



Prestaciones de la Normalización del Rostro en el Reconocimiento Facial

D. Carrero, B. Ruíz, L. Puente y M.J. Poza.

Universidad Carlos III de Madrid. Av. de la Universidad, 30. Leganés. Madrid
dcarrero@di.uc3m.es, bruiz@inf.uc3m.es, lpuente@it.uc3m.es,
mpoza@pa.uc3m.es

Abstract. En este trabajo se estudia la influencia de la normalización facial sobre las prestaciones globales de un sistema de reconocimiento. Para ello se evalúan las prestaciones de un sistema basado en *Principal Component Analysis* (PCA) cuando se aplica normalización facial mediante *Active Shape Models* (ASM) en un escenario de verificación de identidad.

Keywords: Reconocimiento Facial, Eigenfaces, Principal Component Analysis, Active Shape Model.

1 Introducción

El reconocimiento facial ha recibido una considerable atención durante los últimos años debido principalmente a dos motivos [1]: El amplio crecimiento de las aplicaciones comerciales y la rápida evolución de la tecnología. En el campo de la biometría, la imagen facial es uno de los recursos más utilizados en el reconocimiento de personas. Este hecho viene motivado porque la captura de la información biométrica de los sujetos es una técnica no intrusiva que puede realizarse de forma sencilla [2].

El reconocimiento facial presenta dos problemas inherentes que degradan sus prestaciones: por un lado, la variación de la postura y la expresión facial del sujeto [3] y por otro, la variación de las condiciones de iluminación [4]. Así mismo, existen otros muchos factores que influyen en las prestaciones finales del sistema [1]: calidad de la imagen, variabilidad de la escena, número de usuarios y disponibilidad de un criterio de reconocimiento bien definido.

Si bien es cierto que muchas de las aplicaciones biométricas actuales requieren la colaboración de los usuarios, existen otras muchas relacionadas con la vigilancia en las que las muestras biométricas se capturan sin la participación activa o voluntaria de los mismos[5].

1.1 Descripción del Proceso de Reconocimiento

El problema del reconocimiento facial se puede definir como la comparación de una escena o secuencia de imágenes contra una plantilla o modelo facial [5]. De forma

general, la solución al problema del reconocimiento implica una serie de procesos [1]: 1) detección facial, 2) normalización facial y extracción de características y 3) identificación o verificación. La variabilidad del entorno, el ruido, la oclusión de las zonas faciales y los requisitos de velocidad hacen que las tres etapas anteriores puedan llegar a ser altamente complejas.

1.1.1 Detección Facial

La detección del rostro supone el primer paso en cualquier sistema de reconocimiento facial automático. En este primer proceso existen dos objetivos principales [6]: localización de la región facial (si existe) y segmentación de la misma del resto de la escena.

Durante los últimos años han surgido diferentes soluciones a este problema. Viola y Jones [7] proponen un sistema de detección de objetos basado en una cascada de clasificadores AdaBoost. Nilsson et al. [8] ofrecen una solución a la detección facial a partir de características extraídas de la iluminación y clasificadores SNoW (*Sparse Network of Windows*). Como último ejemplo, Xu y Chen [9] proponen un algoritmo de detección basado en diferencias y PCA sobre imágenes en color.

1.1.2 Normalización Facial y Extracción de Características

Es bien sabido que los sistemas de reconocimiento facial necesitan de forma precisa las posiciones de las características faciales (ojos, boca,...) [10]. El objetivo de los métodos de normalización y de extracción de características consiste en localizar de forma precisa las regiones faciales. Estos métodos se pueden agrupar en dos familias [1]:

- Métodos basados en plantillas: Los métodos basados en plantillas utilizan imágenes predefinidas de las regiones faciales para realizar su localización. Los trabajos más relevantes referentes a estos métodos son las propuestas de Hallinan [11] y Yuille et al. [12].
- Métodos estructurales: Los métodos estructurales usan características del contorno y de la textura de la región facial para construir un modelo estadístico de la misma. Un claro ejemplo de modelo estructural es ASM (*Active Shape Model*), propuesto por Cootes et al. [13]. Existen otros métodos estructurales como AAM (*Active Appearance Models*), propuesto también por Cootes et al. [14] y el ajuste de Gráficos Elásticos de Bunch, propuesto por Monzo et al. [15].

1.1.3 Reconocimiento

La última etapa del proceso hace referencia a la propia verificación o identificación del sujeto cuya imagen facial ha sido capturada. Es posible agrupar los métodos de reconocimiento en diferentes familias.

- Métodos holísticos: Los métodos holísticos utilizan la totalidad de la región facial como entrada bruta. En [16], Turk y Pentland proponen la utilización

directa de PCA para construir un subespacio cuya base vectorial (*Eigenfaces*) se corresponde con las direcciones de máxima varianza en el espacio de imágenes. Por su parte, Zhao et al. [17] presentan las *Fisherfaces* como el resultado de la aplicación del discriminante lineal de Fisher sobre el espacio de *Eigenfaces*.

- Métodos locales: Los métodos locales emplean características geométricas y estadísticas como la posición y la apariencia de los ojos, la nariz o la boca en el proceso de reconocimiento. Kanade [18] propone el reconocimiento facial a partir de las distancias entre los puntos nodales de la cara.
- Métodos híbridos: Los métodos híbridos utilizan características tanto globales como locales para realizar el reconocimiento. En [19], Weyrauch et al. presentan un sistema de reconocimiento robusto frente a cambios de iluminación y postura que combina modelos 3D de la geometría facial con características locales. Pentland et al. [20] incorporan al sistema clásico de *Eigenfaces* el modelado de los ojos, la nariz o la boca mediante PCA.

1.2 Objetivos y Organización del Documento

El presente trabajo recoge los resultados de la evaluación de las prestaciones de un sistema basado en *Eigenfaces* cuando se introduce ASM como técnica de normalización facial. Para ello, se propone la evaluación del sistema sobre un escenario de verificación de identidad.

Tras esta introducción, el documento está organizado de la siguiente forma. Las secciones 2 y 3 presentan los fundamentos de PCA y ASM respectivamente. La sección 4 muestra la metodología experimental adoptada. La sección 5 resume los resultados obtenidos tras los experimentos. Finalmente, la sección 6 presenta las principales conclusiones del trabajo.

2 Análisis por Componentes Principales

Una imagen I de dimensiones $K \times K$ puede ser considerada como un vector \mathbf{x} de K^2 dimensiones. Este vector \mathbf{x} se puede expresar a su vez como una combinación lineal de los N vectores de una base vectorial Φ tal que [21]:

$$\mathbf{x} = \sum_{i=1}^N w_i \Phi_i \approx \sum_{i=1}^M w_i \Phi_i, M \ll N \quad (1)$$

La base Φ se puede obtener como la solución al siguiente problema de autovectores sobre la matriz de covarianzas S de las imágenes de entrenamiento:

$$S\Phi = \Phi\Lambda \quad (2)$$

Los M vectores columna de Φ correspondientes a los mayores autovalores se conocen como *Eigenfaces*. Así, dado un conjunto de *Eigenfaces*, cada una de las caras de la base de datos se puede representar como un vector de pesos \mathbf{w} obtenido mediante su proyección sobre los componentes de Φ . Cuando se requiere el reconocimiento de una nueva imagen \mathbf{x}_i , ésta se proyecta sobre Φ para obtener su

vector de pesos \mathbf{w}_i . Por un lado, la identificación se realiza a través de la localización de la imagen en la base de datos cuyos pesos \mathbf{w}_u son los más cercanos a \mathbf{w}_i . Por otro lado, la verificación se realiza en función de la distancia entre el vector \mathbf{w}_u del usuario cuya identidad se pretende y el vector \mathbf{w}_i de la imagen capturada.

3 Active Shape Models

El modelo estructural que define ASM está compuesto por tres elementos [22]: un modelo de contorno, un modelo de apariencia y un algoritmo de búsqueda para ajustar el modelo mediante la minimización de una función de coste.

3.1 Modelo de Contorno

El objetivo del modelo de contorno es asegurar que la segmentación produce contornos válidos para la región facial. El modelo de contorno viene dado por las componentes principales de los vectores definidos por las marcas de referencia. Una región facial queda definida mediante n marcas $(x_1, y_1), \dots, (x_n, y_n)$ que se pueden agrupar para formar vectores de contorno de la forma:

$$\mathbf{x} = (x_1, \dots, x_n, y_1, \dots, y_n)^T \quad (3)$$

Así, un contorno se puede aproximar mediante la siguiente expresión:

$$\mathbf{x} \approx \mathbf{x}_m + \Phi \mathbf{b} \quad (4)$$

En la expresión anterior, \mathbf{x}_m define el contorno medio, Φ es la matriz de t autovectores obtenida como solución al problema de autovectores sobre la matriz de covarianzas de los contornos y \mathbf{b} es un vector de t elementos que contiene los parámetros del modelo tal que:

$$\mathbf{b} = \Phi^T(\mathbf{x} - \mathbf{x}_m) \quad (5)$$

Cuando se ajusta el modelo, los valores de \mathbf{b} se limitan de la siguiente forma:

$$|b_i| < m\sqrt{\lambda_i} \quad (6)$$

De forma general, los valores de m oscilan entre dos y tres. Por su parte, el número de autovalores que se utiliza se elige de forma que el modelo ajuste una cierta proporción f_v de la varianza de los contornos de entrenamiento.

3.2 Modelo de Apariencia

El modelo de apariencia se limita al borde de la región facial sobre la imagen en escala de grises y consiste en la derivada normalizada en la dirección perpendicular al contorno de la misma. Su objetivo es asegurar que la segmentación localiza el rostro en una posición donde la estructura de la imagen a los lados (o dentro) del contorno es similar a la estructura de las imágenes de entrenamiento. El modelo de apariencia que

define la estructura de la imagen alrededor de cada marca se obtiene a partir de un perfil de píxeles muestreados en las direcciones perpendiculares al contorno. La dirección perpendicular a la marca (x_n, y_n) se obtiene rotando 90° el vector formado por las marcas (x_{n-1}, y_{n-1}) y (x_{n+1}, y_{n+1}) .

Una vez se han calculado las direcciones perpendiculares, se muestrean k píxeles a cada lado del contorno en cada una de las marcas para generar perfiles de longitud $2k + 1$. El modelo de apariencia se construye al obtener la primera derivada de cada uno de los perfiles, calculada como la diferencia entre los puntos $j - 1$ y $j + 1$. Una vez calculada la derivada, se normalizan los elementos del perfil de tal forma que la suma de sus valores absolutos sea 1.

A partir de N imágenes de entrenamiento se obtiene una serie de perfiles normalizados g_1, \dots, g_s para cada una de las marcas. Así, si se calculan el perfil medio g_m y la matriz de covarianzas S_g , es posible calcular la distancia de Mahalanobis entre un nuevo perfil y el modelo como:

$$f(g_i) = (g_i - g_m)^T S_g^{-1} (g_i - g_m) \quad (7)$$

3.3 Algoritmo de Búsqueda

El contorno de la región facial se ajusta mediante un proceso iterativo que comienza en el contorno medio. En cada iteración, cada marca se mueve n_s posiciones en la dirección perpendicular al contorno y se coloca en la posición que ofrece la menor distancia de Mahalanobis. Tras la actualización de las marcas, el modelo se desplaza a las nuevas posiciones

4 Metodología Experimental

A lo largo de la siguiente sección se describen tanto la base de datos utilizada como los experimentos realizados sobre la misma.

4.1 Descripción de la Base de Datos

En la realización de los experimentos se ha utilizado la base de datos BIOSEC [23]. Esta base de datos está formada por las imágenes correspondientes a 200 usuarios capturadas en dos sesiones diferentes. Las imágenes están capturadas por una cámara Philips ToUcam PRO II en formato BMP y con una resolución de 480 por 640 píxeles. El principal problema que presenta BIOSEC radica en la diferencias de iluminación entre los rostros y el fondo de la escena. Para solucionar este problema se ha seguido la estrategia ofrecida por Demirel y Anbarjafari [4], consistente en la corrección de la intensidad de cada uno de los canales de color mediante su descomposición en valores singulares.

Para construir el espacio de *Eigenfaces* se ha tomado una imagen (220 por 220 píxeles) de cada uno de los veinte primeros usuarios de la base de datos. De esta

forma, se ha construido una base de diecinueve *Eigenfaces* sobre un espacio de 48.400 dimensiones.



Fig. 1. Ejemplos de la Base de Datos BIOSEC.

La evaluación del sistema se ha realizado a partir de otros veinte usuarios; así mismo, se ha utilizado un tercer grupo de imágenes como usuarios no registrados en el sistema. En el entrenamiento de los modelos se han utilizado dos imágenes por usuario y en la evaluación del sistema se han realizado un total de 120 tests, sesenta correspondientes a usuarios genuinos y otros sesenta correspondientes a usuarios impostores.

4.2 Segmentación de las Imágenes

Para la segmentación de la región facial se ha utilizado el algoritmo propuesto por Nilsson et al. [8], el cual ha demostrado ser bastante robusto en la detección de rostros. Tras el proceso de segmentación, se han obtenido imágenes en escala de grises con una resolución de 300 por 300 píxeles.

4.3 Localización de las Marcas Faciales

La localización de las marcas faciales se ha realizado de acuerdo con la estrategia propuesta por Zhou et al. [24]. Así, se han utilizado un total de cincuenta y ocho marcas faciales para realizar la alineación de los rostros tal y como se puede apreciar en la figura 2. De las cincuenta y ocho marcas, se han utilizado trece para definir el contorno facial exterior y cuarenta y cinco para definir las regiones internas de la cara (nariz, boca, ojos y cejas).

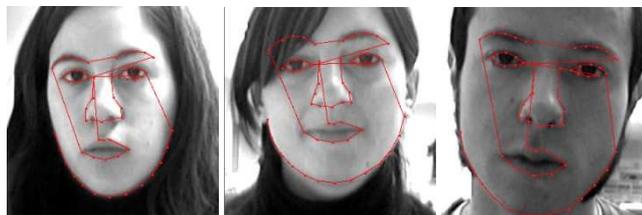


Fig. 2. Localización de las Marcas Faciales.

En los experimentos se ha utilizado la implementación de ASM propuesta por Kroon [25]. En el ajuste del modelo y en la localización de las marcas faciales se han

utilizado los parámetros de configuración listados en la tabla 1. Estos valores se han establecido a partir de las veinte imágenes utilizadas para construir el espacio de *Eigenfaces*.

Tabla 1. Parámetros de Configuración de ASM.

Parámetro	Valor
Longitud del perfil de intensidad (k)	6
Número de píxeles de búsqueda (n_s)	10
Número de escalas de resolución de la imagen	2
Límite del contorno (m)	1
Número de iteraciones del algoritmo de búsqueda	50

4.4 Alineación de los Rostros

Como plantilla de normalización facial se ha utilizado el contorno medio obtenido tras el ajuste del modelo ASM. Así, mediante el ajuste de las marcas localizadas en la imagen a las marcas a la plantilla es posible realizar la normalización de la región facial.

La figura 3 ofrece un ejemplo de normalización facial. La imagen izquierda muestra las marcas faciales del contorno medio superpuestas sobre la imagen capturada. Por su parte, la imagen derecha recoge el resultado de la normalización del rostro. En dicha imagen se puede comprobar cómo la mayoría de las regiones faciales del sujeto coinciden con las marcas de referencia.

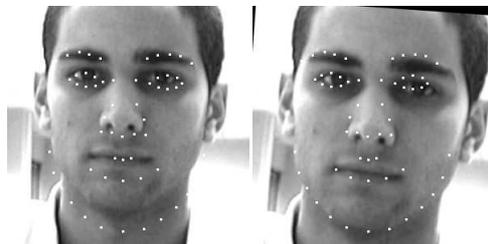


Fig. 3. Ejemplo de Normalización Facial.

La deformación de la imagen se ha realizado mediante transformaciones afines [26] entre las marcas de la imagen facial y las marcas de la plantilla de referencia. Se ha optado por utilizar transformaciones afines porque la deformación que aplican sobre la imagen mantiene las proporciones del rostro y no es tan agresiva como las transformaciones polinómicas. Tras la normalización del rostro, se han descartado 40 píxeles alrededor de la imagen para eliminar las regiones oscuras que puedan aparecer al aplicar la transformación.

5 Resultados

Para medir las prestaciones del sistema se ha utilizado su precisión en la clasificación de los usuarios como genuinos o impostores. El sistema sin normalización facial presenta una precisión del 67.5% en la verificación de la identidad de los sujetos. Al incorporar la normalización facial mediante ASM se comprueba cómo la precisión del sistema aumenta hasta el 73.3%. La mejora en la tasa de acierto se sitúa, por tanto, en un 5.8% con respecto al sistema original.

Si se atiende a la curva ROC, el área bajo la misma se incrementa desde el 0.717 ofrecido por el sistema original hasta el 0.79 conseguido mediante la normalización facial.

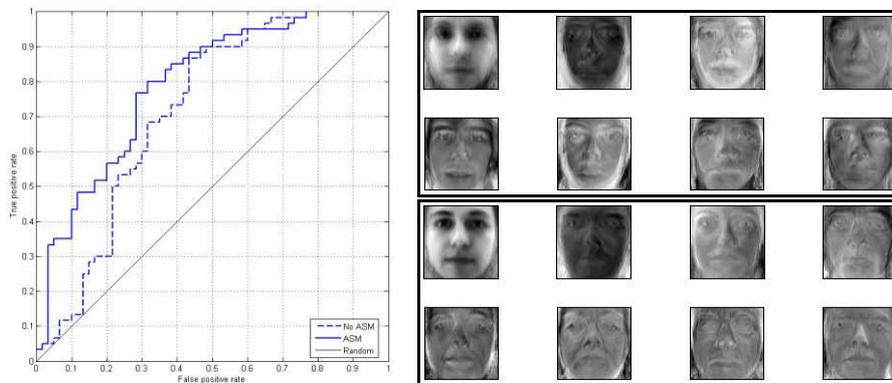


Fig. 4. Izquierda: Curvas ROC obtenidas para el sistema sin normalización (discontinua) y con normalización (continua) facial. Derecha: Comportamiento de las Eigenfaces para el sistema sin normalización (arriba) y con normalización (abajo).

En la figura 4 se observa que las imágenes que componen el espacio de *Eigenfaces* se encuentran difusas debido a la falta de alineación en los rostros. Al introducir la normalización facial, las imágenes quedan definidas con una mayor nitidez debido a la alineación de los rostros.

6 Conclusiones

A lo largo del presente trabajo se ha evaluado el comportamiento de un sistema de reconocimiento facial basado en PCA al introducir técnicas de normalización basadas en ASM.

Tras la evaluación de los resultados es posible extraer una serie de conclusiones. En primer lugar, la normalización de la región facial mejora las tasas de acierto del sistema en el escenario estudiado. La normalización facial mejora la calidad de las imágenes del espacio de *Eigenfaces*, lo que se traduce en una mejora en la calidad del sistema de reconocimiento. Y en segundo lugar, ASM ofrece muy buenas

prestaciones en la localización de las regiones faciales, incluso en escenarios con pocos usuarios de entrenamiento como es el caso.

Agradecimientos

Este trabajo ha sido desarrollado dentro del Proyecto BRAVO (TIN2007-67407-C03-01), financiado por el Ministerio de Ciencia e Innovación.

Referencias

- [1] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, "Face recognition: A Literature Survey," *ACM Computing Surveys*, vol. 35, 2003, pp. 399-458.
- [2] A. Jain, A. Ross, and S. Prabhakar, "An Introduction to Biometric Recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, 2004, pp. 4-20.
- [3] C. Hsieh, S. Lai, and Y. Chen, "Expression-Invariant Face Recognition With Constrained Optical Flow Warping," *IEEE Transactions on Multimedia*, vol. 11, 2009, pp. 600-610.
- [4] H. Demirel and G. Anbarjafari, "Pose Invariant Face Recognition Using Probability Distribution Functions in Different Color Channels," *IEEE Signal Processing Letters*, vol. 15, 2008, pp. 537-540.
- [5] D. Dessimoz, J. Richiardi, C. Champod, and A. Drygajlo, "Multimodal Biometrics for Identity Documents (MBioID)," *Forensic Science International*, vol. 167, 2007, pp. 154-159.
- [6] H. Lee, S. Kim, S. Kim, and S. Lee, "Face Detection using Multi-Modal Features," *2008 IEEE International Conference on Control, Automation and Systems*, IEEE, 2008, pp. 2152-2155.
- [7] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," *2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001)*, IEEE Comput. Soc, 2001, pp. 511-518.
- [8] M. Nilsson, J. Nordberg, and I. Claesson, "Face Detection using Local SMQT Features and Split up Snow Classifier," *2007 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007)*, IEEE, 2007, pp. 589-592.
- [9] Y. Xu and X. Chen, "A New Algorithm of Face Detection Based on Differential Images and PCA in Color Image," *2009 IEEE International Conference on Computer Science and Information Technology*, IEEE, 2009, pp. 172-176.
- [10] A. Martinez, "Recognizing Imprecisely Localized, Partially Occluded, and Expression Variant Faces from a Single Sample per Class," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, 2002, pp. 748-763.

- [11] P.W. Hallinan, "Recognizing Human Eyes," *Proceedings of SPIE*, vol. 1570, 1991, pp. 214-226.
- [12] A. Yuille, P. Hallinan, and D. Cohen, "Feature Extraction from Faces Using Deformable Templates," *International Journal of Computer Vision*, vol. 8, 1992, pp. 99-111.
- [13] T. Cootes, C. Taylor, D. Cooper, and J. Graham, "Active Shape Models. Their Training and Application," *Computer Vision and Image Understanding*, vol. 61, 1995, pp. 18-23.
- [14] T. Cootes and C. Taylor, "Statistical Models of Appearance for Computer Vision," *World Wide Web Publication*, 2004.
- [15] D. Monzo, A. Albiol, J. Sastre, and A. Albiol, "HOG-EBGM vs. Gabor-EBGM," *2008 IEEE International Conference on Image Processing*, 2008, pp. 1636-1639.
- [16] M. Turk and A. Pentland, "Face Recognition using Eigenfaces," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 586-591.
- [17] W. Zhao, R. Chellappa, and A. Krishnaswamy, "Discriminant Analysis of Principal Components for Face Recognition," *3rd IEEE International Conference on Automatic Face and Gesture Recognition*, IEEE Comput. Soc, 1998, pp. 336-341.
- [18] T. Kanade, "Picture Processing System by Computer Complex and Recognition of Human Faces," 1977.
- [19] B. Weyrauch, B. Heisele, J. Huang, and V. Blanz, "Component-Based Face Recognition with 3D Morphable Models," *Conference on Computer Vision and Pattern Recognition Workshop (CVPRW '04)*, 2004, pp. 85 - 85.
- [20] A. Pentland, B. Moghaddam, and T. Starner, "View-Based and Modular Eigenspaces for Face Recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE Comput. Soc. Press, 1994, pp. 84-91.
- [21] M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, vol. 3, 1991, pp. 71-86.
- [22] T. Cootes and C. Taylor, "Statistical Models of Appearance for Medical Image Analysis and Computer Vision," *Proceedings of SPIE Medical Imaging 2001*, 2001, pp. 236-248.
- [23] J. Fierrez, J. Ortega Garcia, D. Torre Toledano, and J. Gonzalez Rodriguez, "Biosec Baseline Corpus: A Multimodal Biometric Database," *Pattern Recognition*, vol. 40, 2007, pp. 1389-1392.
- [24] D. Zhou, D. Petrovska-Delacretaz, and B. Dorizzi, "Automatic Landmark Location with a Combined Active Shape Model," *IEEE International Conference on Biometrics: Theory, Applications, and Systems*, IEEE, 2009, pp. 1-7.
- [25] D. Kroon, "Active Shape Model (ASM) and Active Appearance Model (AAM)," <http://www.mathworks.com/matlabcentral/fileexchange/26706-active-shape-model-asm-and-active-appearance-model-aam>, 2010.
- [26] C. a. Glasbey K. V. Mardia, "A Review of Image-Warping Methods," *Journal of Applied Statistics*, vol. 25, 1998, pp. 155-171.