

Vers une Reconnaissance Faciale 3D Rapide Utilisant les Réseaux d'Ondelettes

Salwa SAID^{1,2}, Boulbaba BEN AMOR², Mourad ZAIED¹, Chokri BEN AMAR¹ et Mohamed DAOUDI²

¹ Ecole Nationale d'Ingénieurs de Sfax,
REsearch Group on Intelligent Machines (DGRST 01/UR/11-02),
Université de Sfax, Sfax, Tunisie.
salwa.said@telecom-lille1.eu

² Institut TELECOM ; Telecom Lille 1,
Laboratoire d'Informatique Fondamentale de Lille (UMR USTL/CNRS 8022),
Lille, France.

Résumé L'émergence, depuis quelques années, de la forme 3D dans la reconnaissance faciale est due à sa robustesse aux changements de pose et d'illumination. Ces avantages séduisants, ne relèvent pas tous les défis pour obtenir des taux de reconnaissance satisfaisants. D'autres verrous comme les expressions faciales et le temps d'exécution des algorithmes de *matching* restent à relever. C'est dans ce contexte que s'inscrivent les travaux décrits dans ce papier. Nous proposons, ainsi, une nouvelle méthode de reconnaissance faciale 3D basée sur les réseaux d'ondelettes. Une image de profondeur est d'abord prétraitée afin d'extraire la partie utile du visage. Une signature biométrique compacte et représentative est, par la suite, extraite par le biais des réseaux d'ondelettes. Le *matching* de deux visages est, en fin, effectué par le calcul de distances entre leurs deux signatures correspondantes.

Mots clés Reconnaissance faciale 3D, réseaux d'ondelettes, expressions faciales.

1 Introduction et motivations

Afin de répondre aux besoins liés à la sécurité, devenue aujourd'hui une préoccupation internationale, la biométrie se présente comme une technologie potentiellement puissante. En effet, les différents moyens biométriques visent à utiliser des caractéristiques comportementales et/ou physiologiques spécifiques à chaque personne. Ces caractéristiques présentent l'avantage d'être universelles, uniques, permanentes et qu'elles ne peuvent être falsifiées à la différence des moyens classiques tels que les mots de passe ou les badges. Les empreintes digitales et l'iris ont montré de bonnes performances. Cependant, ils exigent une coopération des utilisateurs qui les trouvent intrusifs. Par conséquent, la tendance actuelle se dirige vers les biométries qu'on peut collecter à distance, comme le visage. Bien que les approches 2D (utilisant des images couleurs) aient donné de bonnes performances, elles restent très sensibles aux problèmes de changements de conditions d'éclairage et de pose. Avec l'essor du 3D et des techniques d'acquisition de formes

tridimensionnelles, une nouvelle modalité faciale a émergé depuis peu et attire de plus en plus d'attention du fait qu'elle n'est pas influencée par une variation d'illumination et de pose [2]. Elle permet également de déceler des descripteurs de forme importants, essentiellement de courbure et de segmentation de la forme 3D en régions homogènes. Les premières approches proposées sont basées sur le recalage rigide de surfaces [2, 6, 13] en utilisant différentes variantes de l'algorithme ICP (*Iterative Closest Point*) [9]. Les performances de ces approches chutent avec l'amplitude de la déformation faciale causée par les expressions. D'autres approches qui prennent en compte le caractère déformable du visage ont vu alors le jour :

Faltemier et al. propose, dans [12], une approche basée sur le recalage de plusieurs régions de la surface faciale (38 régions) puis d'utiliser des techniques de fusion afin d'obtenir de meilleurs taux de reconnaissance. Leur approche privilégie les régions statiques du visage, celles qui donnent les taux de reconnaissance les plus élevés sur un ensemble d'apprentissage.

Bronstein et al. [1] introduisent, pour la première fois, une hypothèse entièrement empirique qui est la robustesse de la distance géodésique (plus court chemin entre deux points d'une surface tout en restant sur cette surface) par rapport aux changements d'expressions faciales. A l'aide de cette mesure (ou fonction) ils transforment, d'abord, la surface faciale en une forme canonique de \mathbb{R}^3 . Puis, ils appliquent un recalage rigide sur ses nouvelles formes. Ils affirment distinguer, avec leur algorithme, entre des classes de jumeaux en présence d'expressions faciale.

Samir et al. [3] proposent de représenter la surface faciale par une collection de courbes de niveau en utilisant un point de référence (le pic du nez) et une fonction f . L'objectif est de ramener le problème de *matching* de surfaces à un problème de *matching* de courbes faciales plus simple à résoudre. Les auteurs proposent alors une approche *Riemannienne* qui permet de quantifier le chemin géodésique entre deux surfaces faciales dans l'espace des visages.

L'introduction de la forme 3D du visage dans la reconnaissance de personnes est considérée comme une voie très prometteuse. Cependant, les défis ouverts aujourd'hui sont : (i) la robustesse de ces approches 3D aux changements d'expressions faciales et (ii) l'optimisation du temps de calcul nécessaire dans des scénarios d'identification. Nous proposons dans cet article des éléments de réponses à ces deux verrous dans les approches de l'état de l'art. Nous proposons, d'une part, de produire des signatures biométriques compactes à partir des formes 3D des visages en utilisant le codage par les réseaux d'ondelettes. Celles-ci, à la différence de maillages 3D, sont plus rapides à matcher. D'autre part, nous décrivons une nouvelle métrique qui prendra en compte la disposition spatiale dans le visage de régions statiques et non-statiques pour remédier au problème de changement d'expressions faciales.

Le reste de cet article est organisé de la façon suivante : la section 2 décrit une vue d'ensemble de l'approche proposée. Dans la section 3, nous présentons notre solution automatique de prétraitement des données 3D. L'application des réseaux d'ondelettes pour le codage et la reconnaissance faciale 3D est expliquée dans la section 4. Dans la section 5, nous présentons nos résultats sur un ensemble de données de la base FRGC v2 et nous terminons par les conclusions et les perspectives.

2 Vue d'ensemble de l'approche proposée

Nous présentons dans cet article une solution complète de reconnaissance et d'authentification de personnes basées sur les réseaux d'ondelettes. La solution que nous présentons procède, d'abord, par le prétraitement des visages 3D scannés, pour ne garder que la partie utile du visage, boucher les trous et éliminer le bruit présents sous forme de pics sur les nuages de points 3D. Cette étape est commune aux visages qui seront incorporés dans la base de données appelée *Gallery* et les visages de test appartenant à l'ensemble dit *Probe*. Dans une seconde étape, les visages de la *Gallery*

subissent une approximation par réseaux d'ondelettes pour produire une signature biométrique compacte. C'est cette signature, constituée des ondelettes et de leurs coefficients, qui sera utilisée pour matcher un visage de test avec les visages de la *Gallery*.

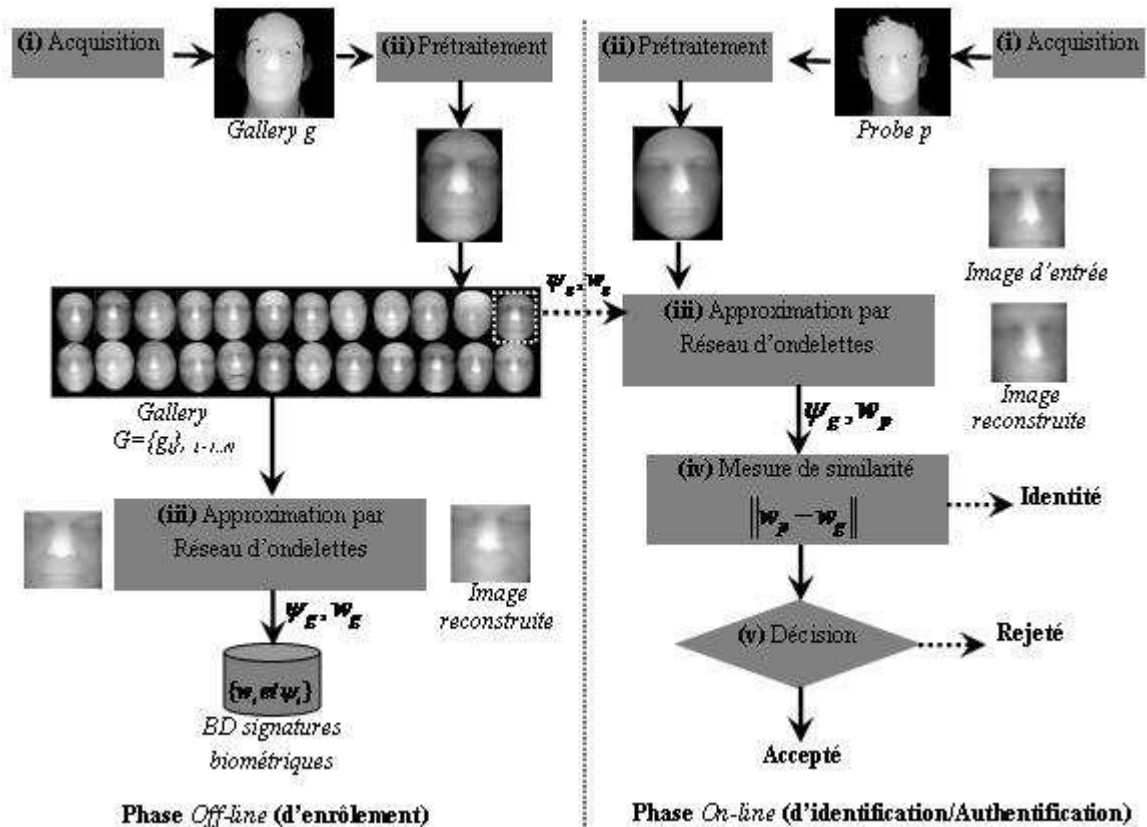


Figure 1 : Framework complet de reconnaissance faciale 3D par réseaux d'ondelettes.

Enfin, au moment de la reconnaissance, le visage de test est projeté sur le réseau d'ondelettes du visage de la *Gallery* et de nouveaux poids (ou coefficients), spécifiques à ce visage, sont produits. La famille d'ondelettes reste alors inchangée (celle du visage de la *Gallery*). Un score de similarité est alors produit en comparant les coefficients du visage *Gallery* avec les coefficients du visage de test (ou *Probe*). Le pipeline de toutes ces étapes est illustré dans la figure 1.

3 Prétraitement des visages 3D

Les données issues de scanners laser 3D sont généralement bruitées (bruit sous forme de pics). Elles comportent, également, des trous dus aux occlusions ou à l'absorption du plan laser par certaines parties du visage comme l'iris ou les sourcils. D'autres parties apparaissent sur les images de profondeur produites comme les cheveux et les vêtements nécessitent d'être supprimées pour ne garder que la partie informative ou utile du visage. Pour toutes ces raisons, nous avons mis en place une procédure simple, efficace et entièrement automatisée de prétraitement d'un visage 3D donné en entrée. Elle consiste, dans un premier temps, à lire et à filtrer avec un filtre médian l'image de profondeur acquise par le biais du scanner afin d'éliminer le bruit sous forme de pics. Dans une seconde étape, elle bouche les trous par l'introduction d'un algorithme d'interpolation

bidimensionnelle. Ensuite, la partie utile du visage est localisée à l'intérieur d'une sphère ayant comme centre le pic du nez et un rayon égal à $100mm$. Enfin, la partie extraite est lissée pour enlever les granularités présentes sur la surface faciale. On obtient alors un masque lisse de la surface faciale prêt à être utilisé dans la reconnaissance. La figure 2 illustre l'algorithme pour produire ce masque et l'image de profondeur du visage.

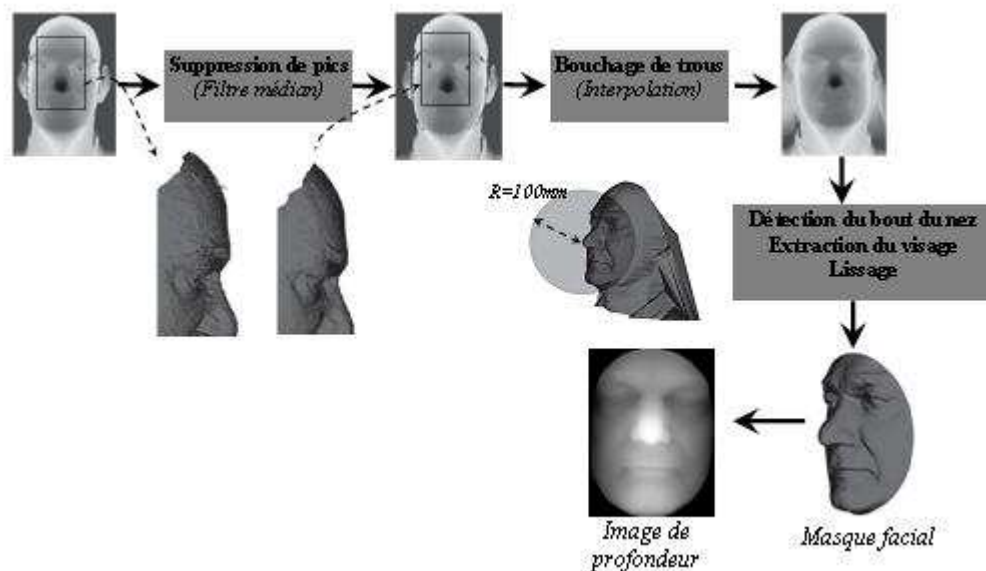


Figure 2 : Procédure de prétraitement de visages 3D.

La détection du bout du nez nécessite une première coupe sagittale au niveau du centre de gravité qui donne un profil horizontal. Le maximum de ce profil sert comme point de référence d'une seconde coupe verticale qui donne naissance à un profil horizontal. Le maximum de ce profil est interprété comme le bout du nez. Notre procédure de prétraitement a été testée sur l'ensemble des 4007 visages de la base FRGC v2, elle a réussi à prétraiter **99.1%** de ces visages et donc à produire des masques faciaux prêts à utiliser pour la reconnaissance.

4 Réseaux d'ondelettes pour le codage et la reconnaissance de visages

La seconde étape dans notre *Framework* de reconnaissance est le codage des visages 3D par réseaux d'ondelettes. L'objectif est d'obtenir une signature biométrique compacte du visage qui permet de retrouver l'identité de son porteur dans une base de visages collectée à l'avance. Il s'agit donc d'extraire en *off-line* les signatures des visages de la base de données (*Gallery*). Dans la phase *on-line* d'identification ou d'authentification, une signature biométrique est extraite et est comparée aux signatures de la base de données. Nous présentons dans cette section, le principe théorique du codage de visages par réseaux d'ondelettes. En suite, nous présentons son application pour le codage des visages de la *Gallery* et, enfin, sur les visages de test ou de l'ensemble *Probe*.

4.1 Généralités sur le codage d'images par réseaux d'ondelettes

Comme décrit dans [11], toute fonction $f \in IL^2(\mathbb{R}^2)$ peut être représentée par un réseau d'ondelettes. Nous interprétons une image de profondeur $f = I(x, y)$ comme une fonction de

$L^2(\mathbb{R}^2)$, par analogie à ce qui a été fait dans [4] pour les images couleurs. Pour trouver le réseau d'ondelettes codant la fonction f , nous minimisons la fonction énergie suivante :

$$E = \min_{n_i, w_i \forall i} \left\| f - \sum_{i=1}^N w_i \psi_{n_i} \right\|_2^2 \quad (1)$$

Avec $\Psi = \{\psi_{n_1}, \dots, \psi_{n_N}\}$ est une famille d'ondelettes, $n_i = (t_x, t_y, \theta, d_x, d_y)^T$ les paramètres des ondelettes (i.e. translation, orientation, et dilatation) et $W = \{w_1, \dots, w_N\}$ le vecteur de poids. Toutes les ondelettes ψ_{n_i} de la famille Ψ découlent de la même ondelette mère ψ en faisant varier les paramètres n_i . L'ondelette mère utilisée dans notre travail est la dérivée première de la fonction Beta β [5, 8], donnée par l'équation (2) :

$$\beta(x) = \begin{cases} \left(\frac{x-x_0}{x_c-x_0} \right)^p \left(\frac{x_1-x}{x_1-x_c} \right)^q & \text{si } x \in [x_0, x_1] \\ 0 & \text{sinon} \end{cases}, \text{ avec } p, q, x_0 < x_1 \in \mathbb{R} \text{ et } x_c = \frac{px_1 + qx_0}{p+q} \quad (2)$$

L'équation (1) montre que les W_i et n_i sont optimisés, c-à-d les translations, les dilatations et la rotation de chaque ondelette sont choisis de façon que l'image est approximée de façon optimale par la somme pondérée des ondelettes ψ_{n_i} . En utilisant le réseau d'ondelettes obtenu (Ψ, W) d'une image f , f peut être reconstruite par une combinaison linéaire des ondelettes du réseau multipliés par les poids correspondants selon (3) :

$$\tilde{f} = \sum_{i=1}^N w_i \psi_{n_i} \quad (3)$$

4.2 Codage *Off-line* des visages 3D de la *Gallery*

Le codage des images de profondeur de la *Gallery* donnera naissance à un ensemble de réseaux d'ondelettes $(\psi_g, w_g)_{1 \leq g \leq m}$. L'algorithme d'apprentissage utilisé pour l'approximation par réseaux d'ondelettes bêta se base sur : l'échantillonnage dyadique afin de construire la bibliothèque des ondelettes candidates à utiliser dans le réseau et la théorie des frames pour le calcul des poids [7]. La figure 3 présente quelques exemples d'une image de la *Gallery* et les images reconstruites après codage par réseaux d'ondelettes Bêta, en faisant varier le nombre d'ondelettes. Il est évident que la qualité de l'image reconstruite croît avec le nombre d'ondelettes utilisées, comme l'illustre la figure 3.

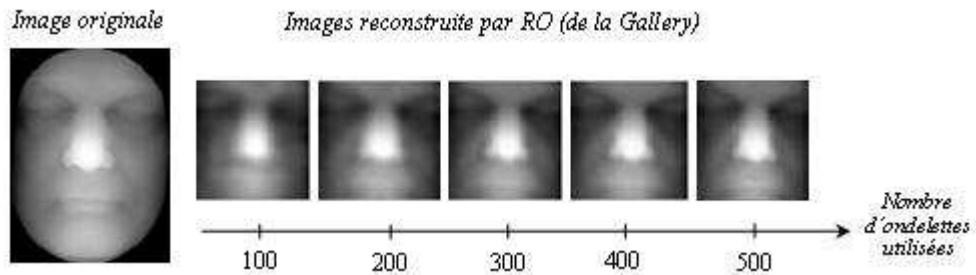


Figure 3 : Exemple d'un visage 3D de la *Gallery* et d'images reconstruites par réseaux d'ondelettes.

4.3 Matching on-line des Probes et reconnaissance

Quand un sujet (*Probe*) se présente pour la reconnaissance ou l'authentification, le codage de son image de profondeur se fera différemment de celui d'une image de la *Gallery*. En effet, sa fonction f est projetée sur le réseau d'ondelettes du visage de la *Gallery* avec lequel se fait le *matching*. Il en résulte de cette projection un nouveau réseau $(\psi_p, w_p)_{1 \leq p \leq m}$ pour lequel $\psi_p = \psi_g$. L'image probe est alors codée en utilisant les mêmes ondelettes de l'image *Gallery*.

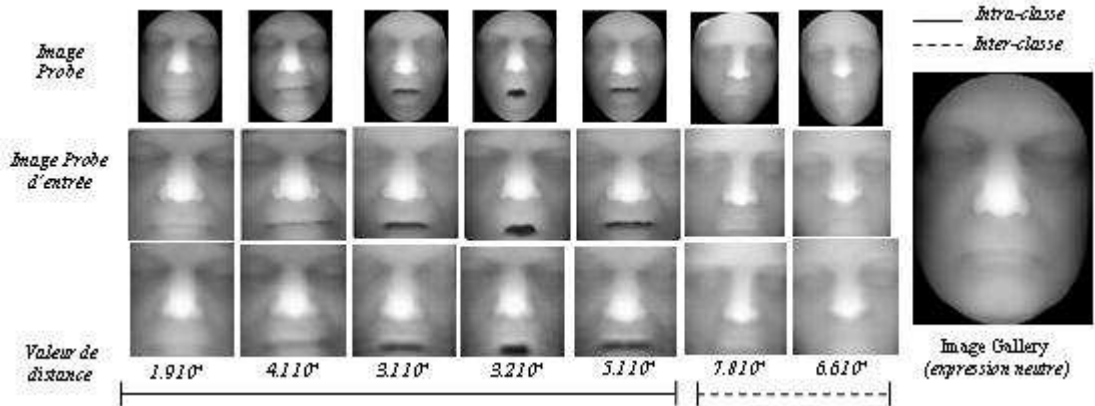


Figure 4 : Exemples de *matching* intra et inter-classes.

La figure 4, montre quelques exemples de projections, sur des exemples appartenant à même sujet et des exemples appartenant à des sujets différents. Nous pouvons constater, sur cette figure, que les reconstructions sont de bonne qualité quand il s'agit de *matching* intra-classes, en comparaison avec les reconstructions à partir de projection inter-classes. En outre, la distance produite entre deux visages en comparant uniquement les vecteurs poids w_p et w_g par une simple distance euclidienne $\|w_p - w_g\|$ confirment ces résultats. En effet, quand les visages sont différents, la distance est, en général, plus importante que dans le cas de deux visages appartenant à la même classe. Cependant, le problème réside dans les *matching* intra-classes en présence d'expressions faciales comme le montre la figure 4. Il est à noter que le temps moyen d'un *matching* est autour de 20ms ce qui permet d'envisager un système automatique de reconnaissance faciale temps réel.

5 Résultats préliminaires et perspectives

Pour évaluer les performances de notre approche, dans le contexte de la biométrie faciale, en présence d'expressions faciales, nous avons pris un échantillon de données de la base FRGC v2. Nous avons défini le protocole suivant : l'ensemble *Gallery* est composé de 122 sessions neutres de 122 sujets différents avec des expressions neutres en revanche l'ensemble de *Probe* est composé de 122 sessions non-neutres des mêmes personnes prises dans la *Gallery*. Ce qui forme une matrice de similarité de taille 122x122 comparaisons dont 122 correspondent à des comparaisons intra-classe (la même personne), situées sur la diagonale de la matrice et le reste des comparaisons correspondent à des comparaisons inter-classes (personnes différentes) ou des accès imposteurs. La figure 5 montre un exemple des images *Gallery* et leurs images *Probes* correspondantes. Nous mentionnant les visages de l'ensemble *Probe* présentent des expressions mais avec les bouches fermés.

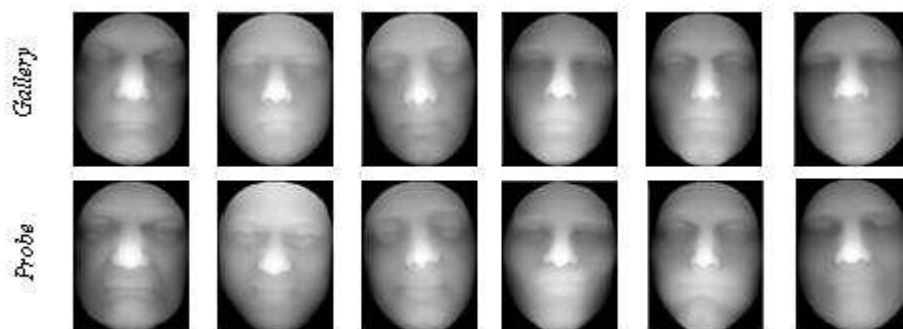


Figure 5 : Exemples de visages utilisés dans l'expérimentation : première ligne des visages de la *Gallery*, seconde ligne visages de l'ensemble *Probe*.

Nous visons dans nos expérimentations les deux scénarios connus en biométrie : l'authentification et d'identification. Nous présentons les résultats du premier scénario sous forme de courbe ROC (*Receiver Operator Characteristic*) produite à partir de la matrice de similarité. En revanche, nous donnons le résultat d'identification sous forme de taux de reconnaissance au premier rang (*Rank-one Recognition Rate*). Dans la courbe ROC présentée dans la figure 6 (a), notre approche affiche un taux de vérification autour de **81%** si l'on accepte **0.1%** (1/1000) des accès imposteur. Elle présente un taux d'égale erreur (*Equal Error Rate*) autour de **5%**. En ce qui concerne le taux d'identification il est autour de **95%**. En effet, sur les 122 images de l'ensemble *Probe*, notre approche ne reconnaît pas 5 personnes, présentées dans la figure 6. Il est clair que la présence de cheveux sur l'image de test gêne le processus de reconnaissance (voir (b), (c) et (d) de la figure 6). Il est clair aussi que les fortes expressions ont une influence sur le taux de reconnaissance (voir figure 4). Quelques erreurs sont dues aussi à l'acquisition (mouvement du visage au moment du passage du laser sur le visage, exemple (e) de la figure 6).

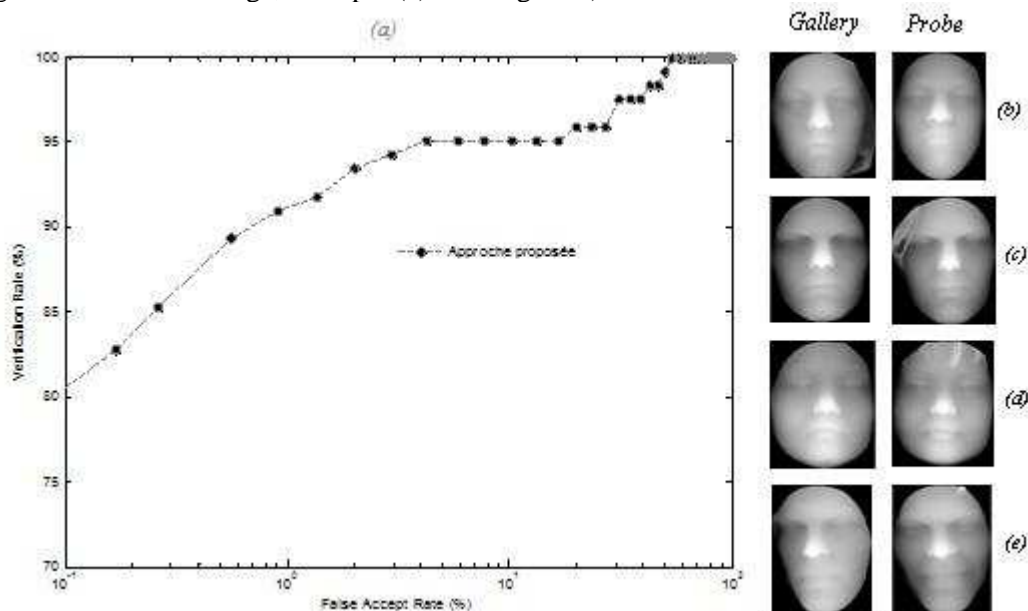


Figure 6 : (a) : courbe ROC obtenue pour l'approche proposée, (b), (c), (d) et (e) : visages non reconnus par notre approche.

Les premiers résultats affichés par notre approche sont prometteurs, en revanche nous devons étendre notre corpus de test. Nous planifions dans nos travaux futurs de faire les tests sur tout le corpus FRGC v2 en suivant le protocole d'évaluation proposé dans [10] afin de comparer notre approche aux approches de l'état de l'art. Nous proposerons également une métrique de comparaison de visages plus approprié et plus adapté pour tenir compte des fortes expressions faciales et de la présence de cheveux sur le visage à reconnaître.

6 Conclusions

Nous avons décrit dans cet article une nouvelle approche de reconnaissance de visage 3D par réseaux d'ondelettes. L'idée de base est d'approximer les images de profondeur prétraitées par des signatures biométriques plus compactes et de faire le *matching* entre les images de la base et les images de test par le calcul de la distance entre leurs signatures correspondantes. Pour valider notre approche en présence d'expressions faciales un ensemble de données extrait de la base FRGC v2 a été utilisé. Notre constat est que notre approche présente un bon taux de reconnaissance en présence de petites expressions faciale, en revanche, elle est plus sensible dans le cas de larges expressions. Nous inscrivons alors ce constat dans nos priorités immédiates. En effet, nous travaillons sur la mise en place d'une métrique de similarité plus adaptée pour tenir compte des déformations faciales dues aux expressions.

Remerciements

Ce travail de recherche est supporté par le projet franco-tunisien CMCU RAVI (Reconnaissance et Analyse de VIages) N° 08G/1406. Les auteurs tiennent à remercier *University of Notre Dame* aux USA qui nous a fourni les données de la base FRGC.

Références

- [1] A. Bronstein, M. Bronstein, R. Kimmel, «Three-dimensional face recognition», *International Journal of Computer Vision (IJCV)*, 64(1):5-30. August 2005.
- [2] B. Ben Amor, M. Ardabilian, L. Chen, «New Experiments on ICP-Based 3D Face Recognition and Authentication», *The 18th International Conference on Pattern Recognition*, 3 :1195-1199, ISSN 1051-4651. 2006.
- [3] C. Samir, A. Srivastava, M. Daoudi, «Three-Dimensional Face Recognition Using Shapes of Facial Curves», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1858-1863, ISSN 0162-8828. 2006.
- [4] G. S. Volker Kruger. «Gabor wavelet networks for object representation», *10th International Workshop on Theoretical Foundations of Computer Vision*, 2000.
- [5] H. Szu, B. Telfer, and S. Kadambe. «Neural network adaptive wavelets for signal representation and classification», *Optical Engineering*, 31:1907–1961, 1992.
- [6] K.I.Chang, K.W.Bowyer, P.J. Flynn, «Adaptive Rigid Multi-region Selection for Handling Expression Variation in 3D Face Recognition», in *the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 157. 2005.
- [7] M. Zaied, O. Jemai and C. Ben Amar, «Training of the Beta wavelet networks by the frames theory: Application to face recognition», *the international Workshops on Image Processing Theory, Tools and Applications*, Tunisia November 2008.
- [8] P. Belhumeur, J. Hespanha, and D. Kriegman « Eigenfaces vs. fisherfaces: Recognition using class specific linear projection», *IEEE T-PAMI*, 19(7), 1997.

- [9] P.J. Besl, N.D. McKay, « A Method for Registration of 3-D Shapes », *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(2):239-256, ISSN 0162-8828. 1992.
- [10] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, «Overview of the Face Recognition Grand Challenge», in *IEEE Conference on Computer Vision and Pattern Recognition*. June 2003.
- [11] Q. Zhang and A. Benviste. «Wavelet networks », *IEEE Trans. Neural Networks*, 3(6): 889-898, Nov. 1992.
- [12] T.C. Faltemier, K.W. Bowyer, P.J.Flynn, «A Region Ensemble for 3-D Face Recognition», *Information Forensics and Security, IEEE Transactions on*, Volume: 3, Issue: (1) On page(s): 62-73, 2008.
- [13] X Lu., A.K Jain., «Automatic Feature Extraction for Multiview 3D Face Recognition», in *FGR '06 : the 7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pp. 585-590. 2006.