



**Universidad**  
Zaragoza

# Trabajo Fin de Grado

Sistema interactivo de identificación de usuarios  
mediante reconocimiento facial

Interactive face recognition system

Autora

Cristina Pascual Navarro

Directora

Ana Cristina Murillo Arnal

Escuela de Ingeniería y Arquitectura  
2016





Escuela de  
Ingeniería y Arquitectura  
Universidad Zaragoza

## DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD

(Este documento debe acompañar al Trabajo Fin de Grado (TFG)/Trabajo Fin de Máster (TFM) cuando sea depositado para su evaluación).

TRABAJOS DE FIN DE GRADO / FIN DE MÁSTER

D./D<sup>ña</sup>. Cristina Pascual Navarro

con nº de DNI 25209450 R en aplicación de lo dispuesto en el art.

14 (Derechos de autor) del Acuerdo de 11 de septiembre de 2014, del Consejo de Gobierno, por el que se aprueba el Reglamento de los TFG y TFM de la Universidad de Zaragoza,

Declaro que el presente Trabajo de Fin de (Grado/Máster) Ingeniería Electrónica y Automática, (Título del Trabajo) Sistema interactivo de identificación de usuarios mediante reconocimiento facial

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

es de mi autoría y es original, no habiéndose utilizado fuente sin ser citada debidamente.

Zaragoza, 24 Junio 2016

Fdo: \_\_\_\_\_

## Agradecimientos

Este Trabajo Fin de Grado no podría haber sido posible sin el apoyo de Ana Cristina Murillo, directora del proyecto, que ha ayudado en todo momento a resolver pacientemente las dudas que pudieran surgir. Tu alegría ha hecho que este trabajo sea mucho más sencillo. También quiero agradecer a ETOPIA el equipamiento proporcionado para desarrollar este trabajo.

A toda mi familia: a mi hermana Begoña, has sido el principal apoyo en los años de estudio y en este tramo final también hemos permanecido unidas. Gracias por ayudarme en todo. A mi madre por darme la paciencia y a mi padre por darme la perseverancia y la fuerza para seguir. Sin vuestra atención y preocupación no hubiera sido posible.

A Daniel, por soportarme, por el cariño y por tus continuas visitas al laboratorio durante todos estos meses. Tus consejos y optimismo hicieron este camino mucho menos difícil.

A todas mis amistades, especialmente a Sofía por todos estos años juntas, por tu ayuda y consejos. A Elena, Irene y Marta, porque juntas hemos formado un equipo magnífico de amigas, que no compañeras. Espero que la amistad que aquí ha surgido dure toda la vida.

Gracias a todos los que me han mostrado su apoyo de principio a fin.

# RESUMEN

## SISTEMA INTERACTIVO DE IDENTIFICACIÓN DE USUARIOS MEDIANTE RECONOCIMIENTO FACIAL

El reconocimiento e identificación de una cara es una parte importante de la capacidad del sistema de percepción humano y una tarea rutinaria para las personas. Sin embargo, la construcción de un sistema informático capaz de realizar este trabajo continúa siendo un área de investigación muy activa, aunque también se trata de una de las que más avances han logrado. Por ello y por la gran cantidad de aplicaciones que tiene esta tarea se trata de uno de los campos más conocidos de la visión por computador.

Como punto de partida se ha realizado un estudio y evaluación de las técnicas de reconocimiento facial existentes seleccionando aquella más apropiada para el sistema interactivo que se propone. Se han diseñado e implementado las etapas correspondientes al sistema con el objetivo de comparar aquellos aspectos relevantes con la precisión de la predicción y la capacidad de hacerlo un sistema realista.

Como existen muchísimas técnicas de reconocimiento facial, en este trabajo se centra la atención en las técnicas que ofrecen las librerías del entorno escogido de OpenCV / C++. Su rendimiento fue evaluado con diversas bases de datos de libre acceso a la comunidad científica. Posteriormente se elaboró una base de datos propia la cual fue empleada para la adaptación del diseño e implementación para su funcionamiento en el sistema real.

En la parte experimental se han llevado a cabo ensayos diferenciados que permiten valorar cada etapa de la que el sistema consta de modo que se puedan evaluar de forma detallada las ventajas y desventajas. Los experimentos constan de un proceso de varias etapas: la detección del rostro en una imagen, la extracción de las características y la identificación/verificación de la cara mediante la clasificación de las características. El objetivo de la evaluación se centra en la extracción de las características y el reconocimiento de personas para lo que se han evaluado los tiempos de procesamiento y las tasas de error respectivamente.

Los resultados obtenidos basados en metodología tradicional ofrecieron buenos resultados dentro de las propias limitaciones que presentan. Con el objetivo de estudiar y evaluar metodologías más actuales basadas en algoritmos más robustos se tomó la determinación de usar alguna de las técnicas más novedosas y actuales basadas en redes neuronales de convolución

El sistema interactivo es el resultado de estos estudios en el que se propone un sistema. El sistema interactivo diseñado utiliza el método *Local Binary Patterns*, que es el que mejores características presentaba para esta aplicación.

# Índice general

|  |    |
|--|----|
| Capítulo 1. Introducción .....                             | 7  |
| 1.1 Introducción y motivación .....                        | 7  |
| 1.2 Objetivos .....  | 8  |
| 1.3 Distribución temporal de las tareas. ....              | 9  |
| 1.4 Contexto de realización .....                          | 9  |
| 1.5 Contenido de la memoria .....                          | 10 |
| Capítulo 2. Trabajo Relacionado .....                      | 11 |
| 2.1 Detección vs Reconocimiento .....                      | 11 |
| 2.2 Reconocimiento de caras .....                          | 11 |
| Capítulo 3. Descriptores y clasificadores .....            | 14 |
| 3.1 Descripción .....                                      | 14 |
| 3.1.1 <i>Eigenfaces</i> .....                              | 14 |
| 3.1.2 <i>Fisherfaces</i> .....                             | 17 |
| 3.1.3 <i>Local Binary Patterns Histograms</i> .....        | 18 |
| 3.1.4 <i>Deep Learning descriptors</i> .....               | 19 |
| 3.2 Clasificación .....                                    | 20 |
| 3.2.1 K-Vecinos más cercanos .....                         | 20 |
| Capítulo 4. Experimentos sobre reconocimiento facial ..... | 22 |
| 4.1 Bases de datos y configuración experimental .....      | 22 |
| 4.2 Experimentos .....                                     | 24 |
| 4.2.1 Experimento 1 .....                                  | 24 |
| 4.2.2 Experimento 2 .....                                  | 24 |
| 4.2.3 Experimento 3 .....                                  | 25 |
| 4.2.4 Experimento 4 .....                                  | 26 |
| 4.2.5 Análisis de los experimentos .....                   | 27 |
| Capítulo 5. Sistema interactivo .....                      | 28 |
| Capítulo 6. Conclusiones finales .....                     | 30 |
| 6.1 Conclusiones .....                                     | 30 |
| 6.2 Líneas futuras .....                                   | 30 |
| Anexo A .....  | 33 |
| Tutorial 1 .....   | 33 |
| Tutorial 2 .....   | 33 |
| Tutorial 3 .....   | 33 |
| Anexo B .....  | 34 |
| Experimento 1 .....  | 34 |
| Experimento 2 .....  | 35 |
| Experimento 3 .....  | 36 |
| Anexo C .....  | 37 |
| Presentación de la Base de Datos EINA .....                | 37 |

# Capítulo 1. Introducción

## 1.1 Introducción y motivación

La inteligencia artificial es una rama de las ciencias de la computación que tiene como objetivo el desarrollo de máquinas inteligentes. Es decir, crear máquinas capaces de aprender y mostrar comportamientos que hasta el momento solo eran propios de los humanos. Esta ciencia consta de importantes campos de estudio siendo uno de los más destacados la visión por computador.

La visión por computador es la ciencia que ofrece a las máquinas la capacidad inherente a los seres humanos de percibir y entender el entorno que les rodea a través del sentido de la vista. El principal objetivo de la visión por computador es la extracción automática, el análisis y la comprensión de la información útil a partir de imágenes.

En los últimos años la visión por computador ha tenido gran explosión y se ha convertido en uno de los campos más atractivos para el desarrollo de nuevas aplicaciones. Este es el principal objetivo de este proyecto, profundizar en una de las áreas más conocidas del reconocimiento de imágenes/patrones: el reconocimiento facial. El reconocimiento facial hace posible la identificación automática de una persona con una imagen digital mediante el análisis de sus características faciales extraídas de la propia imagen y comparándolas con las de una base de datos.



Figura 1.-El reconocimiento facial en la Visión por Computador

Las aplicaciones reales que tienen los sistemas de reconocimiento facial nos acompañan en nuestra vida diaria como las mostradas en la Tabla 1: en el campo de la seguridad destaca el seguimiento de individuos; otro campo es el de la seguridad vial donde encontramos trabajos hacia un coche “inteligente”. Así, la compañía Toyota está desarrollando detectores del sueño para aumentar la seguridad al volante, detectando la cara del conductor y analizando las expresiones del mismo. También, el área del entretenimiento ha empezado a interesarse por la visión artificial y se han integrado diversas áreas de la visión por computador para desarrollar aplicaciones y videojuegos mucho más interactivos con el usuario. Un buen ejemplo son los productos ofrecidos por *Microsoft Project Natal* o *Play Station Eye Sony* que utilizan el reconocimiento de rostros. La idea de la detección y el análisis de personas también se ha trasladado a las redes sociales como *Facebook*, que identifica a los usuarios en las fotografías almacenadas. En la Figura 2 podemos observar algunos de los ejemplos aquí nombrados.

| Aplicaciones                  |   |
|-------------------------------|---|
| <b>Seguridad</b>              | Accesos de seguridad (Bases de Datos)<br>Privacidad de datos (expediente médico)<br>Autenticación de usuarios   |
| <b>Gestión de acceso</b>      | Sistemas basados en la gestión de los permisos de acceso<br>Registro de accesos   |
| <b>Biometría</b>              | Identificación personal (Pasaportes, identificación a nivel internacional, registro de votos, licencias de conducir)<br>Automatizado de verificación de identidad (controles fronterizos) |
| <b>Cumplimiento de la ley</b> | Video vigilancia basada en identidad<br>Seguimiento de personas<br>Reconstrucción forense de caras  |
| <b>Personal de seguridad</b>  | Interpretación de expresiones (sistema de vigilancia en conductores)  |
| <b>Entretenimiento</b>        | Aplicaciones de cámaras de fotos<br>Videojuegos   |

Tabla 1.- Áreas de aplicación

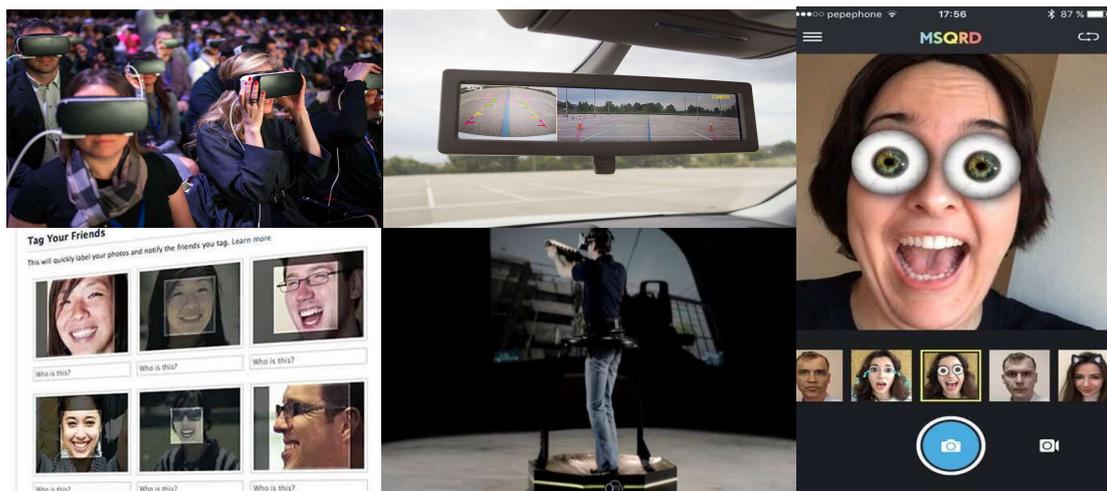


Figura 2.- Áreas de aplicación del reconocimiento facial. De izquierda a derecha y empezando por la hilera superior, se puede observar a personas haciendo uso de gafas de realidad virtual para ocio. En la siguiente, se ve un retrovisor digital en un automóvil que reemplaza a los retrovisores tradicionales. A continuación, aplicación de reconocimiento fácil con fines de entretenimiento. Continuando por la hilera inferior, se observa la identificación de sujetos en redes sociales. Y finalmente, se puede observar a un individuo jugando a un videojuego de realidad aumentada.

Con todo ello queda patente el desarrollo y evolución que ha habido en los últimos años de las nuevas tecnologías sirviéndose de los avances de la visión por computador para mejorar y ofrecer productos innovadores al público.

## 1.2 Objetivos

Los **objetivos** generales de este proyecto son:

-Evaluación de las técnicas de reconocimiento facial automático: *Eigenfaces*, *Fisherfaces*, *Local Binary Patterns Histograms* y reconocimiento basado en Redes Neuronales de convolución.

- Diseñar una aplicación interactiva de reconocimiento de usuarios. Esta aplicación reconoce usuarios que ya están en su base de datos, y si no lo estuvieran, ofrece la opción de incorporarlos de manera incremental.

Para conseguir estos objetivos se han realizado las siguientes **tareas** descritas y detalladas a lo largo de esta memoria:

-Recopilación de las distintas bases de datos de dominio público y la elaboración de una Base de Datos Propia.

-Puesta a punto de los métodos de reconocimiento en el entorno de trabajo.

-Evaluación de dichos métodos con las bases de datos recopiladas. En los experimentos realizados se analizan las ventajas, desventajas y limitaciones de cada algoritmo examinando en detalle el comportamiento de múltiples factores que intervienen en todo reconocimiento visual como la iluminación, la textura, las distancias, las formas...

-En base a las conclusiones extraídas, escoger el más robusto para el sistema interactivo propuesto. Diseñar e implementar las mejoras y/o adaptaciones necesarias para la integración del algoritmo sobre el sistema interactivo. Evaluación del mismo con la base de datos propia. Estudio de las restricciones inherentes a un sistema realista.

### 1.3 Distribución temporal de las tareas.

Las tareas a realizar para conseguir los objetivos fijados se indican de manera aproximada en el siguiente cronograma:

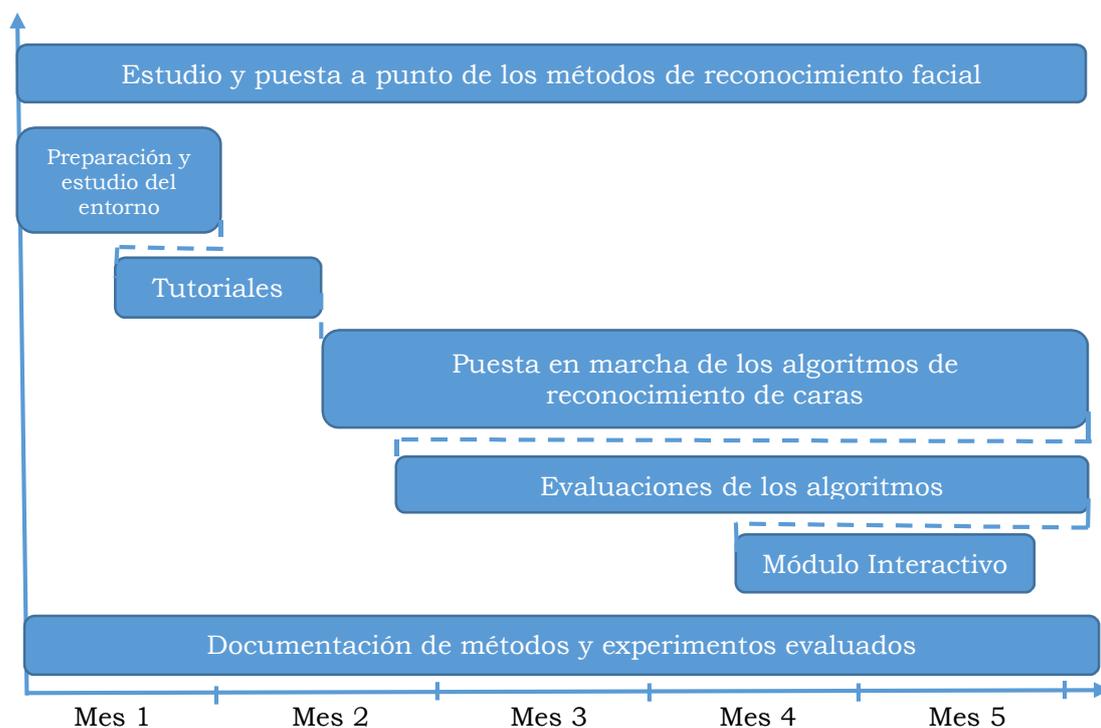


Figura 3.- Cronograma de fases de realización

### 1.4 Contexto de realización

Este trabajo se realiza dentro de las líneas de investigación de reconocimiento visual dentro del grupo de Robótica, Percepción y Tiempo Real, del Instituto de Investigación en Ingeniería en Aragón (i3A). Este proyecto ha usado en parte equipamiento de Unizar del proyecto CESAR (Centro de Supercomputación de Aragón).

El entorno de trabajo en el que se ha desarrollado este proyecto consiste en el sistema operativo Ubuntu versión 14.04, las librerías de OpenCV<sup>1</sup> y el entorno de programación de Eclipse.

<sup>1</sup> <http://opencv.org/>

Dado que el entorno de trabajo era nuevo, cabe mencionar que a principio de este proyecto se estudiaron una serie de tutoriales relacionados con la temática del trabajo para ayudar a la familiarización con el sistema y el entorno elegidos. Dichos tutoriales se describen con más detalle en el Anexo A.

## 1.5 Contenido de la memoria

Además del presente capítulo introductorio, el contenido de la memoria se ha estructurado en los siguientes capítulos.

En el Capítulo 2, se presentan los diferentes estudios relacionados con el reconocimiento facial. Además, se introducen los métodos de reconocimiento facial más utilizados hasta ahora y los emergentes.

En el Capítulo 3 se describen en detalle la base de cada uno de los métodos que se han implementado en este trabajo.

En el Capítulo 4 se exponen los resultados experimentales obtenidos con los diferentes métodos y se escoge el más robusto de ellos para desarrollar el sistema interactivo.

En el Capítulo 5 se presenta y evalúa el sistema interactivo.

En el Capítulo 6 se resumen las conclusiones extraídas de este proyecto y se describen las líneas de trabajo futuro.

# Capítulo 2. Trabajo Relacionado

## 2.1 Detección vs Reconocimiento

A menudo existe confusión entre dos importantes términos que aparecen en el mundo de la visión por computador, y más todavía, dentro del reconocimiento de caras: la detección y el reconocimiento.

Es necesario remarcar sus significados antes de comenzar. La detección facial consiste en dada una determinada imagen, localizar el área de la cara o caras si en esta las hubiera o no, descartando todo lo que sea fondo. Se trata del primer paso frente a un sistema de reconocimiento facial. Sin embargo, el reconocimiento se sirve de una imagen dada en cuyo contenido encontramos una cara que se debe identificar, es decir, indicar a que persona corresponde. Consecuentemente, la detección consiste en una clasificación de dos clases: cara contra no-cara; mientras que el reconocimiento es una clasificación multiclase: una persona contra cualquier otra persona.

Por tanto, la solución al problema que presenta cualquier proceso típico de reconocimiento facial consta por lo general de las siguientes etapas que se muestran en la Figura 4:

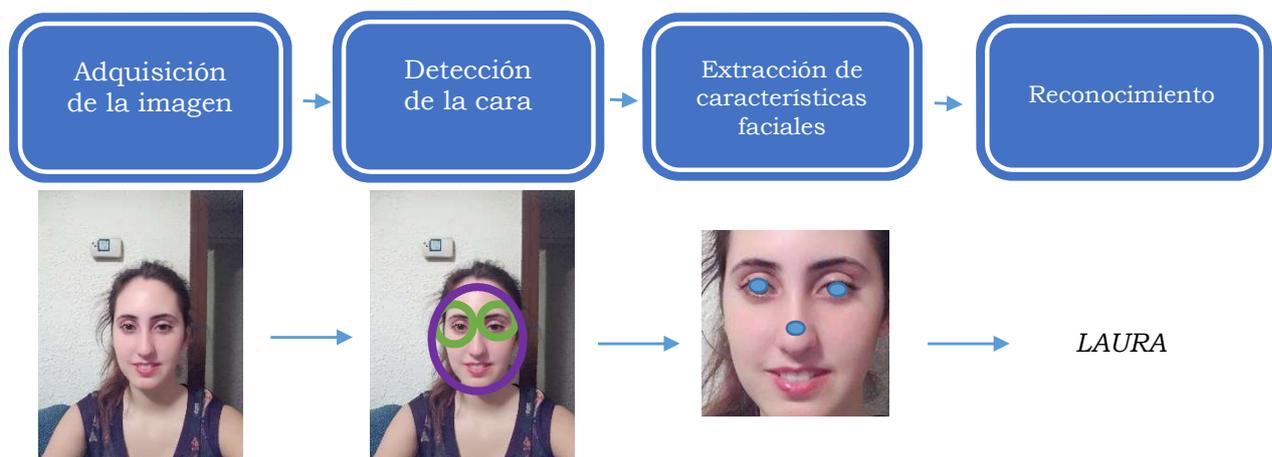


Figura 4.- Etapas de un sistema de reconocimiento facial

## 2.2 Reconocimiento de caras

El reconocimiento facial por computador es un área de investigación que continua en pleno desarrollo. Identificar a una persona por medio de una imagen de su rostro implica emular el proceso cognitivo que realiza un ser humano al reconocer a sus pares.

Su estudio se ha realizado no solo desde la Informática, sino que el interés que suscita ha legado a otras disciplinas científicas como la Psicología, la Medicina Forense y la Antropometría (estudio de las proporciones y las medidas del cuerpo humano.).

Aunque los humanos reconocemos caras familiares con poco o ningún esfuerzo, la automatización de esta tarea para ser llevada a cabo por ordenador es enormemente compleja, por ello muchos estudios han centrado todos sus esfuerzos en intentar alcanzar esta meta. Hay muchas componentes que hacen que una misma cara tenga visualizaciones muy distintas, y que por tanto dificultan su reconocimiento: la posición de la cara, su rotación, la distancia entre la cara y el sensor, la escala, la resolución, el color, el maquillaje y los gestos. Dichas variaciones afectan negativamente a las distintas etapas del reconocimiento automático y por ello las técnicas que han ido apareciendo van proponiendo maneras de obtener algoritmos invariantes o robustos a las mismas.

En psicología, se ha experimentado sobre la metodología que emplea el ser humano para reconocer personas [1][2] y se ha demostrado que en este proceso influyen muchos factores. Los seres humanos utilizamos una amplia colección de estímulos (visuales, auditivos, etc.) recibidos a través de los sentidos, así como la información de contexto, de modo que es bastante complicado con la tecnología actual, aproximarse al modo de reconocer de los seres humanos.

A continuación, se presentan las observaciones y resultados que hasta ahora ha conseguido realizar la psicología sobre el reconocimiento facial:

-El reconocimiento facial se realiza en una parte concreta del cerebro.

-Las caras son más fáciles de recordar por los humanos cuando se encuentran en posición frontal [3].

-La percepción facial es el resultado de un análisis global y también de rasgos locales en los que los rasgos más significativos son el pelo, la línea del borde de la cara, los ojos y la boca en imágenes frontales. Mientras que en perfiles la nariz juega el papel más importante. Por tanto, el punto de vista desde el que se observa la cara es tremendamente importante.

-Los cambios en iluminación hacen muy difícil el reconocimiento facial [4].

Aunque existen multitud de trabajos en otros campos, como la neurociencia, aún no se conoce con certeza cómo funcionan estos procesos internamente en el cerebro humano. Desde la informática los estudios realizados hasta el día de hoy pretenden abordar el problema con objeto de automatizar la tarea de reconocimiento a partir de imágenes digitales utilizando técnicas enmarcadas en el procesamiento de imágenes, la visión por computador y el reconocimiento de patrones.

Los primeros trabajos relacionados con el tema aparecieron en la década de los 60 con el desarrollo de un sistema semi-automático de reconocimiento facial [5]. El sistema requería de un operador humano que localizara y seleccionara las características tales como los ojos, oídos, nariz y boca en cada una de las fotografías para posteriormente un computador proceder al cálculo de distancias y proporciones bajo un punto de referencia. Principalmente, el estudio iba enfocado hacia la Inteligencia Artificial del Departamento de Defensa de EEUU.

Durante los siguientes años los trabajos intentaban buscar la manera de medir automáticamente todas las características. Hasta que en 1973 se desarrolló un sistema de reconocimiento plenamente automático basado en la definición de modelos y patrones geométricos simples.

## Métodos clásicos

No sería hasta el año 1986 cuando apareciera la técnica que se convertiría en las próximas décadas en la técnica dominante: *Eigenfaces* [6] basada en el método estadístico Análisis de componentes principales (PCA). Esta fecha marco un punto de inflexión en el mundo del reconocimiento facial, pues supuso el inicio del esfuerzo incesante por fomentar el desarrollo de algoritmos faciales de reconocimiento y nuevas tecnologías de evaluación de prototipos capaces de enfrentarse a los sistemas de reconocimiento. Es así como en el año 1996 fue presentado uno de los métodos más robustos basado en la técnica *Local Binary Patterns* (LBP) [7] que se describe como un conjunto ordenado de comparaciones binarias de píxeles de intensidad entre el píxel central y sus alrededores. Pocos años después, en 1999 apareció otra potente técnica: *Fisherfaces* [8] basada en otro método estadístico, Análisis discriminante lineal (LDA).

Todos estos métodos basados en la apariencia facial, pese a ser de los primeros en aparecer, son ampliamente utilizados en el campo del reconocimiento y su eficiencia ha sido demostrada en multitud de ocasiones en diversos estudios [9] [10].

Este trabajo se centra en el estudio y evaluación de métodos de reconocimiento facial por lo que se ha decidido comenzar la fase de la experimentación con algoritmos

basados en las técnicas nombradas anteriormente ya que no solo se encuentran disponibles y con libre acceso en las librerías de OpenCV sino por los buenos resultados probados por la comunidad científica [11] [12]. En la siguiente sección se hace un estudio más profundo de cuáles son los fundamentos teóricos y matemáticos sobre los cuales se basan cada una de estas técnicas.

## Métodos recientes

Más recientemente, los mejores resultados en cuanto a reconocimiento facial, se han obtenido con técnicas basadas en Redes Neuronales. El campo de la visión por computador, ha sido testigo de grandes avances en muchas de sus aplicaciones como el reconocimiento de voz, el reconocimiento visual de objetos, detección de objetos... gracias a mejoras recientes de sistemas de aprendizaje basados en *Deep Learning* [13].

Recientemente, la comunidad de aprendizaje automático ha publicado nuevos modelos de redes neuronales, redes neuronales de convolución, que han resultado más sencillas de entrenar con grandes cantidades de datos y generalizan mucho mejor que técnicas anteriores. Se ha conseguido utilizar este tipo de redes para obtener grandes mejoras en muchos problemas de visión por computador. El reconocimiento facial ha sido uno de ellas. Encontramos varias publicaciones recientes de investigadores de alto nivel [14] [15] [16] que demuestran muy buenos resultados en bases de datos muy complicadas mejorando en más de un 10% las tasas de acierto obtenidas hasta la fecha.

El último de estos trabajos, tiene disponible el modelo que permite calcular los descriptores que proponen [15] que son los que hemos utilizado para algunos experimentos de este trabajo.

# Capítulo 3. Descriptores y clasificadores

Como en cualquier tarea de reconocimiento visual, resulta imprescindible la obtención de descriptores de la imagen que proporcionen la información que pueda resultar más útil para la tarea de reconocimiento. El objetivo de un descriptor es extraer la información y características de la imagen, motivo por el cual también es llamado vector de características. Un buen descriptor facial debería presentar invariancia a la pose, a la iluminación, a la expresión y a la calidad de la imagen. Además, sería preferible descriptores compactos, que permitan cálculos eficientes. El tiempo de cómputo para obtenerlo también es uno de los factores importantes. Por ello, para la obtención de los descriptores hemos utilizado algunos métodos que mejor representación facial proporcionan: *Eigenfaces*, *Fisherfaces*, *LBP* y Descriptor basado en *Deep Learning*.



Figura 5. Esquema para clasificación con imágenes de entrenamiento y testeo

Una vez obtenidos los descriptores tendremos que analizarlos y clasificarlos. En general, un sistema de clasificación necesita dos partes: Entrenamiento y Testeo. En la Figura 5, se explica de manera visual en qué consisten los pasos previos a la tarea de clasificación. En el entrenamiento utilizaremos un conjunto de imágenes cuyo número quedará determinado por el tipo de base de datos. De todos los clasificadores existentes nos centramos en uno de los más típicos y utilizados en la literatura: Vecino más próximo (*Nearest Neighbor*).

En este apartado se exponen las bases de los cuatro descriptores y el clasificador que se han utilizado para este trabajo:

## 3.1 Descripción

### 3.1.1 *Eigenfaces*

El método de descripción de caras utilizando *Eigenfaces* [6] está basado en la técnica estadística de componentes principales (PCA) Este modelo estadístico pone fin a un importante problema dimensional reduciendo el tamaño del vector propio de la matriz de covarianza al número de imágenes del entrenamiento.

Dada la imagen de una cara ésta se encuentra distribuida en este alto espacio dimensional definido por el tamaño de la imagen.

$$I_k = \begin{bmatrix} p_{1,1}^k & \cdots & p_{1,N}^k \\ \vdots & & \vdots \\ p_{N,1}^k & \cdots & p_{N,N}^k \end{bmatrix} \quad 3.1.1.1$$

Siendo  $I_k$  las imágenes para  $k=1,2,\dots$  El vector de características para cada imagen corresponde a:

$$T_k = \begin{bmatrix} p_{1,1}^k \\ \vdots \\ p_{N,1}^k \\ \vdots \\ p_{1,N}^k \\ \vdots \\ p_{N,N}^k \end{bmatrix} \quad 3.1.1.2$$

De tal manera que combinando rasgos comunes podemos obtener una cara media como:

$$\Psi = \frac{1}{M} \sum T_k \quad 3.1.1.3$$

De la cara media con el vector de características de cada imagen podemos obtener:

$$\phi_k = T_k - \Psi \quad 3.1.1.4$$

Por tanto la matriz de covarianza queda definida como:  $\mathcal{C} = \frac{1}{M} \sum \phi_k \phi_k^T = AA^T$  3.1.1.5  
Donde  $A = [\phi_1 \ \phi_2 \ \dots \ \phi_N]$

Por lo que las dimensiones de la matriz son  $N^2 \times N^2$  y hallar los vectores y valores propios (*Eigenfaces*) se convierte en una tarea tediosa.

Este método basado en correlaciones pone fin al problema dimensional tomando  $L = A^T A$  de dimensiones  $N \times N$  y los vectores propios quedan definidos como:

$$v_i \text{ con } i = 1, 2 \dots M \Rightarrow Lv_i = \lambda_i v_i \Rightarrow ALv_i = A\lambda_i v_i \Rightarrow AA^T Av_i = \lambda_i Av_i \Rightarrow \mathcal{C} Av_i = \lambda_i Av_i$$

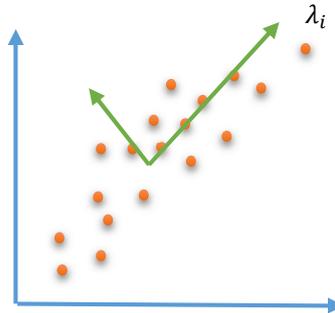


Figura 6.- Representación en el espacio de la dirección principal de un conjunto de muestras

La realidad es que todas las caras presentan básicamente las mismas similitudes unas con otras y contienen los mismos rasgos faciales tales como los ojos, la nariz o la boca que hacen de todas las caras un subconjunto de todo el espacio de la imagen, es decir, las dimensiones de la cara son más pequeñas que las del espacio de la imagen las cuales corresponden a su tamaño.

De esta manera, los *Eigenfaces* son calculados a partir el conjunto de imágenes de entrenamiento. Cada cara podrá ser representada como suma ponderada de cada *Eigenfaces*, los cuales han sido almacenados en un vector de cada uno, es decir, una cara queda definida como una combinación lineal de estos vectores.

De manera simplificada, en el proceso se calcula una cara media en el espacio de las caras para posteriormente computar cada cara como la diferencia con la cara media como se ve en la Figura 7 y Figura 8.



Figura 7. Cara media de la Base de Datos Propia

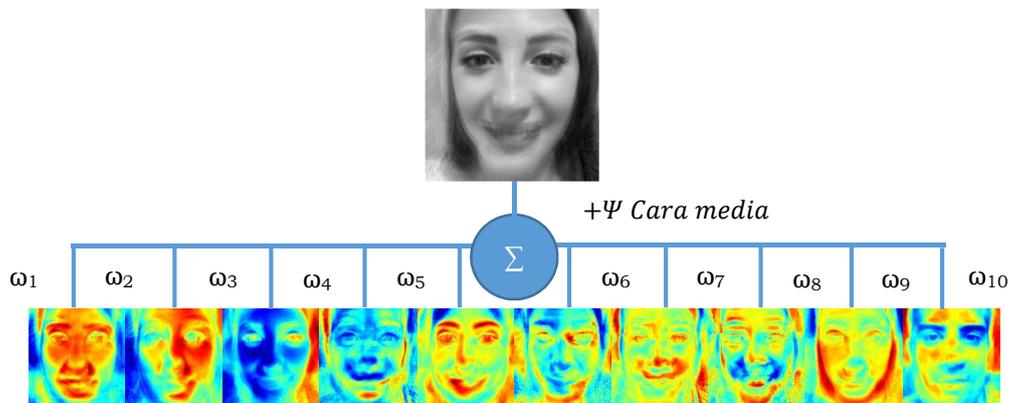


Figura 8.- Diez primeros *Eigenfaces* de la Base de Datos Propia de izquierda a derecha numerados del 1 al 10 contribuyen a la reconstrucción de una imagen dada mediante la suma ponderada de sus pesos.

En los *Eigenfaces* que aparecen en la Figura 8 se ha utilizado un mapa de color para que se pueda observar correctamente como queda distribuida la escala de grises de cada uno. De esta manera, se puede ver que los *Eigenfaces* no solo codifican rasgos faciales, sino que también la iluminación en las imágenes.

Se pueden obtener buenas reconstrucciones de las imágenes usando únicamente algunos *Eigenfaces* como vemos en la siguiente figura:

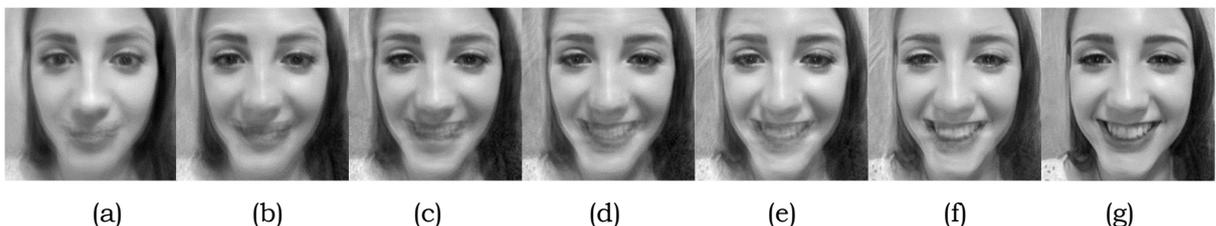


Figura 9.-Reconstrucción de sujeto de la Base de Datos Propia. De izquierda a derecha observamos cómo influye en la reconstrucción el aumento del número de componentes: (a) con 8, (b) con 13, (c) con 18, (d) con 23, (e) con 28, (f) con 33 y (g) con 48.

El número de vectores propios necesarios para una buena reconstrucción depende fuertemente de las imágenes de entrada [17]. Para la base de datos propia hemos podido obtener una reconstrucción con uno número no muy elevado vectores propios tal y como se puede observar en la reconstrucción de la Figura 9.

### 3.1.2 Fisherfaces

El método de descripción de caras utilizado por *Fisherfaces* [8] está basado en el análisis discriminante lineal (LDA). Al igual que el método anterior (3.1.1), reduce la dimensionalidad del espacio de características utilizando la información que extrae de cada clase.

Su objetivo principal es el de maximizar la distancia entre las clases, es decir, entre sujetos, pero al mismo tiempo minimizándola entre los elementos que pertenecen a la misma clase, al mismo sujeto.

Sea un vector aleatorio con muestras extraídas de  $C$  determinadas clases cada una con  $n$  imágenes de muestra:

$$X = \{X_1, X_2, X_3, \dots, X_c\} \quad 3.1.2.1$$

$$X_i = \{x_{i1}, x_{i2}, x_{i3}, \dots, x_{in}\} \quad 3.1.2.2$$

Podemos computar la imagen media de cada clase  $X$  como:

$$\mu_i = \frac{1}{k} \sum x_j \quad 3.1.2.3$$

Donde la cara media  $\mu$  de todas las clases para una base de datos dada puede ser calculada como:

$$\mu = \frac{1}{c} \sum \mu_i \quad 3.1.2.4$$

La matriz de dispersión dentro de una clase queda definida como:

$$S_w = \sum \sum (x_k - \mu_i)(x_k - \mu_i)^T \quad 3.1.2.5$$

Del mismo modo, la matriz de dispersión entre clases se calcula:

$$S_B = \sum \sum N(\mu_i - \mu)(\mu_i - \mu)^T \quad 3.1.2.6$$

Esta técnica busca cual es la dirección dentro del espacio de los descriptores de las caras la en la que la separación entre los sujetos es más alta.

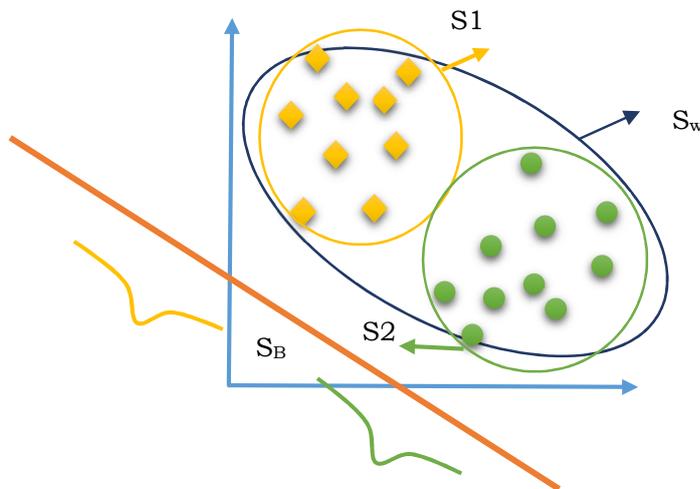


Figura 10- Maximización de la distancia entre clase y minimización dentro de la misma

El método *Fisherfaces* aprende una matriz de transformación para cada clase centrándose únicamente en los rasgos faciales más discriminativos que permitirán identificar a cada sujeto. La iluminación en este método no es capturada por lo que cabe

volver a mencionar que el rendimiento del método volverá a depender fuertemente de los datos de entrada [17].

*Fisherfaces* permite reconstrucción de una imagen proyectando cada imagen en cada uno de los *Fisherfaces* como vemos en la siguiente figura:



Figura 11.- Ejemplo de algunas “Fisherfaces” obtenidas en el conjunto de entrenamiento la Base de Datos Propia



Figura 12.-Reconstrucción sujeto Base de Datos Propia. De izquierda a derecha el número de componentes que intervienen en la reconstrucción va aumento desde cero hasta cinco.

En la Figura 12 se visualiza la influencia del número de *Fisherfaces* que participan en la reconstrucción de la imagen de un sujeto. Como el método está basado en la identificación de sujetos a partir de sus diferencias, no es posible esperar una buena reconstrucción de la imagen original completa.

### 3.1.3 Local Binary Patterns Histograms

Este método consiste en la extracción de características locales a partir de las imágenes. Sigue una metodología basada en Patrones locales binarios (LBP) [7]. Se deshace el concepto que ve la imagen como un vector de alta dimensión para centrarse en las características locales de cada objeto que implícitamente ya tienen baja dimensión. De esta manera, resume la estructura local de la imagen mediante la comparación de cada pixel con sus pixeles vecinos.

El vector de características LBP, de manera simple, es creado de la siguiente manera:

-Divide la ventana examinada en células definidas según la dirección  $x$  y la dirección  $y$ . Cuantas más células contenga la imagen, más fina será la cuadrícula, pero mayor será la dimensionalidad del vector resultante.

-Usando una vecindad circular queda definido el radio y el número de píxeles de la zona. Para un pixel central dado  $(x_c, y_c)$  la posición de un vecino  $(x_p, y_p)$  puede ser calculada como:

$$x_p = x_c + R \cos\left(\frac{2\pi p}{P}\right) \quad 3.1.3.1$$

$$y_p = y_c - R \sin\left(\frac{2\pi p}{P}\right) \quad 3.1.3.2$$

, donde  $R$  es el radio del círculo  $P$  el número de muestras.

Cuanto mayor sea el radio más suave será la imagen de la LBP y detalles más grandes podrán ser capturados, pero menos poder discriminativo tendrá. Por otro lado, cuantos más puntos de muestreo se escojan, más patrones se podrán codificar a costa de aumentar el coste computacional. Por lo tanto, habrá que llegar a un compromiso.

-De manera formal la descripción del operador de LBP puede ser dada como:

$$LBP(x_c, y_c) = \sum 2^P s(i_p - i_c) \quad 3.1.3.3$$

,donde  $i_c$  es la intensidad del pixel central y  $i_p$  la intensidad del pixel vecino.  $s$  es la función signo definida como:

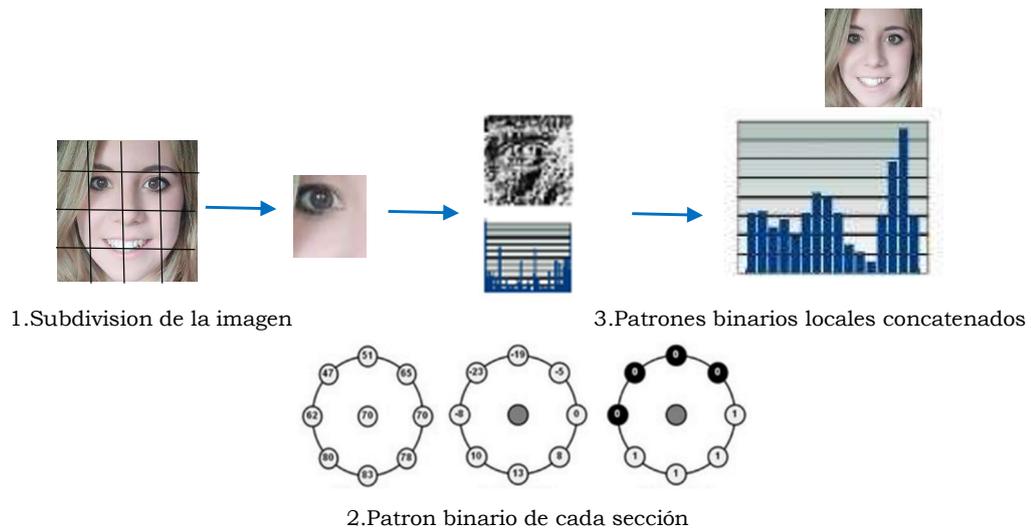
$$s(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases} \quad 3.1.3.4$$

Así se obtendrá un patrón binario local compuesto de 0 y 1 que sigue la siguiente premisa: tomando un pixel como centro y como umbral los vecinos, si la intensidad central es mayor o igual que la del vecino se codifica un '1' y viceversa. De tal manera que acaba por obtenerse una trama cifras por cada pixel. Este proceso es repetitivo con todos y cada uno de los píxeles de la imagen.

-Finalmente se concatenan los patrones de cada celda para obtener así el vector de características.

Por su propia definición el operador LBP es robusto frente a transformaciones de iluminación.

En la Figura 13 se muestra un diagrama explicativo hasta la obtención del vector de características final mediante concatenación de los histogramas locales:



$$1*1+1*2+1*4+1*8+0*16+0*32+0*64+0*128=15$$

Figura 13.- Diagrama de la metodología seguida por LBP

### 3.1.4 Deep Learning descriptors

El descriptor obtenido basado en *Deep Learning* se calcula a partir de una red neuronal de convolución entrenada para el reconocimiento facial. Se trata de una red neuronal de convolución que ha sido entrenada con más 2.6 millones de imágenes y que es capaz de clasificar 5000 identidades distintas. Para este proyecto, en lugar de utilizar la última capa de la red, la cual se encarga de clasificar, hemos tomado el modelo que se propone en [15] y utilizamos como descriptor de imagen una de las últimas capas de la red. Cada una de las capas que componen la red son la respuesta a una serie de convoluciones que combinadas permiten obtener la salida para construir el clasificador final. Cada una de las capas que componen la red podría ser utilizada como descriptor de la imagen, sin embargo, la literatura recomienda utilizar las últimas ya que es en ellas donde más comprimida y completa esta la información referente a contornos, colores, etc.

Concretamente usamos el modelo entrenado y publicado en [15], con las imágenes que constituyen el Base de Datos Propia y nos ubicamos para la extracción del descriptor en la capa denominada como *fully connected (fc7)* [15].

La Figura 14 muestra un esquema simplificado de como se utiliza la red neuronal para calcular el descriptor de una imagen. Cada imagen va siendo procesada en distintos niveles de la red con distintos tipos tamaños de convoluciones lo cual se puede interpretar como distintas capas que permiten extraer todas las características de la cara de cada persona y finalmente combinarlas en un descriptor que las resume.

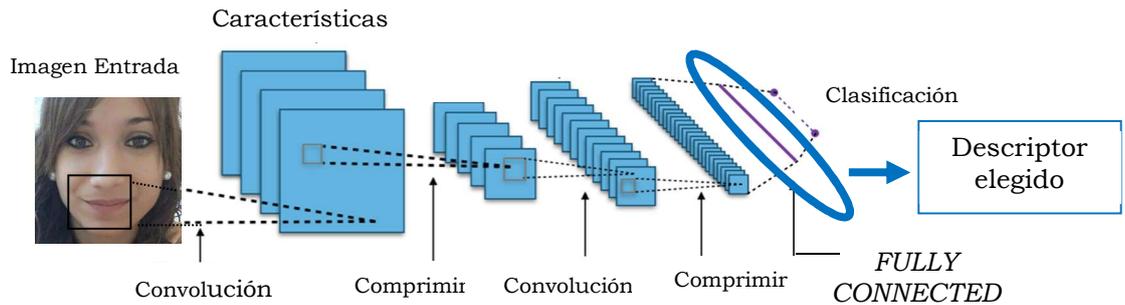


Figura 14.-Esquema de una red neuronal de convolución. Se pueden observar múltiples pasos de compresión y convolución previas a la obtención del descriptor final.

## 3.2 Clasificación

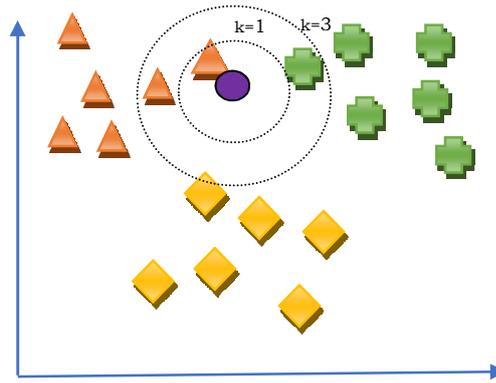
### 3.2.1 K-Vecinos más cercanos

El clasificador que utilizamos con cada uno de los descriptores descritos anteriormente es un clasificador basado en el vecino más cercano (*K-Nearest Neighbor*). Debido a que se trata de uno de los esquemas de clasificación más simple que disponemos, resulta eficaz con el tamaño de la base de datos con la que estamos trabajando y permite un sistema de clasificación que resulta sencillo de actualizar si vamos añadiendo elementos nuevos de manera interactiva. Se trata de un clasificador no paramétrico donde para cada imagen test de una cara se localizan los  $k$  miembros más cercanos pertenecientes al conjunto de entrenamiento con la función de consulta para encontrar el vecino más cercano en el espacio de la imagen.

El algoritmo *Nearest Neighbor* es un algoritmo que dado un conjunto de datos y un dato de entrada devuelve la etiqueta del dato del conjunto al que más próximo se encuentra:

$$x_{nn} = \underset{l^*}{\operatorname{argmin}} D(x^*, x_l) \quad 3.2.1.1$$

Donde es  $D$  es la función de cálculo de la distancia mínima. Esta distancia mínima entre píxeles se calcula con la distancia Euclídea, la cual permite emparejar puntos con máxima proximidad. La medida que proporciona es empleada para calcular como de cerca se encuentra cada miembro del conjunto de entrenamiento de la muestra que está siendo examinada. A partir de estos  $k$  vecinos, determinamos la etiqueta de clase con la que la imagen de prueba guarda mayor semejanza. Las predicciones realizadas por *KNN* se basan en la suposición de que de modo intuitivo los objetos más cercanos en distancia son potencialmente de una clase similar.



*Figura 15.- Diagrama que representa el algoritmo de Vecino más cercano. En la figura observamos la muestra de test (circulo violeta) que debe ser clasificada entre las tres clases que constituyen el espacio. Para  $k=1$  la clase asignada sería aquella en la quede recogida un descriptor de entre las nubes de puntos (triángulo). Siguiendo la misma dinámica para  $k=3$  la clase seguiría siendo la misma (triángulo) ya que son 2 triángulos frente a 1 cruz.*

En este proyecto se ha escogido  $k=1$ . Su elección depende prioritariamente de las características de cada una de las base de datos: la imagen de aprendizaje y conjunto de entrenamiento están reunidos bajo las mismas condiciones de iluminación, por tanto, los puntos correspondientes en el espacio de la imagen estarán fuertemente agrupados además de disponer de conjuntos de entrenamiento suficientemente grandes como para poder asegurar que usamos este método de calificación de modo fiable.

# Capítulo 4. Experimentos sobre reconocimiento facial

## 4.1 Bases de datos y configuración experimental

### **Base de datos**

Cuando de evaluación de algoritmos se trata, se considera recomendable iniciarlo con conjuntos de datos cuya eficiencia haya sido probada por investigadores de tal manera que los resultados de las experimentaciones puedan ser comparados de modo directo [18]. La elección de la base de datos es una tarea que debe realizarse en base a los objetivos que se pretende alcanzar sobre el comportamiento del algoritmo.

Es por ello que para este proyecto han sido seleccionadas varias bases de datos de dominio público y de libre acceso: “*The ORL Database of Faces*” [19] y “*The Yale Face Database B*” [20]. La primera de ellas está constituida por cuarenta sujetos de los que se contiene diez imágenes de cada uno. Las imágenes de cada sujeto son frontales con poca variación en la expresión facial e iluminación constante. Es por ello que esta se trata, prioritariamente, de una base de datos de puesta a punto y comparación de resultados. La segunda base de datos es una de las más apropiadas para el reconocimiento facial dado que contiene veintiocho sujetos con nueve poses bajo 64 condiciones de luz diferentes. Posteriormente, se propuso la elaboración de una base de datos propia que consta de veintiún sujetos con poses y expresiones aleatorias. Para más información sobre la Base de Datos EINA consultar Anexo C.

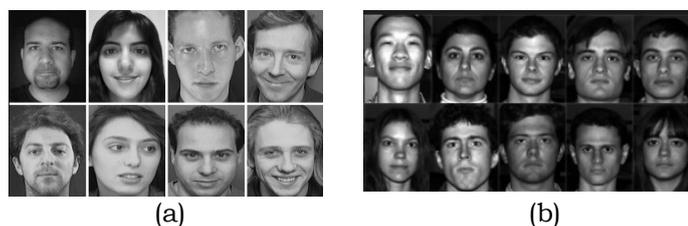


Figura 16. Muestra de las bases de datos públicas de ORL en (a) y Yale (b). Se trata de bases de datos mucho más sencillas que la elaborada. La Base de Datos Propia es más realista que las aquí presentadas porque no son imágenes en escala de grises, no presentan patrones de posición fijos y la iluminación varía entre sujetos.

Los pertinentes experimentos, objetivos y evaluaciones exhaustivas con las bases de datos presentadas anteriormente son explicadas en el siguiente punto.

### **Configuración experimental**

Como se ha explicado, los algoritmos de reconocimiento facial presentan sensibilidad a ciertos factores que pueden reducir su rendimiento de manera significativa: vista frontal o de perfil de la cara, iluminación, rotación en la imagen, expresiones faciales, presencia de gafas, bigotes, barbas, pelo largo, maquillaje... Por ello, previo a los métodos propuestos es extremadamente importante aplicar diversas técnicas de procesamiento de imagen a cada una de las bases de datos, que sean capaces de normalizar las imágenes que se proporcionan al sistema de reconocimiento facial y así ignorar lo redundante para el reconocimiento de cada una. Las técnicas usadas consisten en:

#### *Detección*

-Uso de algoritmo de detección facial capaz de detectar una cara frontal en una imagen mediante la localización de las características más significativas como los ojos situados dentro de la región de interés definida por el mismo rostro. Utilizando la

detección de rostros que prepara la cara, no solo facilitamos el problema de reconocimiento, sino que establecemos un estándar uniforme en cuanto a qué parte de la imagen debe ser utilizada [21] [22].

#### *Redefinición del tamaño y recorte*

-Algoritmo de recorte de imágenes en base al perfil definido por la detección facial. El principal propósito es que solo sean cortadas las imágenes que contienen verdaderamente cara, por eso es de vital importancia que la extracción de las características sea lo más realista posible ya que estos puntos serán la base para el recorte. El modelo de recorte usado ha sido de 200x200 pixel. Con esta redefinición del tamaño se recogen perfectamente los detalles necesarios para el reconocimiento. De este modo, aseguramos que las partes más importantes de la imagen son preservadas.

#### *Almacenamiento*

-Las imágenes de las caras ya cortadas, rotadas y escaladas pasan a formar parte de una biblioteca facial. Dicha biblioteca es dividida en dos conjuntos: datos de entrenamiento y datos de test.

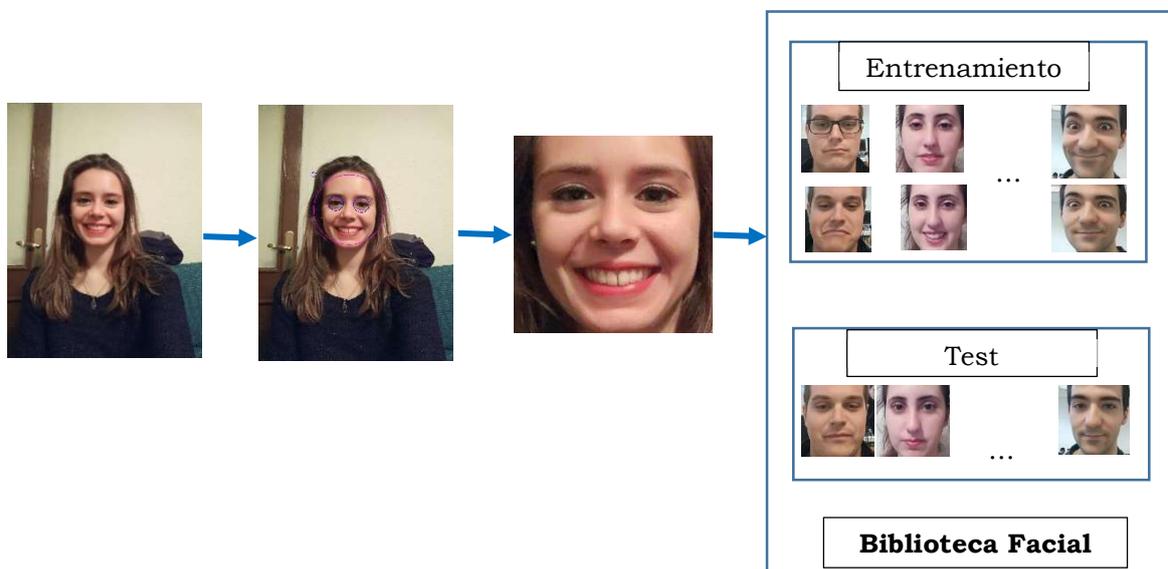


Figura 17.-Diagrama correspondiente a la configuración de las imágenes

Una vez completada esta tarea se podrá proceder a la extracción de las características de la imagen de la cara.

### ***Evaluación de los experimentos***

A lo largo de los experimentos hemos utilizado una serie de técnicas de evaluación de los resultados:

#### *Precisión*

Se trata de una medida de rendimiento cuyo valor corresponde al ratio de entre todas las predicciones realizadas en un experimento y el número de predicciones que han sido correctas.

#### *Matriz de confusión*

Herramienta que permite mostrar los resultados obtenidos de un experimento de manera detallada y visual. Las columnas corresponden al sujeto real mientras que las filas al sujeto predicho. Permite determinar el número de aciertos por sujeto y además distinguir entre que sujetos se ha realizado predicción errónea.

#### *Validación cruzada*

Es la técnica que permite garantizar que los resultados obtenidos son independientes de la partición realizada entre el conjunto de entrenamiento y el

conjunto de test. Ha sido empleada en aquellas bases de datos no muy grandes, de tal manera que para cada subexperimento la división del conjunto de entrenamiento pueda ser maximizado sin perder la opción de testeo. Este proceso es repetitivo para cada subexperimento del experimento y el resultado total permite asegurar que la evaluación es fiable.

#### *Notificación de error correcta*

Se trata de una medida empleada en el sistema interactivo. Dados los sujetos desconocidos que van a pasar a formar parte de la base de datos, cuántos de ellos han sido notificados como nuevos usuarios a incluir.

## 4.2 Experimentos

### 4.2.1 Experimento 1

El primer experimento que se ha llevado a cabo tiene como principal objetivo la puesta a punto de los algoritmos dentro del sistema, así como la evaluación del método para cada uno de los métodos propuestos y cómo se comporta ante una base de datos sencilla. Para ello, la validación cruzada ha constado nueve imágenes de entrenamiento y una de test por sujeto para cada subexperimento (4.1). Normalmente, cuantas más imágenes se puedan utilizar para el entrenamiento más preciso es el modelo aprendido.

Tal y como se esperaba por los resultados publicados en la literatura, las ratios de error son muy bajas. Para cualquiera de los tres métodos las condiciones de iluminación, rotación facial o expresiones no suponen un problema en el reconocimiento. Para este estudio, se han obtenido los valores de precisión y para más detalle se han adjuntado las matrices de confusión en el Anexo B:

|                    | Precisión |
|--------------------|-----------|
| <i>Eigenfaces</i>  | 0.974     |
| <i>Fisherfaces</i> | 0.925     |
| <b>LBP</b>         | 0.986     |

*Tabla 2.-Tabla de precisión*

Con estos resultados queda corroborada la eficiencia del método y su correcta puesta a punto.

### 4.2.2 Experimento 2

En el segundo experimento la evaluación del método se ha centrado en la experimentación con una base de datos que consta de imágenes aleatorias de la base de datos Yale, con el objetivo de observar como para un mismo número de imágenes de entrenamiento los métodos varían su rendimiento y eficiencia ante distintas condiciones de iluminación y expresión facial.

Para este experimento se repite la misma configuración de validación cruzada en el que para cada subconjunto la partición consta de nueve imágenes de entrenamiento y una imagen de test.

#### ***Variaciones con las condiciones de luz***

En este experimento, el objetivo principal es comparar los métodos *Eigenfaces*, *Fisherfaces*, *LBP* bajo variaciones de iluminación. Los resultados experimentales se muestran en las Figura 4, Figura 5, Figuras 6 que se adjuntan en el Anexo B y los resultados básicos se reflejan en la Tabla 3 que se presenta a continuación. Se observa el fuerte impacto de este factor en la tarea de reconocimiento: todos los valores de precisión han descendido significativamente en comparación al Experimento 1 (4.2.1).

En el anterior experimento, nos encontrábamos con un conjunto de imágenes en el cual la dirección de la luz incide frontalmente en el rostro del sujeto y de manera uniforme. Sin embargo, bajo condiciones variables de luz, como es el Experimento 2, la dirección de la iluminación es cambiante sobre la cara del sujeto lo que se traduce en una mayor dificultad en la identificación del sujeto.

Los cambios de iluminación (en cualquier imagen) dan lugar a una percepción distinta de las características de la imagen (forma, en la textura de la cara de cada individuo...). Si centramos nuestra atención en los resultados mostrados en la Tabla 3, Fisherfaces es el método que menor precisión ha tenido. El resultado es el esperado ya que este método solo captura los rasgos faciales y esta base de datos al componerse de escenas con iluminación complicada las características de la cara han sido mal extraídas: *Fisherfaces* no atiende a cambios de la luz.

Para este estudio los valores obtenidos son los que se muestran en la siguiente tabla:

|                    | Precisión |
|--------------------|-----------|
| <i>Eigenfaces</i>  | 0.639     |
| <i>Fisherfaces</i> | 0.619     |
| <b>LBP</b>         | 0.797     |

Tabla 3.-Tabla de precisión

Como era esperado, al ser una base de datos con mayor dificultad las ratios de acierto disminuyen. Como conclusión, los porcentajes bajan, pero el que menos disminuye es el LBP por lo que parece que es el que mejor modelo de aprendizaje ha realizado. Para comprobarlo, el siguiente experimento va a constar de la base de datos completa.

### 4.2.3 Experimento 3

Para este experimento se va a realizar un estudio profundo de la base de datos completa de Yale. Dado que se trata de una base de datos que cuenta con 16146 imágenes se va a prestar especial atención a los tiempos de cómputo de cada método.

El elevado número de imágenes de la base de datos y las limitaciones inherentes al computador forzaron la decisión de formar un conjunto de entrenamiento del 70% de las imágenes y un 30% de las mismas para el conjunto de test. Bajo estas premisas, los resultados básicos obtenidos se muestran en la siguiente tabla:

|                    | Precisión |
|--------------------|-----------|
| <i>Eigenfaces</i>  | 0.6614    |
| <i>Fisherfaces</i> | 0.6188    |
| <b>LBP</b>         | 0.8086    |

Tabla 4.-Tabla de precisión

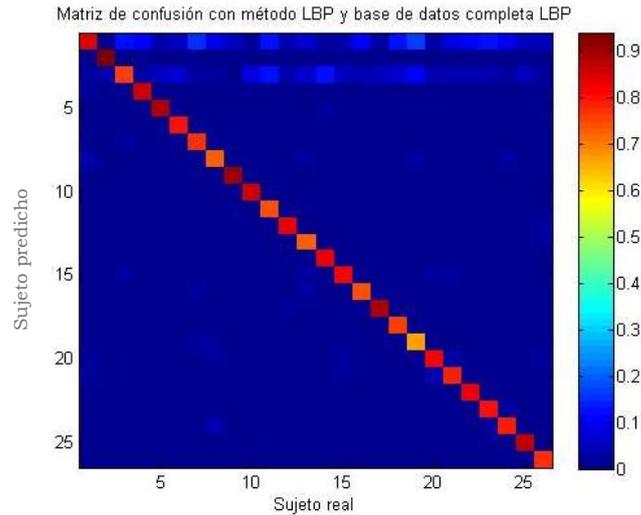
Vemos que el modelo que mejor precisión proporciona continúa siendo el método LBP ya que proporciona una descripción de la imagen invariante a las condiciones del entorno. Dado el elevado número de datos con lo que contábamos para este experimento resulta necesario una valoración de los tiempos de cómputo del entrenamiento:

|                    | Tiempo de cómputo (hh:mm) |
|--------------------|---------------------------|
| <i>Eigenfaces</i>  | 02:57                     |
| <i>Fisherfaces</i> | 02:45                     |
| <b>LBP</b>         | 02:14                     |

Tabla 5.-Tabla de tiempos de cómputo

En la tabla 5, podemos observar los tiempos de computo de cada uno de los métodos correspondientes a la fase de entrenamiento, es decir, a la generación del modelo. Se trata de tiempos muy elevados, del orden de horas, pero debemos recordar que se trata de 11302 imágenes de entrenamiento por lo que el tiempo por imagen correspondería a un segundo aproximadamente. Hasta el momento, ningún experimento ha requerido tanto tiempo para la elaboración del modelo. Este inconveniente está del todo justificado dada la gran cantidad de imágenes que componen la base de datos.

Para mostrar con más detalle cual ha sido el resultado del experimento mostramos la tabla de confusión correspondiente en la Figura 18:



*Figura 18.- Matriz de Confusión con método LBP y base de datos completa de Yale. En esta figura podemos observar que el número de falsos positivos (predicción realizada erróneamente) es bastante bajo. La diagonalización es muy buena y se acerca a los valores máximos lo que indica que el número de verdaderos positivos (predicciones realizadas correctamente) es alto.*

Para más información de los resultados obtenidos por el resto de métodos con los que se realizaron las pruebas para este experimento consultar el Anexo B en donde han sido adjuntadas las figuras correspondientes las matrices de confusión obtenidas.

#### 4.2.4 Experimento 4

Como hemos observado, cuando los datos son más complicados y en entornos menos controlados, los métodos tradicionales empiezan a fallar. En este experimento queremos comprobar si los nuevos métodos son más robustos en este tipo de situaciones. Por ello, el principal objetivo de este experimento es evaluar los métodos del estado del arte basados en redes neuronales de convolución. Como se pretende analizar su rendimiento usando una base de datos realista, se realizó con la base de datos propia, ya que es la que ha sido capturada de manera más natural en un entorno poco controlado y con variaciones de luz y gestos.

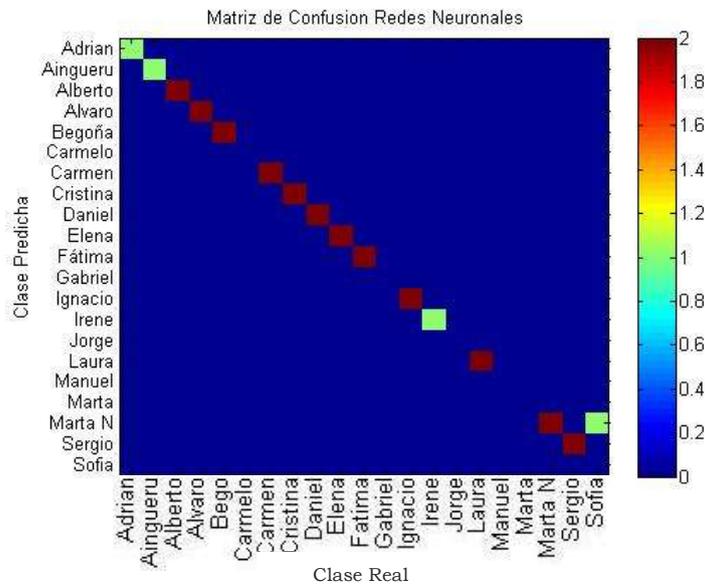


Figura 19.-Matriz de confusión con Redes Neuronales de convolución

|                                    | Precisión   |
|------------------------------------|-------------|
| <b>Red Neuronal de Convolución</b> | <b>0.96</b> |

Tabla 6.-Tabla de precisión

Como podemos observar en la Tabla 6 y en la Figura 19, la precisión de este método es altísima. Podemos observar en el capítulo siguiente como los otros métodos no consiguen tanta precisión. El resultado obtenido para este método era esperado ya que obtenemos una precisión del 96% la cual coincide bastante bien con los porcentajes de acierto publicados en bases de datos realistas [15].

Se trata del resultado más exitoso de entre todos los experimentos realizados hasta el momento. Sin embargo, dada su elevada complejidad, coste y dificultad para integrar dentro de una aplicación en C++ (el modelo utilizado está disponible para Matlab en estos momentos y es muy lento si no se dispone de GPU) no ha sido posible abordar el algoritmo como una posible solución al desarrollo del sistema interactivo.

## 4.2.5 Análisis de los experimentos

Llegado este punto la toma la decisión del método definitivo bajo el cual se implementará el sistema interactivo queda restringida al método LBP. No solo resulta necesario para el desarrollo de los factores que aquí se han tenido en cuenta, sino que además este método incorpora la posibilidad de actualizar el modelo dada su definición. Este aspecto no lo dispone el resto de métodos, y aunque sin ser necesario, permite que la incorporación a la base de datos sea mucho más sencilla. Por estos motivos se trata del método adecuado para su implementación.

El tiempo de procesamiento no es un inconveniente en el sistema interactivo dado que el número de imágenes del sistema interactivo no va a ser tan elevado como en el Experimento 3 (4.2.3). De hecho, una vez proporcionadas las imágenes del sujeto desconocido, la actualización del modelo es instantánea, como veremos en el siguiente capítulo.

# Capítulo 5. Sistema interactivo

El sistema interactivo desarrollado para este proyecto se ha realizado con una base de datos propia, para su consulta acudir a Anexo C. Se trata de una base de datos formada por veintiún sujetos con diez fotos cada uno. La dificultad que presentan estas imágenes es mayor que las anteriores dadas las diferentes condiciones del ambiente y propias de los sujetos: la posición de la cara, la iluminación, la expresión facial (muecas, sonrisas...) y apariencia (gafas, barbas...).

Como se decidió en la sección anterior (4.2.5) el método elegido para la implementación del sistema es el LBP. El sistema pretende la evaluación del *update* por ello, el experimento consta de dos partes. La primera de ellas es la de notificar que el sujeto es desconocido al modelo de partida el cual cuenta con dieciséis sujetos del total. La segunda parte, tras actualizar el modelo, consiste en evaluar el reconocimiento del usuario.

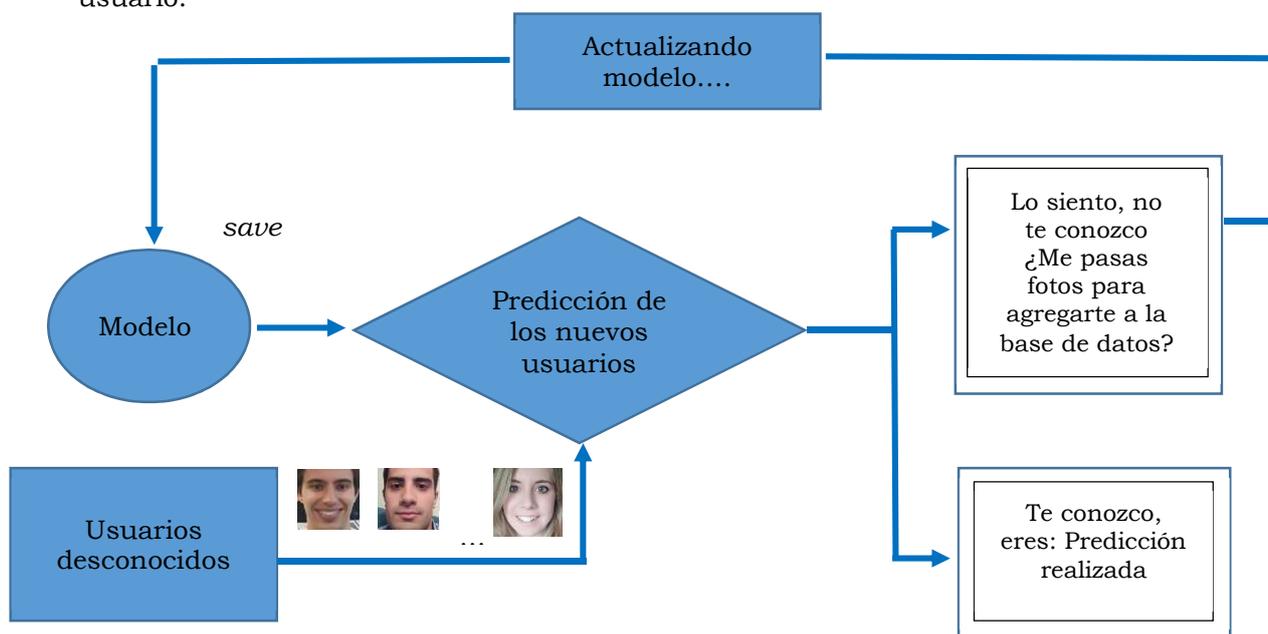


Figura 20.- Diagrama de flujo del funcionamiento del Sistema Interactivo

Los resultados obtenidos en este experimento son mostrados en la Tabla 7. El hecho de que la base de datos propuesta sea más complicada queda patente en los resultados. Cada una de las tres iteraciones corresponde con una nueva imagen para testear de cada uno de los sujetos y como vemos en algunos de los sujetos desconocidos la notificación, y por lo tanto la actualización del modelo, no ha sido posible hasta la segunda iteración. Las causas que lo han imposibilitado corresponden sobre todo a la presencia de gafas o no, barbas en sujetos masculinos o el gran parecido que existe entre algunos de los sujetos. Una vez realizada la actualización completa del modelo, tercera iteración, encontramos que pese a la dificultad la precisión ha sido 0.9047, pero en media del 0.857. Sin embargo, sigue siendo la precisión esperada, bastante alta, y si lo comparamos con el segundo experimento en el que la dificultad de la base de datos también era elevada vemos que incluso la supera.

|                     | Notificación de error correcta<br>(UsuarioDesconocido/TotalUsuariosDesconocidos) | Precisión<br>(Nº Aciertos/Nº Test) |
|---------------------|--|------------------------------------|
| <b>1º Iteración</b> | 2/5  | 17/21(0.8095)                      |
| <b>2º Iteración</b> | 3/3  | 18/21(0.8571)                      |
| <b>3º Iteración</b> | -  | 19/21(0.9047)                      |

Tabla 7.- Tabla Sistema Interactivo

A continuación, mostramos la matriz de confusión correspondiente a este experimento:

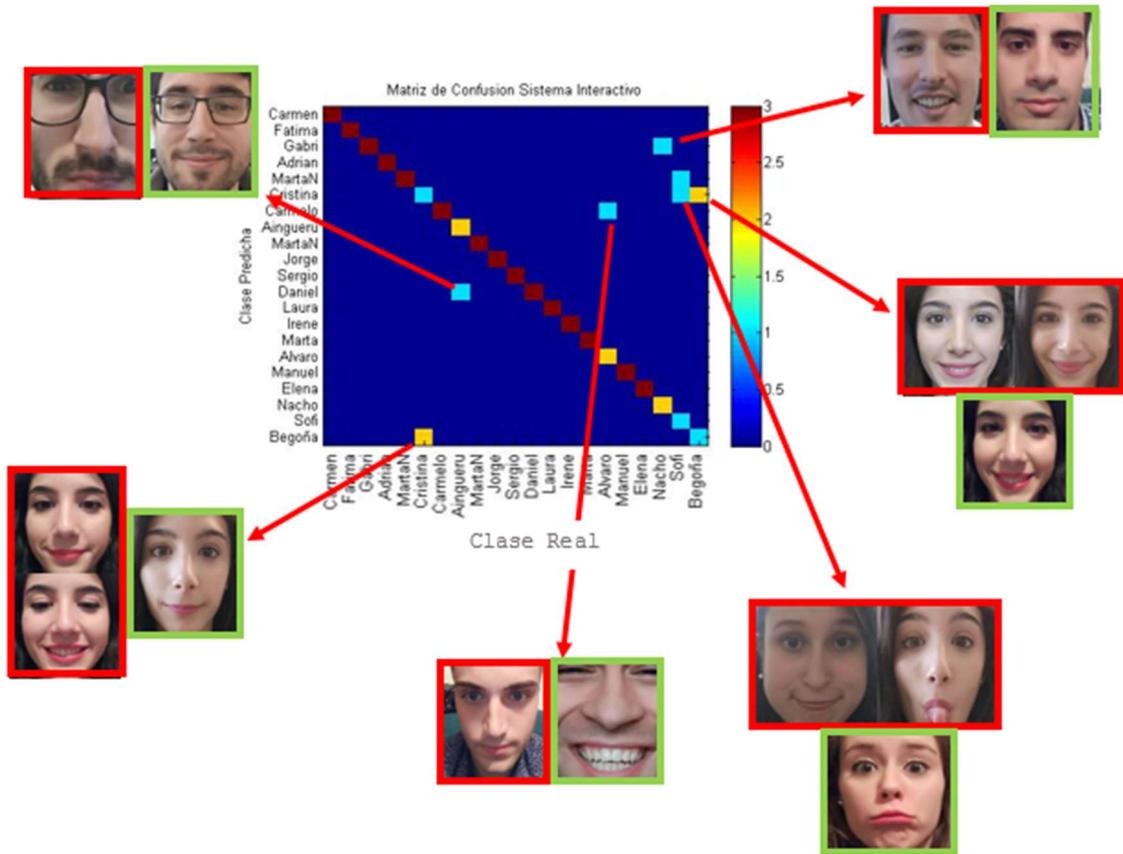


Figura 21.-Matriz de confusión Sistema Interactivo

Tras la experimentación con datos reales se puede concluir que tiene algunos problemas lógicos dentro de las capacidades que ofrece el método y que cuando es un sujeto nuevo lo notifica bastante bien, sin embargo, cuando las imágenes no son controladas (gestos variados, iluminación cambiante...) empieza a fallar.

# Capítulo 6. Conclusiones finales

## 6.1 Conclusiones

El trabajo realizado ha tratado de integrar un sistema capaz de reconocer e identificar rostros de acuerdo a una base de datos. Para ello, se propuso el estudio y evaluación de diferentes técnicas de reconocimiento y posteriormente se realizaron las adaptaciones necesarias para que el algoritmo elegido sea capaz de incorporar nuevos sujetos o identificar a los ya existentes.

En lo relativo a las técnicas de reconocimiento se puede concluir que:

En general, los primeros métodos ofrecen resultados buenos dentro de las restricciones propias. Los cambios de luminosidad en el entorno han modificado las tasas de acierto haciéndolas disminuir en más de un 20% en cada método. Por lo que queda de manifiesto la importancia de este factor en los procesos de reconocimiento. En cuanto al último método utilizado, Redes Neuronales de Convolución, una técnica reciente en las tareas de reconocimiento ha quedado probada la importancia de reunir conjuntos de datos a gran escala. Este trabajo demuestra que esta técnica es capaz de superar las limitaciones de los métodos anteriores superando las tasas de acierto para bases de datos realistas. Su capacidad de representar los rostros queda evidenciada en los resultados obtenidos para esta técnica de manera muy positiva. Sin embargo, su complejidad para ser integrado dentro de un sistema interactivo ha llevado a tomar la decisión de que este no sea el método elegido para su implementación. Pero se consideraba necesario realizar un experimento con las últimas técnicas para estudiar y mostrar la evolución y mejoras no solo en tareas de reconocimiento, sino en todo el campo de la visión por computador.

En cuanto al sistema interactivo, para su implementación se han tenido en cuenta las restricciones propias que aparecen cuando se trata de bases de datos realistas. El sistema es capaz de reconocer usuarios e incluir nuevos. Para ello el sistema cuenta con una interface que en caso de identificar usuario desconocido solicita la dirección a un conjunto de imágenes del usuario en concreto para incluir en su base de datos y actualizar el modelo. Un inconveniente que le aleja de ser considerado un sistema interactivo en tiempo real es que dichas imágenes deben ser adquiridas previamente. Aun así, la evaluación como sistema es positiva dado que bajo estas premisas los resultados obtenidos han sido buenos, aunque presentaba limitaciones dadas las diferentes poses, muecas iluminaciones... en cada sujeto.

## 6.2 Líneas futuras

Como posible trabajo futuro se podría considerar realizar algunos experimentos con más datos, por ejemplo, con la base de datos propuesta en [15] y evaluar cómo afecta un aumento tan significativo de imágenes por sujeto para el reconocimiento facial basado en métodos tradicionales.

Se podría integrar un sistema interactivo basado en redes neuronales de convolución capaz de incluir nuevos usuarios en tiempo real.

Por último, considerar otros clasificadores o descriptores distintos a los empleados y comparar dichos resultados con lo obtenidos en este proyecto.

Otro posible trabajo futuro en cuanto a la aplicación directa del sistema desarrollado, sería incluir el sistema interactivo en un robot. El robot podría recopilar información de las caras de su entorno y reportar información a un centro de control en caso de detectar sujeto desconocido, o ejercer de recepcionista en un museo o centro público en que puede haber usuarios nuevos que quieran suscribirse.

# Bibliografía

- [1] Hay and A. W. Young, "The Human Face", Normality and Pathology in Cognitive Function, A. W. Ellis, Ed. London: Academic, pp. 173-202, 1982
- [2] Ederman and P. Kalocsai, "Neural and Psychophysical Analysis of Object and Face Recognition", in Face Recognition: From Theory to Applications (H. Wechsler, P. J. Phillips, V. Bruce, F. F. Soulie and T. S. Huang, eds.), Berlin: Springer-Verlag, pp. 3-25, 1998.
- [3] EUis 'The role of the right hemisphere in face perception' in Function of the Right Cerebral Hemisphere, A.W. Young, Ed London: Academic, 1983, pp. 33-64.
- [4] Bill, V. Bruce, "Effects of Lighting on Matching Facial Surfaces", Journal of Experimental Psychology: Human Perception and Performance, vol. 22, pp. 986-1004, 1999
- [5] Ian Wallis (2012) 50 Best Business ideas that changed the world. Westminster House, Key Road Richmond, Surrey. TW9 2ND, UK. Recuperado de <https://books.google.es/books?id=WbgQAQAQBAJ&pg=PT49&dq=Woodrow+W.Bledsoe+Facerecognition&hl=es&sa=X&ved=0ahUKEwiOz7f3oazNAhXKlxoKHRC3D6sQ6AEIHDA#v=onepage&q=Woodrow%20W.Bledsoe%20Facerecognition&f=false>
- [6] Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91., IEEE Computer Society Conference on
- [7] Face Description with Local Binary Patterns: Application to Face Recognition Timo Ahonen, Student Member, IEEE, Abdenour Hadid, and Matti Pietikainen, Senior Member, IEEE
- [8] Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection Peter N. Belhumeur, Joao P. Hespanha, and David J. Kriegman
- [9] Martinez, A and Kak, A. *PCA versus LDA* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No.2, pp. 228-233, 2001
- [10] Ahonen, T., Hadid, A., and Pietikainen, M. Face Recognition with Local Binary Patterns. Computer Vision - ECCV 2004 (2004), 469-481.
- [11] Duda, Richard O. and Hart, Peter E. and Stork, David G., *Pattern Classification* (2nd Edition) 2001
- [12] Ahonen, T., Hadid, A., and Pietikainen, M. *Face Recognition with Local Binary Patterns*. Computer Vision - ECCV 2004 (2004), 469-481
- [13] Yann LeCun, Yoshua Bengio & Geoffrey Hinton. (2015) Deep Learning Nature. Recuperado <http://www.nature.com/nature/journal/v521/n7553/pdf/nature14539.pdf>
- [14] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In Proc. CVPR, 2015.
- [15] Omkar M. Parkhi, Andrea Vedaldi, Andrew Zisserman. Visual Geometric Group. Department of Engineering Science University of Oxford. Deep Face Recognition . Recuperado de <https://www.robots.ox.ac.uk/~vgg/publications/2015/Parkhi15/parkhi15.pdf>
- [16] Yaniv Taigman, Ming Yang, Marc Aurelio Ranzato, Lior Wolf, DeepFace: Closing the Gap to Human-Level Performance in Face Verification. USA. [https://www.cs.toronto.edu/~ranzato/publications/taigman\\_cvpr14.pdf](https://www.cs.toronto.edu/~ranzato/publications/taigman_cvpr14.pdf)

- [17] Zhao, W., Chellappa, R., Phillips, P., and Rosenfeld, A. Face recognition: A literature survey. *ACM Computing Surveys (CSUR)* 35, 4 (2003), 399–458
- [18] “From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose”, *PAMI*, 2001 results
- [19] An archive of AT&T Laboratories Cambridge. The Database of Faces. (2002) [http://www.cl.cam.ac.uk/research/dtg/attarchive/pub/data/att\\_faces.zip](http://www.cl.cam.ac.uk/research/dtg/attarchive/pub/data/att_faces.zip).
- [20] Athinodoros Georghiades, Peter Belhumeur, and David Kriegman's.(2001) The extended Yale Face Database <http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html>
- [21] OpenCV 2.4.13.0 documentation 2016. [http://docs.opencv.org/2.4/modules/objdetect/doc/cascade\\_classification.html?highlight=detectmultiscale#cascadeclassifier-detectmultiscale](http://docs.opencv.org/2.4/modules/objdetect/doc/cascade_classification.html?highlight=detectmultiscale#cascadeclassifier-detectmultiscale)
- [22] OpenCV 2.4.13.0 documentation 2016. [http://docs.opencv.org/2.4/modules/objdetect/doc/cascade\\_classification.html?highlight=cascadeclassifier#cascadeclassifier](http://docs.opencv.org/2.4/modules/objdetect/doc/cascade_classification.html?highlight=cascadeclassifier#cascadeclassifier)

# Anexo A

OpenCV proporciona enlaces en los que se describen algunos tutoriales para principiantes. Todos los códigos proporcionados son una parte regular y básica en los lanzamientos de OpenCV. Existe un listado de tutoriales proporcionado por el mismo. Se realizaron y ampliaron aquellos más interesantes para el presente trabajo.

## Tutorial 1

En este tutorial se aprendió a cargar y mostrar una imagen por pantalla. Sus objetivos eran:

- Cargar la imagen usando funciones propias del entorno.
- Crear un entorno de programación en Eclipse.
- Conseguir mostrar cualquier imagen por pantalla.

Además, por motivación propia se mostraron imágenes y vídeo capturadas desde una cámara web.

## Tutorial 2

Con los conocimientos adquiridos en el anterior tutorial, se trató de modificar una imagen y convertirla a escala de grises. Una vez realizada la transformación, se aprendió cómo guardarla en una carpeta.

También se desarrolló un pequeño algoritmo que permitía mostrar simultáneamente el vídeo capturado por la webcam y su transformación en blanco y negro.

## Tutorial 3

Este tutorial es más específico y se aproxima mejor a la dinámica del trabajo. Se trata de un tutorial cuyo objetivo es hacer una clasificación de géneros dados un conjunto de imágenes de rostros. Este tutorial contaba con algunos requisitos como la elaboración de una base de datos de hombres y mujeres. La colección de imágenes se hizo de personajes públicos adquiridos desde el buscador de *Google*.

Para ello, se realizó un etiquetado de las imágenes de acuerdo con el sexo del sujeto. Además, un factor importante para el reconocimiento es alineamiento de las imágenes, por lo que en este tutorial ya se incluyó una etapa de alineamiento de imágenes. Tras la fase de detección, en la que se localizan los rasgos propios de la cara, como los ojos, se utilizó la información que proporcionan estas coordenadas faciales para el alineamiento.

A continuación, se procedió a la fase de entrenamiento en la que se crea un modelo que aprende a discriminar las características entre ambas clases.

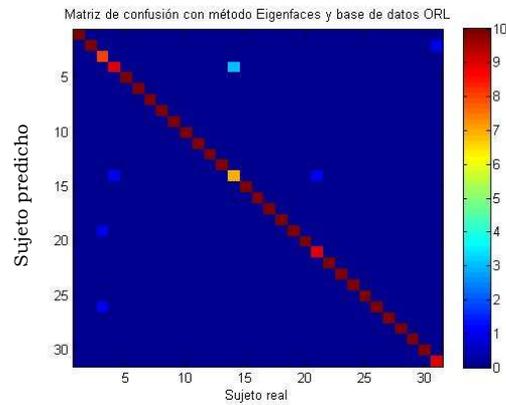
El objetivo principal que se pretendía con este tutorial no era solo el de realizar una tarea de clasificación de géneros sino el de alcanzar un enfoque de la tarea de clasificación

# Anexo B

## Experimento 1

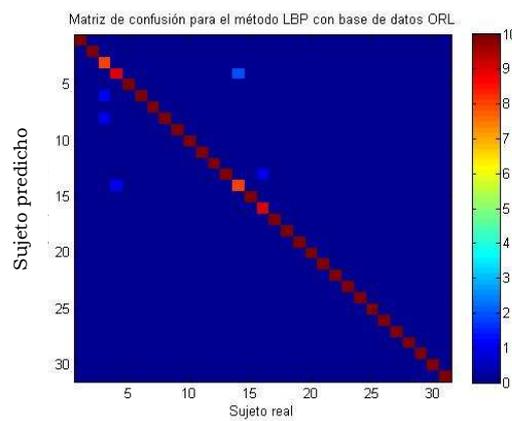
*The ORL Database of Faces*

-Eigenfaces



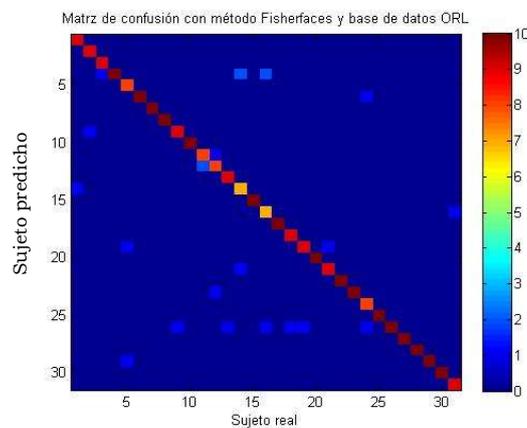
*Figura 1.-Matriz de confusión método Eigenfaces base de datos ORL*

-LBP



*Figura 2.-Matriz de confusión método LBP base de datos ORL*

-Fisherfaces



*Figura 3.-Matriz de confusión método Fisherfaces base de datos ORL*

# Experimento 2

## The Yale Face Database B

-Eigenfaces

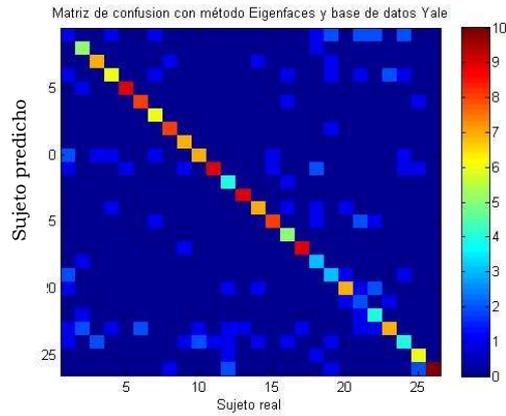


Figura 4.-Matriz de confusión Eigenfaces base de datos Yale

-LBP

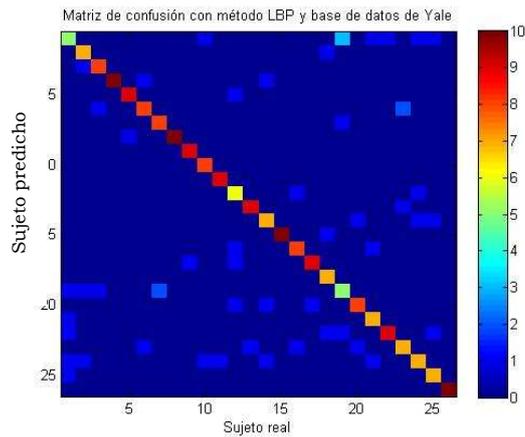


Figura 5.-Matriz de confusión LBP base de datos Yale

-Fisherfaces

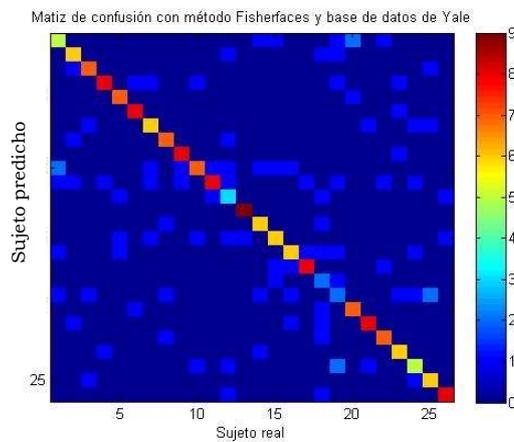
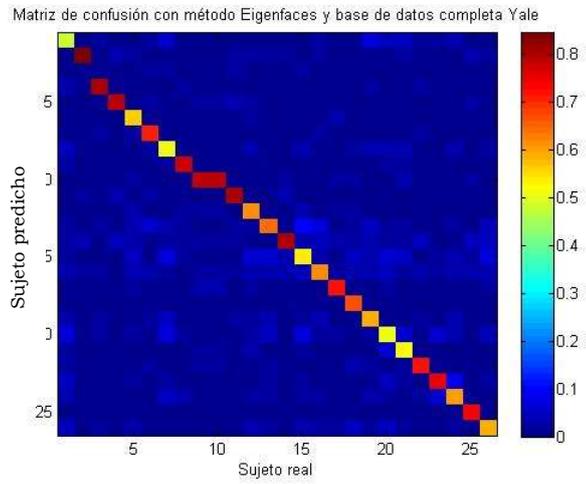


Figura 6.-Matriz de confusión Fisherfaces base de datos Yale

# Experimento 3

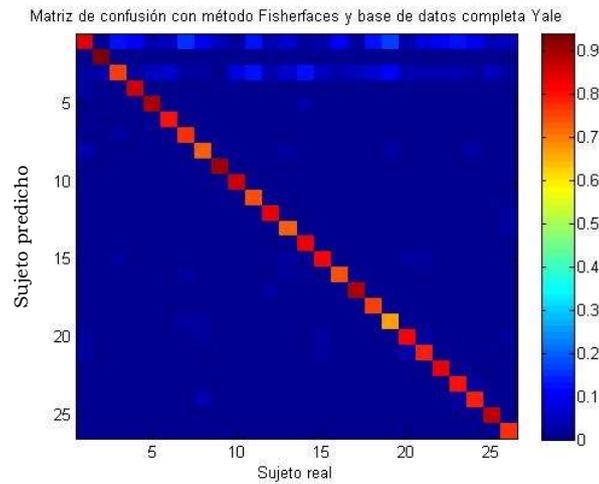
*The Yale Face Database B*

-Eigenfaces



*Figura 7.-Matriz de confusión Eigenfaces base de datos Yale*

-Fisherfaces



*Figura 8.-Matriz de confusión Fisherfaces base de datos Yale*

# Anexo C

## Presentación de la Base de Datos EINA

Se trata de una base de datos realista integrada por veintiún sujetos con diez fotos cada uno. La dificultad que presentan estas imágenes es cualquiera de las bases de datos utilizadas en otros experimentos, dadas las diferentes condiciones del ambiente y propias de los sujetos: la posición de la cara, la iluminación, la expresión facial como muecas, sonrisas... y apariencia tales como maquillajes, gafas, barbas....

