
Comparaison d'approches pour la détection et l'identification de visages 2D

Anis CHAARI^{1,2,3} – Sylvie LELANDAIS^{1,2} – Mohamed SAAIDIA² –
Vincent VIGNERON² - Mohamed BEN AHMED³

(1) IUT d'Evry

Département Qualité, Logistique Industrielle et Organisation
22, allée Jean ROSTAND – 91025 EVRY Cedex

(2) Laboratoire IBISC CNRS FRE 2873

Université d'Evry Val d'Essonne – UFR Sciences et Technologies
40, rue du Pelvoux – 91020 EVRY Cedex

(3) Laboratoire RIADI, Ecole Nationale d'Informatique

Université de la Manouba, 2010 La Manouba, TUNISIE

s.lelandais@iut.univ-evry.fr

Sections de rattachement : 61
Secteur : Secondaire

RÉSUMÉ. Le travail proposé ici s'inscrit dans le contexte de la biométrie dont l'objectif est de l'identification d'une personne à travers des informations biologiques ou comportementales. Nous proposons des travaux liés à deux étapes du processus de reconnaissance d'une personne à partir de l'image 2D de son visage. Dans une première partie, nous proposons une comparaison de deux méthodes de détection du visage dans l'image : l'une utilisant les moments de Zernike et l'autre les « Eigenfaces ». Nous présentons en particulier une mesure de qualité permettant de définir un indicateur de comparaison quantitatif des résultats obtenus par les deux approches proposées. Dans une seconde partie nous nous intéressons au processus d'identification proprement dit. Une approche basée sur l'analyse en composantes principales est opposée à une utilisation de l'analyse discriminante de Fisher. La détection de visage a été testée sur la base de données XM2VTS alors que le processus d'identification est évalué sur un sous-ensemble de la base IV². Les résultats présentés dans chacune de ces parties illustrent la difficulté de ces deux étapes et montrent l'intérêt du développement de nouvelles approches qui permettent d'améliorer la robustesse des traitements et la qualité des résultats.

MOTS-CLÉS : Biométrie, Reconnaissance de visages 2D, Détection de visages, Classification, Eigenfaces ; Moments de Zernike ; Fisherfaces.

1. Introduction

Dans la vie de tous les jours une personne doit être identifiée dans de multiples situations : entrée d'un immeuble, accès au travail, retrait d'argent, paiement par cartes bancaires, connexion à son ordinateur et à sa messagerie,... entraînant ainsi la mémorisation obligatoire de nombreux codes. C'est pourquoi aujourd'hui de nombreuses recherches visent à prouver l'identité d'une personne en utilisant des caractéristiques biologiques ou comportementales qui lui sont propres : iris, empreinte digitale, forme de la main, frappe sur un clavier, vitesse de signature,... La biométrie est la science qui regroupe l'ensemble de ces travaux et dispositifs. Dans le spectre de la biométrie, la reconnaissance de visages revêt un caractère particulier de part le grand nombre de travaux qui lui sont consacrés [1]. En effet capturer une image d'un visage, en particulier à travers une caméra 2D, est simple et non invasif. C'est donc une modalité biométrique facilement tolérée par les utilisateurs, et ceci d'autant plus que les ordinateurs sont équipés de plus en plus souvent de « webcam ». Mais les performances de la reconnaissance faciale sont toujours bien en deçà de ce que l'on pourrait espérer pour de telles applications. L'augmentation des taux de reconnaissance, la diminution des reconnaissances à tort et l'accélération des temps de réponse lors de la recherche dans les grandes bases de données biométriques sont les défis auxquels doivent faire face les algorithmes actuellement développés.

Pour reconnaître une personne à partir de son image, il faut passer par certaines étapes. Tout d'abord il faut localiser le visage dans l'image. Ensuite il faut éventuellement le normaliser pour ramener le visage à une taille standard. Puis il faut aborder la phase de reconnaissance proprement dite. Dans cet article nous proposons une synthèse de travaux permettant d'illustrer deux des étapes clés de cette chaîne d'un processus de reconnaissance faciale. Nous détaillons d'abord deux approches concurrentes pour détecter un visage, l'une basée sur les moments de Zernike et l'autre sur les « Eigenfaces ». Nous présentons ensuite la comparaison de deux approches de reconnaissances, toutes les deux basées sur l'utilisation d'une projection dans un sous-espace. Les images utilisées pour les différents résultats présentés sont issues de la base XM2VTS [2] et de la base IV²[3].

2. Détection du visage dans une image 2D

Selon Hjelmas et Low [4], les méthodes de détection de visages peuvent être classées en « approche globale » dans laquelle on analyse le visage dans son entier, ou en « approche locale », dans laquelle on va essayer de détecter, localiser et regrouper les différents éléments constitutifs du visage : nez, yeux, bouche,... Des informations complémentaires peuvent aussi être utilisées pour détecter les différents composants du visage, telles que la couleur, le mouvement ou encore la forme. Les deux méthodes de détection de visage que nous allons présenter dans cet article relèvent de la catégorie des méthodes globales. Elles diffèrent seulement dans leur première phase de mise en œuvre, comme l'illustre la figure 1, c'est-à-dire en fait par la technique utilisée pour caractériser l'image à traiter. La première méthode utilise les propriétés géométriques à

travers le calcul des moments de Zernike [5]. La seconde est basée sur une projection sur un sous espace. Nommée « Eigenfaces » c'est une technique bien connue en particulier dans le domaine de la reconnaissance de visages [6]. La deuxième étape de ce processus de détection est assurée par un réseau de neurones entraîné par les vecteurs caractéristiques calculés à partir des « Eigenfaces » ou des moments de Zernike. Sa sortie délivre des points de contour supposés délimiter un visage dans l'image d'origine. Pour permettre une évaluation objective des résultats de chacune des deux approches, un critère de qualité est proposé dans le paragraphe 2.3.

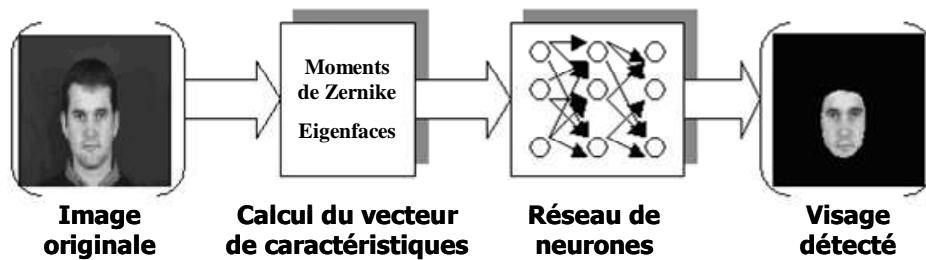


Figure n°1 : Schéma de principe de la méthode de détection du visage

2.1. Présentation des moments de Zernique

Les moments de Zernike ont été introduits en 1934 [5]. Dans le domaine du traitement de l'information, les moments de Zernike ont beaucoup été utilisés pour leur propriété d'orthogonalité qui permet la génération de descripteurs non redondant et leurs propriétés d'invariance en translation, en échelle et en rotation. Ainsi, on retrouve les moments de Zernike dans beaucoup de travaux concernant la reconnaissance d'images de personnes [7], l'indexation d'images dans les bases de données, l'analyse et la description de forme d'objet 2D ou 3D...

2.2. Introduction aux « Eigenfaces »

Les « Eigenfaces » furent le premier type de caractérisation utilisé avec succès dans des traitements faciaux tels que la détection et la reconnaissance du visage [6]. Cette méthode est basée sur la décomposition de l'image traitée selon plusieurs directions de variation autour d'une image moyenne. Nous reviendrons plus en détails sur cette méthode au paragraphe 3.1.

2.3. Définition d'une mesure de qualité de la détection

Pour mesurer l'efficacité d'un algorithme de détection du visage, nous proposons de mettre en œuvre une mesure quantitative basée sur le rapport entre la surface de visage détecté et la surface de visage présente dans l'image originale. Pour permettre de réaliser une telle mesure, l'ensemble des images de la base de test ont été divisées en trois régions (cf. figure 2). La région blanche contient les W pixels jugés essentiels dans la détection du visage (nez, yeux, bouche,...). La région grise contient des pixels éléments

du visage mais jugés non essentiels pour la bonne détection de celui-ci. La région noire contient les B pixels non éléments du visage.



Figure n°2 : Images originales et les trois zones caractéristiques associées à chacune d'elles

La région blanche doit être impérativement contenue dans le contour résultant de l'application de l'algorithme de détection. La détection de pixels de la région grise est considérée comme optionnelle et n'influera pas sur le résultat de la mesure. Par contre aucun pixel contenu dans la région noire ne devrait être détecté, par conséquent, si un pixel élément de cette région est détecté, cela pénalisera le résultat de la mesure.

Le critère de mesure proposé est basé sur le calcul des deux quantités nommées Gdr (Good detection rate) et Qdr (Quality detection rate) définies par les équations (1) et (2) :

$$Gdr = \frac{W_1}{W} \cdot 100 \quad (1) \quad \text{et} \quad Qdr = \left(\frac{W_1}{W} - \frac{B_1}{A - B} \right) 100 \quad (2)$$

W_1 et B_1 représentent respectivement le nombre de pixels correctement et faussement détectés comme faisant partie du visage et A le nombre total des pixels de l'image d'origine. Gdr mesure à quel point les pixels composant l'essentiel du visage ont été correctement détectés. Qdr donne une mesure plus stricte prenant en compte le nombre de pixels faussement détectés comme appartenant au visage dans le calcul de la valeur quantitative de détection. Ces deux mesures sont complémentaires. En effet, si on connaît seulement Gdr on ne possède aucune information sur le taux de pixels attribués à tort au visage. Par ailleurs, la seule connaissance de Qdr ne donne pas d'information sur le nombre de pixels détectés appartenant à l'essentiel du visage et qui ne sont pas inclus dans la région détectée comme visage. En conclusion l'algorithme de détection sera d'autant meilleur que Gdr sera élevé et Qdr proche de Gdr (puisque Qdr est inférieur ou égal à Gdr).

2.4. Résultats obtenus

La figure 3 donne un exemple de résultats obtenus par les deux méthodes de détection introduite dans cet article. Pour obtenir ces résultats, nous avons utilisé 295 images de la base de données XM2VTS [2]. Le tableau 1 illustre le bon comportement des moments

de Zernike face aux Eigenfaces. En effet, on constate que Gdr et Qdr sont plus importants en moyenne lorsque l'image est caractérisée par les moments de Zernike. De même les écart-types plus faibles obtenus permettent d'illustrer le meilleur centrage des résultats autour de ces valeurs moyennes.



Figure n°3 : Résultats de détection et valeurs de Gdr et Qdr obtenus que des images de la base XM2VTS. En haut, images d'origine. Au milieu détection par l'algorithme utilisant les moments de Zernike. En bas détection par les « Eigenfaces »

Tableau n°1 : Résultats des deux mesures de qualité de détection de visage obtenus par les moments de Zernike et les « Eigenfaces »

	Moments de Zernike	Eigenfaces
Moyenne Gdr	93,82%	90,16%
Ecart-type Gdr	9,45%	19,1%
Moyenne Qdr	82,07%	79,82%
Ecart-type Qdr	19,05%	24,23%

3. Définition d'une stratégie de reconnaissance

Les "Eigenfaces" et les "Fisherfaces" sont deux méthodes fondamentales dans les approches proposées pour la reconnaissance de visages. Toutes deux sont basées sur la décomposition de l'image sur un sous espace réduit et sur la recherche d'un vecteur optimal de caractéristiques décrivant l'image du visage à reconnaître.

Dans ces méthodes, une image de taille p par q pixels, est représentée par un vecteur de taille n ($n = p \times q$) dans un espace de grande dimension. Les images de visages sont linéarisées comme des vecteurs x puis sont exprimées en tant que combinaison linéaire de la base orthogonale Φ réduite à une dimension choisie m :

$$\Phi_i : x = \sum_{i=1}^n a_i \phi_i \approx \sum_{i=1}^m a_i \phi_i \quad (3)$$

3.1. Analyse en Composantes Principales ou PCA

Le premier système de reconnaissance de visages qui a permis d'obtenir des résultats significatifs a été réalisé par Turk et Pentland [6] en utilisant la méthode dite des « Eigenfaces ». La base orthogonale Φ définit un espace appelé « facespace » de dimension m inférieure à la taille des vecteurs images ($m \ll n$) en résolvant l'équation

$$C\Phi = \Phi\Lambda \quad (4)$$

où C est la matrice de covariance du vecteur image x : $C = \sum (x - \bar{x})(x - \bar{x})^T$ (5)

$\Phi = [\phi_1, \dots, \phi_m]^T$ est la matrice des vecteurs propres appelés aussi eigenfaces de la matrice de covariance C , et Λ est la matrice des valeurs propres associées.

3.2. Analyse Discriminante Linéaire ou LDA

Belhumeur et col. [8] ont proposé la méthode des « Fisherfaces » pour résoudre le problème de la robustesse face aux variations de pose, d'illumination et d'expressions mettant en difficultés la PCA. Le principe consiste à appliquer l'analyse discriminante linéaire (LDA) pour sélectionner le sous espace linéaire Φ qui maximise le rapport (6) :

$$\frac{|\Phi^T S_b \Phi|}{|\Phi^T S_w \Phi|} \quad (6)$$

où : $S_b = \sum_{i=1}^M (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T$ est la matrice de variance inter-classes,

et $S_w = \sum_{i=1}^M \sum_{x \in X_i} (x - \bar{x}_i)(x - \bar{x}_i)^T$ est la matrice de variance intra-classes.

M étant le nombre d'individus de la base de données pour construire l'espace de projection Φ . Les Fisherfaces tels que proposés dans [8] et [9] commencent par appliquer une analyse en composantes principales (PCA) pour réduire la dimension du sous-espace de projection avant de mettre en œuvre l'analyse discriminante linéaire (LDA). Ceci permet de résoudre le problème de singularité de la matrice de variance intra-classes.

3.3. Processus de reconnaissance

Toutes les images des visages connus (à enrôler dans la base de données) sont projetées dans l'espace des visages obtenu par PCA ou LDA (« face space » Φ) afin de trouver l'ensemble des poids décrivant la contribution de chaque image de visage. Si l'on a plusieurs images pour une même personne, il est possible de calculer plusieurs vecteurs

représentant cette même personne, ou de construire un vecteur unique, qui sera calculé à partir des différents vecteurs obtenus.

Quand une personne se présente au système, son image est projetée dans l'espace des visages construit Φ afin d'obtenir le vecteur des poids le décrivant. En mode, authentification, ce vecteur est comparé au vecteur de la personne proclamée qui a été enrôlée précédemment. Si une personne légitime est rejetée par le système, la performance du système se mesure donc à son taux de faux rejet (*False Rejection Rate ou FRR*). D'autre part, si un imposteur est accepté par le système, la performance du système se mesure à son taux de fausse acceptation (*False Acceptance Rate ou FAR*). Le point d'équivalence des erreurs (*Error Equal Rate ou EER*), correspond au taux $FRR = FAR$.

3.4. Comparaison de résultats entre PCA et LDA sur la base IV²

Des premiers résultats comparatifs de ces deux approches ont été obtenus sur la base de données IV² [3]. Cette base de données biométriques multimodales a été développée dans le cadre d'un projet Techno Vision co-financé par les ministères de la recherche et de la défense. Elle contient environ 1 Tera octets de données comprenant :

- des séquences vidéos de visages en 2D acquises par un caméscope et une webcam durant la lecture d'un texte imposé ;
- des séquences vidéos acquises par des caméras constituant des têtes stéréoscopiques. Deux bancs stéréoscopiques, l'un de face, l'autre de profil, permettent ainsi d'obtenir quatre séquences vidéos en 2D ou, par l'utilisation de paires d'images correspondantes à une acquisition simultanée sur les deux caméras, de remonter à une information 3D ;
- des scans 3D obtenus par un scanner laser. Pour chaque individu, on dispose ainsi de 8 acquisitions différentes ainsi que d'un modèle 3D reconstruit.
- des images des iris des deux yeux à travers deux sessions différentes prises à quelques minutes d'intervalle.

Cette base comporte en final environ 315 identités différentes, réparties dans près de 430 enregistrements à travers une, deux ou exceptionnellement trois sessions d'un même individu. Des variations de pose, luminosité et expressions sont accessibles à travers les différentes séquences vidéos proposées. Une base de développement, et un package correspondant aux données préparées pour une première évaluation sont aussi disponibles.

Les résultats donnés ci-dessous correspondent à une utilisation des données présentes dans la base de développement. Ils ont été calculés à partir de 500 comparaisons tirées au hasard. Sur ces 500 comparaisons, la moitié concernait des comparaisons entre deux images d'un même individu (nommées comparaisons intra). Le tableau 2 donne le détail de ces résultats. La distance de Manhattan (L1), la distance Euclidienne (L2) et la distance cosinus ont été testées avec les deux algorithmes développés. La distance L1 sur les Eigenfaces et la distance cosinus avec les Fisherfaces donnent les meilleurs résultats. Des scores de similarité correspondant à ces distances sont calculés et normalisés entre 0 et 1. Un score proche de 1 indique une grande similarité et vice

versa. La séparabilité des scores intra et inter classes permet de définir un seuil de décision permettant de calculer le taux de fausse acceptation et le taux de faux rejet.

Tableau n°2 : Comparaison des scores de similarités et des taux de reconnaissance
Eigenfaces (PCA) – Fisherfaces(LDA)

Algorithme	PCA	LDA
Distance	L1	Cosinus
Normalisation	Min-Max	Min-Max
Moyenne intra	0,325	0,909
Ecart type intra	0,16	0,101
Moyenne inter	0,089	0,487
Ecart-type inter	0,046	0,196
EER	0,073	0,072
FAR (FRR=0,1)	0,072	0,072

Remarquons toutefois que la faible taille de la base d'apprentissage ne permet pas d'analyse précise des différences de résultats. En effet cette taille de base est un peu limitée dans le cas de la LDA. D'autres résultats obtenus par différents participants et différents algorithmes sur un ensemble de données plus important extrait de la base IV² sont en cours de publication par ailleurs.

4. Conclusion

Nous avons présenté dans cet article des travaux touchant à la détection de visage et à l'authentification d'individus. L'objectif est de montrer la complexité des méthodes pouvant être utilisées dans de telles applications. Si la biométrie est un enjeu important au niveau économique, la recherche, en particulier dans le domaine de la reconnaissance des personnes à partir d'une image 2D de leur visage, offre encore un champ d'investigations très ouvert.

Références bibliographiques

- [1] S. Z. Li, A. K. Jain, "Handbook of Face Recognition", Springer, 2004.
- [2] K. Messer, J. Matas, J. Kittler, J. Luettin, G. Maître, "XM2VTSBD: The Extended M2VTS database", AVBPA, pp. 72-77, 1999.
- [3] Site web du projet IV²: <http://iv2.ibisc.fr/>
- [4] E. Hjelm et B. K. Low. "Face detection : A survey", Computer Vision and Image Understanding, vol. 83, no. 3, pp. 236-274, 2001.
- [5] F. Zernike. "Diffraction theory of the cut procedure and its improved form, the phase contrast method". Physica, 1:pp. 689-704, 1934.
- [6] M. Turk, A. Pentland, *Eigenfaces for recognition*. J. of Cognitive Neuroscience 3, 72–86, 1991.
- [7]. J. Haddadnia, M. Ahmadi, and K. Faez, "An Efficient Feature Extraction Method with Pseudo Zernike Moment in RBF Neural Network Based Human Face Recognition System", EURASIP JASP, vol. 9, pp. 890-891, 2003.
- [8] P. Belhumeur, J. Hespanha, D. Kriegman, *Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection*, IEEE Transactions on Pattern Analysis and Machine Intelligence 19, pp. 711-720, July 1997.
- [9] D. L. Swets, J. Weng, *Using discriminant eigenfeatures for image retrieval*, IEEE Trans. Patt. Anal. Mach. Intell. 18, 831– 836, 1996.