

## **Fusion d'Experts pour une Biométrie Faciale 3D Robuste aux Déformations**

Ben Amor Boulbaba, Drira Hassen, Daoudi Mohamed, Ardabilian Mohsen,  
Ben Soltana Wael, Liming Chen, Pierre Lemaire, Nesli Erdogan, Jean Luc  
Dugelay, Joseph Colineau

► **To cite this version:**

Ben Amor Boulbaba, Drira Hassen, Daoudi Mohamed, Ardabilian Mohsen, Ben Soltana Wael, et al.. Fusion d'Experts pour une Biométrie Faciale 3D Robuste aux Déformations. RFIA 2012 (Reconnaissance des Formes et Intelligence Artificielle), Jan 2012, Lyon, France. pp.978-2-9539515-2-3. hal-00656552

**HAL Id: hal-00656552**

**<https://hal.archives-ouvertes.fr/hal-00656552>**

Submitted on 17 Jan 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Fusion d'Experts pour une Biométrie Faciale 3D Robuste aux Déformations

B. Ben Amor<sup>1</sup>, H. Drira<sup>1</sup>, M. Daoudi<sup>1</sup>, M. Ardabilian<sup>2</sup>, W. Ben Soltana<sup>2</sup>, P. Lemaire<sup>2</sup>, L. Chen<sup>2</sup>  
N. Erdogmus<sup>3</sup>, JL Dugelay<sup>3</sup>, J. Colineau<sup>4</sup>

<sup>1</sup> LIFL (UMR Lille1/CNRS 8022), Telecom Lille1, Université de Lille1, France.

<sup>2</sup> LIRIS (UMR CNRS 5205), Ecole Centrale de Lyon, France.

<sup>3</sup> EURECOM, Sophia Antipolis, France.

<sup>4</sup> THALES-TRT, 91767 Palaiseau, France

## Résumé

Nous étudions dans cet article l'apport de la géométrie tridimensionnelle du visage dans la reconnaissance des individus. La principale contribution est d'associer plusieurs experts (matcheurs) de biométrie faciale 3D afin d'achever de meilleures performances comparées aux performances individuelles de chacun, notamment en présence d'expressions. Les experts utilisés sont : ( $E_1$ ) Courbes radiales élastiques, ( $E_2$ ) MS-eLBP, une version étendue multi-échelle de l'opérateur LBP, ( $E_3$ ) l'algorithme de recalage non-rigide TPS, en plus d'un expert de référence ( $E_{ref}$ ) l'algorithme de recalage rigide connu ICP. Profitant de la complémentarité de chacun des experts, la présente approche affiche un taux d'identification qui dépasse les 99% en présence d'expressions faciales sur la base FRGCv2. Une étude comparative avec l'état de l'art confirme le choix et l'intérêt de combiner plusieurs experts afin d'achever de meilleurs performances.

## Mots Clef

Biométrie, visage 3D, géométrie Riemannienne, LBP, TPS, ICP, fusion.

## Abstract

In this paper, we propose to fuse different 3D-3D matchers to achieve high effectiveness in distinguishing facial differences between distinct individuals even in presence of facial expressions. We present four different matchers: ( $E_1$ ) Elastic radial curves, ( $E_2$ ) MS-eLBP: multi-scale extended LBP, ( $E_3$ ) TPS algorithm, in addition to one baseline matcher ( $E_{ref}$ ) ICP algorithm for rigid registration. Evaluations conducted on FRGCv2 (Face Recognition Grand Challenge) dataset on individual matchers and on their fusion (Multiple matching approach) show interest of the proposed approach. It achieves 98.6% as rank-1 recognition rate in presence of facial expressions which is competitor compared to state-of-the-art results.

## Keywords

Biometrics, 3D face, Riemannian geometry, LBP, TPS, ICP, fusion.

## 1 Introduction et état de l'art

La détection et la reconnaissance automatique de visages est un domaine de recherche très actif en vision par ordinateur. A la différence des autres biométries comme l'iris ou l'empreinte digitale, le visage est connu par son caractère naturel, non-intrusif et sans-contact. Néanmoins, les systèmes basés sur les images 2D (fixes ou vidéo) ont montré leurs faiblesses devant les variabilités causées par le changement de la pose, des conditions d'éclairages ou encore des expressions faciales. Ces verrous encore ouverts ont été soulignés par des campagnes d'évaluation internationales organisées régulièrement par le NIST, comme FRVT<sup>1</sup>, FRGC<sup>2</sup> et MBGC<sup>3</sup>. Depuis quelques années, une nouvelle modalité faciale a émergé vu ses capacités à faire face aux changements de la pose et d'illumination, c'est le forme 3D du visage. Avec le développement considérable de caméras 3D (voire 3D+temps), qui sont basées sur différentes technologies (laser, lumière structurée, ToF, Stéréo, etc.), les bases de test (FRGC, GavabDB, Bosphorus, etc.) ainsi que les approches basées sur cette nouvelle modalité ne cessent d'augmenter. Profitant des avantages des images 3D du visage, les équipes engagées dans le sujet se sont très vite rendu compte que le principal défi à relever réside dans les déformations dues aux expressions faciales. En effet, celles-ci affectent considérablement la forme 3D, notamment quand la bouche s'ouvre. Les expressions engendrent des transformations non-rigides (pliage de la peau), des transformations élastiques (compression, dilatation de la peau) et même des changements de topologie de la surface par l'ouverture de la bouche. Plusieurs approches ont été élaborées et testées. Nous en distinguons trois catégories. D'abord, les **approches globales** qui comparent les surfaces faciales utilisant des algorithmes comme ACP [14], ICP [15] ou TPS [16]. D'autres algorithmes plus sophistiqués ont été récemment proposés. Wang et al. [4] proposent de calculer une carte de différence de formes appelée Signed Shape Difference Map (SSDM). De cette carte sont extraites trois types de caractéristiques afin d'encoder les différences de formes.

<sup>1</sup> <http://www.nist.gov/itl/iad/ig/frvt-home.cfm>

<sup>2</sup> <http://www.nist.gov/itl/iad/ig/frgc.cfm>

<sup>3</sup> <http://www.nist.gov/itl/iad/ig/mbgc.cfm>

Ils utilisent des techniques de *Boosting* pour sélectionner les meilleurs caractéristiques. Ils identifient avec succès 98.39% de sujets de la base FRGCv2. *Kakadiaris et al.* [17] proposent de recalibrer les surfaces faciales à comparer à un modèle déformable annoté. La comparaison est faite par le biais de coefficients d'ondelettes calculées sur l'image de différence des déformations (*gallery-modèle*) et (*probe-modèle*). Ils obtiennent un taux d'identification de 97% sur la même base. La seconde catégorie, que nous appelons **approches locales**, exploitent des parties/régions sur le visage pour faire la comparaison. *Faltemier et al.* [2] définissent, recalent et comparent plusieurs régions chevauchées du visage qu'ils fusionnent par la suite avec des techniques de votes. Ils utilisent l'algorithme ICP de recalage rigide et obtiennent un taux de reconnaissance de 97.2% sur la base FRGCv2. *Berretti et al.* [11] segmentent la surface faciale en bandes géodésiques (en se basant sur une distance intrinsèque) et utilisent une approche d'appariement de graphes pour comparer les visages. L'approche atteint plus de 94% de taux de reconnaissance sur la même base. *ter Haar et al.* [13] utilisent des profils extraits de la surface faciale et calculent des distances euclidiennes entre eux. Leur approche atteint un taux de reconnaissance de 97% sur cette base. *Cook et al.* [1] appliquent la méthode *Log-Gabor* afin de segmenter l'image de profondeur en régions. Une ACP est, par la suite, appliquée sur les caractéristiques *Log-Gabor* pour réduire la dimensionnalité avant la comparaison des signatures. Leur approche achève un taux de reconnaissance de 96.6% sur FRGCv2. La troisième catégorie d'approches comporte **les approches multimodales (2D+3D)**. Nous citons ici l'approche proposée par *Mian et al.* [3] qui proposent un premier classificateur de rejet basé sur une représentation sphérique et SIFT puis ils segmentent le front et le nez pour la reconnaissance. Leur approche affiche un taux d'identification de 96.2% sur FRGCv2. A notre connaissance, aucun travail n'a étudié au part avant l'apport de la fusion d'approches (appelée aussi experts) afin d'améliorer les résultats de reconnaissance. Dans ce papier, nous proposons un schéma de fusion de quatre approches qui utilisent différemment le scan 3D du visage.

## 2 Vue d'ensemble de l'approche proposée

La figure 1 illustre le schéma général de notre approche. Elle est composée de quatre modules : (i) l'acquisition 3D, un scan 3D est pris pour le sujet soit pour la phase d'enrôlement ou alors de reconnaissance par le biais d'une caméra 3D ; (ii) les données brutes passent par une procédure de prétraitement qui sert à extraire la partie utile du visage, à boucher les trous et à filtrer le bruit d'acquisition ; (iii) un module de comparaison de surface faciale qui intègre les différents experts développés ; (iv) la phase de décision d'identification/vérification selon le scénario d'utilisation visé (identification/vérification).

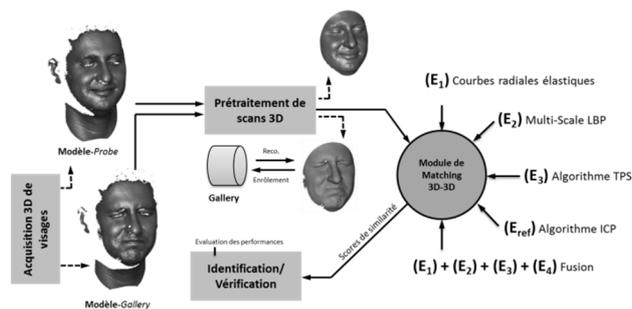


Fig.1 Vue d'ensemble de l'approche proposée.

Le reste de cet article est organisé comme suit. La procédure de prétraitement automatique des scans 3D ainsi que l'ensemble des experts seront détaillés dans la section 3. L'approche de fusion est abordée dans la section 4. La section 5 décrit la base de visages 3D ainsi que le protocole d'évaluation FRGC et discute les performances individuelles des experts et les performances de leur fusion. Nous donnons quelques conclusions et perspectives dans la section 6.

## 3 Présentation des experts

Trois experts de comparaison 3D-3D ont été développés en plus d'un expert de référence basé sur l'algorithme de recalage rigide ICP.

### 3.1 Prétraitement de scans 3D

Les données brutes issues des caméras 3D présentent très souvent des imperfections, telles que la présence de trous ou des points aberrants dus aux bruits d'acquisition. En plus, les scans en question contiennent des parties indésirables pour le procédé de reconnaissance telles que, le cou, les vêtements, etc. C'est pour toutes ces raisons qu'une phase de prétraitement s'avère nécessaire pour entamer la comparaison des surfaces faciales proprement dite.

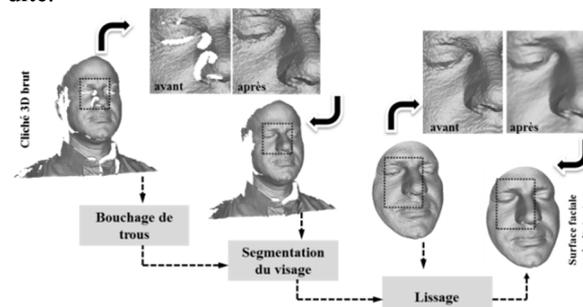


Fig.2 Procédure de prétraitement de scans 3D.

Comme l'illustre la figure 2, un premier filtre de détection et de bouchage de trous est appliqué. Les trous sont d'abord identifiés en localisant les contours fermés du maillage 3D, puis une triangulation est appliquée entre les points des bordures de ces trous. La seconde étape de cette phase consiste à détecter le bout du nez puis à segmenter la partie de la surface se trouvant à l'intérieur

d'une sphère de rayon constant  $R=100mm$  et ayant comme centre le bout du nez. Un filtre de lissage est enfin appliqué afin d'éliminer le bruit d'acquisition sur la surface résultante. Ainsi, nous obtenons des surfaces faciales prêtes à la phase de comparaison. Le pipeline de filtres de cette procédure (bouchage de trous, segmentation, lissage) est rapide et entièrement automatique. Son développement est basé en partie sur les algorithmes implémentés dans la bibliothèque *Visualisation Toolkit* [8].

### 3.2 Expert1: Courbes radiales élastiques,

Cet expert ( $E_1$ ) approxime la surface faciale par une collection indexée de courbes radiales issues du bout du nez et allant dans différentes directions. Le problème de comparaison de surfaces est alors ramené à un problème de comparaisons de courbes. Deux points ont motivé ce choix. Premièrement, cette représentation « locale », garantie une certaine robustesse aux données manquantes dues à l'occlusion par exemple. Deuxièmement, nous disposons d'algorithmes et de métriques plus avancés pour la paramétrisation et la comparaison des courbes comparées aux surfaces. L'expert ( $E_1$ ) est composé donc de deux étapes : l'extraction des courbes radiales et la métrique élastique pour comparer ces courbes modulus les déformations élastiques générées par les expressions.

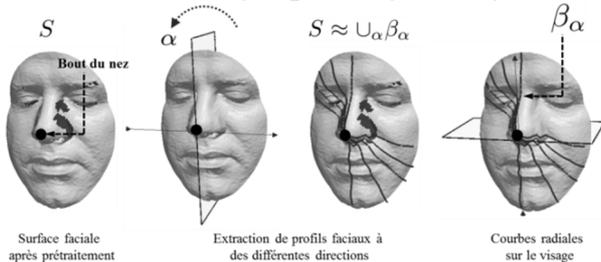


Fig.3 Procédure d'extraction de courbes radiales.

- *Extraction de courbes radiales* : La figure 3 illustre l'essentiel de cette étape. Partant du bout du nez, un plan ayant une normale variable (donc faisant un angle variable  $\alpha$  avec la verticale) coupe la surface faciale et génère ainsi des profils faciaux. Il en résulte une collection indexée de courbes radiales ayant comme origine le bout du nez et partant dans différentes directions.

- *Métrique élastique* : nous nous intéressons à présent à la comparaison des courbes radiales. Le défi à surmonter réside dans les déformations élastiques (compression, dilatation, pliage) que peut subir la peau et par conséquent les courbes. Pour modéliser ces variabilités intra-classes, une nouvelle para-métrisation des courbes est utilisée, la fonction SRVF, *Square Root Velocity Function* [9]. Pour de plus amples détails sur l'approche le lecteur pourra se référer à [5].



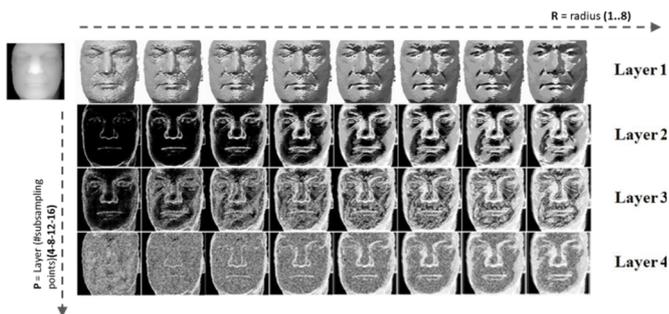
Fig.4 Exemples de géodésiques entre visages.

Si l'on revient maintenant aux surfaces faciales, puisque ces dernières sont approximées pas des collections de courbes, il est possible de construire des géodésiques entre les visages 3D. Les chemins donnés par la figure 4 en sont des exemples. Les deux premières lignes présentent des chemins intra-classe (même sujet avec des expressions différentes), alors que les deux dernières lignes affichent des chemins interclasses (sujets différents). C'est les longueurs de ces chemins qui présentent les distances entre les surfaces faciales.

### 3.3 Expert2 : MS-eLBP,

Le second expert ( $E_2$ ) combine une variante de l'algorithme LBP [18][19], MS-eLBP (*Multi Scale extended Local Binary Pattern*), pour la description de la surface faciale et l'algorithme SIFT pour extraire de cette représentation des points d'intérêt. Ces derniers sont par la suite appariés entre un modèle-*probe* et un modèle-*gallery*. Finalement, la mesure de similarité appliquée tient compte des appariements, des contraintes spatiales et de configuration. L'un des avantages non négligeable de cette approche est qu'il ne nécessite pas de recalage fin entre deux modèles faciaux ; pour la base FRGCv2, il ne nécessite aucun. Aussi, la stratégie adoptée dans cette approche la rend plus efficace aux variations d'expressions et aux occultations partielles.

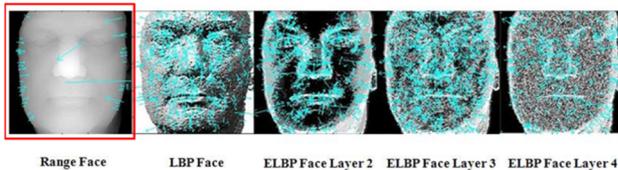
La variante eLBP suit le principe de LBP dans le calcul du gradient entre un pixel et les pixels de son voisinage. Sur des images de profondeur, elle permet d'estimer et d'encoder la courbure en chaque point. Contrairement à LBP, eLBP encode le gradient entre le pixel central et un point de voisinage sur 4 bits. Cet encodage sans perte permet de retenir la différence signée dans l'intervalle de -7 à 8 où le premier bit représente le signe. Finalement, la concaténation par bit des P voisins et la conversion en décimal résulte en 4 cartes (*Layers*).



**Fig.5** Les cartes *MSeLBP-DFs* d'un visage, obtenues avec des rayons de voisinage allant de 1 à 8.

La variante MS-eLBP applique ce principe à des voisinages de rayons différents, accentuant et encodant ainsi les informations discriminantes de la surface faciale. Au final, MS-eLBP comporte deux paramètres. Le premier est le nombre de pixels voisins,  $P$  pour un rayon de voisinage donné. Le second est le nombre  $R$  de rayons. Au total, cet expert produit  $R \times 4$  cartes que nous appelons *MS-eLBP Depth Faces (MSeLBP-DFs)*. La figure 5 présente ces cartes pour 8 rayons différents.

La plupart des techniques existantes, issues de l'état de l'art, nécessitent une étape de recalage très précise et de normalisation de la pose, coûteuse en temps de calcul. Pour pallier ce problème de coût lié au recalage et aussi à une comparaison dense, nous proposons d'utiliser l'approche SIFT [20] pour l'extraction et la comparaison de points d'intérêt. Elle consiste en la localisation, indépendante en échelle/en rotation, de points d'intérêts choisis pour leur saillance. Ici, nous utilisons cette méthode sur les  $R \times 4$  cartes *MSeLBP-DFs*. La figure 6 présente la localisation de points d'intérêt par SIFT sur l'image de profondeur d'un visage, puis sur les cartes *MSeLBP-DFs*.



**Fig.6** Détection de point d'intérêt par SIFT (en cyan) sur une image de profondeur du visage (à gauche) et une image de profondeur traitée par LBP (à droite).

Une fois les points d'intérêt extraits des visages *gallery* et du visage *probe*, il nous reste à les comparer. A cet effet, nous proposons une nouvelle méthode d'association entre points d'intérêt qui établit une mesure de similarité. Elle s'effectue d'une manière hybride, en exécutant une comparaison au niveau local (descripteurs *SIFT*) et à un niveau plus global, en utilisant la configuration spatiale des points d'intérêt. Dans le premier cas, il s'agit d'une méthode connue de l'état de l'art produisant une mesure de similarité  $S_N$  basée sur le nombre de points appariés.

Dans le second cas, nous proposons d'étudier les critères suivants :

(1) *Contrainte spatiale* : il s'agit ici de vérifier que les points associés par *SIFT* se situent bien dans la même région du visage entre le visage *probe* et le visage *gallery*. Pour cela le visage est divisé en  $3 \times 3$  blocs. L'on obtient alors la mesure de similarité  $S_S$ .

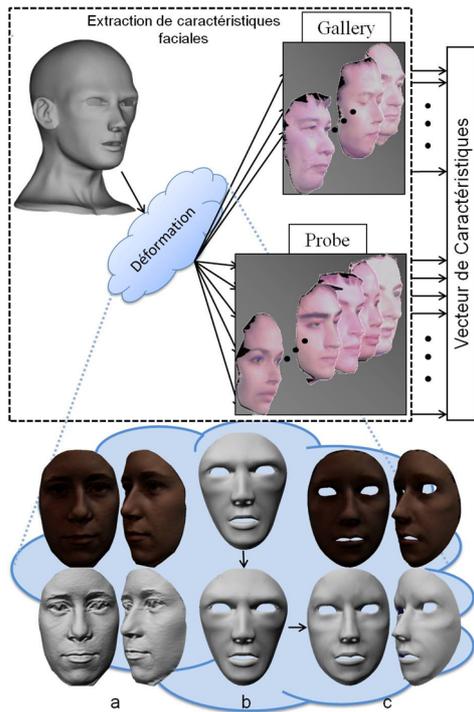
(2) *Contrainte en configuration* : on vérifie que la configuration globale des points d'intérêt sur le visage *probe* respecte la même que celle des visages *gallery*, par le biais de métriques simples. Nous obtenons la mesure de dissimilarité  $S_C$ .

En combinant cet ensemble de mesures d'appariements entre visages (différentes échelles, différentes contraintes d'associations entre points d'intérêt), nous obtenons une série de trois scores de similarité,  $S_N$ ,  $S_S$  et  $S_C$  qui sont normalisés et à leur tour combinés par le biais d'une méthode de fusion.

### 3.4 Expert3 : Reconnaissance à partir des paramètres de déformation TPS,

Le troisième expert ( $E_3$ ) est basé sur l'ajustement d'un modèle générique de visage. La reconnaissance faciale basée sur des modèles déformables a été largement étudiée dans la littérature. Un modèle déformable pour une classe d'objets, élaboré sous la forme d'une combinaison linéaire d'exemples de texture et de forme, a été introduit dans [21]. Son extension en 3D a été proposée dans [22] et a donné des résultats significatifs. Plutôt que de construire un modèle déformable à ajuster, le présent expert utilise ici un modèle générique de visage (comme montré sur la figure 7) fortement déformé/déformable pour s'ajuster sur les visages de la *gallery*, et ce en utilisant l'algorithme TPS (*Thin Plate Spline*) [23]. Ici, l'objectif est d'utiliser les propriétés discriminantes des paramètres d'ajustement et de déformation, obtenus lors du recalage. Quand un modèle générique est déformé pour s'ajuster sur un visage particulier de la base, on obtient une représentation approximative de la surface faciale. Si on considère un modèle générique comme une structure géométrique commune à tous les visages, ces paramètres représentent alors des variations spécifiques (d'une personne donnée par rapport au modèle générique). Ainsi, ces paramètres contiennent des éléments sur la forme du visage examiné permettant l'authentification des individus.

Avant d'appliquer la déformation TPS sur le modèle générique, celui-ci est remis à l'échelle et aligné avec le visage cible en se basant sur les 15 points caractéristiques. Ces 15 points caractéristiques sont repérés manuellement sur à la fois le modèle générique et le visage cible. Pour la mise à l'échelle, la distance entre les coins externes des yeux est utilisée. Pour l'alignement, une transformation rigide est estimée par moindres carrés où la distance au carré entre les ensembles de points est minimisée.



**Fig.7** Schéma général d'extraction de caractéristiques faciales (en haut) et illustration sur un exemple (en bas). (a) Le modèle cible avec et sans texture, (b) Le modèle générique avant et après alignement, (c) Le modèle générique après adaptation/déformation avec et sans texture.

Enfin, les déformations locales sont estimées en utilisant l'approche *TPS* qui, d'une manière générale, s'exprime comme suit : Pour deux surfaces 3D,  $S$  et  $T$ , et un ensemble de points sur chacune des deux surfaces,  $P_i$  et  $M_i$  respectivement, l'algorithme *TPS* estime une fonction d'interpolation  $f(x,y)$  pour calculer  $T'$ , qui approxime  $T$  en déformant  $S$  :

$$T' = \{(x', y', z'), st \forall (x, y, z) \in S, x' = x, y' = y, z' = z + f(x, y)\} \quad (1)$$

$$f(x, y) = a_x x + a_y y + \sum w_i U(|P_i - (x, y)|) \quad (2)$$

avec  $U(\cdot)$ , le noyau de la fonction exprimé comme:

$$U(r) = r^2 \ln(r^2), \quad r = \sqrt{x^2 + y^2} \quad (3)$$

Comme indiqué dans Eq. (2), la fonction d'interpolation  $f(x,y)$  inclut les paramètres de déformations  $w_i$ ,  $i \in \{1, 2, \dots, n\}$  à utiliser. Après l'étape d'alignement, en plus des 15 paires de points repérés manuellement, 136 paires supplémentaires de points sont générées, en couplant chaque point du modèle générique avec son point homologue le plus proche sur la surface cible (Figure 7). La déformation *TPS* est appliquée au modèle générique en utilisant l'ensemble des paires de points mentionnées, c'est-à-dire 151. Les poids  $w_i$  sont obtenus dans les 3 directions et le vecteur de recalage de chaque modèle, qui est une représentation compacte des surfaces faciales, est

trouvé:  $[(w_1^x, w_1^y, w_1^z), (w_2^x, w_2^y, w_2^z) \dots (w_{151}^x, w_{151}^y, w_{151}^z)]$ . La méthode proposée est illustrée sur la figure 7. Avec pour chaque point de contrôle un vecteur  $1 \times 3$  de déformation, une matrice de taille  $151 \times 3$  est obtenue. Les distances euclidiennes et cosinus ont été calculées entre les vecteurs de recalage, conduisant à deux vecteurs de distance de taille  $15 \times 1$ . 10% des points, identifiés par une approche *TM* ('trimmed mean'), sont ignorés dans les calculs car considérés comme erronés.

#### 4 Fusion des experts

En fusion des données, plusieurs approches existent: la fusion précoce qui s'opère sur les caractéristiques, la fusion tardive qui combine les scores de similarité et la fusion multi-niveaux qui exploite les deux précédentes approches. Dans ce travail, une approche de fusion tardive est adoptée vu que les experts utilisés sont hétérogènes. L'idée de base est, d'une part, de favoriser les meilleurs classificateurs en leur assignant les meilleurs poids, et d'autre part, les visages de la *gallery* qui ont obtenu le score le plus élevée. Dans notre cas d'étude, les pondérations associées aux classificateurs (ou les matcheurs)  $P^f$  sont basées sur le taux d'égalité d'erreur (EER) de chaque classificateur au niveau de la phase hors-ligne (phase d'apprentissage).

Soit  $e^f$ ,  $f = 1, 2, \dots, N$ , la valeur EER du  $f$ -ième classificateur.  $N$  étant le nombre total de classificateurs. Le poids  $P^f$ , est calculé comme suit :

$$P^f = \frac{1/e^f}{u}; \quad u = \sum_{f=1}^N \frac{1}{e^f}; \quad \sum_{f=1}^N P^f = 1; \quad 0 \leq P^f \leq 1$$

Pendant la phase en ligne, pour chaque visage «  $g$  » de la *gallery*, on calcule un score de similarité  $S^{g,f}$ , pour chaque descripteur «  $f$  » avec un visage *probe*. Tous ces scores de similarité sont ensuite triés dans un ordre décroissant. Nous assignons à chaque score  $S^{g,f}$ , un poids  $w^{g,f}$  qui est fonction de sa position  $p^{g,f}$ . Plus précisément, le poids  $w^{g,f}$  est définie comme suit:

$$w^{g,f} = f(p) = \ln\left(\frac{N_g}{p^{g,f}}\right)$$

où  $N_g$  est le nombre de personnes dans la *gallery*.

Le score de similarité, final, entre un visage de la *gallery* et le visage *probe* est estimé, dans un scénario d'identification, de la manière suivante :

$$S_{Final}(g) = \sum_{f \in \{Classificateurs\}} p^f * w^{g,f} * S^{g,f}$$

Finalement, le visage *probe* est identifié comme la personne de la *gallery* qui possède le meilleur score  $S_{Final}$ .

#### 5 Résultats expérimentaux

Afin d'évaluer les performances individuelles des experts ainsi que les performances de leur fusion, nous avons suivi dans nos expérimentations le protocole d'évaluation *FRGC* [10]. Ce protocole considéré aujourd'hui comme référence dans le domaine, consiste en une base de scans 3D collectés par le biais d'un scanner laser à l'université

Notre Dame et définissant un certain nombre d'expériences et des critères d'évaluation.

### 5.1 Base FRGC et protocole expérimental

La base FRGCv2 contient 4007 scans 3D de 466 sujets. Chaque sujet dispose d'au moins 1 scan à au plus 22 scans pris à différents intervalles de temps. Une partie de ces scans (frontaux) contient des visages non-neutres affichant une expression aléatoire. Les expériences consistent à calculer des matrices de similarités entre échantillons biométriques et à en extraire les deux types de scores correspondant à des accès clients ou alors à des accès imposteurs. Les deux scénarios classiques de la biométrie à savoir l'identification et la vérification sont visés dans ce protocole. Nous avons suivi ce protocole dans l'évaluation des experts individuellement ainsi que leur fusion, ceci nous permet aussi de situer les résultats de l'approche proposée par rapport à l'état de l'art. Pour le cas de fusion, nous procédons à une validation croisée pour l'évaluation de notre méthode de fusion avec la base de visages FRGCv2. Nous itérons alors l'expérimentation 50-fois et à chaque fois on construit deux parties : Une première partie « P1 » contenant N images pour M personnes sélectionnées aléatoirement. P1 est utilisé pour la génération des poids. Ensuite, on construit une deuxième partie « P2 », qui est composé des images des personnes restantes. On prend les images restantes de la personne 1 puis de la personne 2, etc. P2 est utilisée comme base de test.

### 5.2 Scénario d'identification

Dans ce scénario, les comparaisons sont faites entre les premiers scans neutres des sujets (*Gallery set*) avec le reste des scans (*Probe set*). Afin de quantifier l'effet des expressions faciales, trois comparaisons ont été effectuées (i) Neutre vs. Neutre (ii) Neutre vs. Non-Neutre et (iii) Neutre vs. Tout. Les résultats d'identification (taux de reconnaissance au rang-1) des différents experts sont donnés par la figure 9.

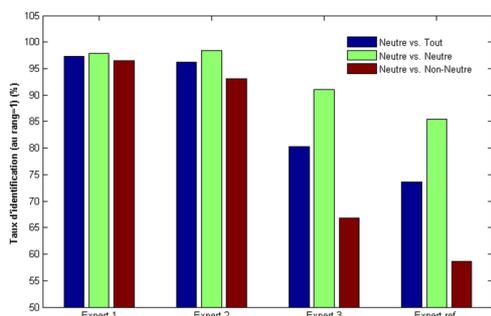


Fig.9 Taux d'identification (rang=1) pour les expérimentations Neutre vs. Tout, Neutre vs. Neutre et Neutre vs. Non-neutre.

Il est clair que les trois experts affichent des améliorations significatives comparées à l'expert de référence. Alors

que ce dernier consiste à aligner rigidement les surfaces faciales, les trois autres algorithmes tentent de modéliser différemment les déformations non-rigides dues aux expressions. Il est à noter aussi que l'expert ( $E_1$ ) affiche les meilleures performances vu la représentation par des courbes radiales et surtout la métrique élastique qui permet de prendre en compte les déformations élastiques. Ceci est encore plus visible dans l'expérience Neutre vs. Non-Neutre où l'on compare des visages expressifs à des visages neutres. La fusion des experts améliore encore plus les performances. En effet, la combinaison des quatre experts affiche un taux de reconnaissance qui dépasse les 99.6%, comparé aux performances des deux meilleurs experts ( $E_1$ ) et ( $E_2$ ) qui affichent respectivement des taux de 97.26% et 95.15%. Cette différence paraît faible, mais nous rappelons qu'un gain de 2% représente plus de 70 sujets supplémentaires reconnus par la fusion. Dans la figure 10 sont affichés les taux d'identification correspondants aux différents schémas de fusion ( $E_1+E_2$ ,  $E_2+E_3$ , ...,  $E_1+E_2+E_3+E_{ref}$ ).

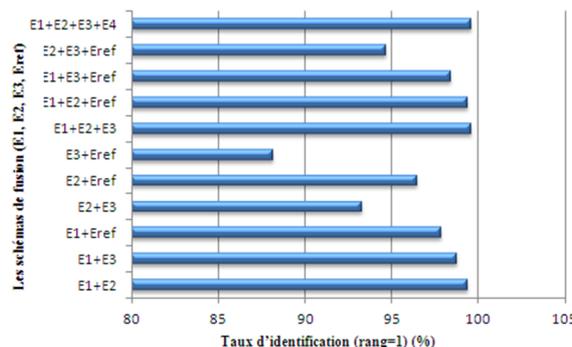


Fig.10 Taux d'identification pour les différents schémas de fusion ( $E_1+E_2$ ,  $E_2+E_3$ , ...,  $E_1+E_2+E_3+E_{ref}$ ).

Tab. 1 Comparaison avec les travaux existants (taux d'identification au rang=1 Neutre vs. Tout).

Approche	Taux d'identification (rang=1)
Mian et al. (3D) [3] (*)	96.2%
Gokberk et al. [24] (-)	96.2%
Cook et al. [1] (-)	94.6%
ter Haar et al. [13] (-)	97%
Berretti et al. [11] (+)	94.1%
Faltemier et al. [2] (-)	97.2%
Wang et al. [4] (*)	98.39%
<b>Expert (<math>E_1</math>) [5] (+)</b>	<b>97.26%</b>
<b>Expert (<math>E_2</math>) [6] (-)</b>	<b>95.15%</b>
<b>Expert (<math>E_3</math>) [7] (?)</b>	<b>80.15%</b>
Expert ( $E_{ref}$ ) (+)	73.08%
Somme simple (-)	98.42%
Produit simple (-)	98.19%
Somme pondérée [25] (-)	95.4%
<b>Fusion (<math>E_1+E_2+E_3+E_{ref}</math>) (-)</b>	<b>99.07%</b>
<b>Fusion (<math>E_1+E_2+E_3+E_{ref}</math>) (*) (-)</b>	<b>99.6%</b>

(+) Pas de fusion ; (-) Fusion en-ligne avec pondération ; (\*) Fusion hors-ligne avec pondération ; (-) Fusion sans pondération.

L'emploi de ce protocole nous permet également de situer les résultats de l'approche proposée par rapport à l'état de

l'art. Le tableau Tab.1 compare les performances des approches existantes avec les performances individuelles des experts et de leur fusion. On remarque, d'une part, la compétitivité affichée par la fusion des différents experts développés par rapport aux méthodes de l'état de l'art. D'autre part, on observe bien l'apport de l'utilisation des deux types de pondération (99.6% par rapport à la méthode de fusion simple basée sur la somme des cores normalisées 98.42%). Finalement, on constate un gain de performance de 0.53% (99.6% vs. 99.07%) en combinant la fusion en-ligne et hors-ligne par rapport à la fusion en-ligne seule.

### 5.3 Scénario de vérification

Sur les mêmes matrices de similarité, nous avons également produit les résultats de vérification. Les courbes DET (*Error trade-off*), traçant les taux d'erreur (faux négatifs et faux positifs) l'un en fonction de l'autre, pour chacun des experts et pour chacune des expériences : Neutre vs. Tout, Neutre vs. Neutre et Neutre vs. Non-Neutre, sont données par les figures 11, 12 et 13, respectivement. Pour les deux scénarios de la biométrie, les algorithmes (experts) de comparaison de visages 3D affichent des performances prometteuses. L'étude comparative avec l'état de l'art en est témoins. La fusion des experts profite de la complémentarité de ces derniers pour afficher des performances encore plus intéressantes.

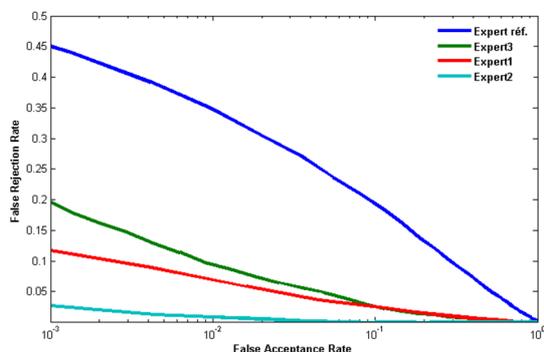


Fig.11 Courbes DET (Neutre vs. Tout).

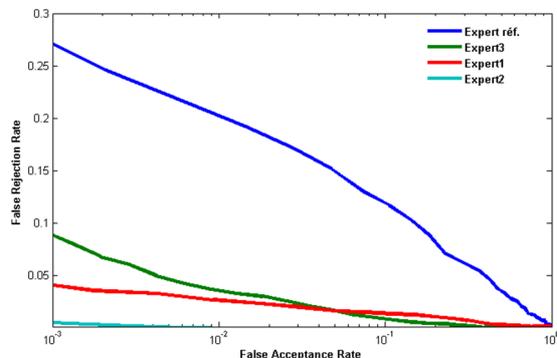


Fig.12 Courbes DET (Neutre vs. Neutre)

De ces trois graphes, nous pouvons constater, d'un côté, les améliorations des trois experts par rapport à l'expert

de référence ; d'un autre côté, les résultats intéressants de l'expert 2 ( $E_2$ ) basé sur la comparaison d'une batterie d'images de profondeurs issues de la version étendue et multi-échelle de l'opérateur LBP. Ceci est dû à son pouvoir de minimiser la zone d'erreur entre les distributions des accès clients et des accès imposteurs. En effet, cet expert est lui-même basé sur la fusion de plusieurs signatures produites par le biais de l'algorithme MS-eLBP.

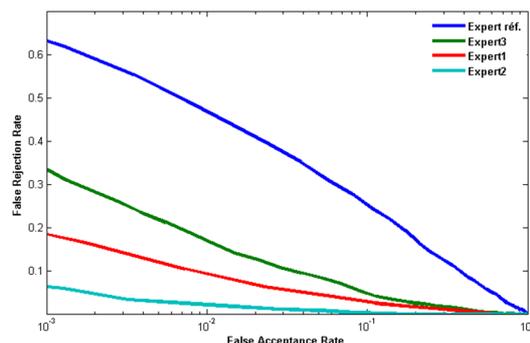


Fig.13 Courbes DET (Neutre vs. Non-Neutre)

## 6 Conclusions et perspectives

Nous avons étudié dans ce papier l'apport de la fusion d'experts de comparaison 3D-3D de visages pour faire face aux déformations causées par les expressions. Trois experts basés sur la forme faciale 3D sont proposés et testés sur un benchmark international de référence. Face aux expressions faciales, les trois experts affichent des performances qui dépassent de façon significative l'expert de référence ICP. Leur fusion améliore encore les performances avec un gain de 2% sur la base FRGCv2 (plus de 70 signatures supplémentaires reconnues par rapport au meilleur expert). Cette amélioration est due essentiellement à la complémentarité des experts. En effet, d'une part l'expert ( $E_1$ ) met l'accent sur les expressions en modélisant les déformations élastiques (compression, dilatation, pliage de la peau en plus des changements de la forme 3D causées par l'ouverture de la bouche). D'autre part, l'expert ( $E_2$ ) est basé sur la fusion d'une batterie de signatures qu'il génère avec l'opérateur eLBP, ce qui le rend plus robuste en scénario de vérification. Une étude comparative avec l'état de l'art montre la compétitivité de notre approche de fusion d'experts (99.6% de taux de reconnaissance comparé à la meilleure approche de l'état de l'art qui achève 98.39%). Les performances des experts de comparaison 3D-3D développés et de leur fusion montrent le niveau de maturité de cette modalité biométrique. Néanmoins, d'autres questions d'ordre théoriques et pratiques restent encore à résoudre. Nous citons dans la première catégorie, le passage à l'échelle (identifier une personne dans une base de millions de signatures) avec tous les verrous scientifiques qui en découlent. Un second problème serait

l'attaque du système par *Spoofing*, les cas des jumeaux ou des sosies en est les meilleurs exemples. Dans la seconde catégorie, nous soulevons d'autres problèmes comme le vieillissement, les interventions chirurgicales faciales, etc.

## Remerciement

Ce travail a reçu le support financier de l'ANR dans le cadre du projet FAR 3D ANR-07-SESU-004.

## Bibliographie

- [1] J. Cook, V. Chandran, and C. Fookes, "3D Face Recognition Using Log-Gabor Templates," *Proc. British Machine Vision Conf.*, pp. 83-83, 2006.
- [2] T. Faltemier, K.W. Bowyer, and P. Flynn, "A Region Ensemble for 3D Face Recognition," *IEEE Trans. Information Forensics and Security*, vol. 3, no. 1, pp. 62-73, Mar. 2008.
- [3] A.S. Mian, M. Bennamoun, and R. Owens, "An Efficient Multimodal 2D-3D Hybrid Approach to Automatic Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1927-1943, Nov. 2007.
- [4] Y. Wang, J. Liu, X. Tang, "Robust 3D Face Recognition by Local Shape Difference Boosting," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol.32, no.10, pp.1858-1870, Oct. 2010.
- [5] H. Drira, B. Ben Amor, M. Daoudi, A. Srivastava: "Pose and Expression-Invariant 3D Face Recognition using Elastic Radial Curves". *BMVC 2010*: 1-11
- [6] D. Huang, M. Ardabilian, Y. Wang, L. Chen: A novel geometric facial representation based on multi-scale extended local binary patterns. *FG 2011*: 1-7.
- [7] N. Erdogmus, R. Ethève, J. Dugelay, "Obtaining MPEG-4 compliant animatable 3D face models by using TPS method", *SIU'10, 18th IEEE Signal Processing and Communications Applications Conference*, April 22-24, 2010, Diyarbakir, Turkey, pp 157-160.
- [8] <http://www.vtk.org/>.
- [9] S. H. Joshi, E. Klassen, A. Srivastava, and I. Jermyn. "A novel representation for riemannian analysis of elastic curves in IR". In *Proceedings of the IEEE CVPR*, 2007.
- [10] 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 (1) 2005*: 947-954
- [11] S. Berretti, A. D. Bimbo, and P. Pala. "3D face recognition using isogeodesic stripes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2162-2177, 2010.
- [12] C. C. Queirolo, L. Silva, O. R. Bellon, and M. P. Segundo, "3D face recognition using simulated annealing and the surface interpenetration measure", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32:206-219, 2010.
- [13] F. ter Haar and R. C. Velkamp, "Expression modeling for expression-invariant face recognition", *Computers and Graphics*, 34(3):231-241, 2010.
- [14] C. Heshner, A. Srivastava, and G. Erlebacher, "Principal component analysis of range images for facial recognition," in *Proc. International Conference on Imaging Science, Systems, and Technology*, Las Vegas, NV, June 2002.
- [15] B. Ben Amor, M. Ardabilian, L. Chen, "New Experiments on ICP-Based 3D Face Recognition and Authentication", *ICPR (3) 2006*: 1195-1199.
- [16] X. Lu and A. K. Jain, "Deformation Analysis for 3D Face Matching", In *Proceedings of the Seventh IEEE Workshops on Application of Computer Vision (WACV/MOTION'05)*.
- [17] 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 Transactions on Pattern Analysis and Machine Intelligence*, 29(4):640-649, 2007.
- [18] T. Ojala, M. Pietikäinen, T. Maenpää. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.24, no.7, pp.971-987, 2002.
- [19] David G. Lowe, "Object recognition from local scale-invariant features", in *Proceedings of the International Conference on Computer Vision*, vol. 2, 1999, p. 1150-1157.
- [20] A. S. Mian, M. Bennamoun, and R. Owens, "Keypoint detection and local feature matching for textured 3D face recognition", *IJCV*, vol. 79, no. 1, pp. 1-12, 2008.
- [21] M. Jones, and T. Poggio, "Model-based matching of line drawings by linear combinations of prototypes", in *Proceedings of the Fifth International Conference on Computer Vision*, pp. 531-536, 1995.
- [22] V. Blanz, T. Vetter, "Face recognition based on fitting a 3D morphable model", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.25, no.9, pp. 1063- 1074, Sept. 2003.
- [23] F. L. Bookstein, "Principal warps: Thin-Plate Splines and Decomposition of Deformations", *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11(6):567-585, June 1989.
- [24] 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.
- [25] 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.