

Reconnaissance de visages en 3D orientée région

P. Lemaire¹

P. Szeptycki¹

M. Ardabilian¹

L. Chen¹

¹ LIRIS (Laboratoire d'InfoRmatique en Image et Systèmes d'information) -

Ecole Centrale de Lyon, Département Math-Info
Bât. E6, 36 avenue Guy de Collongue, 69134 Ecully Cedex – France

{pierre.lemaire, przemyslaw.szeptycki, mohsen.ardabilian, liming.chen}@ec-lyon.fr

Résumé

En reconnaissance faciale 3D, la robustesse des algorithmes aux expressions faciales demeure un défi. Dans cet article, nous abordons ce problème par une approche orientée régions. Le framework présenté est basé sur une nouvelle paramétrisation 3D du visage, invariante aux transformations non élastiques, utilisant les distances géodésiques. Il inclut un algorithme de segmentation automatique du visage en 3D inspirée par l'anatomie faciale, et introduit le concept de potentiel de déformation. Il comprend également une nouvelle méthode de fusion par pondération des scores de similarité. Enfin, nous présentons une évaluation des bénéfices de cette segmentation utilisant la base Bosphorus.

Mots clefs

Reconnaissance faciale, 3D, expressions faciales, région, segmentation, fusion de scores de similarité, distances géodésiques, ICP.

1 Introduction

Dans le domaine de la biométrie, la reconnaissance faciale est une modalités séduisante, car peu contraignante vis-à-vis des conditions d'acquisition et sans contact. Elle est néanmoins confrontée à de nombreux problèmes, parmi lesquels on peut notamment compter les variations d'éclairage, de pose, d'expression, les occultations [1]. La reconnaissance faciale en 3D permet d'en résoudre certains partiellement, notamment en ce qui concerne la pose et l'éclairage ; en revanche, elle accroît le temps de calcul sans pour autant résoudre le problème des expressions faciales [2].

De manière générale, un scénario de reconnaissance faciale classique compare un élément candidat a priori inconnu (probe) à un ensemble d'éléments connus (gallery). Le calcul d'un score de similarité entre le modèle probe et l'ensemble des modèles de la gallery permet d'identifier, ou de rejeter le candidat.

Plusieurs méthodes ont été proposées afin de permettre la reconnaissance automatique de visages en 3D.

La méthode considérée comme référence est une méthode de recalage de surfaces rigides, dite Iterative Closest Point (ICP) [3]. Le principe est de recalculer par itérations successives 2 surfaces rigides en calculant la transformation (translation et rotation) optimale permettant de minimiser la distance point à point entre les 2 surfaces. L'écart restant est considéré comme une mesure de similarité.

Ben Amor et al. [9] proposent une méthode appelée R-ICP inspirée d'ICP sur l'ensemble du visage, mais ne prenant en compte que la partie statique du visage dans l'étape de recalage. Le score de similarité est obtenu en effectuant une somme pondérée sur les distances point à point entre la partie statique et la partie rigide du visage.

Faltemier et al. [8] proposent d'extraire 28 régions sphériques du visage, et d'effectuer un recalage rigide (par méthode ICP) entre probe et gallery sur chacune d'entre elles. La fusion des scores est inspirée de la méthode Borda Count. Pris individuellement, les scores par région indiquent l'importance des régions statiques du visage relativement aux régions mimiques dans la reconnaissance faciale en 3D.

Drira et al. [6] proposent de représenter le visage 3D sous forme de courbes géodésiques, et d'étudier le coût des déformations nécessaires à appliquer à ces courbes pour passer d'un visage probe à un visage gallery, ce qu'ils appellent un chemin géodésique, en tant que score de similarité. Cette méthode les amène à sélectionner et pondérer les meilleures courbes, correspondant globalement à celles subissant le moins de déformations lors des expressions.

Kakadiaris et al. [7] propose de modéliser le visage à l'aide d'un modèle annoté et déformable. A l'aide de ce dernier, on projette des caractéristiques (orientation des normales à la surface notamment) du visage 3D étudié sur un plan 2D, normalisé d'où sont extraits des descripteurs à l'aide de transformées en ondelettes. La fusion des descripteurs est pondérée selon la région du visage d'où sont extraites les caractéristiques, afin d'être moins sensible aux expressions.

Mian et al. [4] et Huang et al. [5] proposent une méthode holistique. Ils extraient un nombre réduit de points caractéristiques de la surface 3D, formant ainsi un graphe,

et en étudiant la similarité en tant que mesure de similarité entre visages 3D. Les auteurs reportent des résultats relativement sensibles aux expressions.

La plupart des méthodes proposées dans l'état de l'art emploie ainsi diverses techniques dont l'objectif est souvent de réduire la sensibilité aux expressions.

Dans cet article, nous nous proposons d'adresser le problème de la reconnaissance faciale en 3D, et plus précisément celui de la robustesse aux expressions, à l'aide du système (*framework*) suivant. Dans un premier temps, le visage en 3D est prétraité et on en extrait des points caractéristiques. Puis on effectue une paramétrisation de la surface 3D à l'aide des distances géodésiques. Nous comparons ensuite le visage 'probe' au visage 'gallery', préalablement et automatiquement segmenté en régions, à l'aide de notre paramétrisation. Enfin, le score de similarité entre 2 visages est obtenu comme une somme pondérée des scores sur chacune des régions. Le score associé à une région est calculé sur la base d'une méthode de recalage rigide.

La suite de cet article est organisée de manière suivante.

La section 2 est consacrée à la présentation des bases de notre théorie et la présentation de notre approche. Dans la troisième partie, nous évaluons les bénéfices de notre approche à l'aide d'une expérimentation sur la base Bosphorus [11]. Nous concluons finalement par la section quatre tout en présentant quelques perspectives.

2 Approche proposée

La notion d'expression faciale étant très subjective, nous privilégions l'étude des déformations causées par les expressions. Ces dernières peuvent être décomposées sous forme de l'activation d'unités d'action (Action Units, AU), telles que décrites par Ekman et al. [10] au sein du Facial Action Coding System (FACS). Chaque AU peut se décrire comme étant la contraction ou la détente d'un ou plusieurs muscles du visage. D'autre part, Ben Amor et al. [9] ont mené une étude quant à l'influence de l'activation des principaux muscles et groupes musculaires faciaux (figure 1) sur les déformations de la surface faciale. Cette étude révèle que l'activation de certains muscles et groupes musculaires engendre des déformations avec une amplitude variant selon la région où ils sont situés sur le visage (figure 2). Plus précisément, l'activation de groupes musculaires sur le haut du visage engendre moins de déformations que l'activation de groupes musculaires sur le bas du visage.

Cette expérimentation est en adéquation avec le point de vue anatomique selon lequel certaines régions du visage sont statiques tandis que d'autres sont mimiques. Cela corrobore également l'idée selon laquelle l'activation d'une AU engendre des déformations locales, tandis que d'autres parties du visage restent statiques ou quasiment statiques. Cette étude anatomique nous confirme en outre l'existence de points de repère anthropométriques tels que les coins intérieurs des yeux et le bout du nez, dont la

présence et la localisation sur la surface du visage est indépendante des expressions et des morphologies.

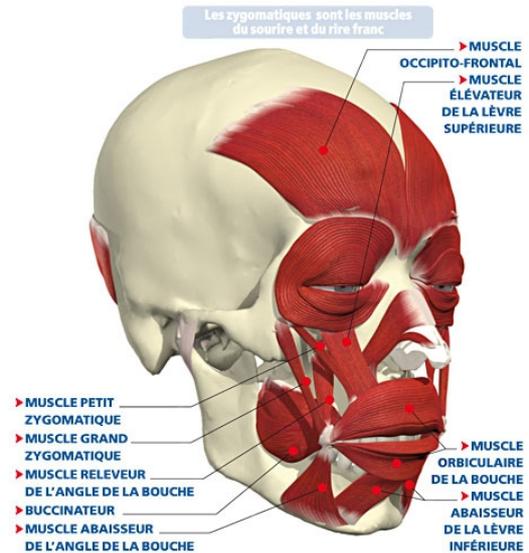


Figure 1 - Anatomie du visage humain

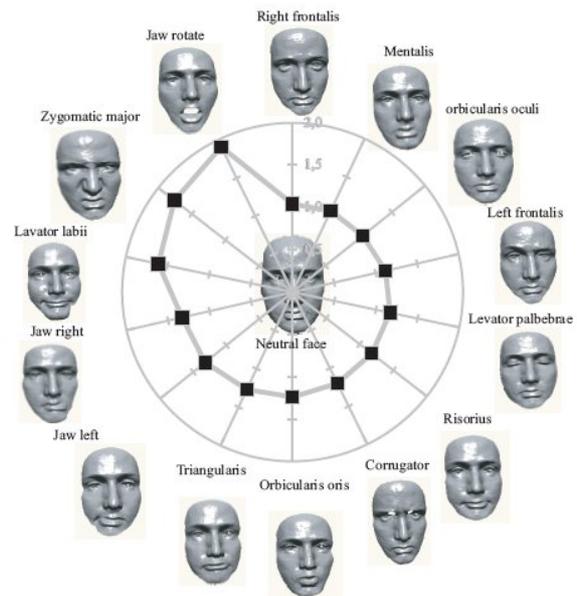


Figure 2 - Classement de l'amplitude des déformations en fonction du groupe musculaire activé

Malheureusement, la modélisation complète du système musculaire facial chez l'humain est difficile à mettre en œuvre. D'autre part, l'étude présentée dans [16] montre que les expressions engendrent des déformations élastiques de la surface du visage.

Nous cherchons donc dans ce travail à segmenter le visage en plusieurs régions, correspondant aux régions déformées ou non par les principales AU. Cette segmentation se doit d'être indépendante autant que possible des déformations élastiques et non élastiques du visage, pour un individu

donné. Elle doit permettre une meilleure robustesse aux expressions dans un scénario de reconnaissance ou d'identification.

2.1 Schéma global

Globalement, l'ensemble de l'approche de reconnaissance basée région peut être scindée en deux phases, l'une hors-ligne et l'autre en-ligne (figure 3). La première phase est consacrée au traitement et à l'analyse des modèles de la gallery. Pendant la phase en-ligne, un modèle probe est apparié avec les modèles de la gallery, et les mesures de similarité sont calculées. Une mesure de similarité entre deux visages est la pondération des mesures de similarité par région, dont les coefficients ont été appris suite à un apprentissage qui sera abordé dans la section 2.5.

Dans la suite nous détaillons les étapes de l'approche proposée.

2.2 Prétraitement

2.2.1 Qualité des modèles

Les objectifs du prétraitement des modèles 3D sont de minimiser l'influence de la qualité de l'acquisition à l'étape de la reconnaissance. Les données sont en effet

généralement des images de profondeur et non pas des modèles 3D complets, ce qui implique que certaines parties du visage peuvent être manquantes. Ils comportent souvent également des pics et des trous, et font presque systématiquement l'objet d'un bruit d'acquisition. Pour assurer l'invariance aux poses, un modèle 3D complet est généré. Afin de supprimer les pics, on applique un filtre médian aux points dont les coordonnées en Z sont détectées comme aberrantes. Afin de corriger les trous, c'est-à-dire de déterminer les coordonnées en Z des points manquant sur l'image de profondeur, on opère une régression linéaire sur l'estimation des coefficients polynomiaux biquadratiques. Le détail de cette partie est disponible dans [12].

2.2.2 Localisation des points de repère

La segmentation du visage, c'est-à-dire la localisation des zones de déformations liées aux AU, repose en grande partie sur une bonne localisation, c'est-à-dire robuste et précise, des points de repère. Cette localisation doit être invariante aux poses et aux conditions d'éclairage. Elle ne doit donc de préférence pas utiliser l'information texturale. Notre approche de localisation de points de repère anthropométriques est basée sur l'utilisation des

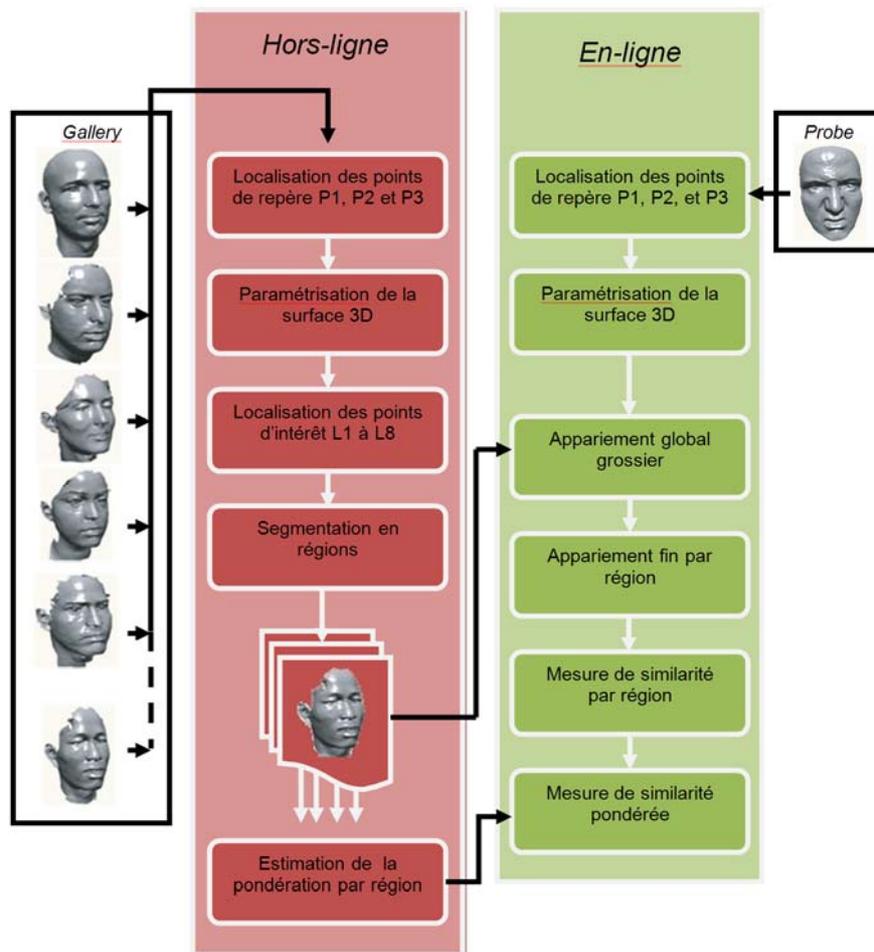


Figure 3 - Schéma global de notre framework

courbures moyenne et gaussienne, et un modèle générique. Par seuillages successifs, et en étudiant les courbures moyennes et gaussiennes à différents rayons, nous localisons les coins intérieurs des yeux (notés P_1 et P_2) et le bout du nez (noté P_3), en affinant progressivement la précision de la localisation. Le détail de cet algorithme est disponible dans [12].

2.3 Paramétrisation de la surface 3D faciale

L'étape de paramétrisation vise à accorder les mêmes coordonnées à un même point physique, dans deux modèles 3D faciaux différents d'un même individu. On suppose que les points de repère anthropométriques P_1 à P_3 (coins intérieurs des yeux et bout du nez) sont localisés de manière précise. Ils sont naturellement distincts. Chaque point p d'un modèle facial 3D est alors décrit de manière unique par ses distances géodésiques (d_1, d_2, d_3) à ces 3 points (figure 4). Cette paramétrisation comporte l'avantage d'être invariante aux déformations non élastiques, c'est-à-dire aux translations et aux rotations. Contrairement à la représentation euclidienne donc, cette paramétrisation n'exige pas de recalage relatif à la pose. Elle est appliquée à chaque visage de la gallery, mais aussi à chaque visage probe. A un point p de coordonnées (d_1, d_2, d_3) d'un visage gallery, on apparie p' de coordonnées (d_1', d_2', d_3') du visage probe tel que la distance D entre (d_1, d_2, d_3) et (d_1', d_2', d_3') soit minimale. Dans nos travaux, D est une distance euclidienne.

2.4 Segmentation en régions

La segmentation du visage constitue l'une des étapes centrales de l'approche de reconnaissance par région. C'est à cette étape que les régions statiques et mimiques du visage sont localisées automatiquement, pour chaque visage de la gallery.

La segmentation d'un visage selon des critères purement anatomiques est délicate, étant donné qu'aucun champ d'action d'un groupe musculaire ou d'une AU ne peut semble-t-il être localisé simplement selon des critères liés à la forme ou à la texture du visage. Nous avons donc choisi de déterminer nos régions de manière empirique, à l'aide de cartes dites de potentiel de déformation. Le potentiel de déformation est ici la mise en correspondance, via la paramétrisation précédemment exposée, d'une mesure de courbure entre 2 visages différents. En l'occurrence, la mesure de courbure dont nous nous sommes servis est le shape-index [13] avec un rayon de 25mm, indiquant la topologie de la surface au voisinage d'un point du modèle 3D. En appariant de la sorte plusieurs visages d'une même personne soumis à l'activation d'une ou plusieurs Action Units (AU), nous pouvons mettre en évidence à l'aide du potentiel de déformation les régions statiques (le shape-index varie peu d'un visage à l'autre, c'est-à-dire que sa variance est faible) et les régions mimiques (le shape-index varie

largement d'un visage à l'autre, c'est-à-dire que sa variance est grande).

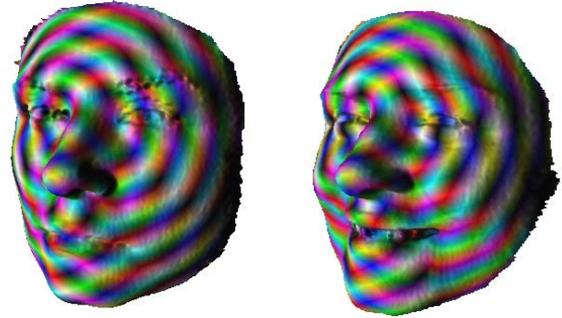


Figure 4 - La paramétrisation du visage à partir des points d'intérêt P_1 à P_3 , utilisant les distances géodésiques. Ici, les composantes RVB correspondent respectivement aux distances géodésiques à P_1, P_2 et P_3 .

A la suite de nos observations sur le potentiel de déformation à l'aide de la base Bosphorus [11] (détaillée dans la partie 3), nous avons décidé de découper le visage en 8 régions, correspondant au nez, aux yeux gauche et droit, aux pommettes gauche et droite, au front ainsi qu'à la mâchoire, répartie en 2 régions séparées (figure 5). Ces régions correspondent à des régions globalement stables, ou globalement mimiques, selon nos observations sur le potentiel de déformation, sur plusieurs visages. La séparation de la bouche en 2 régions distinctes est supposée permettre une meilleure robustesse aux problèmes d'expressions partielles, de modèles partiels ou aux occultations. Afin de segmenter la surface faciale 3D de la sorte, nous localisons d'abord les points de repère L_1 à L_8 correspondant aux représentants des 8 régions. Ceux-ci sont localisés grâce à leurs distances géodésiques à P_1, P_2 et P_3 , à l'aide d'un modèle générique. Ensuite, chaque point du visage est associé respectivement à la région S_i (centrée en L_i) à l'aide de deux paramètres, w_i et $R_i, i \in [1,8]$. Plus précisément, tout point p du visage appartient à la région S_i , représentée par le point de repère L_i , lorsque

$$s_i = \min(s_j) \text{ avec } j \in [1,8]$$

où s_i est calculé de la manière suivante :

$$\text{Si } w_i \times d_i > R_i \text{ alors } s_i = \infty$$

$$\text{sinon } s_i = w_i \times d_i$$

où d_i est la distance géodésique du point p au point de repère L_i .

2.5 Calcul de similarité et pondération par région

Une fois que la segmentation en régions est effectuée, on calcule un score de similarité pour chaque région. Cette étape est relativement indépendante de la segmentation en

régions du visage, et si dans cet article nous présentons des résultats relatifs à l'algorithme de référence ICP, afin d'évaluer les bénéfices de la segmentation du visage, il est envisageable d'utiliser toute autre mesure de similarité dans cette partie.

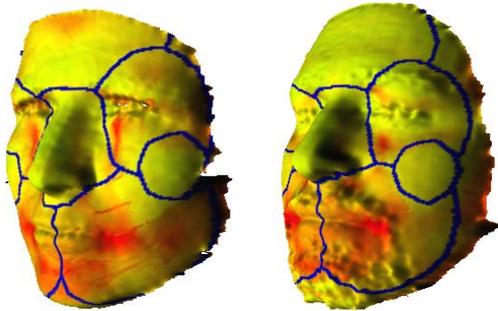


Figure 5 - Exemples de segmentation du visage en régions, sur des cartes de potentiel de déformation

Pour chaque point \mathbf{p} du visage gallery dans une région donnée, nous apparions le point \mathbf{p}' correspondant du visage probe suivant notre paramétrisation. Une itération d'ICP permet de minimiser la distance euclidienne au sens des moindres carrés entre les deux régions. C'est-à-dire qu'on calcule la rotation et la translation qui minimise cette distance. Le score de similarité par région est la moyenne des distances euclidiennes point à point entre le visage gallery et le visage probe au sein de cette région. Finalement, le score de similarité entre deux visages est une mesure de similarité pondérée de la forme suivante :

$$M = \sum_{i=1}^8 W_i \times d(S_i, S'_i)$$

$$\text{avec } d(S_i, S'_i) = \frac{1}{\text{card}(S_i)} \sum_{p_j \in S_i, p'_j \in S'_i} d_{\text{euclid}}(p_j, p'_j)$$

où $\text{card}(S_i)$ correspond au nombre de points \mathbf{p} compris dans la région S_i . La détermination des poids lors de cette fusion de scores est déterminée à l'aide d'une base d'apprentissage. Les poids sont déterminés de sorte à minimiser l'Equal Error Rate (EER). Les détails concernant cet algorithme de fusion sont disponibles dans [14].

3 Expérimentation

L'expérimentation a été menée sur la base de visages Bosphorus [11]. Cette base est constituée des images de profondeur de 105 individus, pour la plupart des acteurs professionnels, avec 1 à 3 modèles neutres par individu. Pour chaque individu, la base comporte 34 modèles avec déformations, dont 28 obtenus par l'activation d'AU isolées et 6 par expression. La base présente également des modèles avec occultations (lunettes ou mains) ou rotations. Nous n'avons pas utilisé ces derniers dans

l'expérimentation présentée dans cet article. Cette base est réputée difficile, étant centrée sur les expressions faciales. Concernant notre expérimentation, 29 individus ont été sélectionnés aléatoirement, avec au moins un modèle neutre par individu. Ces modèles neutres constituent notre gallery. Par la suite, 80 modèles probe dont 74 avec expression et 6 neutres ont été sélectionnés pour évaluer les performances de notre algorithme et les bénéfices de la segmentation en région présentée dans cet article. La moitié de ces modèles est utilisée pour estimer la pondération optimale par région, l'autre constitue les modèles de test.

Nous avons comparé les améliorations incrémentales de notre approche (approche D) avec l'approche baseline ICP. Les améliorations successives de l'approche R-ICP [9] sont obtenues en additionnant simplement les scores par région mais sans apprentissage (approche B), en fusionnant les scores par la méthode Borda Count modifiée présentée dans [8] (approche C), et finalement en utilisant la pondération optimale des scores basée sur l'EER, telle que présentée dans [14].

Les résultats sont présentés dans le tableau 1. On remarque une amélioration progressive du taux de reconnaissance avec le raffinement des méthodes de fusion. Dans tous les cas, notre approche de segmentation du visage en régions ainsi que notre paramétrisation permettent une amélioration du taux de reconnaissance.

L'expérimentation menée jusqu'à présent, même si elle reste à affiner et à approfondir, présente une première tendance. Dans la suite, il est prévu de procéder à des tests plus significatifs en augmentant le nombre d'individus, en utilisant les bases FRGC 1.0, FRGC 2.0 [15] et finalement en procédant à une validation croisée.

Approche	Taux d'identification
A - baseline ICP	78.5%
B - régions sans pondération	82.5%
C - régions Borda Count	87.5%
D - régions avec apprentissage	92.5%

Tableau 1 - Résultats de l'expérimentation : améliorations successives de notre approche comparées à l'approche baseline ICP

4 Conclusion et perspectives

Dans cet article, nous avons présenté nos travaux sur la reconnaissance faciale en 3D orientée régions. Les expérimentations menées montrent des résultats encourageants concernant l'approche région, avec une augmentation de 14% du taux de reconnaissance par rapport à la méthode de référence, en présence d'expressions faciales. Cependant, il reste de nombreux points à améliorer. Notons parmi ceux-ci le score de similarité entre régions, pouvant être amélioré par des mesures d'interpénétration de surface par exemple. Notons également qu'une stratégie de cascade pour la

fusion des scores de similarité peut permettre de rejeter prématurément un visage, réduisant ainsi le temps de calcul. Le cas des occultations est également un travail qu'il nous reste à mener ; une approche orientée région semblerait adaptée pour mieux le traiter.

Références

- [1] W. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, Face Recognition: A Literature Survey, *ACM Computing Surveys*, vol. 35, no. 4, pp. 399-458, 2003
- [2] K.W. Bowyer, K. Chang, P.J. Flynn, A Survey of Approaches and Challenges in 3D and Multi-Modal 3D+2D Face Recognition, *Computer Vision and Image Understanding*, vol. 101, pp. 1-15, 2006
- [3] P.J. Besl and N.D. McKay, A Method for Registration of 3-D Shapes, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239-256, Feb. 1992
- [4] A. Mian, M. Bennamoun, 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, June 2007
- [5] D. Huang, M. Ardabilian, Y. Wang, L. Chen. Asymmetric 3D/2D Face Recognition Based on LBP Facial Representation and Canonical Correlation Analysis, *IEEE International Conference on Image Processing (ICIP 2009)*, Cairo, Egypt, 2009
- [6] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, A Riemannian Analysis of 3D Nose Shapes For Partial Human Biometrics, *ICCV 2009*, Kyoto, Japan, 2009
- [7] I. Kakadiaris, G. Passalis, G. Toderici, N. Murtuza, T. Theoharis, Three-Dimensional Face Recognition in the Presence of Facial Expression: An Annotated Deformable Model Approach, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 640-649, avril 2007
- [8] T. Faltemier, K.W. Bowyer, P.J. Flynn, A Region Ensemble for 3D Face Recognition, *IEEE Trans. Information Forensics and Security*, vol. 3, no. 1, pp. 62-73, Mar. 2008
- [9] B. Ben Amor, Contributions à la reconstruction, la reconnaissance et l'authentification faciale 3D, thèse préparée au sein du laboratoire LIRIS, sous la direction de L. Chen, M. Ardabilian, soutenue le 8 décembre 2006
- [10] P. Ekman, W.V. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement, *Consulting Psychologists Press*, Palo Alto, 1978
- [11] A. Savran, N. Alyüz, H. Dibeklioglu, O. Çeliktutan, B. Gökberk, B. Sankur, L. Akarun, Bosphorus Database for 3D Face Analysis, *Workshop on Biometrics and Identity Management (BIOID)*, Denmark, May 2008
- [12] P. Szeptycki, M. Ardabilian, L. Chen. A coarse-to-fine curvature analysis-based rotation invariant 3D face landmarking, *International Conference on Biometrics: Theory, Applications and Systems*, Washington, 2009
- [13] J.J. Koenderink, A.J. van Doorn, Surface shape and curvature scales, *Image and vision computing*, 1992
- [14] W. Ben Soltana, M. Ardabilian, L. Chen, Comparison of 2D/3D Features and their adaptive Score Level Fusion for 3D Face Recognition, *3DPVT 2010* (soumis, accepté)
- [15] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the Face Recognition Grand Challenge, *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 947-954, 2005
- [16] I. Mpipieris, S. Malassiotis, M.G. Strintzis, Expression Compensation for Face Recognition Using a Polar Geodesic Representation, *3D Data Processing, Visualization, and Transmission (3DPVT'06)*, 2006