

# Reconocimiento de caras: Eigenfaces y Fisherfaces

Guillermo Ottado

## Resumen

En este trabajo se describen dos métodos tradicionales utilizados para el reconocimiento de caras, *Eigenfaces* y *Fisherfaces*. Se compara además el desempeño de dichos métodos utilizando bases de datos con imágenes de caras bajo diferentes condiciones de luz y de pose. Las bases de datos utilizadas son la *Extended Yale Face Database B* y la *Sheffield Face Database*. Los experimentos realizados demuestran que *Fisherfaces* tiene un desempeño superior a *Eigenfaces* para diferentes condiciones de iluminación como de pose.

## 1. Introducción

El reconocimiento de caras es un método biométrico para la identificación de individuos mediante características faciales. Existen diversas aplicaciones prácticas, abarcando numerosas disciplinas, donde puede utilizarse el reconocimiento de caras. Dichas aplicaciones pueden categorizarse en aplicaciones de seguridad o comerciales. Dentro de las aplicaciones relacionadas a la seguridad se encuentra el reconocimiento sobre retratos policiales, vigilancia de video, etc. Ejemplos de aplicaciones comerciales son la identificación en el uso de tarjetas de crédito, tarjetas de débito, pasaportes, licencias de conducir, control de acceso, etc. Todas estas aplicaciones exigen a un sistema de reconocimiento de caras a enfrentar diferentes condiciones de capacidad de procesamiento, memoria, velocidad, sensibilidad a cambios en la cara y variabilidad de las imágenes utilizadas, entre otras. Este trabajo aborda dos de los métodos de reconocimiento facial existentes sin contemplar dichos métodos para su uso sobre una aplicación en particular.

Dos problemas centrales en un sistema automático de reconocimiento de caras son la selección de características para la representación de la cara y su clasificación basada en la representación de características elegido. Dicha selección de características puede ser afectado por variaciones en las imágenes de caras, como la iluminación, expresión y pose.

El reconocimiento de caras es un problema de clasificación utilizando datos de alta dimensión, por lo que para permitir un reconocimiento rápido y robusto es necesario realizar tareas de reducción de dimensión (extracción de características) antes de la clasificación. Los métodos de reducción de dimension puede clasificarse en dos categorías principales. Una de ellas se basa en extraer características faciales de la imagen como por ejemplo, ojos, nariz, etc. El otro enfoque, denominado holístico, toma toda la región de la cara como dato de entrada

para el reconocimiento. Ejemplos de métodos holísticos son *Eigenfaces* y *Fisherfaces*.

*Eigenfaces* es construido sobre técnicas de Análisis de Componentes Principales (PCA) introducido por M.A. Turk and A.P. Pentland [11]. PCA consigue una reducción de la dimensión proyectando los datos sobre la dirección que maximice la distribución total sobre todas las clases. *Fisherfaces* es un método propuesto por Belhumeur et al [2] en el que se utiliza PCA y el Discriminante Lineal de Fisher (FLD) o LDA. Este método maximiza la relación entre la distribución entre clases y la distribución intra-clases.

Existe diversa bibliografía en la que se evalúa alguno de estos métodos de forma independiente, se los compara entre sí o con otros métodos. Algunos de esos artículos han publicado resultados que varían significativamente o son contradictorios. Por ejemplo los trabajos [5] [3] concluyen que los algoritmos basados en PCA superan a los basados en LDA bajo variaciones de iluminación. El artículo [1] afirma que PCA se desempeña mejor que LDA cuando el conjunto de entrenamiento es chico. En cambio los artículos [7] y [2] exponen según sus resultados, que LDA tiene un desempeño mejor que PCA. A su vez [9] concluye que LDA supera a PCA tanto en diferentes condiciones de iluminación como de pose.

En las secciones siguientes se realiza una descripción de los métodos y se los compara empíricamente bajo diferentes condiciones de iluminación y de pose.

## 2. Descripción de los métodos

El proceso de reconocimiento de caras consiste en tomar una imagen de dos dimensiones,  $a$  filas y  $b$  columnas, a la que se transforma en un vector unitario contenido en un espacio de imágenes  $n$ -dimensional ( $n = a \times b$ ). Luego se le subtrae la imagen promedio y se proyecta el vector resultante en un subespacio de menor dimensión utilizando uno de los métodos de reducción de dimension (extracción de características). Esta proyección es comparada con la proyección de un conjunto de imágenes de una base. La clase del vector más similar, utilizando algún criterio de similitud, es el resultado del proceso de reconocimiento.

### 2.1. Eigenfaces

Este método realiza una proyección lineal del espacio de imágenes a un espacio de características de menor dimensión. Esta reducción se realiza utilizando la técnica PCA la cual toma aquella proyección lineal que maximiza la dispersión de todas las imágenes proyectadas.

En primer lugar se considera un conjunto de  $N$  imágenes con valores en el espacio de imágenes  $n$ -dimensional

$$\{x_i\} \quad i = 1, 2, \dots, N \quad (1)$$

Se asume además que cada una de las imágenes pertenece a una de las  $c$  clases  $\{X_1, X_2, \dots, X_c\}$ . Asimismo se considera una transformación lineal que lleva el espacio de imágenes original de  $n$  dimensiones al espacio de características de dimensión  $m$ , donde  $m < n$ . Los nuevos vectores de características  $y_k \in \mathfrak{R}^m$  son definidos por la siguiente transformación lineal

$$y_k = W^T x_k \quad k = 1, 2, \dots, N \quad (2)$$

donde  $W \in \mathfrak{R}^{n \times m}$  es una matriz con columnas ortonormales. Se define además la matriz de distribución total  $S_T$  como

$$S_T = \sum_{k=1}^N (x_k - \mu)(x_k - \mu)^T \quad (3)$$

donde  $\mu \in \mathfrak{R}^n$  es la media de todas las imágenes de (1). Luego de aplicar la transformación lineal  $W^T$ , la distribución de los vectores de características  $\{y_1, y_2, \dots, y_N\}$  es  $W^T S_T W$ . Se toma aquella proyección  $W_{opt}$  que maximiza el determinante de la distribución total de la matriz de las imágenes proyectadas, esto es

$$\begin{aligned} W_{opt} &= \arg \max_W |W^T S_T W| \quad (4) \\ &= [w_1 \ w_2 \ \dots \ w_m] \end{aligned}$$

donde  $\{w_i \mid i = 1, 2, \dots, m\}$  es el conjunto de vectores propios  $n$ -dimensionales de  $S_T$  correspondiente a los mayores  $m$  vectores propios. Dichos vectores propios tienen la misma dimensión que las imágenes originales y se les denomina eigenfaces. En la *Figura 1* se muestra las primeras siete eigenfaces obtenidas con un conjunto de imágenes de la base *Extended Yale Face Database B*.

## 2.2. Fisherfaces

El método *Fisherfaces* [2] utiliza el Discriminante Lineal de Fisher (FLD) para la reducción de dimensión. Este método selecciona el  $W$  de la ecuación (1) de manera que el cociente entre la distribución entre clases y la distribución intra-clases sea máxima. Para esto se define la matriz  $S_B$  de distribución entre clases como

$$S_B = \sum_{i=1}^c N_i (\mu_i - \mu)(\mu_i - \mu)^T \quad (5)$$

y la matriz  $S_W$  de distribución intra-clases

$$S_W = \sum_{i=1}^c \sum_{x_k \in X_i} N_i (x_k - \mu_i)(x_k - \mu_i)^T \quad (6)$$

donde  $\mu_i$  es la imagen media de la clase  $X_i$ , y  $N_i$  es el número de imágenes en la clase  $X_i$ . Si la matriz  $S_W$  es no singular, la proyección  $W_{opt}$  se elige como la matriz con columnas ortonormales que maximiza el cociente del determinante de la matriz de distribución entre clases de las imágenes proyectadas y el determinante de la matriz de la distribución intra-clases de las imágenes proyectadas, esto es

$$\begin{aligned} W_{opt} &= \arg \max_W \frac{|W^T S_B W|}{|W^T S_W W|} \\ &= [w_1 \ w_2 \ \dots \ w_m] \end{aligned} \quad (7)$$

donde  $\{w_i \mid i = 1, 2, \dots, m\}$  es el conjunto de valores propios de  $S_B$  y  $S_W$  correspondiente a los  $m$  mayores valores propios  $\{\lambda_i \mid i = 1, 2, \dots, m\}$ , esto es

$$S_B w_i = \lambda_i S_W w_i \quad i = 1, 2, \dots, m. \quad (8)$$

Se observa entonces, que a lo sumo se tienen  $c - 1$  valores propios distintos de cero, y por lo tanto el límite superior de  $m$  es  $c - 1$ , donde  $c$  es el número de clases. Para el problema de reconocimiento de caras, se tiene que la matriz  $S_W \in \mathbb{R}^{n \times n}$  es siempre singular, dado que el rango de  $S_W$  es a lo sumo  $N - c$ , y en general, el número de imágenes de entrenamiento:  $N$ , es mucho más chico que el número de píxeles de cada imagen:  $n$ . Por lo tanto puede ser posible elegir una matriz  $W$  tal que la distribución intra-clases de las imágenes proyectadas pueda ser exactamente cero. Como alternativa entonces, al criterio establecido en la ecuación (7), se proyecta el conjunto de imágenes a un espacio de menor dimensión, de manera que la matriz resultante de la distribución intra-clases  $S_W$  es no singular. Utilizando PCA se realiza la reducción de dimensiones del espacio de características a  $N - c$  y luego, aplicar FLD definido en (7) para reducir la dimensión a  $c - 1$ . De esta manera  $W_{opt}$  es dado por

$$W_{pca}^T = W_{fld}^T W_{pca}^T \quad (9)$$

donde

$$\begin{aligned} W_{pca} &= \arg \max_W |W^T S_T W| \\ W_{fld} &= \arg \max_W \frac{|W^T W_{pca}^T S_B W_{pca} W|}{|W^T W_{pca}^T S_W W_{pca} W|} \end{aligned}$$

Las columnas de esta matriz se les refiere como fisherfaces. La *Figura 1* muestra las primeras siete fisherfaces obtenidas con un conjunto de imágenes de la base *Extended Yale Face Database B*.



(a)



(b)

Figura 1: Primeras siete Eigenfaces (a) y Fisherfaces (b), obtenidas con un conjunto de imágenes de la base *Extended Yale Face Database B*.

En la *Figura 2* se ilustra la ventaja de FLD sobre PCA para el caso de un problema de clasificación en el que las muestras de cada clase yacen próximas a un subespacio lineal. El problema consiste de dos clases en el cual las muestras de cada clase han sido ligeramente desplazadas en una dirección perpendicular al subespacio lineal, por lo que cada clase yace próxima a una línea que pasa por el origen en el espacio de características de dos dimensiones. Ambos métodos, PCA y FLD, han sido utilizados para proyectar las muestras en un espacio de una dimensión. Se observa que PCA proyecta las muestras de manera que las clases no puedan ser linealmente separables. Por lo tanto FLD consigue una mayor dispersión entre clases.

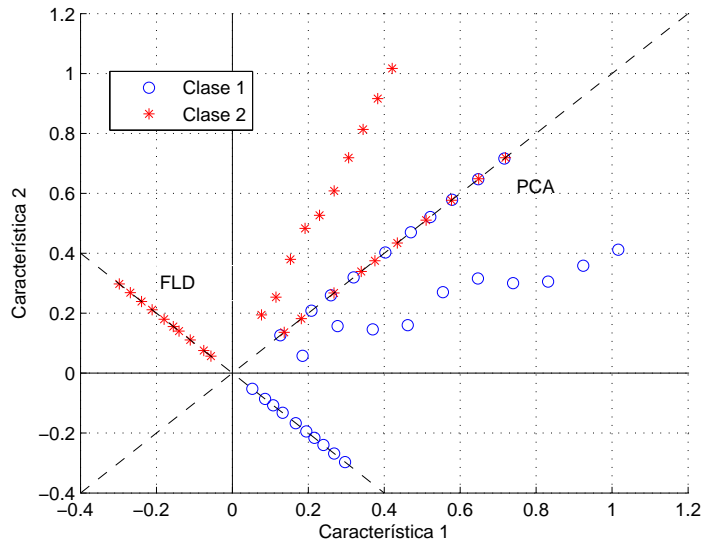


Figura 2: Comparación de PCA y FLD para un problema de dos clases.

### 3. Experimentos

En esta sección se describen los experimentos realizados para comparar el desempeño de los métodos de reconocimiento de caras. Para esto, se han utilizado las bases de imágenes: *Extended Yale Face Database B* [4] y la *Sheffield Face Database* [10]. La primera de las bases es utilizada para evaluar los algoritmos frente a variaciones de iluminación, mientras que la segunda se utiliza para evaluar el desempeño frente a variaciones de pose.

La base de imágenes *Extended Yale B* contiene 2414 imágenes de 38 individuos con pose frontal fija. Por cada individuo se tienen 64 imágenes, cada una capturada en condiciones distintas de iluminación. Estas condiciones se obtienen variando la posición de una única fuente de luz. Para este trabajo se utiliza la base de imágenes recortadas [6], de manera que el fondo de las imágenes no es visible. La *Figura 3* muestra algunas imágenes de un individuo de la base descrita.

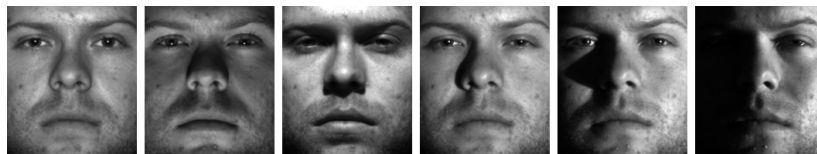


Figura 3: Ejemplo de imágenes de un individuo de la base *Extended Yale B*.

La base de datos *Sheffield Face Database* contiene 564 imágenes de 20 individuos en distintas poses, cubriendo un rango desde una posición frontal a una posición de perfil. Todas estas imágenes son capturadas bajo idénticas condiciones de iluminación. La *Figura 4* muestra algunas imágenes de un individuo de la base.



Figura 4: Ejemplo de imágenes de un individuo de la base *Sheffield Face Database*.

Todas las imágenes utilizadas están en escala de grises, a las que además se les realza el contraste, tal que sus histogramas sean uniformes, y se las transforma a un tamaño de 50 x 50 píxeles.

Para todas las experiencias, además de comparar los métodos descritos, también se evalúa el desempeño de *Eigenfaces* descartando los primeros tres componentes principales, i.e. se considera el mismo número de características para cada experiencia en particular pero tomando los componentes principales a partir del cuarto componente. Existe bibliografía [8] que indica que los primeros tres componentes principales corresponden esencialmente a variaciones de iluminación, y por lo tanto eliminando dichos componentes se incrementa la tasa de reconocimiento. De aquí en adelante, a este método se lo referencia como *Eigenfaces s/3*.

La clasificación de las imágenes se realiza mediante la técnica de vecino más cercano, utilizando la distancia euclidiana.

### 3.1. Iluminación

Se realizaron dos experiencias para evaluar los métodos respecto a las variaciones de iluminación. En la primera de las experiencias se compara el desempeño de los algoritmos en función de la dimensión del número de características. Luego se evalúa el desempeño variando el parámetro  $k$  de la técnica de vecino más cercano. En la segunda de las experiencias se compara el desempeño en función del ángulo de la fuente de luz respecto a la posición frontal, considerando que las imágenes presentan peor iluminación cuanto mayor sea este ángulo.

#### 3.1.1.

El procedimiento seguido en la primera de las experiencias consiste en tomar de forma aleatoria el 50 % de las imágenes por individuo de la base *Extended Yale B*,

(aproximadamente 32 imágenes), como conjunto de entrenamiento, y las restantes imágenes como conjunto de test. Se evalúa la tasa de reconocimiento obtenida para distintos valores de  $m$  (número de características), tomando como resultado el promedio de reiterar este procedimiento 5 veces de manera de asegurar que el resultado obtenido no responda a una elección particular de los conjuntos de entrenamiento y test. Los resultados se muestran en la *Figura 5*.

Se observa en la figura que *Fisherfaces* es superior a los otros dos métodos, haciéndose esta diferencia más evidente cuanto menor sea el número de características que se consideren. Con *Fisherfaces*, por ejemplo, utilizando 10 características se obtiene una tasa de reconocimiento mayor al 95%, superior al máximo reconocimiento obtenido por los otros dos métodos aún empleando en estos 35 características. Se comprueba además que la eliminación de los primeros tres componentes principales, *Eigenfaces s/3* resulta en un desempeño muy superior consiguiendo casi un 40% más que *Eigenfaces* cuando se utilizan 5 características y alcanzando un desempeño superior al 95% cuando se utilizan 35 características.

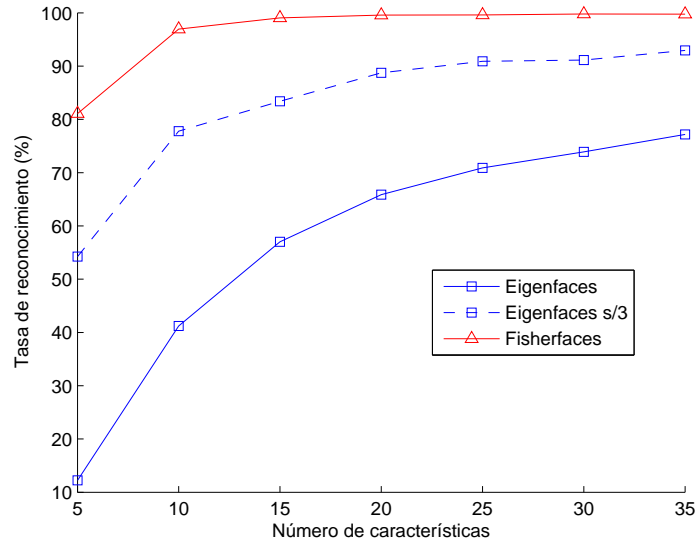


Figura 5: Comparación del desempeño de los algoritmos en función del número de características sobre las imágenes de la base *Extended Yale B*.

Luego de analizar el desempeño de los algoritmos en función del número de características, se estudia el desempeño respecto al parámetro  $k$  de la técnica de clasificación de vecino más cercano. Se analiza tomando los valores de  $k = 1, 3, 5$ . En este caso se sigue los mismos criterios que para el análisis anterior, esto es, se toma aleatoriamente el 50% de la base para obtener los conjuntos de entrenamiento y test, y el resultado es el promedio de reiterar el procedimiento 5 veces. La *Figura 6* y la *Figura 7* muestran los resultados para un número de características  $m = 10$  y  $m = 30$  respectivamente.



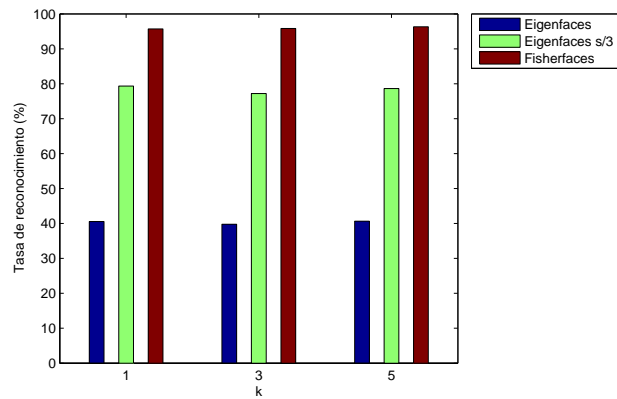


Figura 6: Comparación de los algoritmos para los valores de  $k = 1, 3, 5$  de la técnica de vecino más cercano para  $m = 10$ .

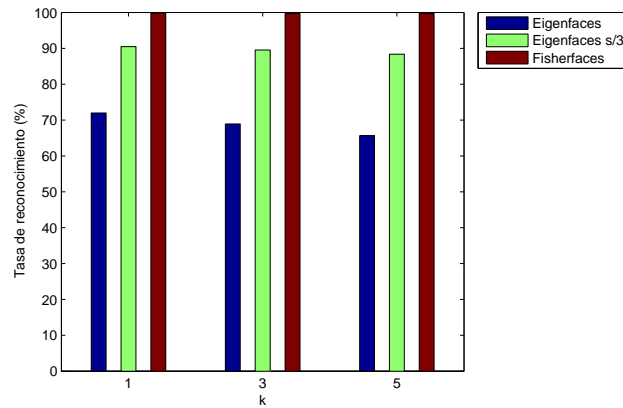


Figura 7: Comparación de los algoritmos para los valores de  $k = 1, 3, 5$  de la técnica de vecino más cercano para  $m = 30$ .

Se observa de las figuras que para  $m = 10$  los algoritmos no presentan cambios significativos en su desempeño al variar el parámetro  $k$ . Para  $m = 30$  el mayor cambio se produce con *Eigenfaces* que presenta un desempeño de un 6% inferior con  $k = 3$  que con  $k = 1$ . Para los otros algoritmos no se observan cambios destacables.

### 3.1.2.

En la segunda de las experiencias se agrupan las imágenes según el ángulo que forma la dirección de la fuente de luz, utilizada en cada imagen, respecto a la posición en la que se ubica la cámara, según se muestra en la *Figura 8*.

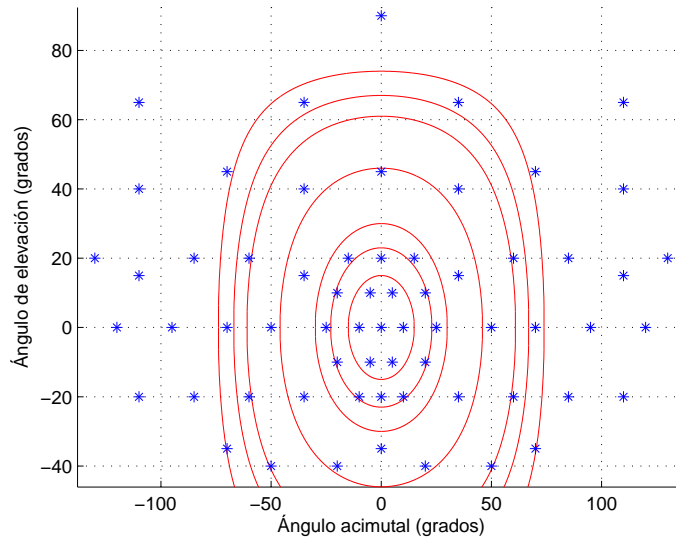


Figura 8: Posición, en coordenadas esféricas, de la fuente de luz correspondiente a cada una de las 64 imágenes por individuo de la base *Extended Yale B*.

En la figura se observa la posición de la fuente de luz utilizada en cada una de las 64 imágenes por individuo. El punto central  $(0^\circ, 0^\circ)$ , es la posición en la que se ubica la cámara. Las imágenes contenidas en cada uno de los siete anillos forman los conjuntos en los que son agrupadas las imágenes. De esta manera, los anillos más externos contienen imágenes capturadas bajo peores condiciones de iluminación. El anillo que contiene al punto  $(0^\circ, 0^\circ)$ , se lo referencia como anillo 1 y los anillos sucesivos son los anillos 2 al 7.

Las imágenes contenidas en el anillo 1, son utilizadas como imágenes de entrenamiento, mientras que las imágenes contenidas en los anillos sucesivos son utilizadas como imágenes de test. Para cada uno de estos anillos (2 al 7), se evalúa la tasa de reconocimiento considerando  $m = 30$ . Los resultados se muestran en la *Figura 9*.

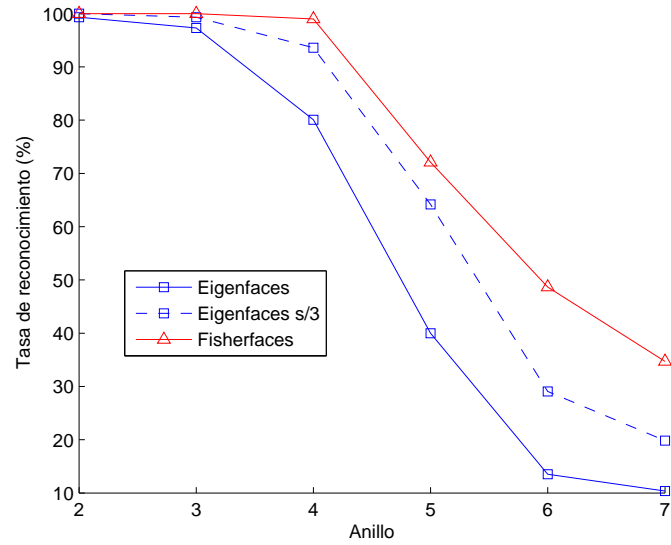


Figura 9: Comparación del desempeño de los algoritmos considerando como imágenes de test, aquellas contenidas en los anillos 2 al 7, y como imágenes de entrenamiento las imágenes del anillo 1.

Si bien todos los algoritmos tienen un buen desempeño para las imágenes de los anillos más centrales, *Fisherfaces* es superior a los otros métodos, especialmente cuando las condiciones de iluminación son peores, es decir para las imágenes contenidas en los anillos más externos. Se comprueba en esta experiencia también que *Eigenfaces s/3* tiene un desempeño superior a *Eigenfaces*.

### 3.2. Pose

Para evaluar el desempeño de los algoritmos utilizando imágenes de caras en distintas poses, se realiza una experiencia en la que se evalúa la tasa de reconocimiento para distintos valores de  $m$  (número de características), utilizando validación cruzada de 10 particiones sobre las imágenes de la base *Sheffield Face Database*. Los resultados se muestran en la *Figura 10*.

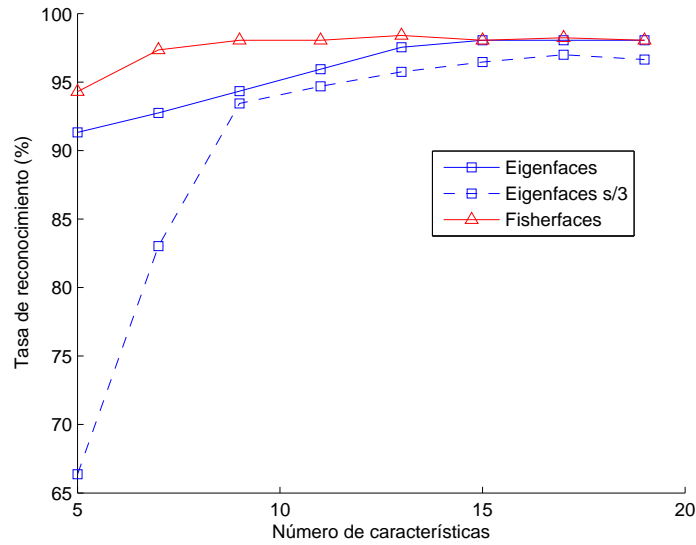


Figura 10: Comparación del desempeño de los algoritmos en función del número de características sobre las imágenes de la base *Sheffield Face Database*.

Se observa en la figura que el método *Fisherfaces* es superior al resto, aunque tanto este método como *Eigenfaces* obtienen un buen desempeño, superior al 90%. Se observa además, que cuanto mayor es el número de características que se consideran, el desempeño de los algoritmos es similar. Para esta experiencia, el método *Eigenfaces s/3* tiene un desempeño inferior a *Eigenfaces*, especialmente cuando se toma un número de características menor a diez.

Luego de analizar el desempeño de los algoritmos en función del número de características, se estudia el desempeño respecto al parámetro  $k$  de la técnica de clasificación de vecino más cercano. Se analiza tomando los valores de  $k = 1, 3, 5$ . En este caso se sigue los mismos criterios que para el análisis anterior, utilizando validación cruzada de 10 particiones. La *Figura 11* y la *Figura 12* muestran los resultados para un número de características  $m = 5$  y  $m = 15$  respectivamente.

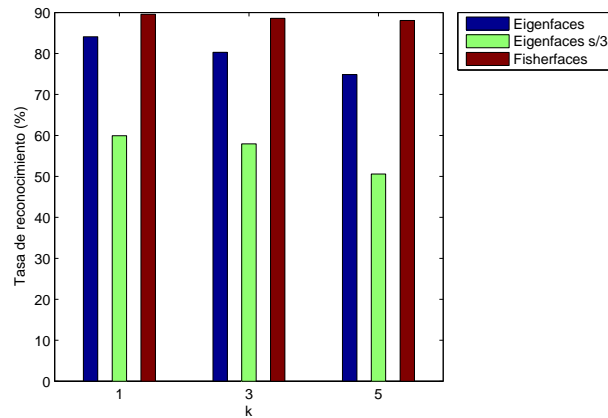


Figura 11: Comparación de los algoritmos para los valores de  $k = 1, 3, 5$  de la técnica de vecino más cercano para  $m = 5$ .

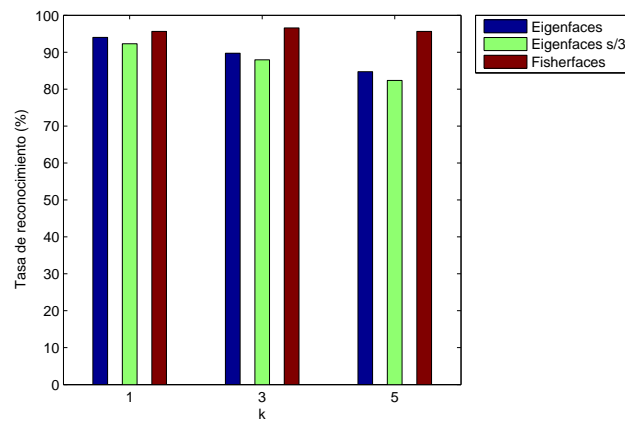


Figura 12: Comparación de los algoritmos para los valores de  $k = 1, 3, 5$  de la técnica de vecino más cercano para  $m = 15$ .

Se observa de las figuras que para valores mayores de  $k$  el desempeño de los algoritmos decrece. Para *Fisherfaces* este cambio no es significativo. Sin embargo tanto para *Eigenfaces* como para *Eigenfaces s/3* el desempeño decrece hasta cerca de 10%.

## 4. Conclusiones

En este trabajo se han comparado dos métodos tradicionales de reconocimiento de caras bajo diferentes condiciones de iluminación y de pose, utilizando para esto las bases de imágenes *Extended Yale Database B* y la *Sheffield Face Database* respectivamente. Las conclusiones generales son:

- *Fisherfaces* es el método que demuestra un desempeño superior tanto para imágenes con variaciones de iluminación como de pose. Estos resultados son coherentes con [9] y con los trabajos [7] y [2] para el problema particular de la iluminación.
- Para el caso de la variación de iluminación esta superioridad se acentúa cuando las condiciones de iluminación son peores.
- El empleo de valores mayores del parámetro  $k$  del clasificador de vecino más cercano ( $k = 3, 5$ ) provoca un desempeño de los algoritmos similar al obtenido con  $k = 1$  o de hasta un 10% inferior.
- La eliminación los primeros tres componentes para *Eigenfaces* resulta en un desempeño superior con imágenes con pose frontal fija y bajo diferentes condiciones de iluminación, verificando los resultados de [11]. Sin embargo para el caso de imágenes en las que varía la pose, el desempeño es inferior.

Si bien, para estos experimentos se ha observado la superioridad de *Fisherfaces*, dicho método resulta computacionalmente más costoso. Por lo tanto la preferencia de uno u otro método depende específicamente de la aplicación en que se utilice los métodos.

## Bibliografía

- [1] Avinash C. Kak Aleix M. Martinez. Pca versus lda. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 2001.
- [2] Peter N. Belhumeur, Joao P. Hespanha, and David J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. In Bernard F. Buxton and Roberto Cipolla, editors, *ECCV (1)*, volume 1064 of *Lecture Notes in Computer Science*, pages 45–58. Springer, 1996.
- [3] Hyunjong Cho and Seungbin Moon. Comparison of pca and lda based face recognition algorithms under illumination variations. 2009.
- [4] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 23(6):643–660, 2001.
- [5] Sonja Grgic Kresimir Delac, Mislav Grgic. A comparative study of pca, ica and lda.
- [6] K.C. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 27(5):684–698, 2005.
- [7] J. Ruiz-del-Solar P. Navarrete. Analysis and comparison of eigenspace-based face recognition approaches. *International Journal of Pattern Recognition and Artificial Intelligence*, 2002.
- [8] Alex Pentland, N. Etcoff, A. Masoiu, O. Oliyide, T. Starner, and M. Turk. Experiments with eigenfaces. In *Looking At People Workshop, IJCAI '93*.

- [9] P.E. Robinson. Comparison of principal component analysis and linear discriminant analysis for face recognition. 2007.
- [10] The University of Sheffield. The Sheffield (previously UMIST) Face Database. <http://www.sheffield.ac.uk/eee/research/iel/research/face>.
- [11] M. Turk and A. Pentland. Eigenfaces for recognition. *J. Cognitive Neuroscience*, 1991.