

Une approche multimodale pour la reconnaissance du visage, inspirée d'une méthodologie anthropométrique

Wael Ben Soltana, Mohsen Ardabilian, Liming Chen et Chokri Ben Amar

Ecole Centrale de Lyon, LIRIS, UMR5205, F-69134, France
{wael.ben-soltana, Mohsen.Ardabilian, Liming.Chen}@ec-lyon.fr,
chokri.benamar@ieee.org

Résumé Cet article présente une nouvelle approche multimodale et hybride de reconnaissance de visage inspirée par une méthodologie anthropométrique. De telles approches sont utilisées principalement par les experts médico-légaux pour identifier l'auteur d'un délit parmi un ensemble de suspects. L'expert possède une vidéo enregistrée au moment du délit ainsi que des clichés des suspects. Une signature invariante à la pose et aux expressions faciales est calculée à partir des points caractéristiques anthropométriques et constitue un descripteur local. Il est combiné avec les informations de couleur et de profondeur qui constituent deux descripteurs globaux. Ces trois descripteurs sont séparément redimensionnés par l'Analyse en Composantes Principales (couleur, profondeur et mesures anthropométriques) pour fournir trois scores qui sont normalisés puis combinés selon une procédure de fusion optimisée pendant la phase d'apprentissage. Les résultats des expérimentations menées sur 116 individus issus de la base FRGC1.0 montrent que les mesures anthropométriques et les informations couleur et profondeur sont complémentaires ; le taux de reconnaissance atteint est de 96.12%.

Mots clés mesures anthropométriques, approche multimodale et hybride, reconnaissance faciale 2D/3D, fusion des scores.

1 Introduction

Nous assistons aujourd'hui à l'émergence de la reconnaissance faciale 3D. Les résultats publiés, à l'issue des campagnes d'évaluation comme FRVT (Facial Recognition Vendor Test 2000/2002), ont montré que les approches de reconnaissance faciales souffrent de plusieurs problèmes liés aux changements de pose, d'éclairage et à l'occultation, etc. La modalité 3D a alors été proposée pour palier les problèmes liés à la pose et à l'éclairage. Néanmoins, la prise en compte de cette nouvelle modalité nécessite des algorithmes de traitement adaptés et éventuellement de fusion 3D et 2D à une étape de reconnaissance.

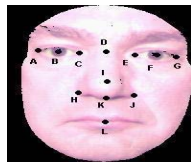
Les approches de reconnaissance faciales peuvent être regroupées en trois catégories ; La reconnaissance 2D utilise les images textures [13], la reconnaissance 3D exploite l'information géométrique, les approches multimodales est une combinaison des deux dernières catégories [2]. Bowyer et al. [5] ont proposé un état de l'art de ces approches. Bien avant l'émergence de la reconnaissance faciale, les anthropologues ont présenté les caractéristiques métriques et descriptives du visage humain. L'approche anthropométrique a permis de traiter avec succès des cas d'identification des individus, d'identification de race,

aussi bien que le diagnostic de paternité [7]. Ces techniques sont toujours valables et utilisées par les experts médico-légaux.

L'approche présentée dans cet article s'inspire d'une méthodologie anthropométrique mise au point par Perrot [9] et propose une technique bimodale de reconnaissance basée sur l'Analyse en Composantes Principales (ACP). Les différents tests réalisés sur la base FRGC1.0 ont permis d'évaluer cette approche. Pour exposer notre méthode, nous avons adopté dans cet article, l'organisation suivante. La section 2 présente les mesures anthropométriques utilisées. La section 3 donne une vue d'ensemble de notre approche. La section 4 introduit les résultats obtenus. Finalement, la section 5 présente nos conclusions.

2 Les mesures anthropométriques

Contrairement à certaines approches automatiques de reconnaissance faciale qui utilise la totalité du visage, la méthodologie anthropométrique préconise souvent l'extraction d'une signature à partir d'un certain nombre de points caractéristiques anthropométriques jugés les plus pertinents ; Ces points doivent être discriminants et stables.



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

Fig 1. Les points caractéristiques du Visage 3D

Les experts des deux domaines arrivent souvent aux mêmes conclusions pertinentes; L'anthropologue explique de surcroît les conclusions. Mian et al. [1] montrent que la partie supérieure du visage est plus statique. Faltemier et al. [11,12] prouvent que la partie du visage qui entoure le Nez est très stable. Arca et al. [10] utilisent des segments 2D du visage autour des yeux et du nez pour en extraire les profils 3D. En se basant sur ses travaux, mais aussi sur le travail de Perrot [9] sur les mesures anthropométriques du visage, nous avons sélectionné un ensemble de points caractéristiques illustrés par la figure 1. 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 3D caractéristiques 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 utilisé est l'indice nasal (1) qui est le rapport entre la largeur horizontale du nez et sa hauteur verticale.

$$Indice\ Nasal = 100 * \left(\frac{dist(H, J)}{dist(D, K)} \right), \quad (1)$$

avec $dist(H, J)$, $dist(D, K)$: distance euclidienne entre les points : H, J et D, K. Or les informations ainsi générées sont corrélés. Pour les décorrélérer et supprimer la redondance, nous utilisons la technique ACP.

3 Vue d'ensemble de l'approche proposée

La figure 2 présente une vue d'ensemble de l'approche proposée. Cette approche opère en

deux phases : une phase hors-ligne et une phase en ligne.

3.1 Phase hors-ligne

Cette phase est une étape d'apprentissage durant la quelle les trois descripteurs, couleur, profondeur et signatures géométriques locales (les mesures anthropométriques) propres à chaque visage sont extraits et organisées sous forme de trois vecteurs caractéristiques. Ensuite, l'ACP est appliquée pour réduire la dimension de ces vecteurs caractéristiques tout en conservant la quasi-totalité de leur énergie, permettant de générer trois espaces de projection. En outre, l'ACP est utilisé comme un classifieur pour chaque descripteur. D'une part les performances de chaque classifieur en termes de taux d'erreur (taux de Faux Rejet, taux de Fausse Acceptation) sont calculées sur l'ensemble de la base d'apprentissage. D'autre part, les distributions (moyenne et variance) des taux de Bonne Acceptation (BA) et de Fausse Acceptation (FA) sont estimées pour chaque identité.

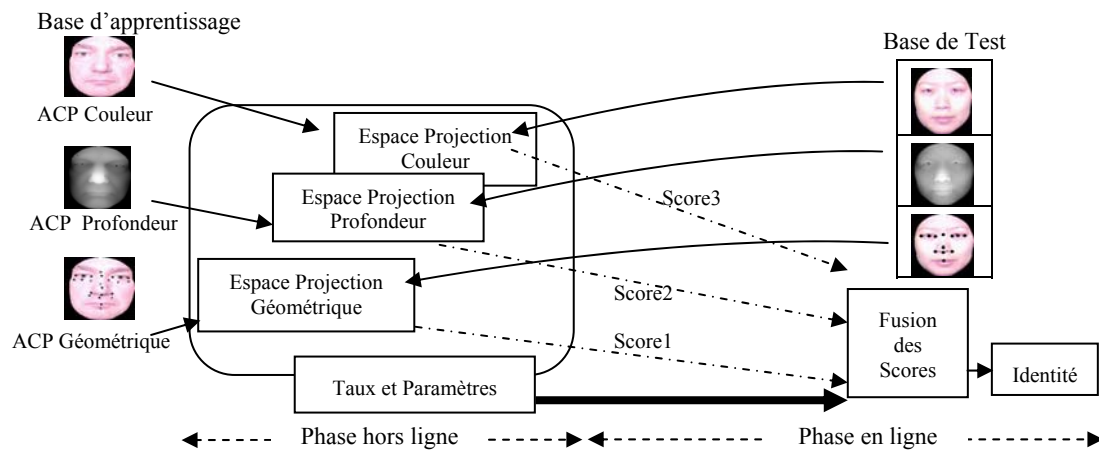


Figure 2 . Vue d'ensemble de la méthode proposée

3.2 Phase en ligne

La deuxième phase permet la génération d'une identité pour le visage requête. Les trois descripteurs sont calculés et projetés sur les espaces respectifs. Trois scores de similarités sont ensuite calculés entre ces projetés et ceux générés pendant la phase hors-ligne pour chaque identité de la base. Puis, ces scores sont normalisés et fusionnés pour générer les scores finals. L'identité la plus proche du visage requête est celle qui possède le score final le plus élevé.

3.2.1 Normalisation des scores

La normalisation des scores est une étape importante avant la procédure de la fusion des différents scores générés par les différents classifieurs. Trois méthodes de normalisation sont évaluées pour ne retenir au final que la plus efficace. Ces méthodes sont : Min-Max (2), Z-score (3), Tanh (4). La normalisation peut être formalisée de la manière suivante. Soit s , le vecteur des scores et n le vecteur des scores normalisés.

- Le Min-Max rend le vecteur des scores appartenant à $[0,1]$, tout en se basant sur la valeur maximale et minimale de s :
$$n = \frac{s - \min(s)}{\max(s) - \min(s)}. \quad (2)$$

- Le Z-score transforme les valeurs de s avec une distribution de moyenne nulle et un écart type égale à 1 :
$$n = \frac{s - \text{moyenne}(s)}{\text{ecart}(s)}. \quad (3)$$

- Le Tanh rend le vecteur des scores appartenant à $[0,1]$:
$$n = \frac{1}{2} [\tanh(0.01 * \frac{s - \text{moyenne}(s)}{\text{ecart}(s)}) + 1]. \quad (4)$$

3.2.2 Fusion des scores

Plusieurs études récentes montrent que la fusion de deux modalités 2D/3D améliore significativement le taux de reconnaissance ou d'authentification [3, 4, 5]. Quatre techniques de fusion sont expérimentées: somme, produit, technique de classifieur pondérée et technique d'utilisateur pondéré. Les deux dernières techniques sont introduites par Snelick [8]. La technique de classifieur pondérée permet de calculer le score de fusion (fs_i) pour chaque personne i en se basant sur le point d'équivalence des erreurs (r^m) de chaque classifieur :

$$fs_i = \sum_{m=1}^M w^m n^m_i, \text{ où } w^m = \frac{1}{r^m}, u = \sum_{m=1}^M \frac{1}{r^m}, \sum_{m=1}^M w^m = 1 \text{ et } 0 \leq w^m \leq 1. \quad (5)$$

où w^m représente le poids du classifieur m ; M le nombre de classifieurs. En effet, on associe à chaque classifieur un poids qui est inversement proportionnel à son taux d'erreur calculés pendant la phase hors-ligne.

La technique d'utilisateur pondéré permet de calculer le score de fusion (fs'_i) en se basant sur l'efficacité de la classification de chaque classifieur pour chaque identité, déterminé pendant la phase hors-ligne :

$$fs'_i = \sum_{m=1}^M w_i^m n^m_i, \quad (6)$$

$$w_i^m = \frac{1}{\sum_{m=1}^M d_i^m} d_i^m, d_i^m = \frac{\mu_i^m(VR) - \mu_i^m(FR)}{\sqrt{(\sigma_i^m(VR))^2 + (\sigma_i^m(FR))^2}}, \sum_{m=1}^M w_i^m = 1, \text{ et } 0 \leq w_i^m \leq 1.$$

où d_i^m représente la mesure de séparation entre les deux distributions Bonne Acceptation (BA) et Fausse Acceptation (FA) pour un classifieur m et une identité i . μ et σ sont respectivement la moyenne et l'écart type.

4 Résultats Expérimentaux

L'approche proposée est évaluée dans différentes conditions d'éclairage et en présence d'expressions faciales. Nous avons utilisé 116 personnes de la base FRGC1 [6]. Cette base inclut, pour chaque identité, 4 images 2D couleur et 4 images 3D. Pour évaluer l'ensemble de l'approche, nous avons utilisé la technique de cross-validation qui consiste à retenir selon chaque modalité 2D/3D 1 image sur 4 pour le test et 3 images pour l'apprentissage [11].

Toutes les combinaisons, image de test et images d'apprentissage sont prises en compte selon cette technique. Les points caractéristiques anthropométriques sont marqués manuellement.

4.1 Classification Individuel

Nous avons trois classifieurs autonomes. Chaque classifieur est caractérisé par une dimensionnalité réduite propre (Voir tableau 1). Pour le choix de mesure de similarité, nous avons utilisé les deux distances L1 (distance de Manhattan) et L2 (distance Euclidienne).

Tableau 1. Application de l'ACP sur les descripteurs

	Dimension	PCA	L1	L2
Intensité	90*90=8100	249	86.64	82.76
Profondeur	142*120=17040	348	50.43	33.62
Caractéristiques Géométriques	117	24	72.41	71.55

D'après le tableau 1, la distance L1 fournit les meilleurs résultats de reconnaissance. De ce fait, nous adoptons cette distance pour la procédure de la fusion. De plus, le descripteur formé par les caractéristiques géométriques nous permet de générer un taux de reconnaissance (72.41%) proche du taux de reconnaissance généré par le descripteur couleur (86.64%). Ceci s'explique par le fait que les mesures anthropométriques possèdent une invariance à la luminosité, à l'expression et à la pose.

4.2 Fusion des scores des classifieurs

Pour améliorer le taux d'identification, il est souvent nécessaire d'intégrer les résultats obtenus des différentes caractéristiques faciales.

Tableau 2. Taux de reconnaissance fourni par les méthodes de fusion via différents techniques de normalisation

	Somme	Produit	Classifieur pondéré	Utilisateur pondéré
Min-Max	87.7125	88.3625	92.14	89.5525
Z-score	95.0425	66.8125	96.12	93.75
Tanh	94.61	95.26	95.905	93.75

Le tableau 2 montre l'influence de chaque méthode de fusion sur les performances de l'approche proposée avec les différentes méthodes de normalisation, Min-Max, Z-score et Tanh. D'après ce tableau, la méthode Z-score fournit les meilleurs résultats sauf pour la méthode produit. Le taux d'identification passe de 86.64% (meilleur classifieur individuel : Intensité) à 96.12%. Ceci s'explique par la complémentarité des caractéristiques anthropométriques et des informations texturales/géométriques.

5 Conclusion et perspectives

Dans cet article, nous avons proposé une approche inspirée d'une méthodologie anthropométrique. Nous l'avons combiné avec la modalité texture 2D pour proposer une approche multimodale. Une étape d'ACP est appliquée d'une part pour redimensionner et décorréliser les données et d'autre part, comme un classifieur. Les scores de similarités obtenus séparément pour chaque modalité sont normalisés et fusionnés suivant des règles établies

pendant une phase d'apprentissage. L'ensemble de l'approche a été évalué sur 116 individus de la base FRGC1.0. Le meilleur résultat est obtenu par le classifieur pondéré, avec un taux de reconnaissance de 96.12 %.

Nous envisageons de nous appuyer sur le résultat de ces travaux pour affiner l'approche anthropométrique qui présente l'avantage d'être invariante à l'orientation et aux expressions. Cette approche nécessite l'extraction d'une signature à partir de points caractéristiques discriminants. Un modèle optimale d'extraction de signature permet de ne calculer que les informations décorréliées et indépendantes. En suite, il est nécessaire d'étudier l'efficacité des signatures partielles obtenues à partir des sous ensembles de points caractéristiques car dans un contexte réel d'identification l'information 2D ou 3D est souvent partielle. Finalement, des expérimentations effectuées, en mode dégradé, sur une base dédiée comme FRGC2.0 permettrait une meilleure évaluation.

Remerciements

Ce travail a bénéficié d'une aide de l'ANR (Agence Nationale de la Recherche) à travers le projet FAR3D (ANR-07-SESU-004).

Références

- [1] A. Mian, M. Bennamoun, and R. Owens. 2D and 3D Multimodal Hybrid Face Recognition. *Computer Vision – ECCV 2006*. vol 3953, pp 344-355, July 26, 2006.
- [2] B. Gokberk, H. Dutagaci, A. Ulas, L. Akarun, B. Sankur, Representation Plurality and Fusion for 3-D Face Recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 38, pp 155-173, 2008.
- [3] G. P. Kusuma and C. S. Chua. Image Level Fusion Method for Multimodal 2D + 3D Face Recognition. *ICIAR*, pp 984-992, 2008.
- [4] K. I. Chang, K. W. Bowyer, and P. J. Flynn. An Evaluation of multimodal 2D+3D Face Biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 27(4), pp 619-624, 2005.
- [5] 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*, vol.101, pp 1-15, Jan 2006.
- [6] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. *CVPR 05*, pp 947-954, 2005.
- [7] P. Sinha. A symmetry perceiving adaptive neural network and facial image recognition. *Forensic science international*, pp 67-89, 1998.
- [8] R. Snelick and U. Uludag and A. Mink and M. Indovina and A. Jain. Large Scale Evaluation of Multimodal Biometric Authentication Using State-of-the-Art Systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.27, pp 450-455, 2005.
- [9] 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.* pp 161-164, 1997.
- [10] S. Arca, R. Lanzarotti, and G. Lipori, Face Recognition Based on 2D and 3D Features. *Knowledge-Based Intelligent Information and Engineering Systems*, pp 455-462, 2007.
- [11] 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* 112(2), pages 114-125, Nov. 2008.
- [12] T. C. Faltemier, K. W. Bowyer and P. J. Flynn. 3D Face Recognition with Region Committee Voting. *Third International Symposium on 3D Data Processing, Visualization, and Transmission*, pp 318-325, June 2006.
- [13] W. Zhao, R. Chellappa, P. J. Phillips, A. Rosenfeld. Face Recognition: A Literature Survey. *ACM Computing Surveys*, vol. 35, no. 4, Dec. 2003.