D et 2D extraits des scans 3D. Ces descripteurs (voir section 3.3) sont obtenus après une étape de prétraitement (voir section 3.2). Dans notre problème d'identification, la recherche d'une stratégie de fusion passe par deux opérations essentielles : la sélection des descripteurs et la sélection des classifieurs. Pour cela, l'encodage de la stratégie de fusion est essentiel. La décision finale est prise par la fusion multi-niveau des décisions des classifieurs sélectionnés dans la stratégie de fusion finale. Dans notre cas, l'algorithme génétique est adapté afin d'encoder la stratégie de fusion à chercher sous forme de chromosomes (voir section 5.2.2.1). Ceci nous a permis de tenir en compte des différents descripteurs et classifieurs mis en jeu. La figure 27 présente l'approche proposée.

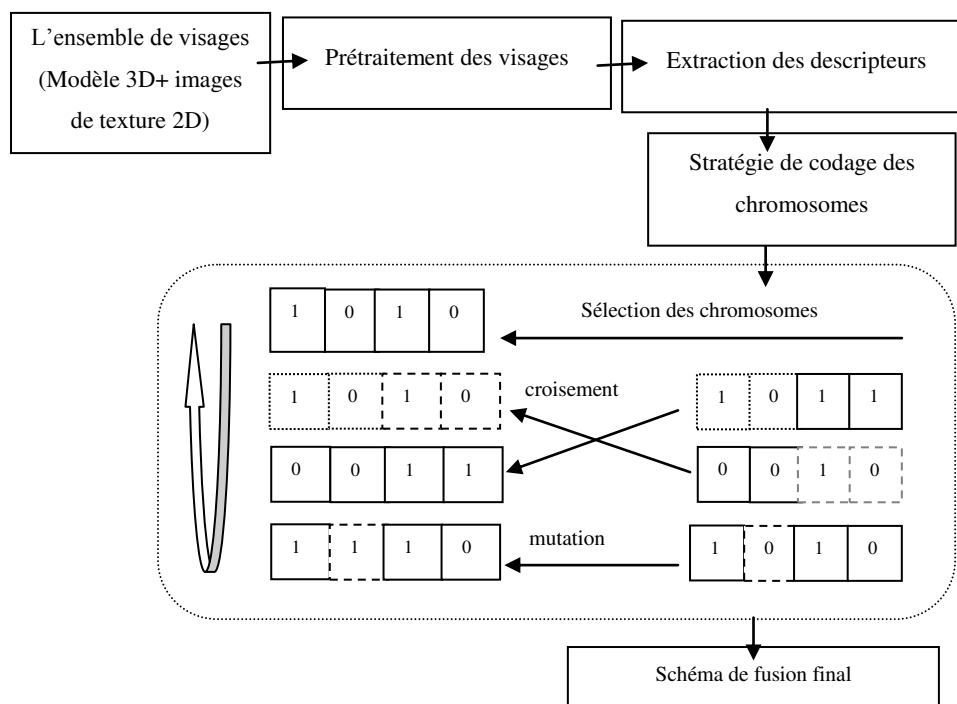


Figure 27. Aperçu général de l'approche proposée avec les différentes étapes de prétraitement des visages, de l'extraction des descripteurs et du codage de la stratégie de fusion. Le schéma de fusion final est généré par l'algorithme génétique.

5.2.1 Principe général de l'algorithme génétique

Les algorithmes génétiques [Gökberk et al., 2007] [Kanan et al., 2008] [Perez et al., 2011] sont des algorithmes d'optimisation heuristiques inspirés de la génétique et des

mécanismes d'évolution de la nature : croisement, mutation et sélection en se basant sur l'idée que seuls les plus forts subsistent dans une population.

Ces algorithmes fonctionnent sur une population composée de chromosomes, tous différents, pouvant être des solutions potentielles du problème à résoudre. Dans notre problème de reconnaissance de visages 3D, les chromosomes représentent les stratégies de fusion. Chaque chromosome est évalué. Cette évaluation permet de juger la pertinence des solutions (chromosomes) par rapport au problème considéré. Ceci conduit à éliminer les solutions jugées inutiles ou très mauvaises. Une fois cette élimination effectuée, les gènes des solutions sélectionnées sont combinés pour obtenir une nouvelle population qui doit être mieux adaptée à la résolution du problème que la population précédente. La population initiale donne ainsi naissance à des générations successives, sous la forme de chromosomes fils, mutés et hybridés à partir des chromosomes initiaux dits chromosomes parents. Ce processus est réitéré jusqu'à ce que l'on atteigne le nombre de génération prédéfini à l'avance.

5.2.2 Application à la reconnaissance de visages 3D

L'utilisation de l'algorithme génétique pour la reconnaissance faciale 3D nécessite que l'on précise l'encodage des stratégies de fusion, le croisement et le critère d'optimalité.

5.2.2.1 Encodage d'une stratégie de fusion

Dans l'algorithme génétique proposé, un nouveau codage est présenté afin de sélectionner simultanément les descripteurs, leurs meilleurs classifieurs et leurs stratégies de fusion. Ce codage consiste à diviser le chromosome en deux parties : partie A et partie B (voir figure 28).

En tenant compte des N descripteurs, la partie A contient N gènes qui correspondent aux différents descripteurs. Ces gènes possèdent des valeurs entières: '1' implique que le descripteur est actif et est utilisé dans la stratégie de fusion au niveau de caractéristique (fusion précoce), '0' implique que le descripteur est actif et est utilisé dans la stratégie de fusion au niveau de score (fusion tardive), et '-1' implique que le descripteur est inactif.

La partie B dépend du nombre de descripteurs sélectionnés dits actifs pour une fusion précoce. Il s'agit bien de la concaténation des descripteurs sélectionnés dans la partie A du chromosome. Cette partie modélise toutes les combinaisons de concaténation M possibles de ces descripteurs actifs. Néanmoins, nous supposons qu'un descripteur seul ne peut pas figurer dans la fusion au niveau de caractéristiques car cela revient à considérer un nouvel expert basé sur ce descripteur au niveau des scores. Nous considérons aussi que les combinaisons obtenues par permutation sont équivalentes.

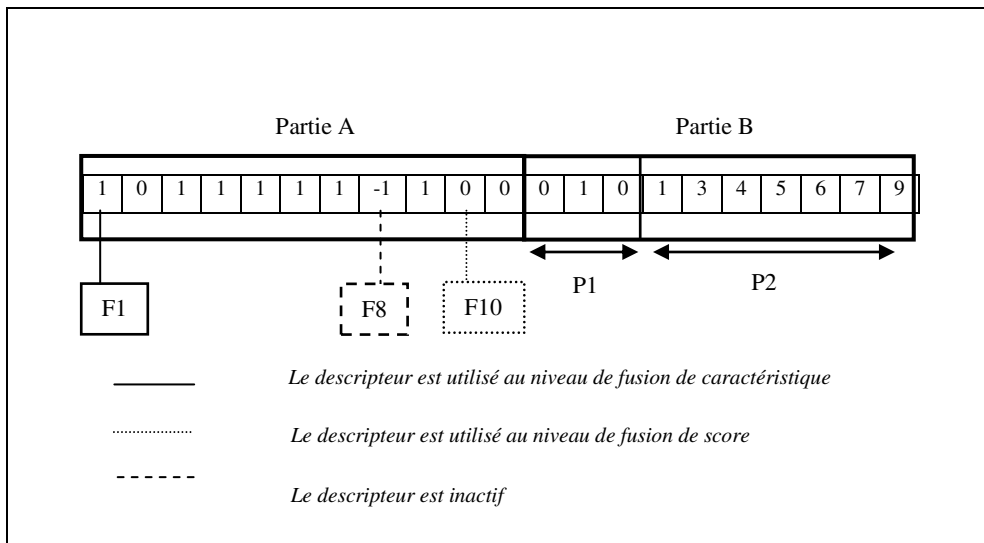


Figure 28. Exemple de stratégie de codage d’un chromosome avec deux parties : Partie A et Partie B. Partie A représente les descripteurs sélectionnés (1 au niveau de caractéristique, 0 au niveau de score et -1 inactif). Partie B modélise les nombres des descripteurs à concaténer parmi l’ensemble sélectionné.

Deux exemples de M combinaisons de concaténation sont illustrés dans la figure 29. Dans le cas, où le nombre de descripteurs sélectionnés par la valeur ‘1’ est égale par exemple à 6, un modèle de concaténation de descripteurs est choisi aléatoirement. Si nous considérons le modèle (2, 2, 2) cela veut dire que les descripteurs seront concaténés deux à deux. Quels sont alors les descripteurs concernés ?

0 0 6	0 0 0 8
0 2 4	0 0 2 6
0 3 3	0 0 3 5
2 2 2	0 0 4 4
	0 2 2 4
	0 2 3 3
	2 2 2 2
$M=6$	$M=8$

Figure 29. Exemples de combinaisons représentées dans la partie P1 pour la concaténation des descripteurs sélectionnés dans la Partie A

Pour répondre à cette question dans notre algorithme génétique, nous avons divisé la partie B du chromosome en deux parties P1 et P2. P1 se réfère au modèle de concaténation M et P2 associe les descripteurs à ce modèle.

Un exemple de chromosome est illustré dans la figure.28. Ce chromosome, représentant une stratégie de fusion, contient deux parties A et B. La partie A contient : 1011111-1100. Nous possédons, dans ce cas, 7 bits à '1'. Nous choisissons aléatoirement un modèle de concaténation parmi les 4 modèles de concaténation définis pour $M = 7$: Modèle 1 = (7, 0), Modèle 2 = (2, 5), Modèle 3 = (3,4), Modèle 4 = (2, 2, 3). Le numéro du modèle sélectionné est représenté dans la partie P1 de la partie B par son code binaire : (010). La partie P2 choisit aléatoirement les descripteurs à concaténer dans le premier niveau de fusion. Ceci implique que le modèle M_2 est sélectionné et deux vecteurs V1, V2 sont créés par concaténation, V1 = [F1, F3] et V2 = [F4, F5, F6, F7, F9] avec F_i représente le descripteur i. Ensuite, la fusion se poursuit au niveau des scores en combinant les scores générés par les classifieurs D_i adéquats pour les vecteurs V1, V2, et F2.

5.2.2.2 Sélection, Croisement et Mutation

Après avoir obtenu la première génération des chromosomes parents selon le modèle décrit dans la section précédente, un algorithme génétique est mis en œuvre pour générer la meilleure solution. Les opérations de sélection, croisement et mutation ont été utilisées pour faire évoluer une nouvelle génération de chromosomes fils à partir de la population initiale.

5.2.2.2.1 Sélection des chromosomes

Cet opérateur est chargé de définir quels seront les chromosomes qui serviront de parents pour la nouvelle génération. Dans notre algorithme, la méthode d'échantillonnage universelle stochastique (Stochastic Universal Sampling ou SUS) proposée par [Baker, 1987] est utilisée pour sélectionner les meilleurs chromosomes. Cette méthode utilise une roulette partagée en autant de portions proportionnellement aux valeurs de qualité des différents chromosomes. Les chromosomes parents sélectionnés sont désignés, par exemple, par un ensemble d'indicateurs équidistants comme le montre la figure 30. Dans ce cas, la probabilité de survie d'un chromosome est reliée directement à son efficacité au sein de la population.

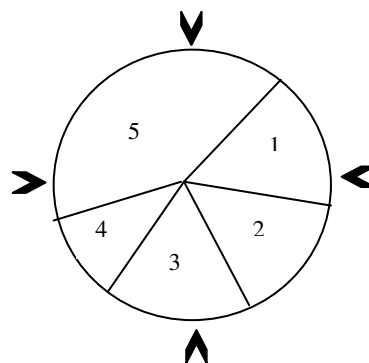


Figure 30. Exemple de sélection SUS pour choisir 4 chromosomes parents

5.2.2.2.2 Croisement uniforme

L'opérateur de croisement favorise l'exploration de l'espace de recherche de la stratégie de fusion. Un croisement uniforme est utilisé uniquement sur la partie A des chromosomes parents. Les chromosomes parents sont alors copiés et recombinaés de façon à former deux chromosomes fils possédant des gènes issus des deux parents.

La mise en œuvre de ce procédé consiste à définir de manière aléatoire un masque, c'est-à-dire une chaîne de bits de même longueur que les chromosomes parents sur lesquels il sera appliqué. Ce masque est destiné à savoir, pour chaque gène, de quel parent le chromosome fils devra hériter. La figure 31 montre un exemple de croisement uniforme entre deux chromosomes parents.

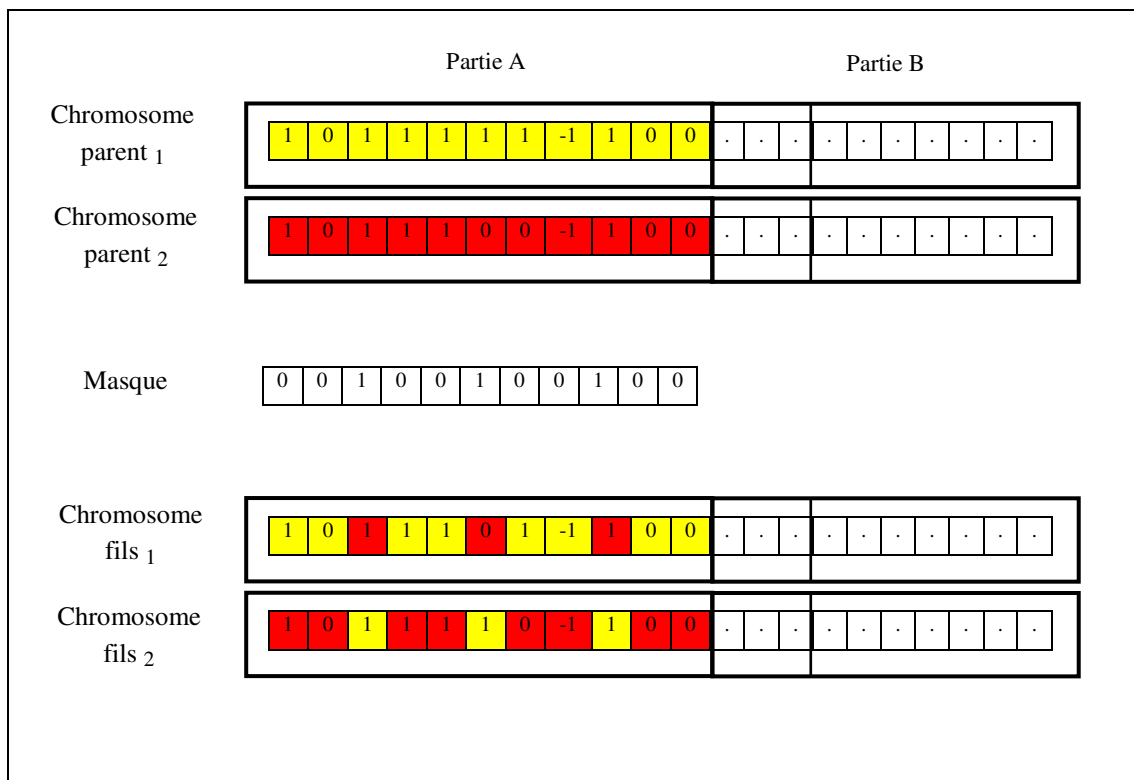


Figure 31. Exemple de croisement entre deux chromosomes parents pour la génération de nouveaux chromosomes fils

5.2.2.2.3 Mutation

L'opérateur de mutation modifie de manière aléatoire les gènes d'un chromosome fils avec une probabilité très faible. Ceci permet de maintenir la diversité au sein de notre population de solutions. Cet opérateur est appliqué sur la partie A d'un chromosome fils. Il s'agit bien d'appliquer un masque de même longueur que le chromosome fils. Ce masque est destiné à savoir, pour chaque gène, si sa valeur sera modifiée. Si la valeur correspondante est

égale à 0, elle sera modifiée aléatoirement en $\{1 \text{ ou } -1\}$. Le même principe est appliqué pour les autres valeurs $\{1, -1\}$ du modèle de codage. La figure 32 montre un exemple de mutation d'un chromosome fils.

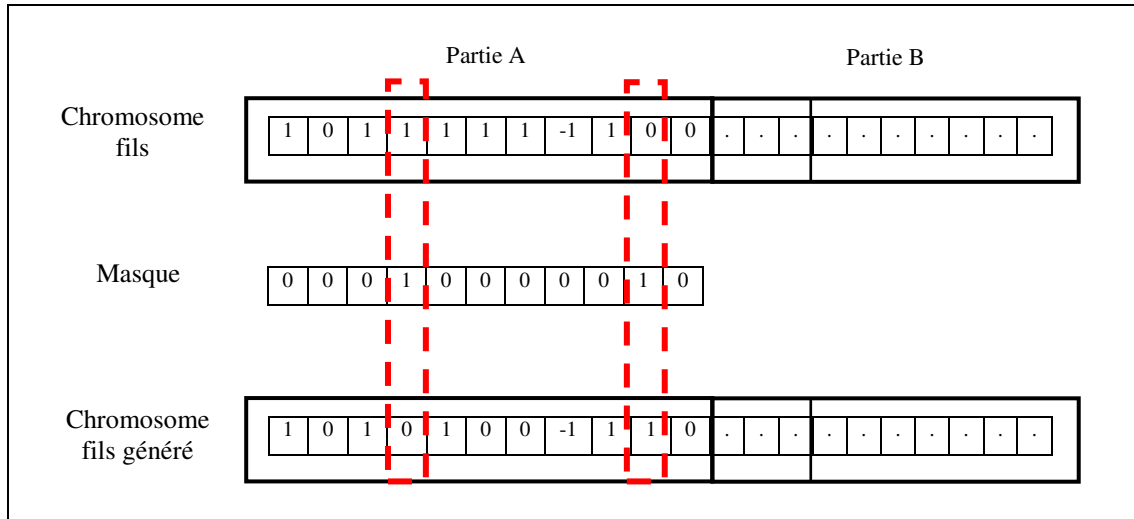


Figure 32. Exemple de mutation pour la génération d'un nouveau chromosome fils

5.2.2.3 Critère de qualité (fitness)

Au niveau de la sélection des chromosomes, la qualité des différents chromosomes parents est évaluée. Plus cette qualité est meilleure, plus les chromosomes correspondants ont plus de chance à être sélectionnés. L'évaluation du degré de satisfaction de cette solution de codage des chromosomes, se base sur le calcul de la valeur du fitness. Dans le scénario de reconnaissance de visages 3D, différentes fonctions peuvent être déployées comme le taux d'identification, le taux d'égalité des erreurs, le taux de fausse acceptation et le taux de faux rejet.

Dans notre travail, le taux d'identification (eq.1) est adopté comme fonction du fitness. Le fitness est calculé pour chaque chromosome en utilisant les différents classifieurs D_i . L'évolution de notre algorithme s'effectue jusqu'à ce que le nombre d'itérations prédéterminé soit atteint, ou quand la stratégie de fusion atteint une valeur stable. La solution finale avec plus de fitness représente la stratégie de fusion finale. Cette stratégie spécifie le sous-ensemble de descripteurs sélectionnés et leurs classifieurs appropriés.

5.3 Descripteurs et classifieurs

5.3.1 Les descripteurs

Les descripteurs utilisés pour la reconnaissance faciale 3D ont été introduits au chapitre 2. Ces descripteurs sont décrits dans la section 3.3. Un ensemble de ces descripteurs est extrait des images de texture 2D : Image en niveau de gris, Banc de filtres de Gabor et Image Local Binary Pattern (LBP). Pour le reste des descripteurs, ils sont extraits de l'information de forme 3D. Ils décrivent les courbures minimum, maximum, moyenne et gaussienne; les vecteurs normaux, les vecteurs tangents, les vecteurs bi-normaux et les mesures anthropométriques.

5.3.2 Les classifieurs

Deux classifieurs sont utilisés dans nos expériences pour générer les différentes distances. Nous déployons le classifieur en représentation parcimonieuse (SRC) (section 3.4) et le classifieur du plus proche voisin (1-ppv) avec quatre distances : la distance euclidienne (NN-ED), la distance de Mahalanobis (NN-MD), la distance cosinus (NN-COSD) et la mesure de corrélation (NN-CorrD).

Les quatre distances : distance euclidienne (eq.21), distance de Mahalanobis (eq.37), distance cosinus (eq.38) et la distance déduite de la corrélation (eq.39) sont introduites par la suite :

$$d(x, y)^2 = (x - y)C^{-1}(x - y)', \quad (\text{eq.37})$$

$$d(x, y) = 1 - \frac{(x y')}{(x'x)^{1/2}(y'y)^{1/2}}, \quad (\text{eq.38})$$

$$d(x, y) = 1 - \frac{(x-x_M)(y-y_M)'}{((x-x_M)(x-x_M)')^{1/2}((y-y_M)(y-y_M)')^{1/2}}. \quad (\text{eq.39})$$

Avec x et y sont deux vecteurs de descripteurs extraits du visage à comparer, C est la matrice de covariance, x_m est la valeur moyenne de x , et y_m est la valeur moyenne de y .

Une méthode de normalisation Min-Max (eq.26) est employée sur les différentes distances avant la fusion des différents scores. Pour la combinaison, une règle de somme simple est utilisée pour la génération du score final.

5.4 Résultats expérimentaux

5.4.1 Protocole expérimental

Dans nos expérimentations, nous avons utilisé la base de visage FRGC v2.0 avec la même base de galerie (410 sujets) utilisée dans les chapitres précédents. Les 3541 modèles de visages

3D restants avec leurs images de texture correspondantes sont séparés pour construire la base d'apprentissage (2332 modèles de visage) et la base de test (1209 modèles de visage). Une première série d'expérience « Neutre vs Tout » est réalisée avec ces deux bases (base de galerie et base de test) afin d'étudier les performances de notre méthode de génération de la meilleure stratégie de fusion.

Une deuxième série d'expérience a été élaborée pour analyser le comportement de notre système par rapport à l'expression du visage. Par conséquent, la base de test est divisée en deux sous-ensembles en fonction de l'expression du visage dans la base FRGC v.2.0. Le premier sous-ensemble contient les modèles de visage caractérisés par une expression neutre avec leurs images de texture (713 modèles). Le deuxième sous-ensemble regroupe le reste des modèles et leurs images de texture présentant différentes expressions faciales (496 modèles). Deux expériences sont alors effectuées : « Neutre vs Neutre », « Neutre vs Non-neutre ». Le Tableau.18 présente l'organisation de la base FRGC v.2.0.

Tableau 18. Organisation de la base FRGC v.2.0

	Nombre de sujets	Nombre de modèles 3D avec leurs images de texture	Nombre de modèles avec leurs images de texture sans présence d'expression	Nombre de modèles avec leurs images de texture en présence d'expression
Base de galerie	410	410	410	0
Base d'apprentissage	410	2332	-	-
Base de test	410	1209	713	496

5.4.2 Analyse des résultats expérimentaux

Dans nos expériences, la structure des chromosomes manipulés par l'algorithme génétique est organisée comme suit : F1- vecteurs tangents (VT), F2- vecteurs bi-normaux (VB), F3- vecteurs normaux (VN), F4- courbure gaussienne (CG), F5- courbure maximum (CMax), F6- courbure moyenne (CMoy), F7- courbure minimum (CMin), F8- LBP (LBP), F9- filtres de Gabor (FG), F10- Mesures anthropométriques (MA), F11- Image niveau de gris (ING).

Dans la phase d'apprentissage, l'algorithme génétique a été utilisé avec une population de 50 chromosomes qui se reproduisent pendant 100 générations. La stratégie de fusion finale générée est codée comme suit : la première partie du chromosome Partie A : [0, 0, -1, 1, -1, 1, -1, 0, 0, 0, 0], et la deuxième partie du chromosome Partie B : [P1 : 001, P2 :4, 6].

Cette stratégie consiste à concaténer tout d'abord les deux descripteurs courbure gaussienne et courbure moyenne dans un seul vecteur $N_V = [CG-CMoy]$. Ensuite, le sous-ensemble de

descripteurs {N_V, VT, VB, LBP, FG, MA, ING} est utilisé dans le schéma de fusion au niveau de scores. Les classifieurs SRC et NN-CorrD ont permis de générer respectivement les meilleures performances pour les caractéristiques {VB, LBP, FG, MA, ING} et {V1, VT}. Les descripteurs vecteurs normaux (VN), courbure maximum (CMax) et courbure minimum (CMin) ne sont pas sélectionnés dans la stratégie de fusion finale. La performance de chaque classifieur propre à chaque descripteur, qui a été sélectionné par l'algorithme génétique, est affichée dans le Tableau.19.

Tableau 19. Les taux d'identification des classifieurs sélectionnés par l'algorithme génétique de la meilleure stratégie de fusion

Classifieur	Descripteurs			
SRC	FG	81.97	VB	73.12
	LBP	77.01	MA	58.97
	ING	53.35		
NN-CorrD	VT	82.22		
	CG-CMoy	74.16		

Tableau 20. Taux d'identification avec l'expérience « Neutre vs. Tous »

Méthode	Phase d'apprentissage	Phase de Test
Règle Somme Simple	–	94.13%
Gökberk et al., 2008	–	95.28%
Mian et al., 2008	–	94.71%
Notre Approche	95.67%	97.27%

Comme nous venons de constater, les descripteurs sélectionnés par l'algorithme génétique (Tableau 19) ont permis d'obtenir un taux d'identification final de l'ordre de 97.27 % avec l'expérience « Neutre vs. Tous » (Tableau 20).

Pour comparer avec les méthodes de fusion, nous avons choisis trois méthodes. La première méthode est considérée comme la méthode de fusion de base appliquant la règle de somme simple. La deuxième méthode de [Mian et al., 2008] (voir section. 3.6.3) utilise une règle de somme pondérée. La dernière méthode est la méthode de [Gökberk et al., 2008] (voir section. 3.6.3) qui applique une sélection des meilleures matrices de similarité couplé avec la règle de somme simple. Toutes ces méthodes sont évaluées avec les matrices de similarité générées avec le classifieur SRC de notre base de test.

Nous pouvons constater que notre méthode de fusion affiche de meilleures performances. Un gain de 2% par rapport au meilleur résultat obtenue par la méthode [Gökberk et al., 2008]. Ceci nous permet de souligner surtout l'apport de notre méthode de sélection des descripteurs pertinents et de leurs classifieurs adéquats.

Pour comparer la méthode opérant avec l'algorithme génétique (AG) à celle utilisant l'algorithme du recuit simulé (RS), nous avons choisi la première expérience SOM (section 4.5.1). Ce choix est unique vu que l'algorithme génétique opère avec une règle de fusion de somme simple. La comparaison est faite à travers l'observation de l'amélioration qu'apporte chaque méthode par rapport à la méthode de base. Rappelons que la méthode AG permet d'avoir un gain de 3%. Tandis que, RS affiche un gain de performance de l'ordre de 6%. Sur cette base, la comparaison peut favoriser l'algorithme du recuit simulé mais elle n'exclut pas la généralité au niveau du codage de la solution par l'algorithme génétique.

Par la suite, cette stratégie de fusion finale est appliquée dans les deux séries d'expérimentation « Neutre vs Neutre » et « Neutre vs Non neutre ». Les résultats expérimentaux sont présentés dans le Tableau 21.

Tableau 21. Taux d'identification avec les expériences « Neutre vs. Neutre (NN) », « Neutre vs. Non-neutres (N-nonN) »

Méthode	N-nonN	NN
Règle Somme Simple	91.33%	97.76%
Gökberk et al., 2008	92.75%	97.88%
Mian et al., 2008	90.73%	97.76%
Notre Approche	95.16%	99.16%

D'après les résultats obtenus, il est clair que notre méthode de fusion permet d'avoir un gain entre 2% et 3% par rapport aux autres méthodes de fusion pour l'expérience «Neutre vs Tous». Ce gain diminue légèrement par rapport au meilleur résultat de fusion [Gökberk et al., 2008] dans les deux expériences «Neutre vs Neutre», «Neutre vs Non neutre». Malgré cela, la performance de notre méthode reste plus élevée par rapport aux autres méthodes de fusion.

Dans ce cas, en plus de l'amélioration de la performance du système de reconnaissance de visages 3D, ceci nous permet de mettre l'accent sur l'importance des trois opérations suivantes: la sélection des descripteurs les plus discriminants, la sélection de leurs classifieurs appropriés et l'obtention d'une stratégie de fusion presque-optimale.

5.5 Conclusion

Dans ce chapitre, nous avons analysé un cas plus général au niveau de la combinaison des données de visages 3D. Deux niveaux de fusion ont été présentés: la fusion au niveau des caractéristiques et la fusion au niveau des scores. Cependant, la cardinalité de l'espace de recherche de la stratégie de fusion optimale est importante. Nous avons adopté une méthode d'optimisation heuristique qui permet de profiter de la puissance exploratoire des algorithmes génétiques. Pour cela un nouveau codage est adapté pour la représentation des solutions (chromosomes) au niveau de l'algorithme génétique.

Des expériences ont été menées sur la base de données FRGC v.2.0 et ont montré les performances de la méthode proposée. Cette nouvelle approche proposée ne contient pas seulement la stratégie de fusion pour combiner les descripteurs les plus pertinents, mais elle possède aussi la capacité étendue pour sélectionner les descripteurs les plus discriminants et leurs classifieurs appropriés.

CONCLUSION GENERALE

6. Conclusion générale

La reconnaissance faciale (RF) est un domaine de recherche très actif en vision par ordinateur en général et en biométrie en particulier. Cet intérêt est motivé au moins pour trois raisons. D'abord, le visage est le moyen le plus naturel par les êtres humains d'identifier les uns des autres. Le visage en tant que modalité biométrique est donc bien accepté de tous et présente un caractère non intrusif, ce qui le distingue d'autres modalités biométriques comme l'iris ou l'emprunte digitale. Ensuite, ces applications potentielles de reconnaissance sont aussi nombreuses que diversifiées. Enfin, la RF représente aussi des défis scientifiques importants parce que tous les visages humains ont des configurations similaires, et qu'avec les images faciales en 2D que l'on peut acquérir facilement, la variation intra-classe, due à des facteurs comme le changement de poses et de conditions d'éclairage, les variations d'expressions faciales, le vieillissement, est bien plus importante que la variation inter-classe.

Depuis quelques années et conjointement avec l'arrivée de systèmes d'acquisition en 3D capables de capturer la profondeur d'objets, la reconnaissance faciale 3D (RF 3D) a émergé comme une voie prometteuse pour traiter les deux problèmes non résolus en 2D, à savoir les variations en pose et en éclairage. Ainsi, les modèles 3D de visages ont la particularité de contenir les deux informations de forme et de texture alignés. Une solution en RF 3D peut donc tirer parti d'une fusion avisée d'informations en 3D et celles de texture en 2D.

Cependant, certains problèmes se posent avec la présence à la fois des surfaces faciales pour la modalité 3D pure et les images de texture 2D alignées. Ceci produit un nombre de possibilités de fusion trop important. L'état d'art reflète cette situation avec l'existence de diverses stratégies de fusion pour la reconnaissance de visages en 3D, allant de la fusion précoce "early fusion" opérant au niveau de caractéristiques à la fusion tardive "late fusion" sur les sorties de classifieurs, en passant par de nombreuses stratégies intermédiaires. Pour les stratégies de fusion tardive, on distingue encore des combinaisons en parallèle, en cascade ou multi-niveaux. Une exploration exhaustive d'un tel espace étant impossible, il faut donc recourir à des solutions heuristiques. En plus, en s'inscrivant dans un cadre de systèmes biométriques, les critères d'optimalité des stratégies de fusion restent des questions primordiales. En effet, une stratégie de fusion est dite optimisée si elle est capable d'intégrer et de tirer parti des différentes modalités et, plus largement, des différentes informations extraites lors du processus de reconnaissance quelque soit leur niveau d'abstraction.

6.1 Contribution

Pour surmonter toutes ces difficultés et proposer une solution optimisée, une première étape a été réalisée en analysant le comportement des descripteurs extraits de l'information 2D

et 3D du visage, ainsi que leurs résistances par rapport aux changements d'expressions. Nous avons montré sur la base FRGC v2.0 qu'aucun des descripteurs 2D ou 3D sont suffisamment discriminants pour une reconnaissance fiable du visage, notamment en présence d'expressions faciales. Si les descripteurs 3D, en capturant précisément la forme 3D de visage, permettent une meilleure reconnaissance faciale que leurs homologues 2D dans le cas des expressions neutres, ils s'avèrent bien plus sensibles que les descripteurs 2D en présence d'expressions faciales. Ces avantages et défauts différenciés entre les descripteurs 2D et 3D nous ont conduit à proposer une méthode de fusion adaptative au niveau de scores pour prendre en compte cette complémentarité entre les différents classifieurs propres à chaque descripteur. Cette méthode utilise les scores de similarité et se base sur une règle de somme pondérée avec deux types de poids. Un premier poids hors ligne, qui est inversement proportionnel au taux d'égalité d'erreur, est calculé lors d'une phase d'apprentissage. Il évalue la performance de chaque classifieur et définit la relation inter-classifieur. Un deuxième poids en ligne, intra-classifieur, est calculé lors de la phase de reconnaissance utilisant les différents scores de similarité. Il évalue la qualité de décision de chaque classifieur.

Dans une deuxième étape, nous nous sommes intéressés en premier lieu à la présence d'un nombre N important d'experts. Dans ce cas, la cardinalité de l'espace de recherche de la stratégie de fusion presque-optimale est importante. Une recherche exhaustive de la meilleure stratégie de fusion dans cet espace est quasi-impossible. Ce qui nous a incités à recourir à une sorte de technique d'optimisation heuristique. Une approche, qui se base sur le couplage de l'algorithme du recuit simulé et d'un schéma de pondération adaptatif (SPA), a été proposée. Quatre séries d'expérimentation ont été mises en œuvre avec différentes variantes du SPA. Ceci a permis de valider certaines propositions. La première proposition confirme que la procédure de pondération a un impact positif sur la performance du système d'identification. En effet, le choix de deux poids couplés (le poids de la qualité de l'expert et le poids de la décision de l'expert) est le meilleur choix par rapport à l'absence du poids, ou l'utilisation seule de l'un des poids. La deuxième proposition montre que l'utilisation d'une méthode de sélection d'expert permet d'avoir un système d'identification plus performant. La troisième proposition prouve que l'utilisation d'une partie de la base d'apprentissage est considérée suffisante pour prédire une stratégie de fusion stable et (preque)-optimisé.

En troisième lieu, nous avons analysé un cas plus général qui est la combinaison en présence de deux niveaux de fusion: niveau des scores et niveau des caractéristiques. Une nouvelle stratégie envisageant cette présence a été développée. Elle utilise comme technique l'algorithme génétique avec un nouveau codage afin d'assurer deux aspects : la sélection des meilleurs descripteurs et la recherche d'une stratégie de fusion presque-optimale. Comme une méthode de fusion, il ne contient pas seulement la stratégie de fusion pour combiner les

descripteurs les plus pertinents, mais a aussi la capacité étendue pour sélectionner les descripteurs les plus discriminants et leurs classifieurs appropriés. Des expériences sont menées sur la base de données FRGC v.2.0 et ont montré la performance de la méthode proposée.

6.2 Perspectives

Dans cette section, nous présentons certaines perspectives qui sont possibles pour notre thèse :

- Nous envisageons d'étudier d'autres métriques qui permettent de quantifier la performance du système de reconnaissance de visages 3D comme le taux de reconnaissance, le taux de vérification, le taux de fausse acceptation ou le taux de faux rejet. Un premier travail a été élaboré dans [Ben Soltana et al., 2012].
- La performance de notre algorithme génétique dépend des métriques utilisées. Dans notre cas, les métriques ont été sélectionnées à partir des données de visages 3D. Ce chemin de réflexion joint un nouveau domaine maintenant qualifié d'apprentissage de métrique (metric-learning). Il s'intéresse à l'apprentissage de distance ou de similarités à partir des données. Il est donc important d'apprendre correctement les métriques à utiliser pour notre problème afin d'améliorer la sélection des classifieurs adéquats. En conséquence, une analyse plus fine des caractéristiques et de leurs comportements est assurée.
- Un modèle plus complet, dans l'algorithme génétique, peut être étendu afin de tenir compte d'autres niveaux de fusion comme la fusion en cascade, la fusion au niveau des rangs et la fusion au niveau des décisions. Aussi, d'autres règles de fusion peuvent être mises en œuvre comme le schéma de pondération adaptatif.
- Nous étudions davantage la fusion au niveau des caractéristiques, puisque dans notre travail nous n'avons utilisé que la concaténation des caractéristiques. L'objectif est d'analyser d'avantage la pertinence des vecteurs nouvellement construits.
- Nous pouvons exploiter notre système afin de valider l'optimalité des stratégies de fusion de certains systèmes. Aussi, nous pouvons étendre l'exploitation de notre système pour d'autres applications par exemple la reconnaissance de l'empreinte digitale, la reconnaissance de la main, etc.

BIBLIOGRAPHIE

7. Bibliographie

- [Arca et al., 2007] S. Arca, R. Lanzarotti, and G. Lipori, "Face Recognition Based on 2D and 3D Features," Knowledge-Based Intelligent Information and Engineering Systems (KES' 07), 455-462, 2007.
- [Abate et al. 2006] A. F. Abate, M. Nappi, D. Riccio and G. Sabatino, "3D face recognition using normal sphere and general Fourier descriptor," Int. Conf. On Pattern Recognition ICPR 2006, Hong Kong, 20-24 August 2006.
- [Achermann et al., 1997] B. Achermann, X. Jiang, H. Bunke, "Face Recognition Using Range Images," Proc. Of the International Conference on the Virtual Systems and Multimedia (VSMM97), pp: 129-136, September 1997.
- [Ahonen et al., 2004] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face Recognition with Local Binary Patterns," European Conference on Computer Vision, 2004.
- [Agresti, 1990] A. Agresti, "Categorical Data Analysis," New York: John Wiley & Sons, 1990.
- [Baker, 1987] J. E. Baker, "Reducing Bias and Inefficiency in the Selection Algorithm," Proceedings of the Second International Conference on Genetic Algorithms and their Application, 14-21, 1987.
- [Bartlett et al. 2002] M. S. Bartlett, J. R. Movellan and T. J. Sejnowski, "Face recognition by independent component analysis," IEEE Transactions on Neural Networks, vol. 13, pages 1450-1464, 2002.
- [Ben Amor et al., 2007] B. Ben Amor, K. Ouji, M. Ardabilian, L. Chen, "R-ICP: une nouvelle approche d'appariement 3D orientée régions pour la reconnaissance faciale," Dans TAIMA'2007, cinquième édition des ateliers de travail sur le traitement et l'analyse de l'information, Tunisie. pp. 33-40, 2007.
- [Ben Amor et al., 2006] B. Ben Amor, M. Ardabilian, L. Chen, "New experiments on ICP-based 3D face recognition and authentication," International Conference on Pattern Recognition, 2006.
- [Beumier and Acheroy 2001] C. Beumier and M. Acheroy. Face verification from 3D and grey level clues. Pattern Recognition Letters, vol. 22, no. 12, pages 1321-1329, 2001.
- [Bowyer et al., 2006] K. W. Bowyer, K. Chang, P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," Computer Vision and Image Understanding, 2006.
- [Belhumeur et al., 1997] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pages 711-720, Jul. 1997.
- [Besl and McKay, 1992] P.J. Besl, N.D. McKay, "A Method for Registration of 3-D Shapes," IEEE Trans. Pattern Anal. Mach. Intell., 14(2) :239-256, 1992.
- [Brunelli & Poggio, 1993] R. Brunelli, T. Poggio, Face Recognition: Features versus Templates, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15, No. 10, October 1993, pp. 1042-1052.

[Bronstein et al., 2005] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Three-dimensional face recognition," *International Journal of Computer Vision*, 64(1) : 5–30, 2005.

[Candes et al., 2006] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. on Information Theory*, 2006.

[Chang et al., 2005] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "An Evaluation of multimodal 2D+3D Face Biometrics," *IEEE Trans. Pattern Anal. Mach. Intell.* 27(4): 619-624 (2005).

[Chang et al., 2003] K. I. Chang, K. W. Bowyer and P. J. Flynn, "Face Recognition Using 2D and 3D Facial Data," In *ACM Workshop on Multimodal User Authentication*, pages 25-32, Santa Barbara, California, 2003.

[Chua et al., 1997] C.S. Chua, R. Jarvis, " Point signature: A new representation for 3d object recognition," *Internat. J. Computer Vision* 25 (1), 63–85, 1997.

[Cook et al., 2007] J. Cook, M. Cox, V. Chandran, and S. Sridharan, "Robust 3D Face Recognition from Expression Categorisation," *International Conference on Biometrics*, 271-280, 2007.

[Datta et al., 2011] A. Datta, R. Feris, and D. Vaquero, "Hierarchical Ranking of Facial Attributes," *IEEE Conference on Automatic Face and Gesture Recognition (FG 2011)*, Santa Barbara, California, March 22-24, 2011.

[Daugman, 1993] J. G. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE Trans. PAMI*, 15(11) :1148-1161, Nov 1993.

[Donoho, 2006] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, 2006.

[Dorizzi and García-Mateo, 2007] B. Dorizzi, C. García-Mateo, " Multimodal Biometrics," *Annales des Télécommunications* 62(1-2): 2-10 (2007).

[Drira et al., 2009] H. Drira, B. Ben Amor, A. Srivastava and M. Daoudi A Riemannian Analysis of 3D Nose Shapes For Partial Human Biometrics, accepted *ICCV 2009* September 29 - October 2, 2009 Kyoto Japan.

[Faltemier et al., 2008] T. C. Faltemier, K. W. Bowyer, and P. J. Flynn, "Using multi-instance enrollment to improve performance 3D face recognition," *Computer Vision and Image Understanding*, 2008.

[Faltemier et al., 2006] T. C. Faltemier, K. W. Bowyer, P. J. Flynn, "3D face recognition with region committee voting," *International Symposium on 3D Data Processing, Visualization, and Transmission*, 2006: 318-325.

[Figueiredo and Jain, 2002] M. Figueiredo and A. K. Jain, "Unsupervised Learning of Finite Mixture Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 381–396, March 2002.

[Gabrys & Ruta, 2006] B. Gabrys, D. Ruta, " Genetic algorithms in classifier fusion," *Appl. Soft Comput.* 6(4), 337-347 (2006).

[Gökberk et al., 2008] B. Gökberk, H. Dutağacı, A. Ulaş, L. Akarun, and B. Sankur, "Representation plurality and fusion for 3D face recognition," *IEEE Trans. on Systems Man and Cybernetics-Part B: Cybernetics*, 38(1):155–173, 2008.

[Gökberk et al., 2007] B. Gökberk, M. O. Irfanoglu, L. Akarun, E. Alpaydin, "Learning the best subset of local features for face recognition, " *Pattern Recognition (PR)* 40(5):1520-1532 (2007).

[Gökberk et al., 2006] B. Gökberk, L. Akarun, "Comparative Analysis of decision level fusion algorithms for 3D face recognition," *ICPR 2006*:1018-1021.

[Gökberk et al., 2005] B. Gökberk, A. A. Salah, and L. Akarun, "Rank-Based Decision Fusion for 3D Shape-Based Face Recognition," *International Conference on Audio- and Video-Based Biometric Person Authentication*, 1019-1028, 2005.

[Godil et al., 2005] A. Godil, S. Ressler, and P. Grother, "Face recognition using 3D facial shape and color map information: comparison and combination," *Biometric Technology for Human Identification, SPIE*, 5404:351–361, 2005.

[Griffin, 2004] P. Griffin, "Optimal Biometric Fusion for Identity Verification," *Identix Research, Tech. Rep. RDNJ-03-0064*, 2004.

[Hesher et al., 2003] C. Hesher, A. Srivastava and G. Erlebacher. A novel technique for face recognition using range imaging. In *International Symposium on Signal Processing and Its Applications*, volume 2, pages 201-204, Jul. 2003.

[Heseltine et al., 2004] T. Heseltine, N. Pears and J. Austin. Three-dimensional face recognition: an eigensurface approach. In *IEEE International Conference on Image Processing*, volume 2, pages 1421-1424, Oct. 2004.

[Huang et al., 2011b] D. Huang, M. Ardabilian, Y. Wang, L. Chen, "A Novel Geometric Facial Representation based on Multi-Scale Extended Local Binary Patterns," *IEEE International Conference on Automatic Face and Gesture Recognition*, Santa Barbara, CA, USA. 2011.

[Huang et al. 2011c] D. Huang, C. Shan, M. Ardabilian, Y. Wang and L. Chen, "Local binary patterns and its application to facial image analysis: a survey, " *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 41, no. 4, pages 1-17, 2011.

[Huang et al., 2010] D. Huang, G. Zhang, M. Ardabilian, Y. Wang, L. Chen, "3D Face Recognition using Distinctiveness Enhanced Facial Representations and Local Feature Hybrid Matching," *International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, Washington D.C., USA 2010.

[Huang et al. 1994] Y.S. Huang, K. Liu and C.Y. Suen, "The combination of multiple classifiers by neural network approach," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 9, no. 3, pp. 579-597, 1994.

[Huang & Aviyente, 2006] K. Huang and S. Aviyente, "Sparse representation for signal classification," *International Conference on Neural Information Processing Systems*, 609-616, 2006.

[Hüsken et al., 2005] M. Hüsken, M. Brauckmann, S. Gehlen, C. Von der Malsburg, "Strategies and benefits of fusion of 2D and 3D face recognition," *Computer Vision and Pattern*

Recognition - Workshops, 2005. IEEE Computer Society Conference on Volume, Issue, 25-25 June 2005 Page(s):174 – 174.

[Hubel and Wiesel, 1977] D. H. Hubel and T. N. Wiesel, Ferrier Lecture: Functional Architecture of Macaque Monkey Visual Cortex. In Proceedings of the Royal Society of London. Series B, Biological Sciences, Vol. 198, No. 1130 (May 19, 1977), pp. 1-59.

[Ho et al., 1994] T.K. Ho, J.J. Hull and S.N. Srihari, " Decision combination in multiple classifier systems", IEEE Trans. PAMI, vol. 16, no. 1, pp. 66-75, January 1994.

[Jain et al., 2000] A. K. Jain, R. P. W. Duin, J. Mao,"Statistical Pattern Recognition: A Review," IEEE Trans. PAMI., 22(1), 4-37, 2000.

[Jain et al., 2005] A. Jain, K. Nandakumara, A. Ross, " Score normalization in multimodal biometric systems," Pattern Recognition Society, Elsevier, pp. 2270 - 2285, 2005.

[Jacobs et al., 1991] R.A. Jacobs, M.I. Jordan, S.J. Nowlan, G.E. Hinton, "Adaptive mixtures of local experts," Neural Computation 3(1), 79-87, 1991.

[Kusuma & Chua, 2008] G. P. Kusuma and C. S. Chua, "Image Level Fusion Method for Multimodal 2D + 3D Face Recognition," ICIAR 2008, 984-992, 2008.

[Kakadiaris et al., 2007] I. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis, "Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach," IEEE Trans. on Pattern Analysis and Machine Intelligence, 2007.

[Kanan et al., 2008] H. R. Kanan, K. Faez, "GA-based optimal selection of PZMI features for face recognition," Applied Mathematics and Computation (AMC) 205(2):706-715 (2008).

[Kirkpatrick et al., 1983] S. Kirkpatrick, C.D. Gelatt Jr., M.P. Vecchi, "Optimization by simulated annealing," Science, 220, 4598, 671-680, 1983.

[Kittler et al., 1998] J. Kittler, M. Hatef, R. P.W. Duin, and J. Matas, "On Combining Classifiers," IEEE Trans. PAMI, vol. 20, no. 3, March 1998.

[Lao et al., 2000] S. Lao, Y. Sumi, M. Kawade, and F. Tomita, "3D template matching for pose invariant face recognition using 3D facial model built with iso-luminance line based stereo vision," International Conference on Pattern Recognition, 2000.

[Lades et al., 1993] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Wurtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," IEEE Trans. on Computers, 42(3) :300-311, 1993.

[Lee and Milios, 1990] J.C. Lee, E. Milios, "Matching Range Images of Human Faces, " in Proc. IEEE International Conference on Computer Vision, pp. 722– 726. 1990.

[Li et al., 2009] X. Li, T. Jia, H. Zhang, "Expression-insensitive 3D face recognition using sparse representation," CVPR 2009: 2575-2582.

[Li et al., 2007] C. Li, G. Su, Y. Shang, Y. Li, and Y. Xiang, "Face Recognition Based on Pose-Variant Image Synthesis and Multi-level Multi-feature Fusion," AMFG 2007: 261-275.

[Li et al., 2005] S. Z. Li, C. Zhao, M. Ao, Z. Lei, " Learning to Fuse 3D+2D Based Face Recognition at Both Feature and Decision Levels ," In AMFG, 44-54, 2005.

[Li & Jain, 2005] S. Z. Li and A. K. Jain, Handbook of Face Recognition, Springer-Verlag, February 2005.

[Li et al., 1999] J. Q. Li and A. Barron, "Mixture Density Estimation," in Advances in Neural Information Processings Systems 12, S. A. Solla, T. K. Leen, and K.-R. Muller, Eds. San Mateo, USA: Morgan Kaufmann Publishers, 1999.

[Lin et al., 2006] W. Y. Lin, K. C. Wong, N. Boston, and Y. H. Hu, "3D human face recognition using summation invariants," in Proc. IEEE Conf. on ICASSP, 2006, vol. 2, pp. 341 –344.

[Lin et al., 2007] W. Y. Lin, K. C. Wong, N. Boston, Y. H. Hu, "3D Face Recognition Under Expression Variations using Similarity Metrics Fusion," ICME 2007: 727-730.

[Lu et al., 2006] X. Lu, A. K. Jain, and D. Colbry, "Matching 2.5D Face scans to 3D models," IEEE Trans. on Pattern Analysis and Machine Intelligence, 28(1):31-43, 2006.

[Lu & Jain, 2005] X. Lu and A. K. Jain, "Integrating Range and texture Information for 3D face Recognition," IEEE Workshop on Applications of Computer Vision, 155-163, 2005.

[Maurer et al., 2005] T. Maurer, D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev, D. West, and G. Medioni. "Performance of geometrix activeidtm 3D face recognition engine on the FRGC data," IEEE Workshop on Face Recognition Grand Challenge Experiments, 2005.

[Marcelja, 1980] S. Marcelja. Mathematical description of the responses of simple cortical cells. Journal of Optical Society of America A, 70(11):1297–1300, 1980.

[Metropolis et al., 1953] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, " Equation of state calculations by fast computing machines," J. Chem. Phys. ,21, 6, 1087-1092, 1953.

[Mian et al., 2008] A. S. Mian, M. Bennamoun, and R. Owens, "Keypoint detection and local feature matching for textured 3D face recognition," International Journal of Computer Vision, 2008.

[Mian et al., 2007] A. S. Mian, M. Bennamoun, and R. A. Owens, "An efficient multimodal 2D-3D hybrid approach to automatic face recognition," IEEE Trans. on Pattern Analysis and Machine Intelligence, 29(11):1927-1943, 2007.

[Mian et al., 2006] A. Mian, M. Bennamoun and R. Owens, "Face recognition using 2D and 3D multimodal local features," International Symposium on Visual Computing, 2006.

[Monwar & Gavrilova, 2008] Md. M. Monwar, M. Gavrilova, "FES: a system for combining face, ear, and signature biometrics using rank level fusion," International Conference on Information Technology: New Generations, 922-927, 2008.

[Moreno et al., 2003] A.B. Moreno, A. Sanchez, J.F.Velez and F.J Diaz, "Face recognition using 3D surface-extracted descriptors," In: Proc. Irish Machine Vision and Image, (IMVIP'03), September, 2003.

[Medioni and Waupotitsch, 2003] G. Medioni, R. Waupotitsch, "Face recognition and modeling in 3D," in Proc. IEEE International Workshop on Analysis and Modeling of Faces and Gestures, pp. 232–233. 2003.

[Ouji et al., 2011] K. Ouji, M. Ardabilian, L. Chen, F. Ghorbel, "Pattern-based Face Localization and Online Projector Parameterization for Multi-Camera 3D Scanning," International Conference on 3D Body Scanning Technologies, Lugano, Switzerland, 2011.

[Ojala et al., 1996] T. Ojala, M. Pietikainen., and D. Harwood, "A comparative study of texture measures with classification based on feature distributions", Pattern Recognition 29 (1), 1996, pp. 51-59.

[Ojala et al. 2002] T. Ojala, M. Pietikainen and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pages 971-987, 2002.

[Papatheodorou & Rueckert, 2004] T. Papatheodorou, D. Rueckert, "Evaluation of Automatic 4D Face Recognition Using Surface and Texture Registration," Automatic Face and Gesture Recognition, 2004. Proceedings Sixth IEEE International Conference on Volume, Issue, 17-19 May 2004 Page(s): 321 – 326.

[Prabhakar and Jain, 2002] S. Prabhakar and A. K. Jain, "Decision-level Fusion in Fingerprint Verification," Pattern Recognition, vol. 35, no. 4, pp.861–874, April 2002.

[Phillips et al., 2005] P. J. Phillips, P. J. Flynn, W. T. Scruggs, K. W. Bowyer, J. Chang, K. J. Hoffman, J. Marques, J. Min, W. J. Worek, "Overview of the Face Recognition Grand Challenge," CVPR, 947-954, 2005.

[Perez et al., 2011] C. A. Perez, L. A. Cament, L. E. Castillo, "Methodological improvement on local Gabor face recognition based on feature selection and enhanced Borda count," Pattern Recognition (PR) 44(4):951-963 (2011).

[Perrot, 1997] R. Perrot, "Use of anthropological methods in the identification of unknown individuals: human remains and armed robbers," The Forensic Scientist Online Journal, Shundersen Communications Inc, 161-164, 1997.

[Rayward-Smith et al., 1996] V. J. Rayward-Smith, I. H. Osman, C. R. Reeves, and G. D. Smith, "Modern Heuristic Search Methods", Wiley, 1996.

[Ruta & Gabrys, 2005] D. Ruta, B. Gabrys, "Classifier selection for majority voting," J. Inf. Fusion, 6 (1), 63–81, (2005).

[Rakhlin et al., 2005] A. Rakhlin, D. Panchenko, and S. Mukherjee, "Risk Bounds for Mixture Density Estimation," ESAIM: Probability and Statistics, vol. 9, pp. 220–229, June 2005.

[Raghavendra et al., 2011] R. Raghavendra, B. Dorizzi, A. Rao, G. H. Kumar, "Particle swarm optimization based fusion of near infrared and visible images for improved face verification," Pattern Recognition 44(2): 401-411 (2011).

[Samir et al., 2006] C. Samir, A. Srivastava and M. Daoudi. Three-dimensional face recognition using shapes of facial curves. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 11, pages 1858-1863, Nov 2006.

[Samir et al., 2005] C. Samir, M. Daoudi and J.P Vandeborre, "Automatic 3D face recognition using topological techniques," IEEE ICME 2005.

[Savran et al. 2008] A. Savran, N. Alyuz, H. Dibeklioglu, O. Celiktutan, B. Gokberk, B. Sankur and L. Akarun, "Bosphorus database for 3D face analysis," In COST Workshop on Biometrics and Identity Management, 2008.

[Scheenstra et al., 2005] A. Scheenstra, A. Ruifrok, and R. C. Veltkamp, "A survey of 3D face recognition methods," International Conference on Audio- and Video- based Biometric Person Authentication, 2005.

[Sedaaghi et al., 2007] M. Sedaaghi, C. Kotropoulos, D. Ververidis, "Improving speech emotion recognition using adaptive genetic algorithms," Proc. EUSIPCO, Poland, 2007.

[Sadegheih, 2011] A. Sadegheih, "New methods for system planning," Applied Soft Computing, Vol. 11, no. 1, pp. 1098-1107, 2011.

[Snelick et al., 2005] R. Snelick, U. Uludag, A. Mink, M. Indovina, and A. Jain, "Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems," IEEE Trans. on Pattern Analysis and Machine Intelligence, 2005.

[Szeptycki et al., 2009] P. Szeptycki, M. Ardabilian, and L. Chen, "A coarse-to-fine curvature analysis-based rotation invariant 3D face landmarking," International Conference on Biometrics: Theory, Applications and Systems, 2009.

[Shan et al. 2009] C. Shan, S. Gong and P. W. Mcowan, "Facial expression recognition based on local binary patterns: a comprehensive study," Image and Vision Computing, vol. 27, no. 6, pages 803-816, 2009.

[Tanaka et al., 1998] H. Tanaka, M. Ikeda, and H. Chiaki, "Curvature-based face surface recognition using spherical correlation," IEEE International Conference on Automatic Face and Gesture Recognition, 372-377, 1998.

[Tsalakanidou et al., 2007] F. Tsalakanidou, C. Dimitriadis, S. Malassiotis, "A secure and privacy friendly 2D+3D face authentication system robust under pose and illumination variations," Pattern Recognition Letters (PRL) 28(16):2238-2249 (2007).

[Tsalakanidou et al., 2004] F. Tsalakanidou, D. Tzovaras, M.G. Strintzis, "Integration 2D and 3D images for enhanced face authentication," FGR 2004:266-271.

[Tsalakanidou et al., 2003] F. Tsalakanidou, D. Tzovaras, M.G. Strintzis, "Use of depth and colour Eigen faces for face recognition," Pattern Recognition Letters (PRL) 24(9-10):1427-1435 (2003).

[Toh et al., 2004] K.-A. Toh, X. Jiang, and W.-Y. Yau, "Exploiting Global and Local Decisions for Multimodal Biometrics Verification," IEEE Transactions on Signal Processing, (Supplement on Secure Media), vol. 52, no. 10, pp. 3059-3072, October 2004.

[Turk and Pentland, 1991] M. Turk and A. Pentland. Eigenfaces for face recognition. Journal of Cognitive Neuroscience, vol. 3, no. 1, pages 71-86, 1991.

[Wang et al., 2007] J.-G. Wang, K.-A. Toh, E. Sung and W.-Y. Yau, "A Feature-level Fusion of Appearance and Passive Depth Information for Face Recognition," Face Recognition, Chapter

28, Face Recognition, book, ISBN 978-3-902613-03-5, I-Tech Education and Publishing, Vienna, Austria, April 2007.

[Wang et al., 2005] J.-G. Wang.; K.-A. Toh; R. Venkateswarlu, "Fusion of Appearance and Depth Information for Face Recognition," Audio- and Video-based Biometric Person Authentication AVBPA 2005, LNCS 3546, pp. 919-928, 2005.

[Wang et al., 2002] Y. Wang, C. S. Chua, Y. K. Ho, "Facial feature detection and face recognition from 2D and 3D images," Pattern Recognition Letters, 23:1191- 1202, 2002.

[Wright et al., 2009] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," IEEE Trans. on Pattern Analysis and Machine intelligence, 2009.

[Wiskott et al., 1997] L. Wiskott, J. M. Fellous, N. Kruger, and C. v. d. Malsburg, "Face recognition by elastic bunch graph matching," IEEE Trans. on Pattern Analysis and Machine Intelligence, 1997.

[Xie & Feng, 2009] Z. Xie, J. Feng, "KFCE: a dictionary generation algorithm for sparse representation," Signal Processing, 2009.

[Yang & Honavar, 1998] J.H. Yang and V. Honavar, "Feature subset selection using a genetic algorithm," IEEE Intelligent Systems, vol. 13, no. 2, pages 44–49, 1998.

[Zhao et al., 2003] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: a literature survey," ACM Computing Survey, 399–458, 2003.

[Zhao, 2011] X. Zhao, "Simulated annealing algorithm with adaptive neighborhood," Applied Soft Computing, vol. 11, no. 2, pp. 1827-1836, 2011.

[Zucker, 2005] S.W. Zucker, "Differential geometry from the frenet point of view: boundary detection, stereo, texture and color," Handbook of Mathematical Models in Computer Vision, N. Paragios, Y. Chen, and O. Faugeras, eds., Springer, 2005.

Liste de publications

Articles acceptés :

[Ben Soltana et al., 2012] W. Ben Soltana, M. Ardabilian, B. Ben Amor, L. Chen, J. Colineau, L. Daniel, M. Daoudi, H. Drira, J.L. Dugelay, N. Erdogmus, D. Huang, P. Lemaire, Przemyslaw Szeptycki, "3D Face Recognition: A Robust Multi-matcher Approach to Data Degradations," ICB 2012, the 5th IAPR International Conference on Biometrics, March 29 - April 1, 2012 New Delhi.

[Huang et al, 2011] D. Huang, W. Ben Soltana, M. Ardabilian, Y. Wang, L. Chen, "Textured 3D Face Recognition using Biological Vision-based Facial Representation and Optimized Weighted Sum Fusion, " IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshop on Biometrics, Colorado Springs, USA. 2011.

[Ben Soltana et al., 2011b] W. Ben Soltana, D. Huang, M. Ardabilian, L. Chen, C. Ben Amar, "A mixture of gated experts optimized using Simulated Annealing for 3D Face Recognition," International Conference on Image Processing (ICIP), Brussels, Septembre 11 -14, 2011.

[Ben Soltana et al., 2010a] W. Ben Soltana, D. Huang, M. Ardabilian, L. Chen, C. Ben Amar, "Comparison of 2D/3D Features and Their Adaptive Score Level Fusion for 3D Face Recognition," 3D Data Processing, Visualization and Transmission (3DPVT), Paris, May 17-20, 2010.

[Ben Soltana et al., 2010b] W. Ben Soltana, M. Ardabilian, L. Chen, C. Ben Amar, "Adaptive Feature and Score Level Fusion Strategy using Genetic Algorithms," International Conference on Pattern recognition (ICPR), Istanbul, August 2010.

[Ben Amor et al., 2012] B. Ben Amor, H. Drira, M. Daoudi, M. Ardabilian, W. Ben Soltana, P. Lemaire, L. Chen, N. Erdogmus, J.L. Dugelay, J. Colineau, "Fusion d'Experts pour une Biométrie Faciale 3D Robuste aux Déformations," 18ème édition du congrès francophone sur la Reconnaissance des Formes et l'Intelligence Artificielle (RFIA'12), 24 au 27 janvier 2012 à Lyon.

[Ben Soltana et al., 2011a] W. Ben Soltana, D. Huang, M. Ardabilian, L. Chen, C. Ben Amar, "Modèle de fusion de classifieurs basé sur le recuit simulé et son application à la vérification de visages en 3D," TAIMA'2011, Traitement et Analyse de l'information Méthodes et Applications, Octobre 2011 Hammamet.

[Ben Soltana et al., 2010c] W. Ben Soltana, M. Ardabilian, L. Chen, C. Ben Amar, "Optimal fusion scheme selection framework based on genetic algorithms for multimodal face recognition," Compression et Représentation des signaux audiovisuels (CORESA), Lyon, Octobre 2010.

[Ben Soltana et al., 2009] W. Ben Soltana, M. Ardabilian, L. Chen, C. Ben Amar, "Une approche multimodale pour la reconnaissance du visage inspirée d'une méthodologie anthropométrique," Traitement et Analyse de l'information Méthodes et Applications, 2009, Hammamet.

Articles soumis:

W. Ben Soltana, D. Huang, M. Ardabilian, C. Ben Amar, and L. Chen: A Mixture of Gated Experts Optimized using Simulated Annealing for 3D Face Recognition, IEEE Transactions on Information Forensics and Security (TIFS).