

AUTHENTIFICATION D'INDIVIDUS PAR RECONNAISSANCE DE VISAGES

D. SAIGAA⁽¹⁾ N. BENOUDJIT⁽²⁾ K. BENMAHAMED⁽³⁾ S. LELANDAIS⁽⁴⁾

⁽¹⁾ Département d'Automatique Université Mohamed Khider Biskra (e-mail : d_saigaa@yahoo.fr) Algérie

⁽²⁾ Département d'électronique Université Elhadj Lakhdar Batna (e-mail : benoudjit_nabil@yahoo.fr) Algérie

⁽³⁾ Département d'électronique Université Ferhat Abbas Sétif Algérie

⁽⁴⁾ Laboratoire des systèmes complexes LSC université D'Evry France

RESUME

La vérification de visage est un outil important pour l'authentification d'un individu. Elle peut être de valeur significative dans la sécurité et les applications de commerce. Le but assigné à ce travail est de faire l'authentification d'individus. Pour cela, nous avons utilisé un modèle Biométrique. La biométrie est la science qui étudie les méthodes de vérification d'identité (authentification), identification, ou même de chiffrement basées sur la reconnaissance de caractéristiques physiologiques de l'individu. Pour être efficaces dans leur exploitation Ces caractéristiques doivent bien entendu posséder certaines qualités intrinsèques pour permettre le développement de systèmes fiables et robustes. Les qualités indispensables pour chaque caractéristique sont les suivantes : l'universalité, unicité, permanence, collectabilité et mesurabilité. Celles-ci assurent que chaque personne possède la caractéristique considérée, qu'elle est unique pour chaque individu, qu'elle ne change pas ou peu dans le temps, qu'il est possible d'en récolter un échantillon et de l'analyser. Pour notre application, nous avons choisi pour l'extraction des caractéristiques la méthode de ACP (analyse en composantes principales) [1]. Une fois que le vecteur caractéristique du visage est extrait, l'étape suivante consiste à le comparer avec le vecteur caractéristique de l'identité proclamée. Ici, il s'agit de classer l'utilisateur comme un vrai utilisateur ou un imposteur. Par la suite le taux d'erreur est calculé dans les deux ensembles, de validation et de test pour la base de données XM2VTS [2] selon le protocole de Lausanne [3].

1 INTRODUCTION

L'authentification de visage a gagné une attention considérable ces dernières années, par le besoin croissant de systèmes de vérification d'accès en utilisant plusieurs modalités (voix, image de visage, empreintes digitales, cartes à puce, etc.). De tels systèmes sont employés pour la vérification de l'identité d'un utilisateur en net, en utilisant un automate de banque, en entrant dans un lieu sécurisé, etc..

L'authentification de visage est différente de l'identification de visage (ou de la classification): dans des tâches d'authentification, le système sait a priori l'identité de l'utilisateur (par exemple par le code de sa carte à puce), et on doit vérifier cette identité; en d'autres termes, le système doit décider si l'utilisateur à priori est un imposteur ou pas. Dans l'identification de visage, l'identité a priori n'est pas connue : le système doit décider lesquelles des images stockées dans une base de données ressemblent plus à l'image à identifier; la décision n'est plus binaire.

Travailler avec des images approximatives n'est pas efficace: dans l'authentification de visage, plusieurs images d'une même personne peuvent être nettement différentes, en raison des changements de point de vue, en couleurs et l'illumination, ou tout simplement parce que le look du visage de ce même personne est différent de jour en jour. Pour cette raison, l'une des méthodes souvent employées afin d'extraire les caractéristiques appropriées dans l'authentification de visage est l'Analyse en Composante

Principale (ACP) [1][5].

Dans cet article, nous montrons comment la technique ACP extrait les caractéristiques discriminantes. Nous donnons aussi le taux d'erreur pour différentes techniques de mesure de similarité. Le contenu de cet article est organisé comme suit. La section 2 présente le problème de l'authentification de visage. La section 3 montre comment extraire les caractéristiques, à partir des images approximatives, et présente le procédé basé sur des expositions ACP. Dans la section 4 les résultats expérimentaux sont présentés. Enfin dans la section 5 nous donnons les conclusions et les perspectives.

2 AUTHENTIFICATION DE VISAGE

Les systèmes d'authentification de visage comparent typiquement un vecteur X de caractéristiques extrait à partir de l'image du visage pour vérifier avec un modèle du client, consistant en vecteurs semblables Y_i de caractéristiques extraites à partir des images de la personne réclamée stockés dans une base de données ($1 \leq i \leq m$ où m est le nombre d'images de cette personne dans l'ensemble d'apprentissage). La mesure de similarité peut se faire de différentes manières (distance euclidienne, la mesure en angle, la mesure de corrélation) : par exemple si la distance euclidienne entre X et Y_i est inférieure qu'un seuil, le visage à partir dont X est extrait sera considéré pour

correspondre au visage duquel Y_i est extrait. Le choix du meilleur seuil est une phase importante du problème: un seuil trop petit mènera à un taux *élevé* de *faux rejet* (TFR), alors qu'un seuil trop grand mènera à un taux *élevé* de *fausse acceptation* (TFA).

3 EXTRACTION DE CARACTERISTIQUES

La prise des décisions sur des images approximatives s'est avérée [6] nettement sensible aux conditions d'illumination, aux points de vue, à l'expression et aux différences de jour en jour dans un visage de la même personne, au point que deux images très semblables (à l'œil humain) pourraient être extrêmement différentes si elles sont comparées pixel par pixel. Il est donc nécessaire d'extraire les caractéristiques appropriées et discriminantes à partir des images du visage et de comparer ces caractéristiques au lieu de comparer des images approximatives. Naturellement, plus sont les caractéristiques discriminantes, plus l'authentification sera facile.

3.1 Analyse en composante principale (ACP)

L'ACP est une méthode linéaire qui à pour but de trouver les axes maximisant la variance des données (minimisant l'écart au carré par rapport aux axes)[4][5]. Dans cette approche, l'image a deux dimensions d'un visage est transformée en un vecteur obtenu en enchaînant les lignes (ou colonnes) de l'image. Les coefficients de projection du vecteur visage aux vecteurs de base sont employés comme une représentation (features) de chaque image du visage.

Soit $A = (X_1, X_2, \dots, X_i, \dots, X_N)$ représente une matrice de données de dimension $n \times N$, où chaque X_i est un vecteur visage de dimension n . Ici n représente le nombre de pixel dans l'image du visage et N est le nombre d'images du visage dans l'ensemble d'apprentissage.

Le vecteur moyen de l'ensemble d'apprentissage est donné par :

$$\mu = \frac{1}{N} \sum_{i=1}^N X_i \quad (1)$$

Ce vecteur est soustrait de chaque vecteur image. Le vecteur image centré est donné par :

$$\overline{X}_i = X_i - \mu \quad (2)$$

La transformation linéaire d'une image original est donnée par :

$$Y_i = W^T \overline{X}_i \quad (3)$$

où Y_i est un vecteur caractéristique de dimension $d \times 1$ et qui contient les coefficients de projection du vecteur image

\overline{X}_i dans l'espace de transformation et W est la matrice de transformation. Noter bien que d est très inférieur à n ($d \ll n$).

Les vecteurs de base de transformation en utilisant l'ACP sont définies par les vecteurs propres de la matrice de covariance S_T donnée par [1][5][9]:

$$S_T = \frac{1}{N} \sum_{i=1}^N (X_i - \mu)(X_i - \mu)^T, \quad (4)$$

L'ACP d'une matrice de données met en facteurs sa matrice covariance dans la forme suivante :

$$S_T = \Phi \Lambda \Phi^T, \quad \Phi = [\phi_1 \phi_2 \dots \phi_i \dots \phi_n],$$

$$\Lambda = \text{diag} \{ \lambda_1, \lambda_2, \dots, \lambda_n \} \quad (5)$$

où $\Phi \in \mathbb{R}^{n \times n}$ est une matrice orthogonale des vecteurs propres et $\Lambda \in \mathbb{R}^{n \times n}$ est une matrice diagonale des valeurs propres.

La matrice W_{ACP} de transformation se compose de vecteurs propres correspondant aux d plus grandes valeurs propres. Après application de la projection, le vecteur d'entrée (visage) dans l'espace de dimension n est réduit à un vecteur caractéristique dans un sous espace de dimension d .

4 RESULTATS EXPERIMENTAUX

4.1 Base de données

Nos expériences ont été exécutées sur des images frontales de visage de la base de données XM2VTS [8]. XM2VTS est une base de données multimodale publiquement disponible enregistrée spécifiquement pour évaluer les exécutions des approches biométriques à la vérification d'identité. Elle contient 8 images par visage de 295 personnes. Les sujets ont été enregistrés en quatre sessions séparées distribuées pendant 5 mois. Le protocole expérimental standard lié à la base de données divise la base de données en 200 clients et 95 imposteurs. Les sujets sont des deux sexes, proviennent de diverses ethnies et de catégories d'âge différentes. Les photos sont en couleur, de haute résolution (format ppm : 256x256 pixels).

Il n'est pas suffisant d'utiliser la même base de données pour pouvoir honnêtement comparer des résultats. Il est nécessaire également de définir un protocole de test. Dans le protocole de Lausanne [3] la base de données est scindée en trois ensembles : ensemble d'apprentissage, ensemble de validation et ensemble de test. L'ensemble d'apprentissage est utilisé comme ensemble de référence. Il sert d'ensemble de base, maintenant ainsi l'information concernant les

personnes connues du système. L'ensemble d'évaluation permet de fixer les paramètres du système de reconnaissance de visages. Enfin, l'ensemble de test permet de tester le système en lui présentant des images de personnes lui étant totalement inconnues.

En fait, la base de données est divisée en deux classes : clients et imposteurs. L'ensemble d'apprentissage ne contient que des clients. Les imposteurs sont répartis dans les deux autres ensembles, à raison de 25 pour l'ensemble d'évaluation et 70 pour l'ensemble de test.

La répartition des images dans les différents ensembles est décrite par la figure 1 [3].

Session	Shot	Clients	Impostors	
1	1	Training	Evaluation	Test
	2	Evaluation		
2	1	Training		
	2	Evaluation		
3	1	Training		
	2	Evaluation		
4	1	Test		
	2			

Figure 1 : Configuration de la base de données

Les tailles des différents ensembles sont reprises dans le tableau 1.

Tableau 1 : Répartition des photos dans les différents ensembles

Ensemble	clients	imposteurs
apprentissage	600(3 par personne)	0
évaluation	600(3 par personne)	200(8 par personne)
test	400(2 par personne)	400(8 par personnes)

4.2 Mesure de qualité

Le problème qui nous occupe contient deux classes, à savoir d'une part les *clients* et d'autre part les *imposteurs*. La matrice de confusion m est donc carrée de dimension deux. Si chaque client doit être *accepté* et chaque imposteur *rejeté*, on peut écrire :

$$m = \begin{pmatrix} TBA & TFA \\ TFR & TBR \end{pmatrix} \quad (7)$$

Les TFR et TFA sont respectivement les Taux de Faux Rejet et Taux de Fausse Acceptation ; ils doivent être bien sûr le plus faible possible. Les TBA et TBR sont respectivement le taux de bonne acceptation et le taux de bon rejet. Ces chiffres caractérisent ce que l'on appelle en reconnaissance de visage le pouvoir d'identification.

Nous supposons que le système de vérification soit impitoyable, et extrêmement strict. Le TFA sera bien sûr faible, mais le TFR sera fort élevé. Au contraire, un système laxiste sera caractérisé par un TFA élevé et un TFR plutôt bas. Le juste milieu se situe quelque part entre les deux, et, si les coûts des erreurs sont égaux, il se trouvera au Taux d'Égale Erreur ou TEE illustré par la figure 2.

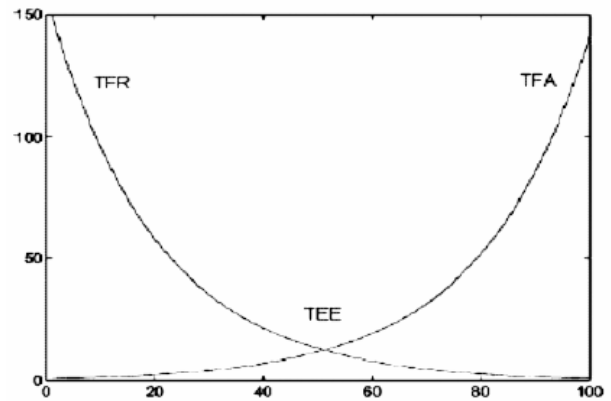


Figure 2 : Les TFA, TFR et TEE en fonction du laxisme du système.

Tous ces taux d'erreurs sont calculés dans deux ensembles : d'abord dans un ensemble d'évaluation, qui va permettre de fixer plus ou moins le TEE en faisant varier les paramètres d'acceptation et de rejet du système. Ensuite dans un ensemble de test, en utilisant les paramètres fixés précédemment. Ainsi, on peut vérifier la robustesse du système.

4.3 Pré traitement

Dans l'étape du pré traitement, nous regroupons tous les modules qui vont modifier la représentation sources. L'objectif est de trouver une représentation qui facilite le travail des modules suivants. Deux types de transformation sont utilisés. Elles permettent souvent une première réduction des données et elles atténuent les effets dus à une différence de conditions lors des prises de vues.

En regardant les images nous constatons clairement l'apparition de particularités non souhaitées au niveau du cou comme les cols de chemise, les polos, etc. Par ailleurs, les cheveux sont également une caractéristique changeante au cours du temps (changement de coupe, teinte, calvitie,...). L'arrière plan apparaît sur les images ; il ne sert à rien, et gonfle inutilement la taille des données. Enfin les oreilles posent aussi un problème. En effet, si la personne se présente légèrement différemment devant la caméra (rotation), nous ne pouvons apercevoir qu'une seule oreille. C'est pourquoi nous avons décidé de couper l'image verticalement et horizontalement et ne garder qu'une fenêtre de dimension 132x120 centrée sur le visage en utilisant la technique inspirée de [10]. Ensuite nous passons les images par un filtre passe bas pour pouvoir effectuer

une décimation de facteur 2. Ce qui réduit par 4 la taille de l'image découpée. Les images passeront ainsi d'une dimension 65536 à une dimension 66x60=3960 (après découpage et décimation), voir figure 3.

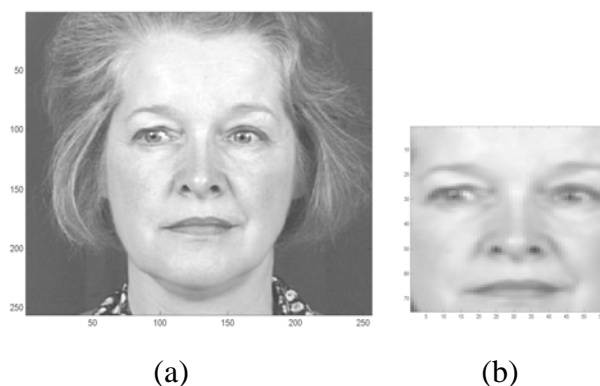


Figure 3 : L'image avant (a) et après (b) découpage et décimation

Après, nous appliquons une photonormalisation. Cela veut tout simplement dire que pour chaque image, nous soustrayons à chaque pixel la valeur moyenne de ceux-ci sur l'image, et que nous divisons ceux-ci par leur déviation standard. Enfin nous faisons une normalisation. La photonormalisation agit sur une image alors que la normalisation agit sur un groupe d'images (pour chaque composante, nous retirons la moyenne de cette composante pour toutes les images et nous divisons par la déviation standard).

4.4 L'extraction des caractéristiques

L'algorithme utilisé pour calculer l'ACP comprend les étapes suivantes :

- Pré traitement
- créer la matrice des données X
- calculer la matrice de covariance $\overline{X.X^T}$
- calculer les valeurs et vecteurs propres de la matrice de covariance
- ordonner les vecteurs propres par ordre de valeur propre décroissante.

4.5 Choix des coefficients

Etant donné que nous avons 600 images de clients dans l'ensemble d'apprentissage et que la dimension des images est supérieure à ce nombre, nous aurons au maximum 600 vecteurs propres. Nous gardons uniquement les premiers vecteurs propres, étant donné que ce sont ceux qui minimisent l'erreur. Ceci est bien justifié en compression ou en visualisation [7][8].

4.6 Classification

Une fois les caractéristiques sont extraites, le problème de vérification d'identité, devient à chercher, pour chaque personne, ou de manière globale, un seuil. Ce seuil va déterminer le minimum de ressemblance entre deux images pour admettre qu'il s'agit de la même personne. Ce minimum de ressemblance va s'exprimer s'en doute, comme une distance maximale entre les caractéristiques des deux images. Pour fixer ce seuil nous avons utilisé l'ensemble de validation.

4.7 Comparaison

Afin de mieux pouvoir étudier l'apport d'une méthode particulière, nous allons présenter les résultats obtenus avec une méthode très basique. Ces résultats de base serviront par la suite de comparaison.

Les paramètres de la méthode de base sont :

- Preprocessing sans la photonormalisation.
- Composante couleur : luminance (niveaux de gris).
- Coefficients : les coefficients de projection sur les 100 premiers vecteurs propres (triés suivant les valeurs propres décroissantes).
- Mesure de score : Distance euclidienne.
- Seuillage : global.

Remarque : dans toutes les expériences qui vont suivre, nous avons fixé le seuil de telle manière à obtenir un TEE dans l'ensemble d'évaluation.

Les résultats obtenus avec ces paramètres sont repris dans le tableau 2.

Tableau 2 : Les résultats de la méthode de base

	Evaluation	Test
TFA	0.1510	0.1327
TFR	0.1517	0.1525
TEE	0.1513	0.1426

Nous remarquons que les différents taux sont très stables dans les différents ensembles (évaluation et test). Ceci est une propriété très importante.

Les résultats de la méthode de base avec photonormalisation sont repris dans le tableau.3.

Nous remarquons donc, que la photonormalisation est en accord avec nos espérances et améliore sensiblement les résultats.

Les résultats de la méthode de base avec seuillage individuel sont repris dans le tableau 4.

Tableau 3 : Les résultats de la méthode de base avec photonormalisation

	Evaluation	Test
TFA	0.1171	0.1299
TFR	0.1167	0.1025
TEE	0.1169	0.1162

Tableau 4 : Les résultats de la méthode de base avec seuillage individuel

	Evaluation	Test
TFA	0.1475	0.2438
TFR	0.1425	0.1200
TEE	0.1463	0.1819

Nous remarquons que le seuillage individuelle est décevant. Certes, il améliore quelque peu les résultats dans l'ensemble d'évaluation mais dégrade fortement ceux-ci dans l'ensemble de test.

Pour fixer le seuil individuel, nous nous sommes basés sur les images imposteurs de l'ensemble d'évaluation. Mais celles-ci ne représentent que 200 images. En fait, ces 200 images sont issues de 25 personnes différentes. C'est probablement ce qui rend la méthode instable. Si nous possédions un nombre plus grand d'imposteurs dans l'ensemble d'évaluation, le problème ne se poserait certainement pas.

Le seuillage individuel ne fonctionne pas bien couplé à une norme euclidienne.

Les résultats de corrélation appliquée à la méthode de base et à celle avec photonormalisation sont repris dans le tableau 5.

Tableau 5 : Les résultats de corrélation appliquée à la méthode de base et à celle avec photonormalisation

		Evaluation	Test
Avec la méthode de base	TFA	0.0993	0.1015
	TFR	0.1000	0.0975
	TEE	0.0996	0.0995
Avec photo normalisation	TFA	0.0610	0.0657
	TFR	0.0617	0.0550
	TEE	0.0613	0.0603

Nous remarquons que la corrélation améliore sensiblement les résultats. Elle est mieux adaptée que la norme euclidienne à des données en grande dimension.

Nous avons également essayé la mesure en angle, mais celle-ci donne des résultats légèrement bons que la corrélation. Les résultats de mesure de similarité en angle

appliquée à la méthode de base et à celle avec photonormalisation sont repris dans le tableau 6.

Tableau 6 : Les résultats de la mesure en angle appliquée à la méthode de base et à celle avec photonormalisation

		Evaluation	Test
Avec la méthode de base	TFA	0.0975	0.1003
	TFR	0.0965	0.0975
	TEE	0.0970	0.0989
Avec photo normalisation	TFA	0.0609	0.0657
	TFR	0.0617	0.0525
	TEE	0.0613	0.0591

5 CONCLUSIONS ET PERSPECTIVES

L'ACP donne d'assez bons résultats. Elle est d'autant plus performante avec une mesure en angle. La mesure en angle apporte un gain en performance non négligeable (TEE_{angle} présente une amélioration de 40% par rapport à la norme euclidienne).

Nous avons vu plusieurs moyens pour augmenter les performances du système (photonormalisation, choix de la norme, choix des meilleurs coefficients,...). Toute la difficulté réside dans le fait que ces différentes améliorations ne sont pas indépendantes; elle interfèrent entre elles, améliorant parfois les résultats, parfois les dégradant.

Nous pensons également que nous avons exploré beaucoup de voies pour améliorer les résultats de l'ACP et que continuer dans cette direction n'apportait plus beaucoup de gain.

Dans les travaux futur nous proposons l'utilisation de la couleur comme information dans une approche ACP pour différents espaces de couleurs existant et de voir quel espace choisir. On peut aussi proposer la fusion des résultats de différents espaces ou d'utiliser un poste ACP par exemple LDA (analyse en composantes discriminantes) ou ICA (analyse en composante indépendante).

REFERENCES BIBLIOGRAPHIQUES

- [1] M. Turk and A. Pentland, Eigenfaces for recognition. Journal of cognitive neuroscience, **3**(1), 1991.
- [2] K.Messer, J.Matas, J. Kittler, J. Luetin and G. Maître, XM2VTSBD: The Extended M2VTS database Int'l Conf. on Audio- and Video-based Biometric Authentication (AVBPA 99), Washington D.C.,1999.
- [3] J.Luetin and G.Maître, Evaluation Protocol for the extended M2VTS database (XM2VTSDB) IDIAP tech. report, Martigny, July 98.
- [4] C. Havran, L. Hupet, J. Czyz, J. Lee, L.

Vandendorpe, M. Verleysen "Independent Component Analysis for face authentication" KES'2002 proceedings – Knowledge- Based Intelligent Information and Engineering Systems, Crema (Italy),16- 18 September 2002

[5] Wendy S. Yambor "analysis of PCA and Fisher discriminant-based image recognition algorithms. Technical rapport, Colorado State University July 2000.

[6] P. Belhumeur, J.P. Hespanha, D.J. Kriegman , Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1997, pp. 711-720.

[7] A. Spataru Fondements de la théorie de la transmission de l'information presses polytechniques romandes 1987

[8] J.Pontier, A. B. Dufour, M. Normand Le model Euclidien en Analyse des Données Editions Ellipses 1990.

[9] L. Lorente, L. Torres, "A global eigen approach for face recognition", international Workshop on Very Low Bit-rate Video Coding, Urbana, Illinois, October 8-9, 1998.

[10] R. Brunelli and T. Poggio. "Face recognition: features versus templates". IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.15,no.10,pp 1042-1052, October 1993.



Exemples d'images frontales de la base de données XM2VTS utilisées pour nos expériences