



Escola Tècnica Superior d'Enginyeria
de Telecomunicació de Barcelona

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Departamento de Procesado de Señal y Comunicaciones

ESTUDIO DE TÉCNICAS DE RECONOCIMIENTO FACIAL

Autor: Roger Gimeno Hernández

Director: Josep Ramon Morros i Rubió

Barcelona, Mayo 2010

Agradecimientos

Me gustaría agradecer a mi familia y mi pareja el apoyo que me han dado durante todos estos años, al Dr. Ramon Morros sus consejos y guiado durante todo este proyecto, sin el no habría sido posible, y a Albert Gil que siempre estuvo ahí para ayudarme con el entorno de trabajo del departamento y los problemas de código.

0 Índice

0	ÍNDICE	5
1	INTRODUCCIÓN	7
2	ESTADO DEL ARTE	15
2.1	RECONOCIMIENTO DE IMÁGENES FIJAS	17
2.2	DESCRIPCIÓN DE LAS PRINCIPALES TÉCNICAS	19
2.2.1	<i>PCA (Principal Component Analysis)</i>	19
2.2.2	<i>LDA (Linear Discriminant Analysis)</i>	22
2.2.3	<i>LPP (Locality Preserving Projections)</i>	24
2.2.4	<i>DCT (Discrete Cosine Transform)</i>	26
2.2.5	<i>DCT por Bloques</i>	28
3	ARQUITECTURA DEL SISTEMA	31
3.1	ANÁLISIS LINEAL	32
3.2	PCA	33
3.2.1	<i>Creación de la matriz de proyección</i>	33
3.2.2	<i>Reconocimiento</i>	34
3.3	DCT	34
3.3.1	<i>Reconocimiento</i>	35
3.4	LPP	36
3.4.1	<i>Creación de la matriz de proyección</i>	36
3.4.2	<i>Reconocimiento</i>	39
4	BASES DE DATOS	41
4.1	BASE DE DATOS OLIVETTI (ORL DATABASE)	41
4.2	BASE DE DATOS XM2VTS	42
4.3	BASE DE DATOS BANCA	42
4.4	BASE DE DATOS YALE	44
5	RESULTADOS	45
5.1	FUNCIONAMIENTO DEL SISTEMA EN UN ENTORNO CONTROLADO	46
5.2	INFLUENCIA DEL NÚMERO DE INDIVIDUOS EN EL SISTEMA	49

5.3	INFLUENCIA DE LOS CAMBIOS DE ILUMINACIÓN	52
5.3.1	<i>Normalización</i>	53
5.3.2	<i>Descarte de coeficientes</i>	55
5.4	EFEECTO DE LA VARIACIÓN DEL ASPECTO DE LAS CARAS.....	58
5.5	INFLUENCIA DE ENTRENAR EL SISTEMA CON IMÁGENES DISTINTAS	60
5.6	EVALUACIÓN DE UN ENTORNO REAL.....	64
5.6.1	<i>Entrenamiento con imágenes parecidas al set de referencia</i>	65
5.6.2	<i>Entrenamiento con imágenes parecidas al set de test</i>	68
5.7	EFICIENCIA COMPUTACIONAL	71
6	APLICACIÓN I. QUERY BY EXAMPLE	73
7	APLICACIÓN II. SISTEMA DE RECONOCIMIENTO EN TIEMPO REAL	75
8	CONCLUSIONES Y TRABAJO FUTURO	79
9	BIBLIOGRAFÍA	83

1 Introducción

Durante los últimos años, el reconocimiento facial se ha convertido en uno de las aplicaciones más estudiadas en campos como la biometría, el procesado de imagen o el reconocimiento de patrones. Una de las razones que ha llevado a este crecimiento son las necesidades cada vez mayores de aplicaciones de seguridad y vigilancia utilizadas en diferentes ámbitos.

El creciente interés en el reconocimiento facial y los sistemas biométricos, lo podemos ver reflejado de 3 maneras distintas:

- En primer lugar, se puede ver en la comunidad científica donde por ejemplo se pasó de la publicación de unos 140 artículos en 2007, relacionados con el reconocimiento y diferentes herramientas para llevarlo a cabo, a través de publicaciones en diferentes organismos como el IEEE, a unos 180 en 2008, incluyendo conferencias como el “*IEEE internacional conference on automatic face and gesture recognition*” (IEEE FG 2008).
- En segundo lugar, se puede observar en el aumento en el número de soluciones comerciales y el crecimiento de este mercado. Ejemplo de ello son los sistemas basados en el reconocimiento para identificación civil/criminal y los sistemas de video vigilancia desarrollados por empresas como Cognitec [1], Eyematic [2], Viisaje [3] y Identix [4] (tabla 1). Desde el punto de vista económico [5], los beneficios asociados a la industria dedicada al reconocimiento facial, crecieron de 50 millones de dólares en 2003 a 802 millones en 2008 y se espera que abarque un 17% del mercado dedicado a sistemas biométricos (figura 1 y figura 2).
- En tercer y último lugar, se ve reflejado este aumento en la industria del reconocimiento facial desde el punto de vista de los clientes, los cuales cada vez más solicitan productos funcionales basados en esta tecnología.

Descripción del proyecto	Situación	Empresa	Sector vertical	Sector Horizontal	Descripción aplicación	Descripción adicional
Manchester, NH. Visaje	US-NH	Viisage	viajes y transportes	Vigilancia/Seguimiento	Seguimiento	4th US solución para aeropuerto
Cognitec, 'Smart Gate' Aeropuerto de Sydney	Australia	Cognitec	viajes y transportes	ACC/T&A	Acceso	Basado en lectura de pasaporte
Vigilancia playa de Virginia	US-YA	Identix	Aplicación Ley	Identificación Criminal	Vigilancia	Base de datos de 600 imágenes, 10 sujetos
Aeropuerto de Berlín	Germany	ZN	viajes y transportes	ACC/T&A	Acceso	Terminal de reconocimiento facial, muestra guardada en CS
Programa VISA de diversidad	US-MA	Viisage	Gobierno	Identificación civil	Identificación	Imagen a la vez que se introduce la tarjeta y prevenir 2x aplic. tarjeta y prevenir 2x aplic.
CODL	US-CO	Identix	Gobierno	Identificación civil	DL	Detección de inscripción duplicada
Caras aeropuerto Zurich	Suiza	C-VIS	viajes y transportes	Vigilancia/Seguimiento	Seguimiento	Sistema de la policía de Zurich, usado para encontrar inmigrantes ilegales de S.Africa, Asia y P.Este
Departamento de policía de Brentwood	US-CA	Imagis	Cumpliment o Ley	Identificación Criminal	Forense	Sistema integrado con la base de datos del servicio 911

Tabla 1. Tabla con algunas empresas y aplicaciones del sector.

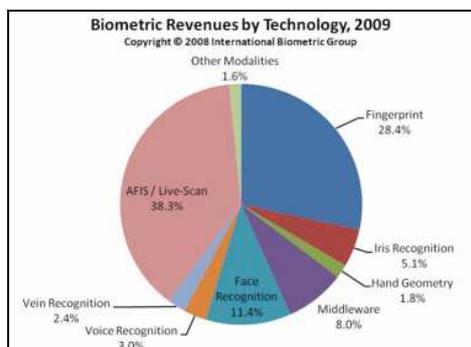


Figura 1. Desglose del porcentaje que ocupa cada técnica biométrica dentro del mercado.

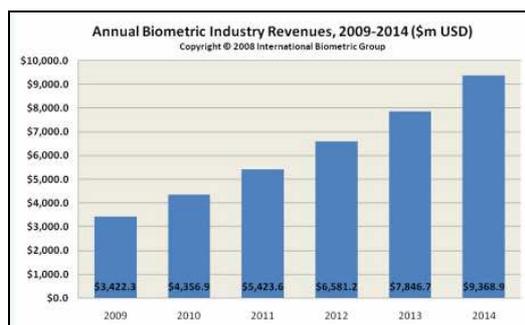


Figura 2. Previsión económica de la industria biométrica 2009-2014.

Estos sistemas en su objetivo de determinar la identidad de una persona, pueden tener en cuenta o analizar muchas características. Estas características pueden ser tanto físicas, como sería el caso de las pupilas, la cara, el iris, etc., o se puede tratar de características psicológicas como serían los gestos. Todas estas propiedades, presentan sus propias peculiaridades, las cuales pueden ser analizadas siguiendo entre otros, los siguientes criterios:

- *Universalidad*, indica como de común es encontrar esta característica en todas las personas u objetos a reconocer.
- *Carácter distintivo*, indica si dicha propiedad, es suficientemente diferente entre un conjunto de personas u objetos diferentes.
- *Permanencia*, indica la estabilidad en el tiempo de dicha característica.
- *Colectividad*, indica si la característica es fácilmente adquirida y medida por el sistema.
- *Rendimiento*, indica la precisión, velocidad y coste (recursos) necesarios para llevar a cabo el reconocimiento.
- *Aceptabilidad*, indica en que medida está la gente preparada para aceptar el uso de esta técnica.
- *Elusión*, indica la respuesta del sistema cuando alguien está tratando de engañarlo.

Gracias a estas propiedades que se han enumerado, se puede describir bien las características que va a utilizar el sistema y así poder ver si es adecuado o no para este tipo de aplicaciones. Se pueden encontrar otros modos de evaluar estas características, uno de ellos es el análisis de Zephyr [5] gracias al cual se pueden clasificar diferentes indicadores biométricos según su carácter intrusivo, precisión, coste y dificultad (figura 3). Según otras escalas [6], considerando 6 parámetros biométricos, las características faciales obtienen la mayor puntuación para sistemas de lectura de documentación de viaje MRTD [6] (figura 4).

Según todos estos criterios que se acaban de ver, el reconocimiento facial destaca por ser una técnica con una alta capacidad de respuesta frente a

múltiples características biométricas y una buena aceptación lo cual hace de los sistemas basados en estos indicadores útiles para el desarrollo de aplicaciones no intrusivas.

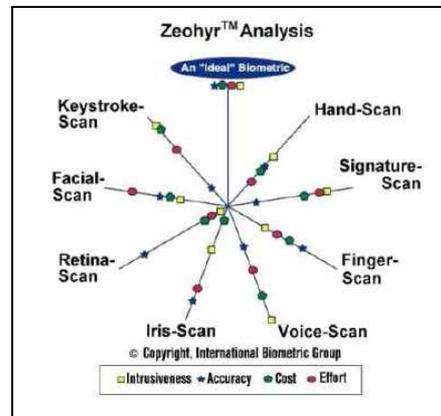


Figura 3. Clasificación de características biométricas según análisis de Zephyr.

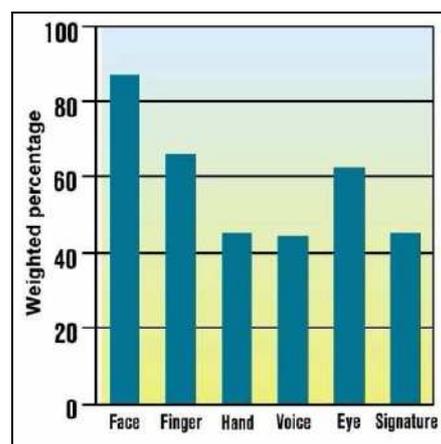


Figura 4. Clasificación de características biométricas según compatibilidad MRTD.

Por otro lado, las aplicaciones basadas en el reconocimiento facial, presentan algunas limitaciones o inconvenientes como pueden ser:

- Orientación del rostro.
- Ruido.
- Iluminación (incluyendo interior / exterior).
- Expresión facial.
- Oclusión debido a objetos o accesorios tales como gafas de sol, sombreros...
- Vello facial.
- Envejecimiento.

Aun así algunas de estas barreras, se puede solucionar de una forma eficaz como sería el caso del envejecimiento. Para ello bastaría con una actualización periódica de la base de datos de la aplicación. Un ejemplo de imágenes de un mismo individuo para evaluar el sistema frente a estas limitaciones, lo podemos ver en la figura 5 y figura 6.



Figura 5. Ejemplo de 10 imágenes de un individuo de la base de datos Yale [7] con diferente iluminación, expresión y postura.

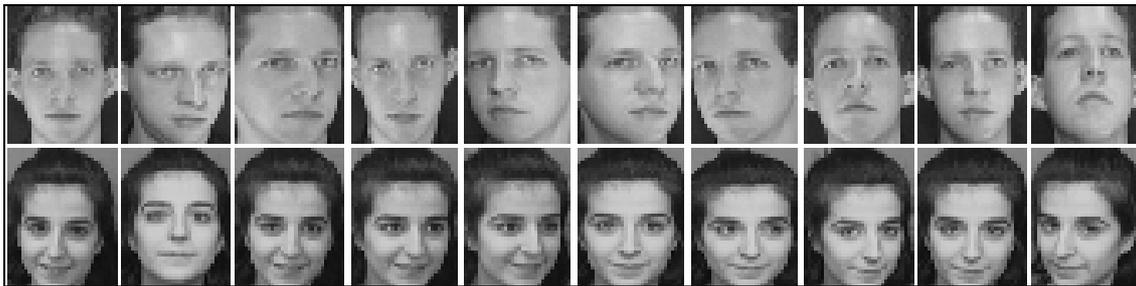


Figura 6. Ejemplos de imágenes de la base de datos ORL [18].

Debido a que las aplicaciones basadas en el reconocimiento facial [26], [27] no necesitan de la interacción del individuo, son aplicaciones de mucho interés para el entorno de video vigilancia, aeropuertos, comisarías, etc. De igual modo son aplicaciones necesarias y de utilidad en entornos controlados donde por ejemplo se pueda llevar a cabo el testeo de las aplicaciones antes de salir al mercado.

En toda aplicación de reconocimiento facial, se debe tener en cuenta diferentes etapas. Una de ellas es la etapa de detección facial, la cual en multitud de ocasiones se trata de manera conjunta con el reconocimiento. En este proyecto se tratarán por separado estos 2 pasos, en concreto se centrará en la etapa de

reconocimiento. De este modo el sistema de reconocimiento recibirá como entrada un conjunto de imágenes 'cara', ya detectadas dentro de una imagen en diferentes posiciones y condiciones ambientales, las cuales serán procesadas por nuestro sistema para después hacer el reconocimiento (figura 7).

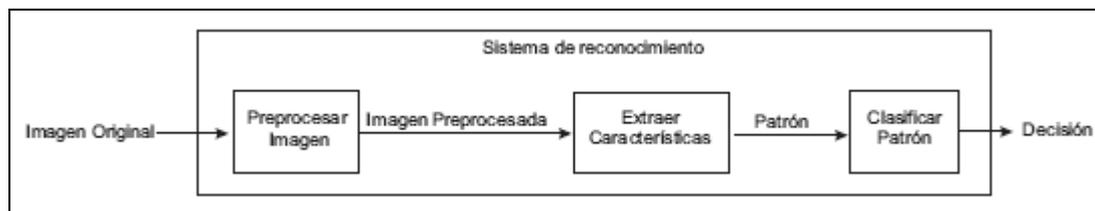


Figura 7. Diagrama de bloques general en un sistema de reconocimiento facial.

Finalmente se puede decir que estamos ante un problema de reconocimiento de patrones, pues el objetivo será el de clasificar y reconocer imágenes cara dependiendo de sus características. Dentro de esta rama se pueden encontrar 2 grandes categorías, sistemas supervisados y no supervisados. La diferencia entre ambos casos reside en que el caso supervisado conoce con antelación a quien pertenece cada imagen y asigna una etiqueta identificatoria a cada imagen, mientras que los no supervisados no lo hacen. En el caso de este proyecto estaremos frente a un caso no supervisado, pero cuando se hable de LPP (Locality Preserving Projections), si se tendrá la posibilidad de hacer uso de cierto etiquetado de las imágenes en el entrenamiento del sistema.

Por lo tanto en este proyecto se estudiarán diferentes técnicas de extracción de características en una imagen de una cara. Las tareas que desarrollaremos se pueden enumerar cronológicamente del siguiente modo:

- Estudio de las bases de reconocimiento de patrones.
- Estudio del entorno de trabajo.
- Estudio de las diferentes técnicas de selección de características, prestando especial atención a PCA (Principal Component Analysis), DCT (Discrete Cosine Transform) y LPP.
- Evaluación de los principales problemas como pueden ser los cambios en la iluminación y en la expresión o posición de la cara.

- Implementación del algoritmo en C++ de forma coherente con el entorno de trabajo proporcionado por las librerías del departamento de imagen de la universidad, *ImaGepplus*.
- Testeo de todo el sistema utilizando imágenes pertenecientes a diferentes bases de datos para simular diferentes situaciones.
- Creación de aplicaciones para ver el funcionamiento del sistema.

Durante este proyecto se compararán diferentes técnicas de reducción dimensional y clasificación de las imágenes utilizadas para el reconocimiento facial. Estos resultados serán útiles como punto de partida para estudios que se realicen posteriormente. Otro objetivo que se tendrá es el de crear un software modular usando la plataforma *ImaGepplus* del Grupo de Procesado de Imagen y Video de la UPC, el cual contiene diferentes librerías con multitud de herramientas. Esta plataforma está constantemente actualizada con nuevas contribuciones de otros miembros del grupo de imagen de la UPC, de manera que al final de este proyecto, ciertas partes del código que se desarrolle formará parte de esta plataforma.

Este proyecto está organizado del siguiente modo. En el capítulo 2 se estudiará el estado del arte relacionado con técnicas de reconocimiento facial. Después en el capítulo 3 se detallará como funcionan los métodos que se han seleccionado para la comparativa. En el capítulo 4 se hará una breve descripción de las diferentes bases de datos que se han utilizado. En el capítulo 5 se mostrarán los resultados obtenidos frente a diferentes situaciones. En el capítulo 6 y 7 se presentarán dos aplicaciones haciendo uso de los algoritmos estudiados. Finalmente, en el capítulo 8, se mostrarán las conclusiones y se enfocará el trabajo futuro.

2 Estado del arte

Hay dos familias de técnicas de reconocimiento facial: técnicas basadas en la apariencia y técnicas basadas en modelos [8]. En cada una de estas familias, se encuentran varios métodos para caracterizar la imagen, aunque en este estudio solo se tratarán algunos de los métodos basados en la apariencia (figura 8).

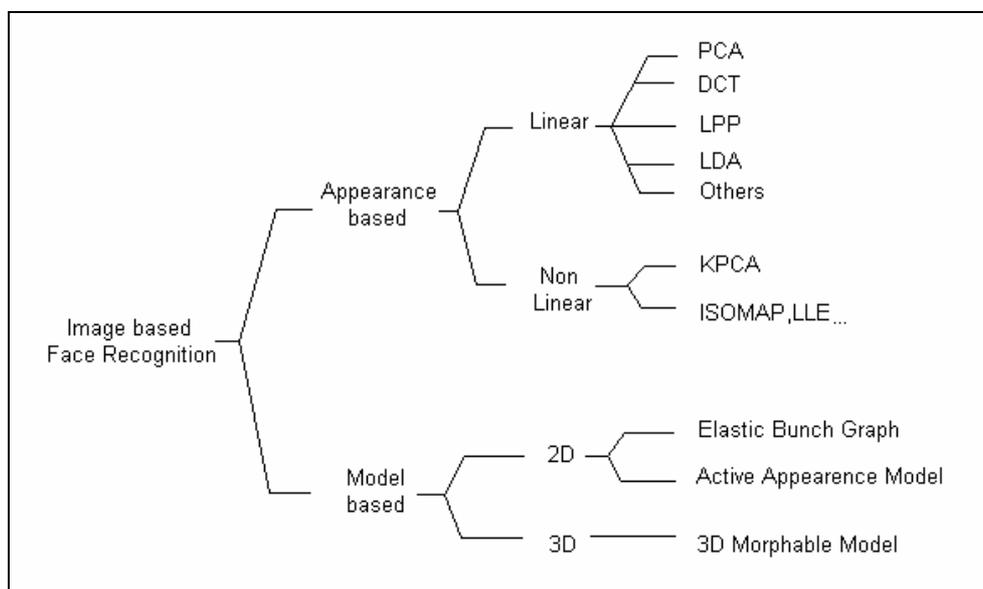


Figura 8. Algunos métodos de clasificación

Los sistemas basados en la apariencia se utilizan directamente sobre las imágenes sin hacer uso de modelos 3D. Estos tipos de sistemas representan un objeto en función de diferentes vistas del mismo. En estos sistemas cada imagen se representa como un punto en un subespacio vectorial, de forma que la comparación entre la imagen de test y las imágenes de referencia se realiza en el subespacio vectorial caras. El objetivo de estos algoritmos es clasificar las diferentes caras en el nuevo subespacio, pero para ello será necesario entrenar previamente el sistema con imágenes de diferentes caras con diferentes vistas. Estas técnicas serán explicadas con más detalles en el capítulo 2.2 donde se expondrán los métodos de clasificación basados en la apariencia que se han utilizado.

Por otro lado están los sistemas basados en modelos (figura 9), los cuales intentan construir un modelo lo más descriptivo posible de la cara humana capaz de detectar con precisión las variaciones faciales.

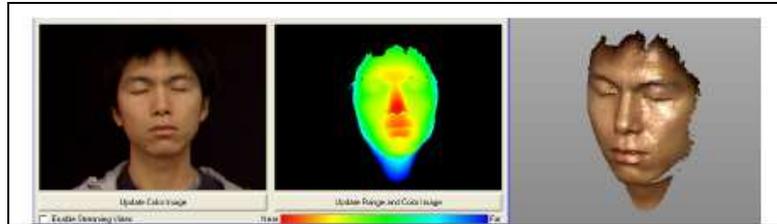


Figura 9. Imagen 2D, mapa de profundidad y representación 3D del modelo.

Estos sistemas tratan de obtener características biométricas de las imágenes para realizar el reconocimiento (distancia entre ojos (figura 10), grosor de la nariz...). Habitualmente estas técnicas requieren de imágenes de gran resolución. Cuando se utilizan estos sistemas, el algoritmo sabe con antelación el objeto que ha de representar y lo que intenta hacer es que corresponda la cara real con el modelo. El proceso que se suele seguir cuando se usan estas técnicas está formado por tres pasos [9]:

- Construcción del modelo.
- Ajustar el modelo a la imagen de test.
- Utilizar los parámetros del modelo ajustado para calcular la similitud entre la imagen de test y las imágenes de referencia para realizar el reconocimiento.

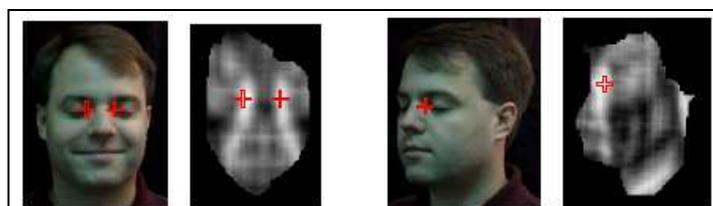


Figura 10. Detección de la posición y distancia de los ojos a través del método de convolución

Algunas de las ventajas e inconvenientes de los sistemas basados en la apariencia frente a los basados en modelos son las siguientes:

- Ventajas:
 - Más rápidos.

- Requieren de un menor tamaño de las imágenes.
- Menor complejidad.
- No requieren de un conocimiento previo de las imágenes.
- Inconvenientes:
 - Más afectados por cambios en la orientación o expresión de la cara.
 - Más dificultades frente a cambios en la iluminación.

A continuación se hará una pequeña descripción de la literatura que existe sobre algunos de los métodos más relevantes de reconocimiento facial y una pequeña explicación de cómo se va a realizar el reconocimiento de las imágenes.

2.1 Reconocimiento de imágenes fijas

En este estudio se hará uso de métodos basados en la apariencia, los cuales para el reconocimiento de imágenes requiere de los bloques que podemos encontrar en la figura 11.

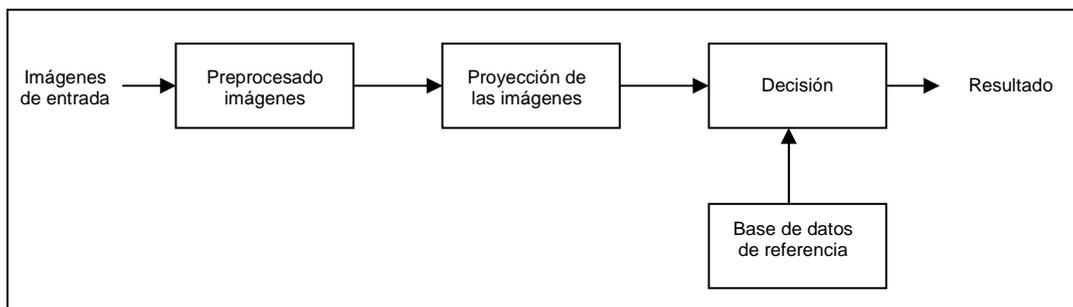


Figura 11. Diagrama de bloques del sistema.

Para realizar los diferentes experimentos que se realizan en este estudio el sistema hace uso de los siguientes conjuntos de imágenes:

- *Imágenes de referencia*: son las que están almacenadas y son conocidas por el sistema y son utilizadas para saber si una imagen de entrada pertenece o no a algún sujeto registrado en el sistema.

- *Imágenes de test*: son las que recibiremos en el sistema y se tienen que reconocer.
- *Imágenes de entrenamiento*: son utilizadas en los métodos PCA y LPP para conseguir las matrices de proyección.

Los pasos que se llevan a cabo para este reconocimiento, sin tener en cuenta el preprocesado previo de las imágenes, son los que se detallan a continuación [34]:

- 1) El sistema necesita de un set de imágenes de referencia las cuales son procesadas y proyectadas ($\underline{y} = \underline{W} \cdot \underline{x}$) en el nuevo subespacio para después ser utilizadas para conocer si una imagen pertenece o no a un individuo registrado en el sistema (figura 12).

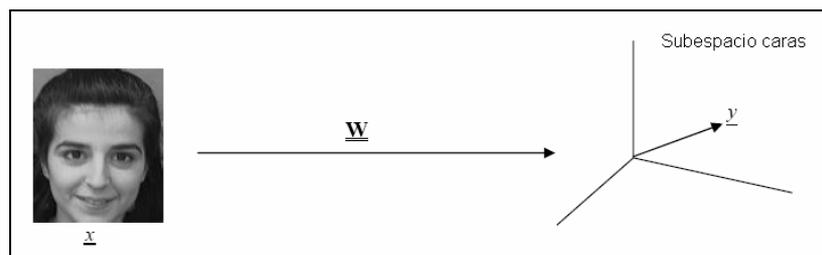


Figura 12. Esquema de proyección de una imagen de referencia al nuevo subespacio [34].

- 2) Cuando las imágenes de referencia están correctamente proyectadas, se realiza el mismo procedimiento con las imágenes de test (figura 13). Estas pueden pertenecer a personas que estaban incluidas entre las imágenes de referencia pero mostrando diferentes expresiones faciales o diferentes condiciones del entorno.

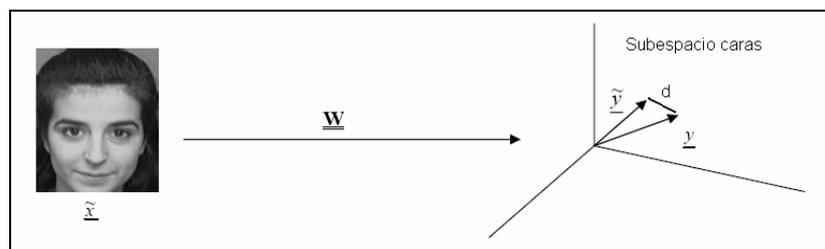


Figura 13. Ejemplo de proyección de una imagen de test [34].

- 3) Finalmente se dirá que ha habido reconocimiento si se cumple alguna condición, en general se considera que hay un reconocimiento positivo si se cumple una condición:

$$d \Rightarrow \left| \underline{\tilde{y}} - y \right| < d_{th}, \text{ donde } d_{th} \text{ es una distancia threshold. (1)}$$

En concreto para el caso de este estudio, una imagen de test pertenece a otra del conjunto de imágenes de referencia si es la que cumple distancia mínima de manera que siempre se tiene que asignar un sujeto a una imagen de entrada.

2.2 Descripción de las principales técnicas

Como en este estudio se va a hacer uso de técnicas basadas en la apariencia, a continuación se describen brevemente los principales métodos que se pueden encontrar.

2.2.1 PCA (Principal Component Analysis)

PCA es un método que transforma un número de variables posiblemente correladas en un pequeño número de variables incorreladas llamadas componentes principales, es decir, es un algoritmo de reducción dimensional que permite encontrar los vectores que mejor representan la distribución de un grupo de imágenes.

PCA esta basado en la Transformada de Karhunen-Loeve (KLT) [10], que consiste en la representación de un proceso estocástico no periódico a través de una base de vectores obtenidos completamente del proceso en sí mismo, es decir, PCA permite representar una imagen de una cara usando una base que se ha conseguido a partir de muchas observaciones de diferentes caras.

El objetivo de este método consiste en representar una imagen en términos de un sistema de coordenadas óptimo reduciendo el número final de componentes que tendrá la imagen. Un ejemplo de PCA lo podemos ver en la figura 14.

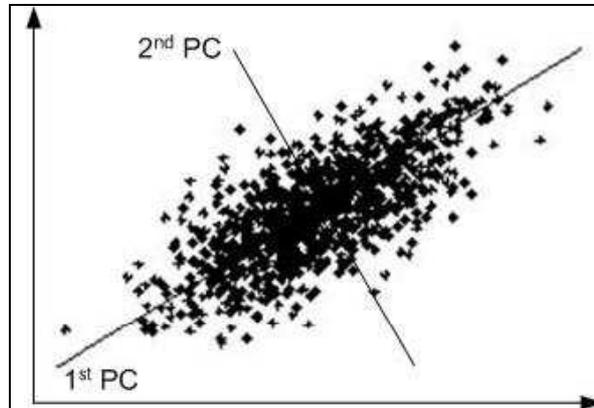


Figura 14. Componentes principales de un conjunto de puntos bidimensional. La primera componente ofrece una reducción lineal óptima de dimensión de 2D a 1D en cuanto a error cuadrático medio se refiere.

Esta base de transformación depende de las observaciones y se forma a partir de los d autovalores más significativos de la matriz de covarianza, los cuales representan las componentes principales que son más comunes en imágenes de diferentes caras.

La reducción dimensional realizada por PCA es equivalente al número de autovectores que se utilicen. Por lo tanto la imagen proyectada por PCA tendrá una dimensión de valor d , como se puede ver en la figura 15.

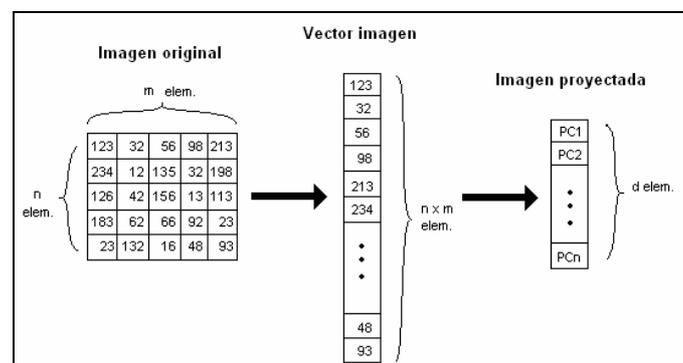


Figura 15. Ejemplo de reducción dimensional al aplicar PCA.

Estos autovectores representan las componentes principales que son más comunes en imágenes de diferentes caras, otra forma de verlo, es que son los

ejes del subespacio donde proyectaremos las imágenes de caras humanas los cuales se corresponden con las direcciones de máxima varianza de los datos. La matriz de transformación, esta formada por los autovectores correspondientes a los d autovalores más significativos. En la figura 16 se muestran los primeros 7 autovectores que derivan del uso de la base de datos ORL [18]. La imagen media correspondiente a estos 7 autovectores la podemos ver en la figura 17 y en la figura 18 podemos ver las imágenes correspondientes a los 7 autovalores más bajos. Estos últimos se corresponden con autovalores muy pequeños y son considerados como ruido de manera que no se tienen en cuenta para el reconocimiento.

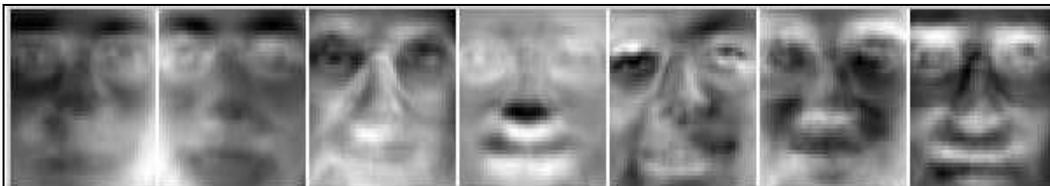


Figura 16. Autovectores correspondientes a los 7 mayores autovalores mostrados como imágenes $pxp = n$ (Obtenidos de la base de datos de caras ORL [18]).



Figura 17. Cara media de los 7 primeros autovectores (Obtenidos de la base de datos de caras ORL [18]).



Figura 18. Autovectores correspondientes a los 7 menores autovalores mostrados como imágenes $pxp = n$ (Obtenidos de la base de datos de caras ORL [18]).

Otra de las cualidades de PCA, es que realiza una discriminación entre clases, es decir, todas las caras correspondientes a un mismo individuo están cerca entre si, mientras que las imágenes correspondientes a individuos diferentes están más alejadas. Esta propiedad es más destacable en las primeras componentes de las imágenes proyectadas. En la figura 19 se muestra un

ejemplo de la proyección de tres imágenes de personas diferentes mostrando diferentes componentes.

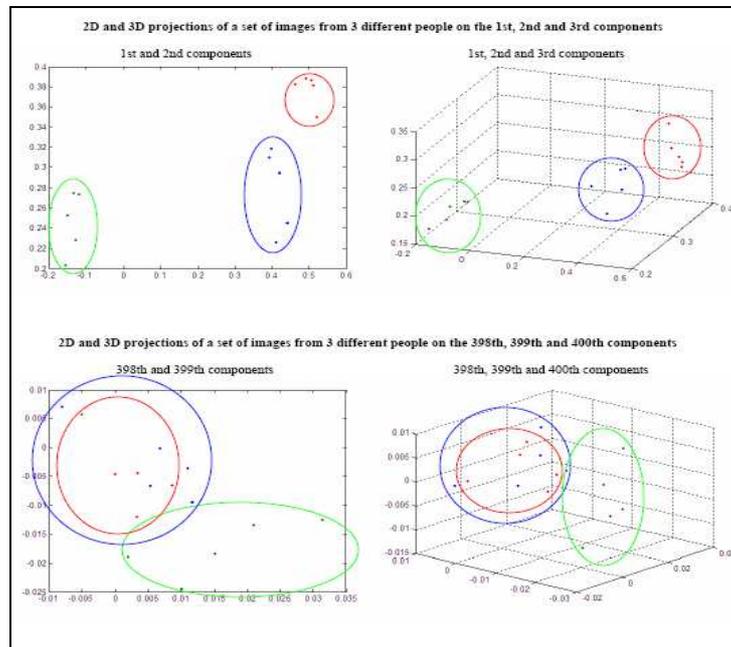


Figura 19. Representación de diferentes componentes de tres imágenes de individuos distintos proyectadas.

2.2.2 LDA (Linear Discriminant Analysis)

LDA tiene como objetivo convertir un problema de alta dimensionalidad en uno de baja [15, 16]. Para ello LDA proyecta los datos (imágenes) en un espacio vectorial de baja dimensionalidad de manera que la ratio entre la distancia entre clases y la distancia dentro de la clase se maximiza. De este modo se garantiza una máxima discriminación entre las clases.

A diferencia de PCA, que es un método de reducción dimensional sin supervisión, LDA si que es supervisado y utiliza información de los datos. Para ello lo que se hace es maximizar la siguiente expresión:

$$J(w) = \frac{w^T \cdot S_B \cdot w}{w^T \cdot S_W \cdot w} \quad (2)$$

Donde S_B es la “matriz de dispersión entre clases” mientras que S_W es la “matriz de dispersión intra clase” y se definen como:

$$\begin{aligned}
 S_B &= \sum_c N_c \cdot (\mu - \bar{x}) \cdot (\mu_c - \bar{x})^T \\
 S_W &= \sum_c \sum_{i \in c} (x_i - \mu_c) \cdot (x_i - \mu_c)^T \\
 \mu_c &= \frac{1}{N_c} \cdot \sum_{i \in c} x_i \\
 \bar{x} &= \frac{1}{N} \cdot \sum_i x_i = \frac{1}{N} \cdot \sum_c N_c \cdot \mu_c
 \end{aligned} \tag{3}$$

y N_c es el número de casos dentro de la clase c .

Una vez esta resuelto este problema, se obtiene la matriz de proyección que permite pasar de un problema de alta dimensionalidad a uno de baja. Cada una de las columnas que forman esta matriz serán las bases del nuevo subespacio, las cuales se conocen como FisherFaces (figura 20). Estas bases son aquellas que garantizan una máxima discriminación entre clases por encima de ser las que mejor describen los datos, a diferencia de PCA [17] (figura 21).

Una de las limitaciones que tiene LDA, es que se requiere de matrices de dispersión no-singulares. Esto en aplicaciones de reconocimiento facial es complicado debido a que en muchas ocasiones la alta dimensionalidad de las imágenes es mucho mayor que el número de imágenes lo que desemboca en un problema de matrices singulares, algo que como se verá más adelante también puede suceder en LPP, conocido como *undersampled* o *singularity problem*. Para evitar este efecto, se utilizan diferentes aproximaciones como pseudo-inverse LDA [16], que utiliza la pseudo-inversa de la matriz de covarianza en lugar de su inversa, y PCA + LDA [16], que realiza un reducción previa de los datos utilizando PCA.

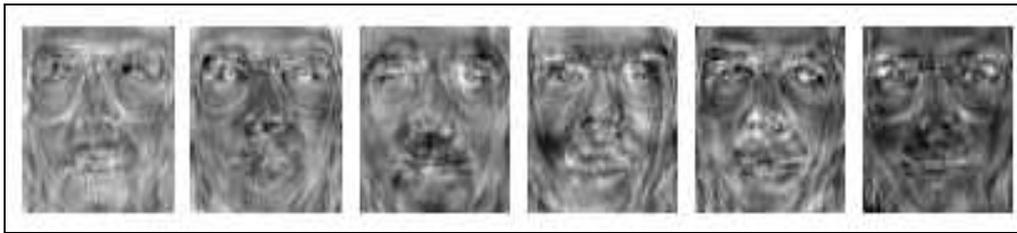


Figura 20. Las 6 Primeras FisherFaces calculadas en la base de datos ORL

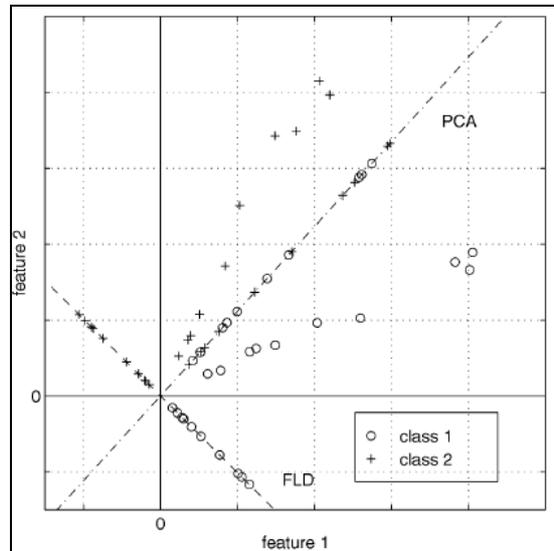


Figura 21. Comparación de la proyección de 2 clases en PCA y LDA.

2.2.3 LPP (Locality Preserving Projections)

LPP [14] es un algoritmo lineal que del mismo modo que PCA realiza una reducción dimensional de los datos. Al tratarse de un algoritmo lineal es rápido y útil para aplicaciones prácticas.

Una de las propiedades que lo diferencian de PCA es que en lugar de conservar la estructura global de los datos, conserva la estructura local (figura 22). De este modo los 'vecinos' para un dato en concreto serán los mismos en el espacio original, de alta dimensionalidad, y en el nuevo subespacio de baja dimensionalidad [13]. Al conservarse la estructura local de los datos, las imágenes pertenecientes a un mismo individuo estarán cercanas entre sí y alejadas de las de otros individuos, es decir, hay una discriminación entre clases (figura 23).

Para conservar la estructura local de los datos se hace uso de un grafo de adyacencias que incluye información de la estructura de los datos. Este grafo, como se verá con más detalle en el apartado 3.4.1, consiste en la creación de una matriz de tamaño $N \times N$, donde N es el número de imágenes, que tiene asignados unos pesos dependiendo de si los elementos i y j son vecinos o no. En el momento de crear este grafo cabe la opción de crearlo de manera automática (caso no supervisado), utilizando métodos de búsqueda de ‘vecinos’ como *K nearest neighbors* o *ϵ -neighborhoods*, o hacerlo de manera manual (caso supervisado) asignando manualmente quienes son o no vecinos. Todas estas configuraciones serán vistas con más detalle en el apartado 3.4.

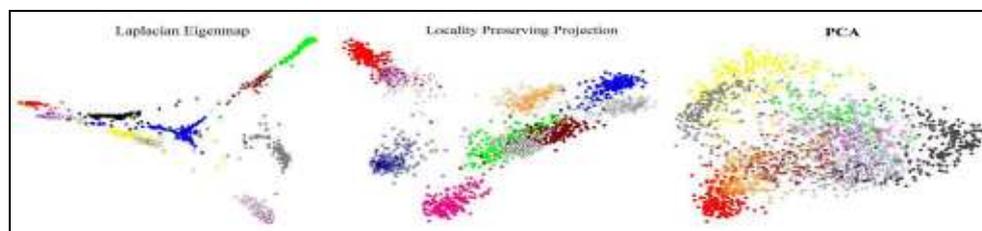


Figura 22. En esta imagen podemos ver como a diferencia de PCA, en LPP se conserva la estructura local de los datos.

Este método presenta algunos inconvenientes. Uno de ellos, aunque para el caso de este estudio no afecta, es la dificultad de recuperar los datos originales a partir de los datos proyectados al nuevo subespacio, esto se debe al hecho de hacer uso de bases no ortogonales.

Otro inconveniente que se puede dar es el de las matrices singulares, que veremos con más detalle en el apartado 3.4.1, el cual se produce cuando la dimensionalidad de los datos es mucho más elevada que el número de muestras ($n \gg N$). Para solucionar este problema se hace uso de técnicas de reducción dimensional de los datos antes de utilizarlos de manera que $n = N$ o $n > N$.

Por otro lado, este método es interesante por diferentes perspectivas:

- Los mapas están diseñados para minimizar algún criterio objetivo procedente de las técnicas lineales clásicas.

- Como LPP preserva la estructura local de los datos, el sistema tendrá los mismos 'vecinos' en el espacio de baja dimensionalidad que en el de alta.
- LPP es un método lineal, lo que hace que sea rápido y conveniente para aplicaciones prácticas.
- LPP esta definida para todos los casos, a diferencia de las técnicas de reducción no lineales que solo están definidos para el conjunto de datos de entrenamiento, de manera que al tener un nuevo dato LPP es capaz de representarlo en el nuevo subespacio sin tener que recalcularse las matrices de proyección.

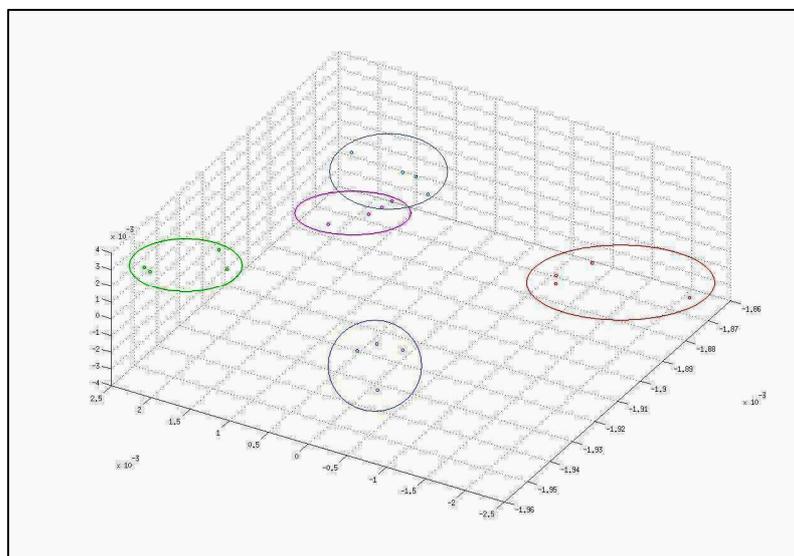


Figura 23. Representación de la estructura de los datos en el nuevo subespacio. Este ejemplo esta formado por 5 personas y cuatro imágenes por persona.

2.2.4 DCT (Discrete Cosine Transform)

La DCT es una transformación que representa una secuencia finita de datos como la suma de una serie de funciones cosenoidales oscilando a diferentes frecuencias. Esta técnica es muy utilizada en aplicaciones de procesado de señal, desde compresión de audio e imágenes hasta métodos espectrales para la solución numérica de ecuaciones diferenciales. Una de estas aplicaciones es el reconocimiento facial [11].

En este trabajo, la DCT será utilizada para obtener características de la imagen y luego usarlas para poderlas clasificar, como se hace con PCA [12]. A diferencia de PCA, este método no necesita ser entrenado con imágenes del mismo tipo a las que se van a usar sino que simplemente se transforman directamente las imágenes, es decir, la base de la transformación es independiente de las imágenes. Otra de sus ventajas es su bajo coste computacional en relación con PCA [12].

Se definen 8 variantes de DCT, de las cuales las más relevantes son: DCT-I, DCT-II, DCT-III y DCT-IV. En lo que se refiere al reconocimiento facial, la transformación más utilizada es DCT-II, que se define como:

$$v(k_1, k_2) = \alpha(k_1) \cdot \alpha(k_2) \cdot \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} u(n, m) \cdot \cos\left(\frac{(2 \cdot n + 1) \cdot \pi \cdot k_1}{2 \cdot N}\right) \cdot \cos\left(\frac{(2 \cdot m + 1) \cdot \pi \cdot k_2}{2 \cdot M}\right)$$

$$0 \leq k_1 \leq N - 1$$

$$0 \leq k_2 \leq M - 1$$
(4)

Donde $u(n)$ es una secuencia de N elementos, $v(k)$ es su transformada y:

$$\alpha(0) = \sqrt{\frac{1}{N \cdot M}}, \quad \alpha(k_1) = \sqrt{\frac{2}{N}}, \quad \alpha(k_2) = \sqrt{\frac{2}{M}} \quad 1 \leq k_1 \leq N - 1, \quad 1 \leq k_2 \leq M - 1$$
(5)

Si la interpretación de esta fórmula es que la entrada al sistema es una imagen \underline{u} y la DCT es la matriz de transformación para obtener \underline{w} , se tiene una expresión como la expuesta en el capítulo 2.1 donde $\underline{w} = c(k, \xi) = c(k_1, n) \cdot c(k_2, m)$ y k y ξ son las filas y columnas:

$$c(k, \xi) = \begin{cases} \frac{1}{\sqrt{N \cdot M}} & k=0, 0 \leq n \leq N-1, 0 \leq m \leq M-1 \\ \sqrt{\frac{2}{N}} \cdot \sqrt{\frac{2}{M}} \cdot \cos\left(\frac{(2 \cdot n + 1) \cdot \pi \cdot k_1}{2 \cdot N}\right) \cdot \cos\left(\frac{(2 \cdot m + 1) \cdot \pi \cdot k_2}{2 \cdot M}\right) & 1 \leq k_1 \leq N-1, 0 \leq n \leq N-1, 1 \leq k_2 \leq M-1, 0 \leq m \leq M-1 \end{cases}$$
(6)

De esta manera se puede tratar este algoritmo del mismo modo que PCA o LPP, es decir, como un sistema que tiene una matriz de proyección (en este

caso única para cada imagen) gracias a la cual se puede proyectar las imágenes a un subespacio de baja dimensionalidad de igual modo que se hace en PCA (figura 15).

2.2.5 DCT por Bloques

Este método realiza una mezcla de las técnicas basadas en la apariencia y las basadas en modelos, en concreto, hace uso de la misma metodología que el método DCT pero aplicado de forma distinta.

Lo que se hace es aplicar DCT-II sobre bloques de 8x8 píxeles de la imagen [19]. Una vez calculada la DCT sobre cada uno de estos bloques de 8x8 píxeles, se selecciona un determinado número de coeficientes por bloque con el fin de crear la nueva representación de la imagen (figura 24).

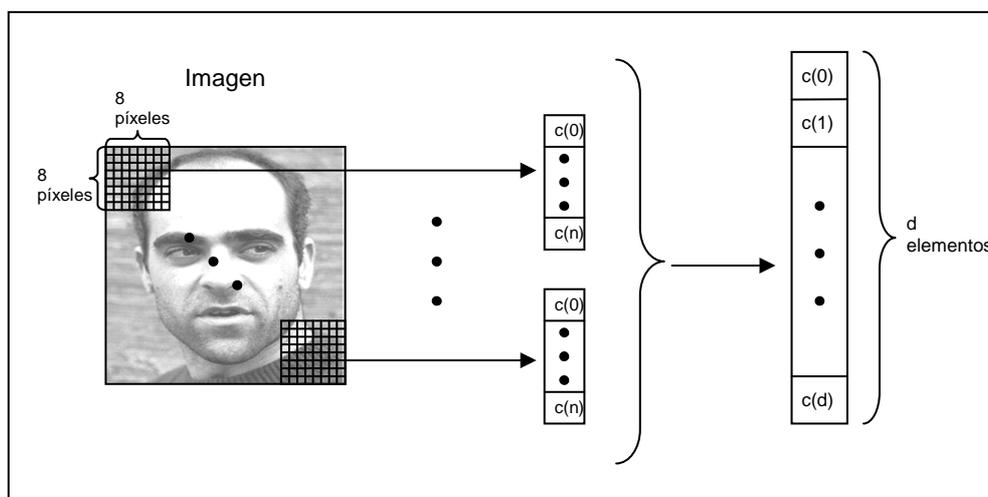


Figura 24. Esquema de funcionamiento del sistema DCT por bloques.

Este sistema hace uso parcial de los sistemas basados en modelos porque requiere de la localización de alguna característica facial. En este caso para el buen funcionamiento del sistema se necesita que las imágenes estén recortadas y totalmente alineadas [12]. Por esta razón previamente a aplicar el algoritmo, se debe detectar la posición de los ojos y alinear la imagen según un criterio (figura 25). En caso de que no se realice esta alineación previa de las imágenes, el sistema no funcionará correctamente. En la figura 26 se puede ver el efecto causado al no realizar una correcta alineación de las caras.

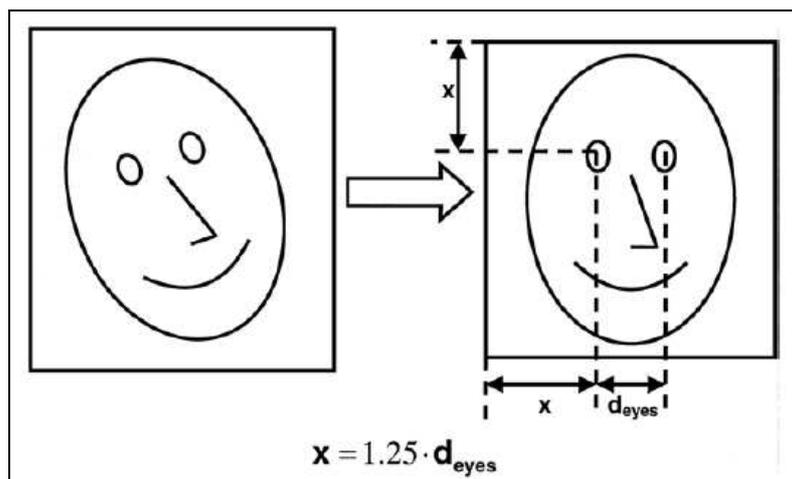


Figura 25. Alineación de las caras usando DCT por bloques [12].

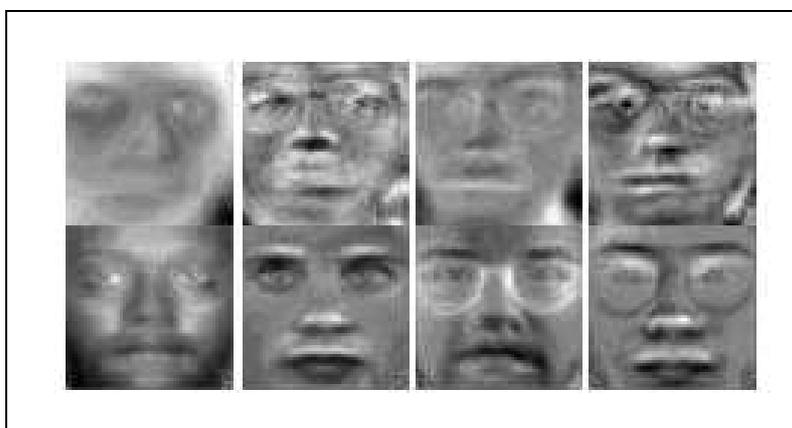


Figura 26. Autocaras calculadas sin alinear (superior) y alineadas (inferior)

3 Arquitectura del sistema

En este capítulo se darán a conocer los diferentes procedimientos que se han realizado para hacer uso de cada uno de los algoritmos así como del tratamiento previo que se hará sobre las imágenes.

Durante el resto de este documento, se tratarán todas las imágenes de manera similar al protocolo FERET [35], pero en nuestro caso las imágenes no se alinean previamente ni se enmascaran, es decir, dada una imagen (matriz de $N \times M$ valores) se concatenarán las filas formando un vector de $N \cdot M$ elementos con los valores de los píxeles (en escala de grises de 256 niveles). En general se usa un conjunto de imágenes de golpe, por lo que el procedimiento será el siguiente.

Se tiene $X = \{x_1, x_2, \dots, x_i, \dots, x_N\}$ que representa la matriz de datos $n \times N$ donde cada x_i es un vector cara de dimensión n que ha sido construido a partir de la imagen original de tamaño $p \times q$ ($n = p \times q$). En este caso n representa el total de píxeles que contenían las imágenes originales, mientras que N se corresponde con el número de imágenes que vamos a utilizar.

Como se ha dicho en la introducción, en este proyecto se hará una comparación de diferentes métodos de reconocimiento basados en la apariencia. El esquema correspondiente al sistema utilizado durante el análisis de las diferentes capacidades de los métodos, es el siguiente (figura 27):

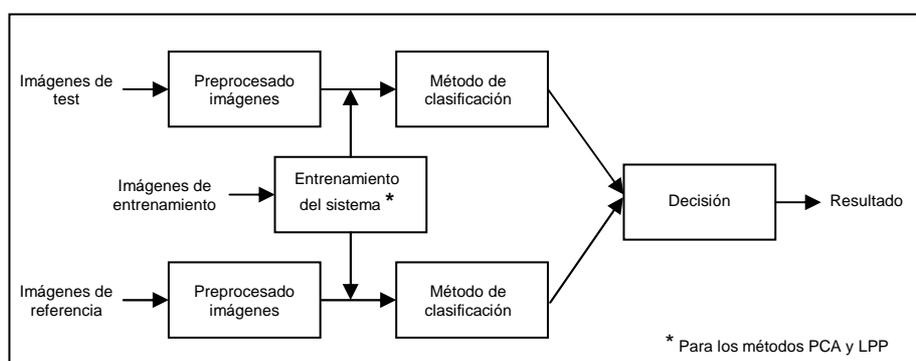


Figura 27. Esquema de funcionamiento del sistema.

3.1 Análisis Lineal

Todos los sistemas basados en la apariencia que se van a estudiar en este proyecto (PCA, LPP y DCT que lo hace de una forma indirecta), realizan un procedimiento parecido con el que a través de una matriz de proyección transformarán las imágenes de entrada a un nuevo subespacio. Cada uno de estos clasificadores, tiene su propia representación (vectores base) en un espacio vectorial carac de alta dimensionalidad basado en diferentes estadísticas.

Al proyectar las imágenes de test en el nuevo espacio vectorial, se utilizan los coeficientes de dichas proyecciones como la nueva representación de cada una de las imágenes. Para determinar si una imagen pertenece o no a un individuo en concreto de la base de datos, se aplica algún método de medidas (por ejemplo, la distancia Euclídea) sobre estas imágenes proyectadas.

Todas estas representaciones son consideradas como transformaciones lineales del vector imagen original a un vector proyectado de características:

$$\underline{y} = \underline{W}^T \cdot \underline{x} \quad (7)$$

Donde y es el vector de características (imagen proyectada) de dimensión d , d se corresponde con el número de componentes que se tendrán en cuenta para el reconocimiento, W es la matriz de transformación que se usará para cada caso, sea PCA, DCT o LPP y x es la imagen de entrada.

3.2 PCA

PCA es un proceso que permite realizar una reducción dimensional, de manera que nuestro análisis se reduce a unas pocas componentes. Para ello se necesita tener la matriz de proyección PCA. A continuación se detallarán los pasos a seguir para preparar esta matriz y el procedimiento a seguir para el reconocimiento.

3.2.1 Creación de la matriz de proyección

Para la creación de la matriz de proyección, se entrena PCA con N imágenes (\underline{x}_i), que son de diferentes personas con diferentes vistas (frontales, de perfil, con cambios de iluminación, con diferentes gestos faciales, con vello facial y sin el, con y sin gafas...).

El procedimiento que se realiza para calcular \underline{W} (matriz de proyección) consta de los siguientes pasos:

- 1) Calcular la media del vector:

$$\underline{\mu} = \sum_{i=1}^N \underline{x}_i \quad (8)$$

- 2) Estimar la matriz de covarianza:

$$\underline{S}_T = \sum_{i=1}^N (\underline{x}_i - \underline{\mu}) \cdot (\underline{x}_i - \underline{\mu})^T \quad (9)$$

- 3) Calcular los autovectores y autovalores de \underline{S}_T y generar \underline{W} de forma que sus columnas son los autovectores correspondientes a los autovalores más significativos (con mayor energía):

$$\underline{S}_T \Rightarrow \underline{W} = (e_1 \ e_2 \ \dots \ e_M) \quad (10)$$

3.2.2 Reconocimiento

Cuando se tiene una nueva imagen de test en el sistema, el procedimiento que se realiza es el siguiente:

- 1) Primero se redimensionan la imagen de test al tamaño que se ha utilizado para la creación de la matriz de proyección PCA (este valor esta almacenado conjuntamente con toda la información referente a la clase PCA). Después de analizar diferentes estudios, se ha decidido utilizar imágenes de tamaño 32 x 32 píxeles [33].
- 2) Se normaliza en media 0 y norma 1.
- 3) Se quita la media de la clase PCA por filas.

Una vez realizados estos tres pasos la imagen está lista para ser proyectada al nuevo subespacio. En este procedimiento toda imagen de entrada al sistema \underline{x} (imagen facial), es proyectada al subespacio 'caras' por el método de PCA, tal como hemos visto en la ecuación 7, donde \underline{W} es la matriz de proyección PCA, \underline{y} es la imagen proyectada en el nuevo subespacio de baja dimensionalidad y \underline{x} es la imagen de entrada. Después de realizar esta proyección, el vector (imagen) n-dimensional de entrada se reducirá a un vector d-dimensional en el subespacio 'caras' (figura 15). Este nuevo tamaño del vector dependerá del número de componentes principales que se utilicen.

3.3 DCT

DCT puede ser utilizada como otro sistema de clasificación basado en la apariencia, aunque en este caso la matriz de transformación es fija e independiente de la imagen. A diferencia de PCA, este método no necesita de un entrenamiento previo del sistema con otras imágenes.

3.3.1 Reconocimiento

Cuando el sistema recibe una nueva imagen de test se realizan los siguientes pasos para llevar a cabo el reconocimiento:

- 1) Se escala la imagen a un cierto tamaño predefinido, en este sistema se ha seleccionado un tamaño de 32 x 32 píxeles [33].
- 2) Se calcula la DCT-II de la imagen original escalada (figura 28).

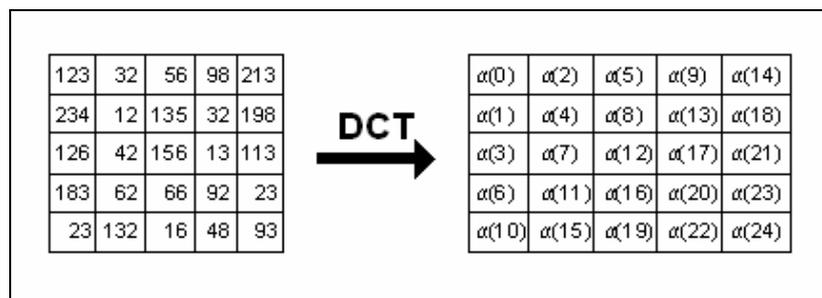


Figura 28. Ejemplo de aplicación de la DCT sobre una imagen.

- 3) Se seleccionan en zig-zag los d coeficientes que se utilizarán para el reconocimiento y se forma un vector columna (figura 29). Para hacer esta selección, no se tiene en cuenta $\alpha(0)$ que representa la componente continua y se pueden descartar los primeros coeficientes para tener más control sobre los cambios de luminosidad ya que estos son los que atesoran una mayor energía y son los que se ven más afectados por los cambios de iluminación. El efecto de descartar estos primeros coeficientes lo podremos ver en el apartado 5.2.2. Como sucedía con las últimas componentes principales de PCA, los últimos coeficientes obtenidos de aplicar la DCT, tampoco son muy relevantes.

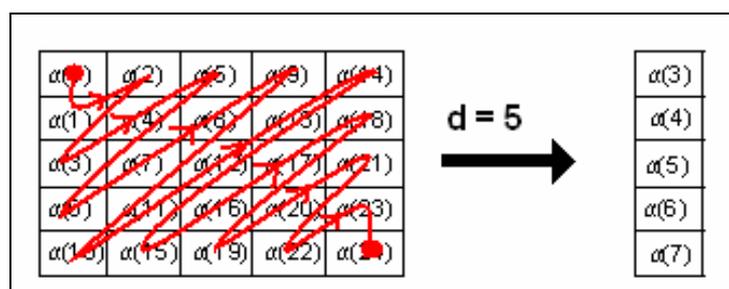


Figura 29. Ejemplo de selección en zigzag de los d elementos.

- 4) Se compara la imagen de test con la imagen de referencia y se decide. Para ello se vuelve a utilizar un criterio de distancias, en este caso la distancia Euclídea. Nuevamente el criterio de decisión será el de distancia mínima.

3.4 LPP

Este método realiza una reducción dimensional y requiere, como PCA, de un entrenamiento previo del sistema con otras imágenes de caras para crear una matriz de proyección.

3.4.1 Creación de la matriz de proyección

Para la creación de esta matriz de transformación, como se hace en PCA, se entrenará el sistema con un conjunto de \mathbf{N} imágenes \mathbf{x}_i de personas diferentes tomadas bajo condiciones diferentes. El procedimiento que se sigue para encontrar \mathbf{A} , matriz de proyección, consta de los siguientes pasos:

- 1) Se preprocesan las imágenes del mismo modo que se hace para PCA en la etapa de reconocimiento (pasos 1, 2 y 3 del capítulo 3.2.2).
- 2) Se proyectan las imágenes haciendo uso de la matriz de proyección PCA para evitar las matrices singulares.
- 3) Una vez se tienen proyectadas las imágenes mediante PCA, se construye un grafo de adyacencia. Este grafo se forma de manera que las imágenes consideradas como vecinas tienen un conector entre ellas. Se pueden diferenciar dos casos:
 - a. ε -neighborhoods. (El parámetro ε pertenece a \mathbf{R}) Los nodos i y j están conectados si $\|\mathbf{x}_i - \mathbf{x}_j\|^2 < \varepsilon$, donde la norma es la norma Euclídea usual en \mathbf{R}^n .
 - b. K-nearest neighbors. (k pertenece a \mathbf{N}). Los nodos i y j están conectados entre si en el caso en que i sea de los k vecinos más cercanos a j o j sea de los k vecinos más cercanos de i .

Se debe tener en cuenta que también es posible asignar de forma manual estos conectores en el grafo, caso supervisado, si se dispone de esta información con anterioridad, entonces se asignará un conector entre las imágenes que pertenezcan a mismos individuos. En el caso concreto de este proyecto se hace uso de esta variante supervisada para la creación del grafo de adyacencia.

- 4) Escoger los pesos. Hay 2 métodos para asignar estos pesos. La matriz que se crea, \mathbf{W} , es una matriz simétrica de $d \times d$ elementos que tendrá un valor, dependiendo del método que se utilice, si hay una conexión entre los elementos i y j en el grafo de adyacencia o cero en caso de no conexión. Los métodos que se usan son:

- a. Heat kernel. Necesita de un parámetro t que dará el usuario. En caso de que los nodos i y j estén conectados se asigna a ese elemento el valor:

$$W_{ij} = e^{-\frac{\|x_i - x_j\|^2}{t}} \quad (11)$$

- b. Simple-minded. Este sería el caso binario, es decir, si hay conexión entre los nodos i y j entonces se asigna un 1 a ese elemento (W_{ij}), 0 en caso contrario (figura 30). Como se verá en el capítulo 5 este es el método utilizado para la realización del estudio.

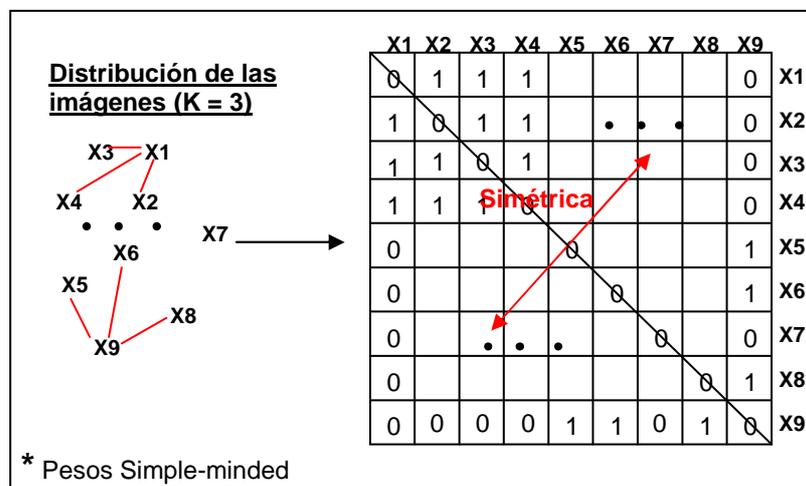


Figura 30. Ejemplo de matriz de pesos utilizando método Simple-minded.

5) Solucionar la ecuación de autovalores. Se calculan los autovectores y autovalores correspondientes a la solución de la ecuación general de autovectores. Para ello se hace uso de una matriz diagonal (**D**), que indica la relevancia que tendrá el dato y_i , formada a partir de la suma por columnas de **W**, o por filas al ser **W** simétrica, y de la matriz Laplaciana (**L**). Una vez solucionada la ecuación, se obtienen los autovectores a_0, \dots, a_{l-1} ordenados de menos a mayor autovalor ($\lambda_0 < \dots < \lambda_{l-1}$), como es posible observar a diferencia de PCA se utilizarán los autovectores correspondientes a los menores autovalores.

$$D_{ii} = \sum_j W_{ji} \quad (12)$$

$$L = D - W \quad (13)$$

$$X \cdot L \cdot X^T \cdot a = \lambda \cdot X \cdot D \cdot X^T \cdot a \quad (14)$$

Finalmente se obtiene lo siguiente:

$$x_i \rightarrow y_i = A^T \cdot x_i, \quad A = (a_0, a_1, \dots, a_{l-1}) \quad (15)$$

Donde **y** es un vector de dimensión l elementos, y **A** es una matriz de $n \times l$ elementos.

Como ya se explicó brevemente en el estado del arte, capítulo 2.2.3, en LPP existen algunos inconvenientes. El principal inconveniente que se puede encontrar es el de las matrices singulares [20]. Las matrices singulares se forman cuando al intentar resolver la solución mínima de autovalores la dimensionalidad de los datos es mucho mayor al número de muestras (imágenes), es decir $\mathbf{n} \gg \mathbf{N}$. Algunas soluciones a este problema, son *kernel-based LPP* [21] [22], *two-dimensional LPP* [23] y *orthogonal LPP* [24]. Para evitar esta situación, todos estos métodos realizan una reducción del vector imagen de forma que $\mathbf{n} > \mathbf{N}$, o $\mathbf{n} = \mathbf{N}$. En el caso de este estudio, también se realiza esta reducción, y se hace como hemos explicado haciendo uso de PCA.

3.4.2 Reconocimiento

Para realizar el reconocimiento de imágenes con LPP se realizan unos pasos muy parecidos a los que se siguen con el método PCA. Para una nueva imagen de test en el sistema el procedimiento a seguir es el siguiente:

- 1) Se realiza el mismo preprocesado que en PCA (pasos 1, 2 y 3 del capítulo 3.2.2).
- 2) Se proyectan las imágenes. Para realizar este paso, se hace uso de la matriz de proyección de PCA y de la de LPP de manera que la imagen en el nuevo subespacio sigue la siguiente expresión:

$$\underline{y} = \left(\underline{W}_{PCA} \cdot \underline{A}_{LPP} \right)^T \cdot \underline{x} \quad (16)$$

- 3) Se aplica un método de decisión. En el caso de este estudio, del mismo modo que en los otros métodos, se aplicará el criterio de distancia mínima.

4 Bases de datos

En este apartado se van a mostrar y comentar las propiedades de las bases de datos de imágenes que se han utilizado durante este proyecto. Cada una de estas bases de datos está tomada bajo condiciones diferentes. Por otra parte, cada una de estas bases de datos, está formada por un número distinto de imágenes y de sujetos.

4.1 Base de datos Olivetti (ORL Database)

Se trata de una base de datos formada por 40 individuos, a razón de 10 imágenes por individuo tomada entre 1992 y 1994 [27]. En esta base de datos se pueden encontrar tanto imágenes de hombres como de mujeres.

Como es posible observar en la figura 31, en esta base de datos se pueden encontrar imágenes con diferentes posiciones y expresiones faciales (ojos abiertos/cerrados, sonriendo o no,...), lo cual será una dificultad añadida para el sistema. Del mismo modo se puede observar como en algunos sujetos, se tendrán imágenes con distintos detalles faciales (gafas o no, bigote o no,...). Por último decir que todas estas fotos han estado tomadas bajo unas condiciones de iluminación y de fondo de imagen controladas. El tamaño de todas estas imágenes es de 92x112 píxeles, con 256 niveles en la escala de grises para cada píxel.



Figura 31. Ejemplo base de datos ORL.

4.2 Base de datos XM2VTS

Está formada por imágenes pertenecientes a 295 sujetos, con 4 imágenes por individuo tomadas durante 4 meses. Son imágenes de un tamaño de 286x350 píxeles y en color (4:2:2). Previamente a su uso, han sido convertidas a escala de grises (256 niveles) y se han recortado de manera que la imagen solo contenga la cara de los individuos.

Como se observa en la figura 32, se trata de imágenes tomadas en un entorno controlado, bajo condiciones estables de iluminación y con una variación en la posición facial y de expresión muy leves.

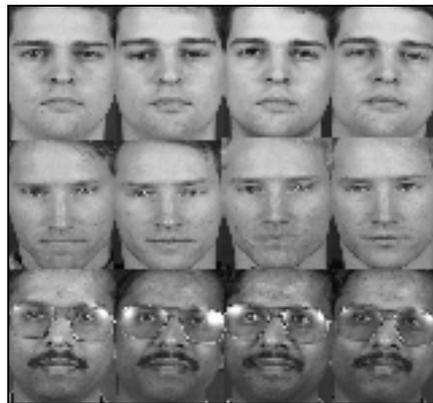


Figura 32. Ejemplo base de datos XM2VTS.

4.3 Base de datos BANCA

Se trata de una base de datos muy amplia. Dentro de la base de datos BANCA [28], se pueden encontrar diferentes grupos de imagen. Por un lado hay una primera diferenciación en cuanto al origen de los individuos de las imágenes (España, Inglaterra,...) y por otro se clasifican las imágenes entre las tomadas en entornos controlados, y las tomadas en entornos no controlados. Durante el resto de experimentos en que se utilice esta base de datos, se hará uso de las imágenes correspondientes a individuos de Inglaterra.

Las imágenes tomadas en entorno controlado, que en adelante se describirán como imágenes de la base BANCA *Controlled* (figura 33), presentan unas

condiciones de iluminación constantes a la vez que los cambios de posición y expresión facial son pequeños.

Las imágenes pertenecientes al conjunto no controlado, en adelante *BANCA Degraded* (figura 34), presentan unas condiciones de iluminación variable. En cuanto a la posición de las caras se refiere, estas siguen siendo bastante frontales y con pequeños cambios en la expresión.

Como es posible observar en la figura 33 y 34, todas las imágenes han sido previamente convertidas a escala de grises y recortadas de modo que las imágenes solo contengan la cara.



Figura 33. Ejemplo base de datos BANCA Controlled.



Figura 34. Ejemplo base de datos BANCA Degraded.

4.4 Base de datos YALE

Está formada por 150 imágenes pertenecientes a 15 sujetos con 10 imágenes por individuo que han sido tomadas bajo diferentes condiciones. Estas imágenes tienen un tamaño de 320x243 píxeles y no incluyen solamente el rostro sino que se ve de fondo una pared granulada la cual es muy visible o no dependiendo de cómo incida la luz sobre la escena, por lo tanto es adecuado realizar un recorte de estas imágenes de manera que solo tengamos la cara para el reconocimiento.

Estas imágenes son muy cambiantes y presentan situaciones de todo tipo como pueden ser cambios en la expresión facial, presencia o no de vello facial, gafas y cambios en la iluminación no uniformes (figura 35).



Figura 35. Ejemplo base de datos YALE.

5 Resultados

En este capítulo, se realizarán experimentos para evaluar el rendimiento de los métodos de reconocimiento facial que se han estudiado en este proyecto. Cada uno de estos experimentos utilizará una determinada base de datos ya que cada una de éstas presenta unas características diferentes que serán útiles para tener una idea más clara y global del funcionamiento de cada uno de los métodos de reconocimiento facial frente a diferentes condiciones.

Del mismo modo que serán evaluadas diferentes bases de datos, se tendrá en cuenta la capacidad de las diferentes técnicas que hacen uso de matrices de proyección para funcionar con unos buenos resultados si se utilizan matrices creadas con imágenes diferentes a las que se utilizaran para el reconocimiento.

Durante los siguientes resultados y para el caso concreto del método LPP, se va a utilizar la siguiente configuración ya que los resultados aunque son bastante similares para todos los casos, son levemente mejores utilizando estos parámetros (figura 36):

- Método de elección de los pesos Simple-minded.
- Construcción del gráfico de adyacencia de manera supervisada.
- Parámetro t no tiene efecto al no utilizar Heat-Kernel.

De entre los experimentos que se realizarán, los habrá enfocados a evaluar el tiempo necesario para el reconocimiento o la capacidad de cada una de las técnicas de funcionar frente a condiciones adversas como son los cambios de luminosidad o los cambios en la posición de las imágenes. Para evaluar cada una de estas situaciones se hará uso de una determinada base de datos, de las expuestas en el capítulo 4, utilizando una selección de imágenes como imágenes de entrenamiento y el resto, en general, como imágenes de referencia y test al mismo tiempo, aunque nunca se comparará una imagen con ella misma. En todos estos experimentos se realizará identificación de las

imágenes utilizando el criterio de distancia mínima [36], sólo una imagen es la correcta.

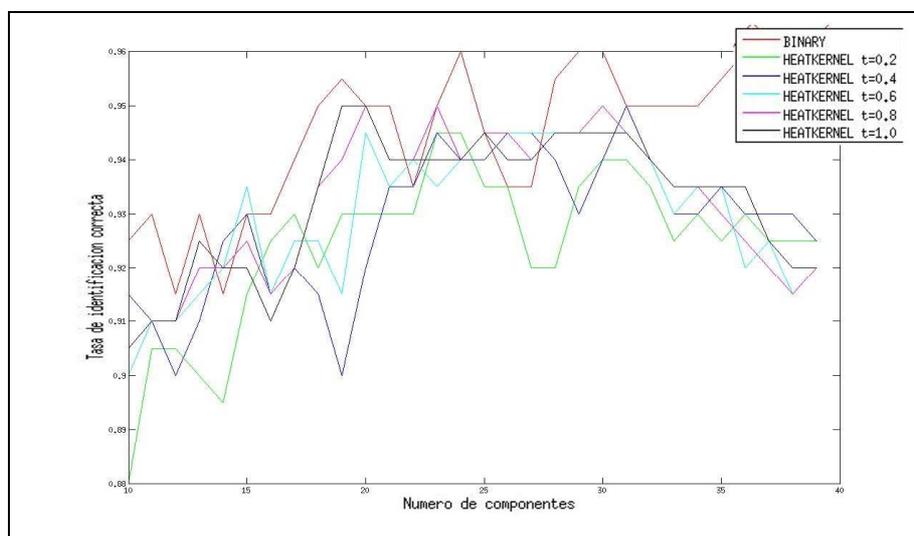


Figura 36. Tasa de identificación correcta obtenida para diferentes configuraciones en la creación de la matriz de pesos LPP, utilizando pesos *simple-minded* y *Heatkernel* con diferentes valores de t , para el caso de utilizar imágenes de la base de datos ORL.

A la finalización de cada experimento se mostrará una tabla resumen con los resultados obtenidos mostrando el número de componentes para el que se obtiene la máxima tasa de reconocimiento, la tasa de reconocimiento máxima, y la tasa de reconocimiento para el número componentes escogido en las conclusiones como valor óptimo para una aplicación general.

5.1 Funcionamiento del sistema en un entorno controlado

En esta sección, se estudia el funcionamiento del sistema en situaciones controladas. Estas situaciones son ideales de forma que no es usual encontrarlas en un entorno real.

Para este experimento se hará uso de las bases de datos XM2VTS y BANCA *Controlled* ya que se trata de conjuntos de imágenes tomadas bajo unas condiciones de iluminación y fondo controladas a la vez que presentan pequeñas variaciones en cuanto a la posición de la cara y las expresiones faciales se refiere.

Los parámetros que se van a utilizar para la realización de este experimento, son:

- Base de datos XM2VTS, con 590 imágenes (2 imágenes por persona y 295 individuos) y la base de datos BANCA *Controlled*, con 1223 imágenes (52 individuos distintos y con aproximadamente 23 imágenes para cada individuo).
- Número de componentes variable de 10 a 100 componentes.
- Los sistemas de reconocimiento PCA, LPP y DCT.
- Matrices de proyección creadas con imágenes de la misma base de datos (590 para XM2VTS y 832 para BANCA *Controlled*) para los métodos PCA y LPP.
- Distancia Euclídea.

Como es posible observar en la figura 37, los resultados en el caso de utilizar las imágenes BANCA *Controlled* son muy elevados como ya esperábamos por la poca variabilidad de las imágenes pertenecientes a esta base de datos. La tasa de identificación es muy alta para todos los métodos, aunque cada método presenta su respuesta particular al sistema. Una razón para este alto rendimiento es que tenemos unas 23 imágenes por individuo lo cual hace más fácil para el sistema reconocer una imagen de entrada, ya que al tener muchas imágenes por individuo es más sencillo que una de ellas cumpla la condición.

El método LPP es el que ofrece unos mejores resultados cuando se utilizan pocas componentes para el reconocimiento, mientras que los otros dos aumentan su rendimiento con el número de componentes hasta estabilizarse para un cierto valor.

Para el caso de utilizar las imágenes XM2VTS, los resultados se pueden ver en la figura 38. Como se puede observar, la tasa de identificación correcta es muy inferior a la obtenida para el caso anterior. Esto es debido en gran medida a que el experimento es más complejo para el sistema ya que se tienen muchos más individuos tanto en el set de referencia como en el de test y además ahora

solo hay 2 imágenes representando a cada individuo de manera que una imagen de entrada solo podrá corresponderse con otra del set de referencia. Esto afecta de una manera muy importante ya que cuantas más imágenes de un mismo individuo tengamos, más fácil es para el sistema decidir cual es la imagen correcta. Las imágenes que componen esta base de datos están afectadas por pequeñas variaciones de posición, expresión e iluminación, pero al ser estos prácticamente despreciables no se consideran un factor relevante para el bajo rendimiento del sistema.

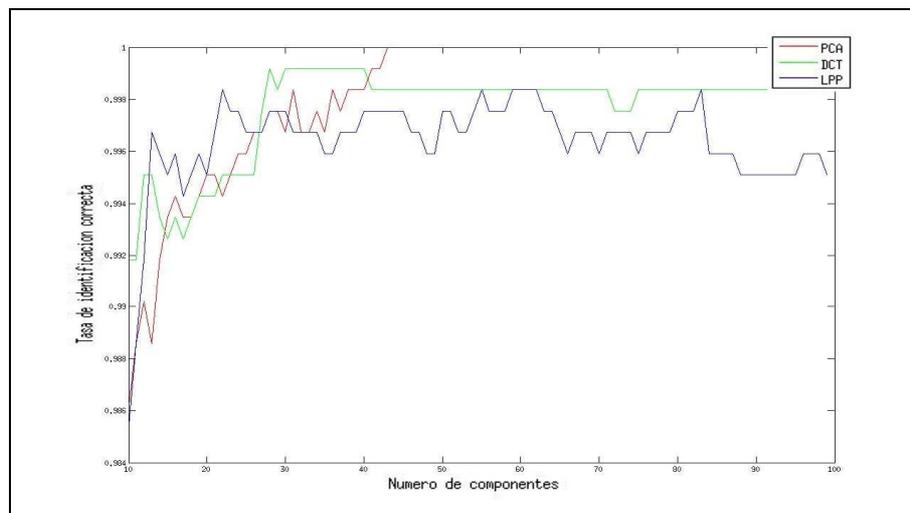


Figura 37. Tasa de identificación correcta para la base de datos BANCA Controlled

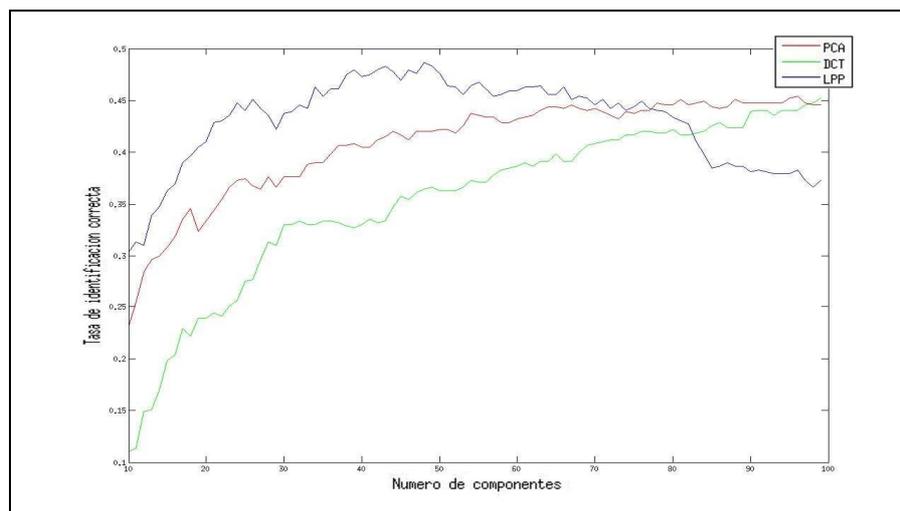


Figura 38. Tasa de identificación correcta para la base de datos XM2VTS

Nuevamente LPP ofrece unos mejores resultados para número de componentes bajo mientras que los otros dos métodos mejoran cuantas más

componentes se utilizan. En este caso LPP ofrece la mejor tasa de identificación correcta.

Como conclusión de este primer experimento, se puede decir que el uso de un número reducido de imágenes por sujeto, en bases de datos muy grandes, puede afectar negativamente al sistema, es decir, el funcionamiento del sistema se incrementará cuanto mayor sea el ratio entre imágenes totales y el número de sujetos, esto lo veremos con más detalle en el siguiente apartado (5.2).

BANCA Controlled			
Método	Número de componentes max. Rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	44	100	100% (55)
DCT	29	99,9182	99.8365% (80)
LPP	23	99,8365	99.6729% (40)
XM2VTS			
Método	Número de componentes max. Rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	97	45,4237	43.7288% (55)
DCT	100	45,2542	41.8644% (80)
LPP	49	48,6441	47.9661% (40)

Tabla 2. Detalle de resultados Capítulo 5.1. En la primera columna: número de componentes máxima tasa de identificación, segunda columna: máxima tasa de identificación, tercera columna: tasa de identificación para el número de componentes elegido para el caso general.

5.2 Influencia del número de individuos en el sistema

Para verificar los resultados obtenidos, en el apartado anterior, utilizando las imágenes XM2VTS, se va a analizar como varía la tasa de identificación correcta en relación al número de individuos utilizado, en concreto se va a realizar esta prueba para 50, 100, 200 y 295 (todos los de la base de datos) individuos.

Los parámetros del sistema que se van a utilizar son los siguientes

- Base de datos XM2VTS, con 590 imágenes (2 imágenes por persona y 295 individuos), de esta base se utilizarán diferentes configuraciones con diferente número de individuos (50, 100, 200, 295).
- Número de componentes variable de 10 a 100 componentes.

- Los sistemas de reconocimiento PCA, DCT y LPP.
- Matrices de proyección creadas con imágenes de la misma base de datos (590).
- Distancia Euclídea.

Como es posible observar en las figuras 39a, 39b, 39c y 39d, la tasa de identificación correcta mejora a medida que se utilizan menos individuos para el reconocimiento obteniendo su mejor rendimiento para 50 individuos. De este modo se verifica la conclusión del apartado anterior donde se dijo que el funcionamiento del sistema mejora cuanto mayor es el ratio entre número de imágenes y el número de sujetos en las imágenes de referencia. Con esto se obtiene que cuando se haga uso de muchos individuos en el set de referencia el sistema requerirá muchas imágenes por cada uno de ellos, de no ser así el rendimiento se verá afectado.

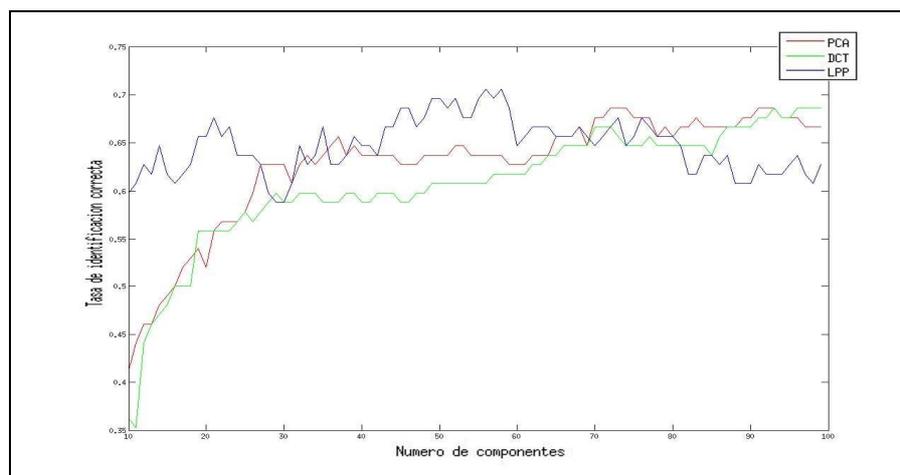


Figura 39a. Tasa de identificación correcta con 50 individuos en el sistema.

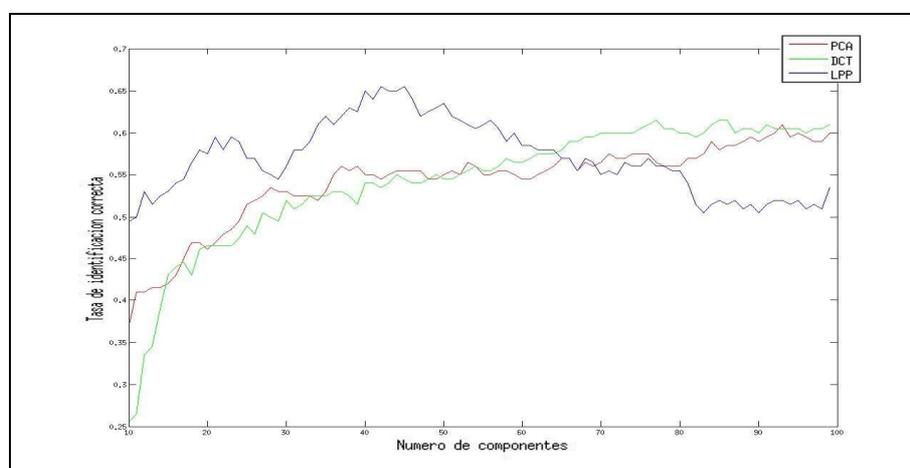


Figura 39b. Tasa de identificación correcta con 100 individuos en el sistema.

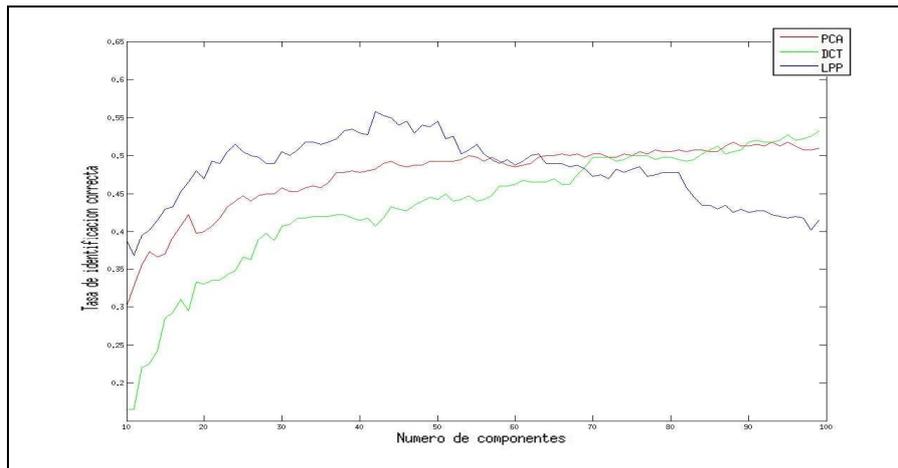


Figura 39c. Tasa de identificación correcta con 200 individuos en el sistema.

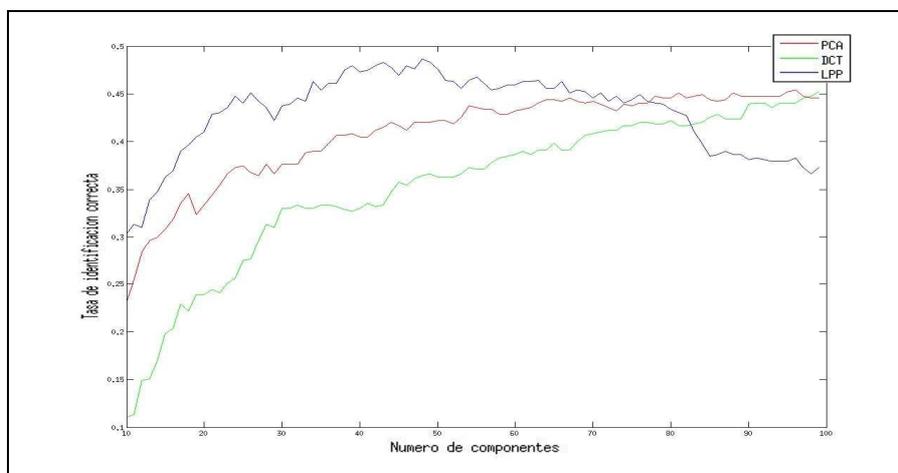


Figura 39d. Tasa de identificación correcta con 295 individuos en el sistema.

Los resultados obtenidos son parecidos a los del apartado anterior, es decir, LPP obtiene una tasa de identificación mayor para un número de componentes bajo y a partir de un cierto punto se ve superado por los otros métodos.

XM2VTS 50 individuos			
Método	Número de componentes max. rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	73	68,6275	63.72.55% (55)
DCT	94	68,6275	64.7059% (80)
LPP	57	70,5882	65.6863% (40)
XM2VTS 100 individuos			
Método	Número de componentes max. rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	94	61	56% (55)
DCT	78	61,5	60.5% (80)
LPP	43	65,5	62.5% (40)

Tabla 3.1. Detalle de resultados Capítulo 5.2. En la primera columna: número de componentes máxima tasa de identificación, segunda columna: máxima tasa de identificación, tercera columna: tasa de identificación para el número de componentes elegido para el caso general.

XM2VTS 200 individuos			
Método	Número de componentes max. rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	89	51,75	50% (55)
DCT	100	53,25	49.75% (80)
LPP	43	55,75	53.5% (40)
XM2VTS 295 individuos			
Método	Número de componentes max. rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	97	45,4237	43.7288% (55)
DCT	100	45,2542	41.8644% (80)
LPP	49	48,6441	47.9661% (40)

Tabla 3.2. Detalle de resultados Capítulo 5.2. En la primera columna: número de componentes máxima tasa de identificación, segunda columna: máxima tasa de identificación, tercera columna: tasa de identificación para el número de componentes elegido para el caso general.

5.3 Influencia de los cambios de iluminación

Para evaluar como se ve afectado el sistema bajo diferentes condiciones de iluminación, se va a hacer uso de imágenes que presentan unas condiciones de iluminación variables.

Se va a contemplar dos casos, uno en el que se intentará minimizar los problemas producidos por los cambios en la iluminación normalizando las imágenes en la etapa de preprocesado (procedimiento utilizado en el experimento anterior), y otro caso en el que lo que se hará será descartar las primeras componentes de las imágenes proyectadas, que son las que más se ven afectadas por los cambios de iluminación en PCA y DCT, para ver como afecta a LPP. En el caso de DCT nunca realizamos esta normalización previa de las imágenes sino que se descarta el primer coeficiente correspondiente a la componente continua, como consecuencia, al analizar el primer caso, normalización, se mostrarán los resultados obtenidos de descartar el primer coeficiente.

Se hará uso de la base de datos BANCA *Degraded*, para ver el efecto de cambios en la iluminación uniformes, y YALE, para ver como responde el sistema cuando estos cambios de iluminación no son uniformes.

5.3.1 Normalización

El procedimiento consiste en normalizar las imágenes en la etapa de preprocesado de modo que se atenúen los efectos de los cambios de iluminación. Este preprocesado se ha explicado en los capítulos 3.2 y 3.4.

Los parámetros utilizados para la realización de este experimento, son:

- BANCA *Degraded*, con 1158 imágenes (52 individuos distintos y con aproximadamente 23 imágenes para cada individuo) y YALE, con 45 imágenes (15 individuos y 3 imágenes por persona).
- Número de componentes variable de 10 a 100 componentes.
- Los sistemas de reconocimiento PCA, LPP y DCT.
- Matrices de proyección creadas con imágenes de la misma base de datos para los métodos PCA y LPP (770 para BANCA *Degraded* y 105 para YALE).
- Distancia Euclídea.

Los resultados obtenidos, para cambios de iluminación uniformes, son los que se pueden ver en la figura 40, donde se puede observar que aunque se están usando imágenes tomadas en un entorno no controlado, la tasa de identificación correcta es muy elevada. Las razones por las que se tiene un rendimiento tan alto del sistema, son varias.

En primer lugar, está el hecho de que nuevamente, igual que con BANCA *Controlled*, se utilizan muchas imágenes de referencia para cada individuo haciendo así más fácil para el sistema el reconocimiento.

En segundo lugar, aunque son imágenes tomadas en un entorno no controlado, la iluminación solo cambia de individuo a individuo y no entre imágenes de un mismo individuo, y estos cambios de iluminación son uniformes.

En tercer y último lugar, se debe tener en cuenta que las imágenes utilizadas son de buena calidad y con pocos cambios de expresión y posición facial a pesar de ser tomadas en un entorno no controlado.

Los resultados obtenidos para todos los métodos son parecidos y ninguno destaca sobre los demás.

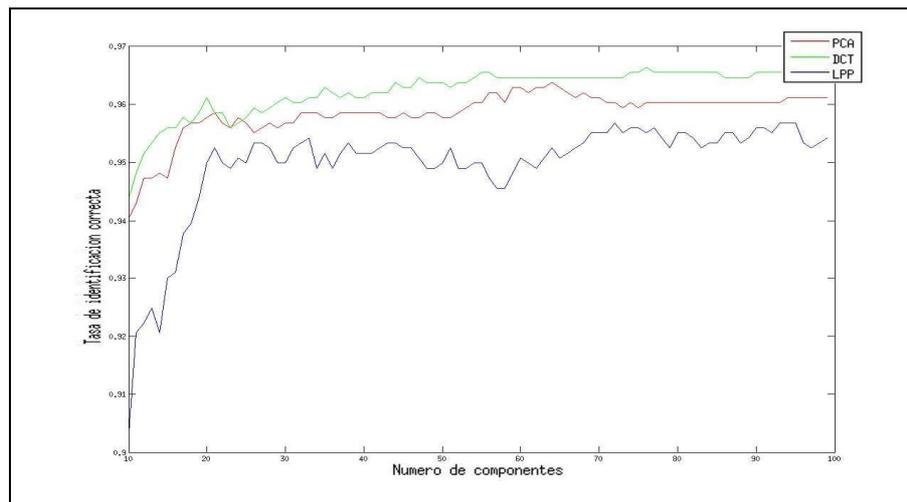


Figura 40. Tasa de identificación correcta para el caso de cambios de iluminación uniformes utilizando normalización.

Cuando los cambios de iluminación no son uniformes, figura 41, la tasa de identificación correcta del sistema se ve más afectada y los resultados son inferiores a los obtenidos para los experimentos anteriores. Como es posible observar en el gráfico, DCT mejora su rendimiento cuanto mayor es el número de componentes utilizado hasta estabilizarse entorno a las 60 componentes. En el caso PCA y LPP se obtienen las mejores tasas de identificación correcta para un número de componentes bajo y a medida que éste aumenta su rendimiento se ve reducido. La mejor tasa de reconocimiento la tenemos utilizando el método LPP.

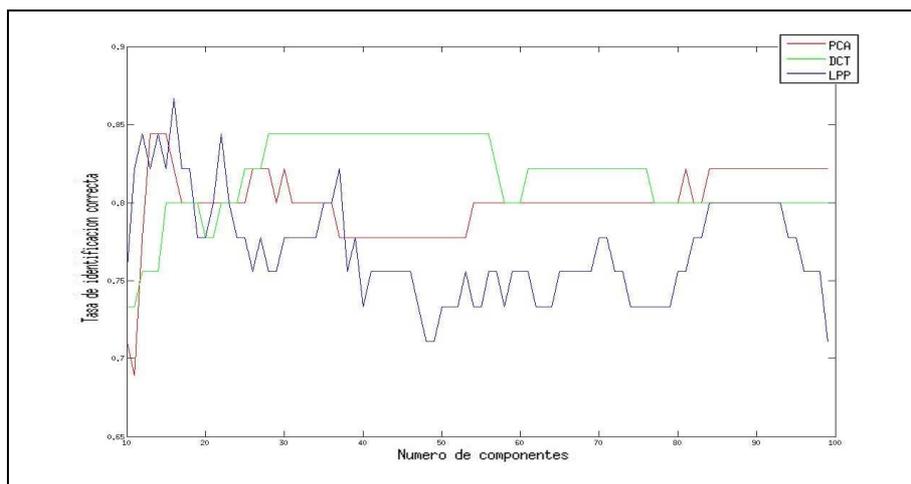


Figura 41. Tasa de identificación correcta para el caso de cambios de iluminación no uniformes utilizando normalización.

BANCA Degraded			
Método	Número de componentes max. rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	65	96,3731	96.0276% (55)
DCT	100	96,7185	96.5458% (80)
LPP	73	95,6822	95.1641% (40)
YALE			
Método	Número de componentes max. rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	14	84.4444	80% (55)
DCT	29	84.4444	80% (80)
LPP	17	86.6667	77.7778% (40)

Tabla 4. Detalle de resultados capítulo 5.3.1. En la primera columna: número de componentes máxima tasa de identificación, segunda columna: máxima tasa de identificación, tercera columna: tasa de identificación para el número de componentes elegido para el caso general.

5.3.2 Descarte de coeficientes

En este caso, en lugar de realizar una normalización en la etapa de preprocesado se descartarán los primeros coeficientes de las imágenes proyectadas, de manera que se atenúe el efecto de los cambios de iluminación.

Se van a analizar tres casos que se corresponden con el número de componentes descartados en cada caso (1, 3 y 5 componentes), de manera que podamos ver bien el efecto de este procedimiento.

Los parámetros que se van a utilizar para la realización de este experimento, son los mismos que en el apartado anterior. Al descartar 1 coeficiente, los

resultados utilizando DCT son los mismos que en el caso de normalizar las imágenes ya que en DCT siempre se descarta el primer coeficiente en lugar de hacer una normalización. Los parámetros utilizados son:

- BANCA *Degraded*, con 1158 imágenes (52 individuos distintos y con aproximadamente 23 imágenes para cada individuo) y YALE, con 45 imágenes (15 individuos y 3 imágenes por persona).
- Número de componentes variable de 10 a 100 componentes.
- Los sistemas de reconocimiento PCA, LPP y DCT.
- Matrices de proyección creadas con imágenes de la misma base de datos para los métodos PCA y LPP (770 para el caso de utilizar la base BANCA *Degraded* y 105 en el caso de YALE).
- Distancia Euclídea.

En la figura 42, se pueden observar los resultados de los experimentos para el caso de cambios de iluminación uniformes. Como es posible observar, los resultados para el caso de descartar 1 coeficiente, sin tener en cuenta DCT, son parecidos a los del caso normalizado aunque levemente por debajo. En los otros dos casos los resultados son inferiores respecto utilizar normalización en la etapa de preprocesado.

El efecto de este procedimiento sobre cada método es distinto, se puede ver que el método LPP se ve ligeramente más afectado que los otros dos. Esto puede deberse a que los coeficientes utilizados en LPP son los obtenidos de proyectar las imágenes con los autovectores correspondientes a los autovalores más bajos y éstos están menos afectados por estos cambios de iluminación.

En el caso de tener cambios de iluminación no uniformes, figura 43, el efecto es parecido al caso anterior. La mejor tasa de reconocimiento se vuelve a obtener en el caso de descartar un solo coeficiente pero a diferencia del caso anterior, la tasa de reconocimiento no se ve reducida al descartar 1 coeficiente e incluso obtenemos un mejor rendimiento para LPP. A medida que vamos

descartando más coeficientes la tasa de reconocimiento se reduce de manera muy notoria para LPP, de manera suave para PCA y prácticamente nada en el caso de DCT.

Como conclusión de este apartado se puede decir que el método con el que funciona mejor esta técnica es con DCT, mientras que para el caso LPP este procedimiento afecta de manera muy negativa en los resultados sobretodo si descartamos más de 1 coeficiente.

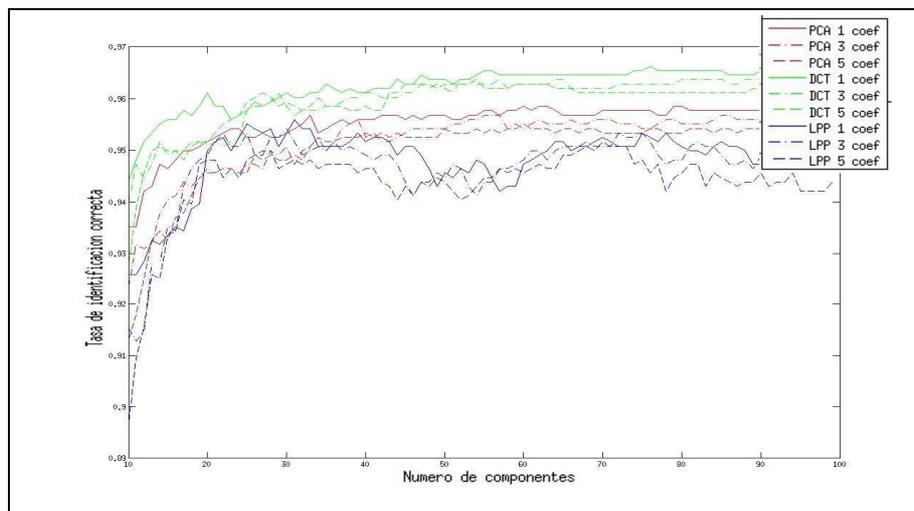


Figura 42. Tasa de identificación correcta para el caso de cambios de iluminación uniformes descartando coeficientes.

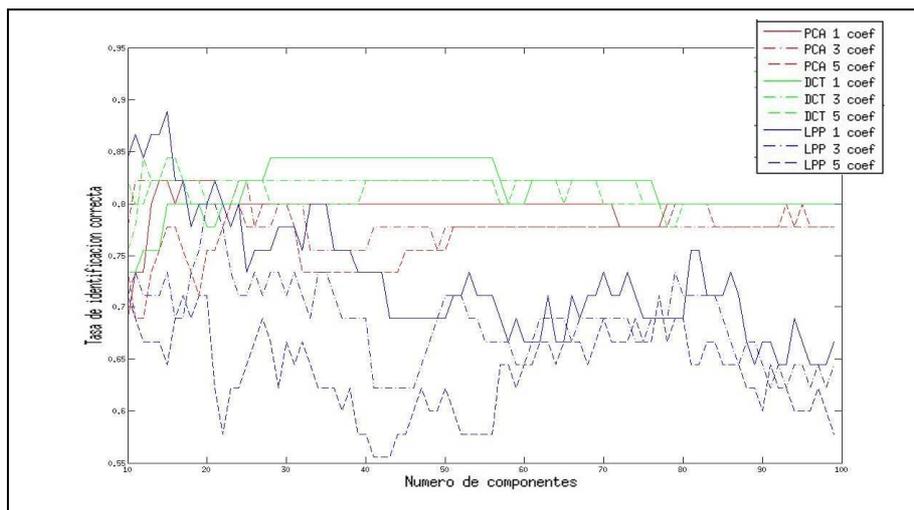


Figura 43. Tasa de identificación correcta para el caso de cambios de iluminación no uniformes descartando coeficientes.

BANCA Degraded descartando 1 coeficiente			
Método	Número de componentes max. Rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	61	95,8549	95.6822% (55)
DCT	100	96,7185	96.5458% (80)
LPP	32	95,5959	95.3368% (40)
BANCA Degraded descartando 3 coeficientes			
Método	Número de componentes max. Rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	56	95,8622	95.5959% (55)
DCT	92	96,4594	96.2867% (80)
LPP	31	95,4231	94.9914% (40)
BANCA Degraded descartando 5 coeficientes			
Método	Número de componentes max. Rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	61	95,5095	95.4231% (55)
DCT	55	96,3731	96.1140% (80)
LPP	71	95,1641	94.5596% (40)
YALE descartando 1 coeficiente			
Método	Número de componentes max. Rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	15	82.2222	80% (55)
DCT	29	84.4444	80% (80)
LPP	16	88.8889	73.3333% (40)
YALE descartando 3 coeficientes			
Método	Número de componentes max. Rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	12	82.2222	77.7778% (55)
DCT	13	84.4444	80% (80)
LPP	21	80	68.8889% (40)
YALE descartando 5 coeficientes			
Método	Número de componentes max. Rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	24	80	77.7778% (55)
DCT	16	84.4444	77.7778% (80)
LPP	11	71.1111	57.7778% (40)

Tabla 5. Detalle de resultados Capítulo 5.3.2. En la primera columna: número de componentes máxima tasa de identificación, segunda columna: máxima tasa de identificación, tercera columna: tasa de identificación para el número de componentes elegido para el caso general.

5.4 Efecto de la variación del aspecto de las caras

Para ver como afecta la variación del aspecto de las caras al rendimiento del sistema, se va a hacer uso de la base de datos Olivetti formada por imágenes que a parte de presentar diferentes gestos y posiciones faciales para un mismo individuo, presentan aparición o no de objetos en los rostros, vello facial, etc.

Los parámetros que se van a utilizar para la realización de este experimento, son los siguientes:

- Base de datos Olivetti, 200 imágenes (5 imágenes por persona y 40 individuos).
- Número de componentes variable de 10 a 100 componentes.
- Los sistemas de reconocimiento PCA, LPP y DCT.
- Matrices de proyección creadas con imágenes de la misma base de datos (200 imágenes).
- Distancia Euclídea.

Como es posible ver en los resultados de la figura 44, la respuesta obtenida para cada método es distinta.

Para PCA, el sistema mejora su tasa de identificación correcta a medida que aumentamos el número de componentes hasta que el sistema se estabiliza. Este método es el que ofrece la menor tasa de identificación correcta.

En segundo lugar, DCT se comporta de una manera muy parecida a PCA pero obteniendo unos niveles de funcionamiento alrededor al 3% mejores. Como sucede con PCA el sistema aumenta su rendimiento hasta estabilizarse.

Por último, LPP es el método que ofrece un rendimiento más alto al utilizar imágenes que presentan una alta variabilidad en el aspecto. La tasa de identificación correcta, como ya sucedía en el apartado 5.3.2 con cambios de iluminación no uniformes y variación en el aspecto de las caras (YALE), es más elevada para un número de componentes bajo y luego a partir de unas 45 se reduce. Esto puede ser interesante para aplicaciones que requieran de un reconocimiento muy rápido ya que como menor es el número de componentes, menor es el tiempo de respuesta del sistema.

Como conclusión de este apartado, podemos decir que LPP es el método que mejor funciona frente a variaciones en el aspecto de las caras. Los otros

métodos también ofrecen una tasa de identificación elevada aún tratándose de imágenes con cambios de posición facial bastante destacables, pero se debe tener en cuenta que las imágenes de referencia solo contienen a 40 individuos con 5 imágenes para cada uno de ellos, echo que hace que sea mas fácil el reconocimiento que por ejemplo en el caso de usar la base de datos XM2VTS.

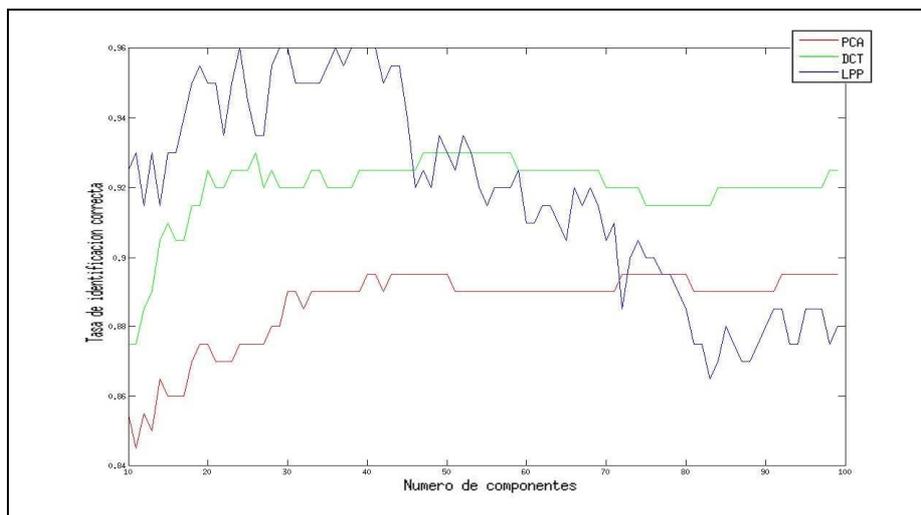


Figura 44. Tasa de identificación correcta para imágenes con variaciones en el aspecto de las caras.

Olivetti (ORL)			
Método	Número de componentes max. rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	41	89,5	89% (55)
DCT	27	93	91.5% (80)
LPP	25	96	96% (40)

Tabla 6. Detalle de resultados Capítulo 5.4. En la primera columna: número de componentes máxima tasa de identificación, segunda columna: máxima tasa de identificación, tercera columna: tasa de identificación para el número de componentes elegido para el caso general.

5.5 Influencia de entrenar el sistema con imágenes distintas

En este experimento se realizará el entrenamiento, para PCA y LPP, con imágenes de una base de datos distinta a la utilizada para el reconocimiento. Al estar estudiando el comportamiento del sistema frente a diferentes tipos de entrenamiento, no tiene sentido evaluar en este apartado el método DCT ya que este método no lo requiere, pero se incluirá el resultado para compararlo.

Se realizarán tres tipos de entrenamiento. Primero se entrenará el sistema con imágenes de una base de datos totalmente distinta a la que se utilizará en el reconocimiento (ORL para el reconocimiento y BANCA *Degraded* para el entrenamiento). En el segundo caso, se utilizarán imágenes de los mismos individuos pero tomadas bajo unas condiciones diferentes (BANCA *Controlled* para el reconocimiento y BANCA *Degraded* para el entrenamiento). Por último, se realizará el entrenamiento con todas las imágenes utilizadas para tal efecto de todas las bases de datos (2393 imágenes) para poder ver el funcionamiento de estos métodos con una matriz de proyección generalizada.

Teóricamente, la utilización de imágenes distintas para el entrenamiento, debería implicar un reto para estos métodos ya que la matriz de proyección que se utilizará no definirá bien el nuevo subespacio.

Para este experimento, se han utilizado los siguientes parámetros del sistema:

- Base de datos Olivetti para el reconocimiento, formada por 200 imágenes (5 imágenes por persona con 40 individuos) y base de datos BANCA *Controlled*, formada por 1223 imágenes (52 individuos distintos con aproximadamente 23 imágenes por sujeto).
- Base de datos BANCA *Degraded* para el entrenamiento del sistema (770 imágenes) y conjunto de imágenes de todas las bases (2393 imágenes).
- Número de componentes variable de 10 a 100 componentes.
- Los sistemas de reconocimiento PCA y LPP.
- Distancia Euclídea.

Los resultados obtenidos son los de la figura 45, para el caso de utilizar las imágenes de la base de datos Olivetti para el reconocimiento, la figura 46 para el caso BANCA *Controlled* y la figura 47 y 48 para el caso de utilizar la matriz de proyección generalizada.

Como se puede observar en las gráficas, los resultados para el segundo caso son muy parecidos a los obtenidos en el capítulo 5.1 cuando se entrenaba el

sistema con imágenes tomadas bajo las mismas condiciones, mientras que para el primer caso el rendimiento ha variado por completo.

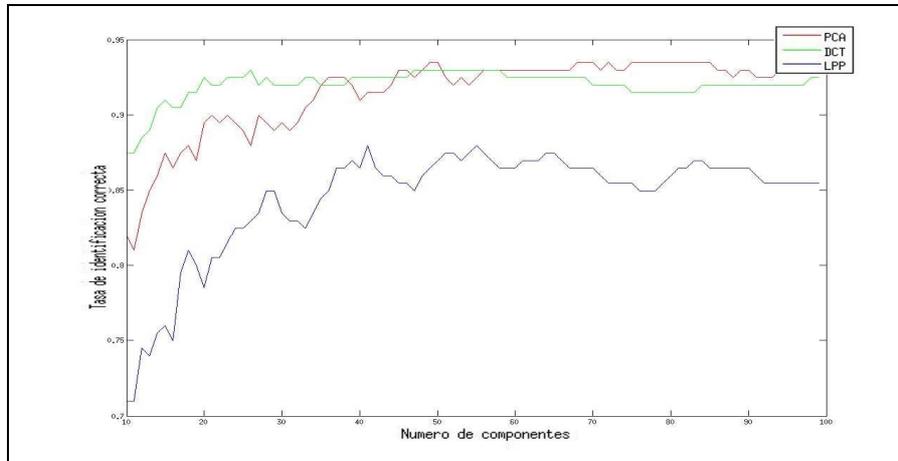


Figura 45. Tasa de identificación correcta obtenida utilizando para el reconocimiento imágenes de la base de datos ORL y entrenando el sistema con imágenes de la base de datos BANCA *Degraded*.

En el primer caso, a diferencia del apartado 5.4, LPP incrementa su tasa de identificación con el número de componentes utilizado hasta estabilizarse obteniendo una tasa máxima inferior mientras que PCA sigue mejorando su tasa de identificación cuantas más componentes se utilizan pero en este caso mejora su tasa máxima de identificación.

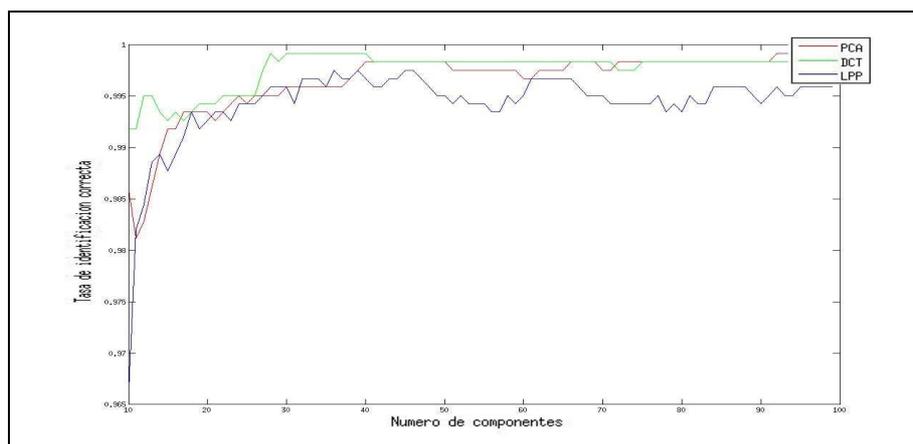


Figura 46. Tasa de identificación correcta obtenida utilizando para el reconocimiento imágenes de la base de datos BANCA *Controlled* y entrenando el sistema con imágenes de la base de datos BANCA *Degraded*.

En el segundo caso, utilizando BANCA *Controlled* para el reconocimiento, los resultados son muy parecidos a los obtenidos en el apartado 5.1. La única diferencia es una leve disminución en la tasa de identificación correcta para

ambos casos pero esta es poco apreciable al ser el funcionamiento del sistema para este caso casi perfecto.

En el tercer caso, utilizando una matriz de proyección generalizada, la tasa de reconocimiento se ve reducida considerablemente para las imágenes ORL mientras que para BANCA *Controlled* los resultados son prácticamente los mismos. La diferencia entre un caso y el otro es que aunque se hace uso de imágenes de todas las bases de datos para el entrenamiento, el número de imágenes utilizadas de cada una es distinto, y mientras ORL aporta 200 imágenes, BANCA *Controlled* junto BANCA *Degraded*, los dos grupos son imágenes de los mismos sujetos, aportan 1600. Esto hará que el sistema este mejor definido para las bases de datos que aporten más imágenes al entrenamiento.

Como conclusión de este apartado se puede decir que el método LPP se ve más afectado al ser entrenado con imágenes de distintas personas a las que se van a reconocer, mientras que PCA lo que realmente necesita es un elevado número de imágenes para el entrenamiento sean o no de las mismas personas. En caso de utilizar una matriz de proyección generalizada, se ha observado que el funcionamiento del sistema se ve reducido si los sujetos que se van a reconocer no aportan un número de imágenes predominante en la fase de entrenamiento.

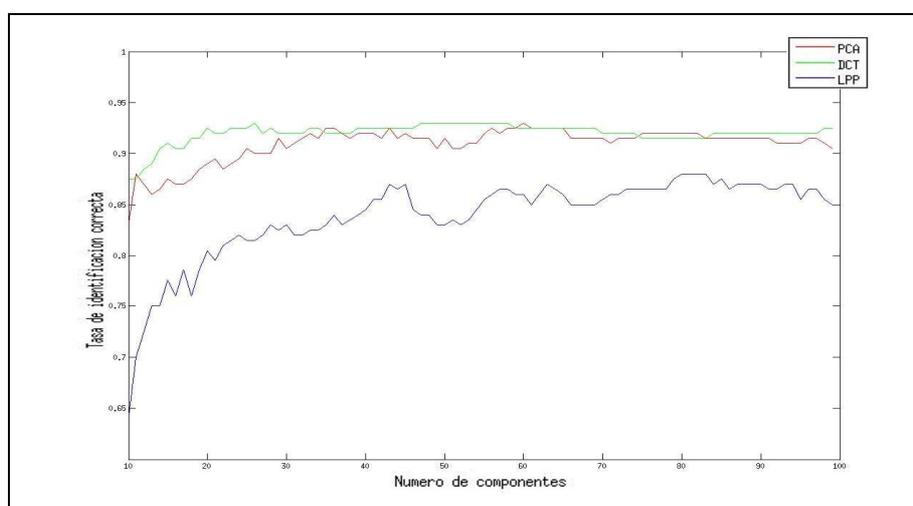


Figura 47. Tasa de identificación correcta obtenida utilizando para el reconocimiento imágenes de la base de datos ORL y entrenando el sistema con imágenes de todas las bases de datos.

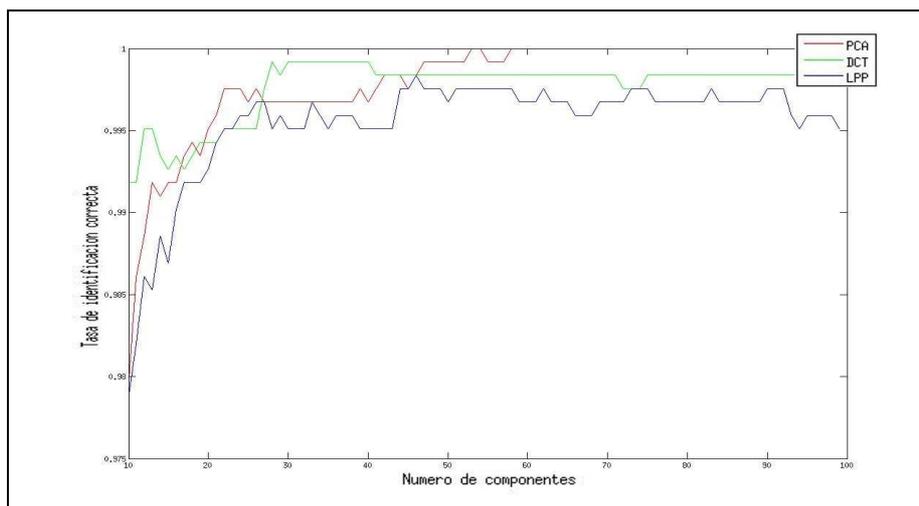


Figura 48. Tasa de identificación correcta obtenida utilizando para el reconocimiento imágenes de la base de datos BANCA *Controlled* y entrenando el sistema con imágenes de todas las bases de datos

ORL entrenando con BANCA Degraded			
Método	Número de componentes max. rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	50	95,8549	92% (55)
LPP	42	88	87% (40)
BANCA Controlled entrenando con BANCA Degraded			
Método	Número de componentes max. rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	93	99,9182	99.7547% (55)
LPP	37	99,7547	99.7547% (40)
ORL entrenando con Matriz de proyección generalizada			
Método	Número de componentes max. rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	61	93	91%(55)
LPP	81	88	84% (40)
BANCA Controlled entrenando con Matriz de proyección generalizada			
Método	Número de componentes max. rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	54	100	100% (55)
LPP	47	99,8365	99.5094% (40)

Tabla 7. Detalle de resultados Capítulo 5.5. En la primera columna: número de componentes máxima tasa de identificación, segunda columna: máxima tasa de identificación, tercera columna: tasa de identificación para el número de componentes elegido para el caso general.

5.6 Evaluación de un entorno real

Para ver el funcionamiento del sistema bajo unas condiciones un poco más reales y parecidas a las que se presentan en una aplicación final se utilizará distintos conjuntos de imágenes para el set de referencia y de test. De este modo, las imágenes que se tendrán en el set de referencia y las que se usarán

en el de test, no presentarán unas mismas condiciones de iluminación del mismo modo que sucedería utilizando este tipo de sistemas en la calle o un entorno no controlado.

Para la realización de este experimento, se van a utilizar nuevamente las imágenes de la base de datos BANCA. En concreto se utilizarán o bien imágenes del grupo *Controlled* como set de referencia y las del grupo *Degraded* en el de test o a la inversa.

Se contemplan 2 situaciones posibles, una primera en la que se hará uso de un conjunto de imágenes de la misma base de datos que el set de referencia para el entrenamiento del sistema, y otra en la que se utilizarán imágenes de la misma base de datos que el set de test. Cuando se presenten estas 2 situaciones, en el primer caso se mostrarán y compararán los resultados de los tres métodos (PCA, DCT y LPP), pero en el segundo ya no se comparan los referentes a DCT ya que cambiar las imágenes que se utilicen para entrenar el sistema no variará los resultados de este caso.

5.6.1 Entrenamiento con imágenes parecidas al set de referencia

En este primer caso, se analiza la situación más usual que se puede encontrar en una aplicación funcional, se utilizará un conjunto de imágenes de la base de datos utilizada como set de referencia, pero diferentes a estas, para entrenar el sistema. Se describe como situación más usual ya que lo que se hace normalmente es ir actualizando el sistema de manera periódica y para ello es mejor utilizar para el entrenamiento, como se ha podido ver en los apartados anteriores, imágenes que estén relacionadas con el sistema y bajo condiciones controladas.

Para este experimento, se han utilizado los siguientes parámetros del sistema:

- Base de datos BANCA *Controlled*, con 1223 imágenes (52 individuos distintos y con aproximadamente 23 imágenes para cada individuo).

- Base de datos BANCA *Degraded*, con 1158 imágenes (52 individuos distintos y con aproximadamente 23 imágenes para cada individuo).
- Imágenes de la base de referencia para el entrenamiento (832 para el caso *Controlled* y 770 para el caso *Degraded*).
- Número de componentes variable de 10 a 100 componentes.
- Los sistemas de reconocimiento PCA, DCT y LPP.
- Distancia Euclídea.

Se estudiarán dos situaciones, una en el que se utilizarán las imágenes con condiciones de luz controladas (BANCA *Controlled*) como set de referencia y las imágenes con condiciones de luz no controladas (BANCA *Degraded*) como set de test (figura 49), y otra en el que se hará a la inversa (figura 50).

La primera situación es la configuración ideal para una aplicación futura ya que utiliza como set de referencia imágenes tomadas bajo condiciones controladas y imágenes bajo las mismas condiciones para entrenar el sistema, mientras que como imágenes de test se tienen las no controladas, que serán más parecidas a las que recibe el sistema en una situación práctica.

Como resultado a destacar, LPP es el que presenta unos mejores resultados hasta 80 componentes, aunque su máxima tasa de identificación correcta no se ve superada por ninguno de los otros métodos. Su mejor rendimiento se obtiene para un número de componentes bajo.

Los otros dos métodos tienen un rendimiento muy discreto. En el caso de DCT, es comprensible que frente a cambios en las condiciones de iluminación, entre las imágenes del set de test y referencia, su rendimiento empeore cualitativamente ya que es un método sin entrenamiento previo. PCA tiene una tasa de identificación ligeramente superior a DCT pero muy inferior a LPP.

En el segundo caso se observa como afecta al sistema utilizar como referencia imágenes tomadas bajo condiciones no controladas. El comportamiento del sistema para este caso es algo inferior que para el caso anterior. En la figura

50 se ve como de nuevo el sistema obtiene su mejor tasa de identificación utilizando LPP. A diferencia de antes LPP tiene una mejor tasa de reconocimiento para cualquier número de componentes. PCA y DCT presentan nuevamente un rendimiento muy inferior a LPP, y si se compara con el caso anterior este es inferior, perdiendo aproximadamente un 10% en la tasa de identificación correcta.

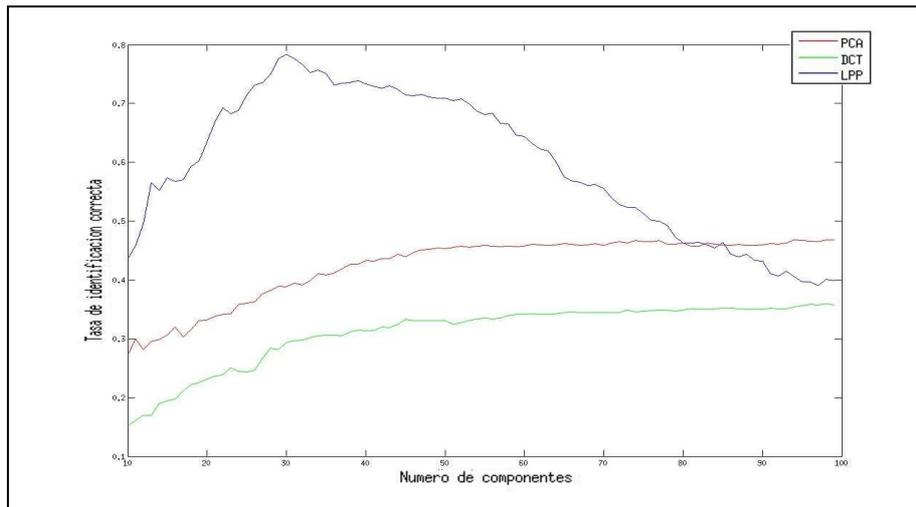


Figura 49. Tasa de identificación correcta obtenida utilizando la base de datos BANCA Controlled (referencia), Degraded (test) y entrenando el sistema con imágenes del grupo Controlled para los diferentes métodos de estudio.

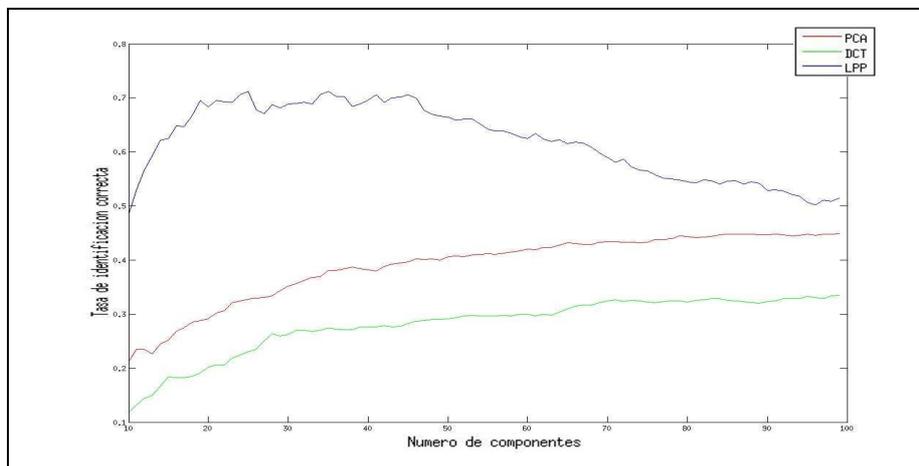


Figura 50. Tasa de identificación correcta obtenida utilizando la base de datos BANCA Degraded (referencia), Controlled (test) y entrenando el sistema con imágenes del grupo Degraded para los diferentes métodos de estudio.

5.6.2 Entrenamiento con imágenes parecidas al set de test

Se repetirá el experimento del apartado anterior pero en lugar de entrenar los sistemas con imágenes de la base de datos utilizada como set de referencia, se hará con las de la base de datos utilizada como set de test. Este tipo de entrenamiento no es el que se encuentra en una aplicación real, pues a priori el sistema no sabe que imágenes van a entrar en el sistema.

Durante este apartado, como se dijo al principio, no se compararán los resultados obtenidos con DCT, aunque se mostraran en los gráficos como referencia, ya que estos no se ven afectados por el entrenamiento.

Para el estudio de este caso, se ha utilizado la siguiente configuración del sistema:

- Base de datos BANCA *Controlled*, con 1223 imágenes (52 individuos distintos y con aproximadamente 23 imágenes para cada individuo).
- Base de datos BANCA *Degraded*, con 1158 imágenes (52 individuos distintos y con aproximadamente 23 imágenes para cada individuo).
- Imágenes de la base de test para el entrenamiento (832 para el caso *Controlled* y 770 para el caso *Degraded*).
- Número de componentes variable de 10 a 100 componentes.
- Los sistemas de reconocimiento PCA y LPP.
- Distancia Euclídea.

Los resultados que hemos obtenido, son los que se pueden ver en las figuras 51 y 52. Como se puede observar, los resultados son inferiores a los obtenidos cuando se entrenaba al sistema con imágenes de la base de datos utilizada como set de referencia.

En la primera gráfica, figura 51, en la que se utiliza como set de referencia imágenes del grupo *Controlled*, la tasa de identificación correcta es algo inferior para LPP y superior para PCA que en el segundo caso (figura 52). LPP

sigue mostrando una mejor tasa de identificación para casi cualquier número de componentes viéndose esta ligeramente reducida en comparación al apartado anterior mientras que PCA presenta prácticamente los mismos resultados no viéndose afectado por el nuevo entrenamiento.

La segunda gráfica, figura 52, LPP presenta unos resultados inferiores a los obtenidos en el segundo caso del apartado 5.6.1 mientras que para PCA son prácticamente los mismos. Nuevamente LPP es el método que presenta una mejor tasa de identificación correcta.

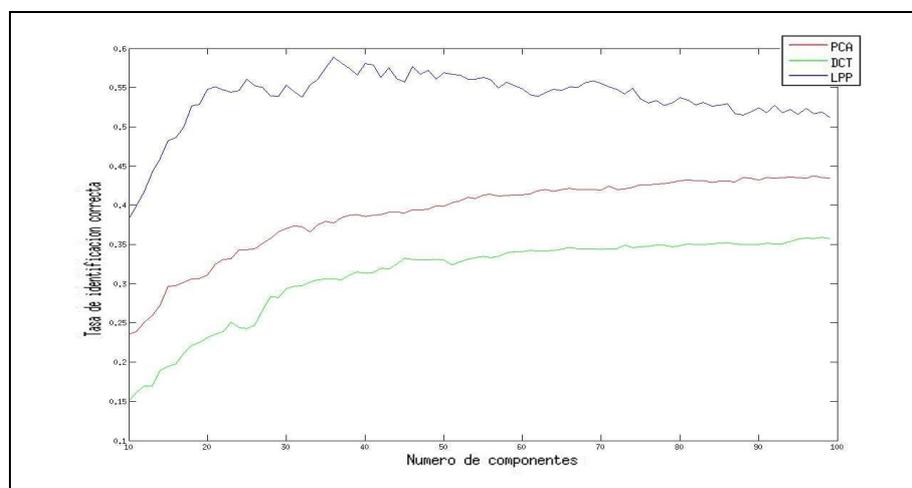


Figura 51. Tasa de identificación correcta obtenida utilizando la base de datos BANCA Controlled (referencia), Degraded (test) y entrenando el sistema con imágenes del grupo Degraded para los diferentes métodos de estudio.

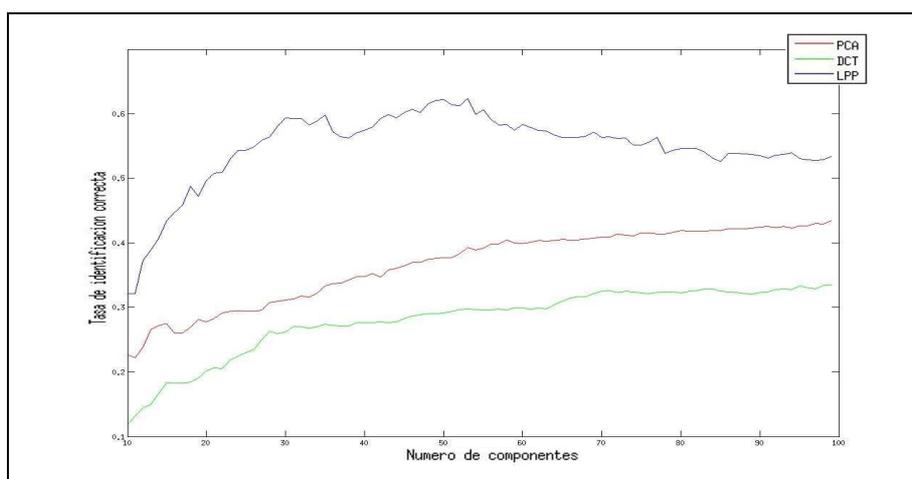


Figura 52. Tasa de identificación correcta obtenida utilizando la base de datos BANCA Degraded (referencia), Controlled (test) y entrenando el sistema con imágenes del grupo Controlled para los diferentes métodos de estudio.

Como conclusión de este apartado hemos podido ver que el sistema tiene su mejor rendimiento si se utilizan como set de referencia imágenes con buena calidad y tomadas en un entorno controlado a la vez que se utilicen imágenes tomadas bajo las mismas condiciones para realizar el entrenamiento del sistema. Por otro lado, hemos podido ver que el método que ofrece unos mejores resultados para este tipo de situaciones es LPP.

Entrenamiento (Controlled), Ref. (Controlled), Test (Degraded)			
Método	Número de componentes max. rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	95	46,852	45.7073% (55)
DCT	99	35,8953	34.6888% (80)
LPP	31	78,332	73.9166% (40)
Entrenamiento (Degraded), Ref. (Degraded), Test (Controlled)			
Método	Número de componentes max. rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	100	44,9050	41.1054% (55)
DCT	100	33,5060	32.3834% (80)
LPP	26	71,1572	68.8256% (40)
Entrenamiento (Degraded), Ref. (Controlled), Test (Degraded)			
Método	Número de componentes max. rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	99	43,7449	40.8831% (55)
LPP	37	58,8716	56.5822% (40)
Entrenamiento (Controlled), Ref. (Degraded), Test (Controlled)			
Método	Número de componentes max. rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	100	43,437	38.9465% (55)
LPP	54	62,3489	57.0812% (40)
Entrenamiento (Controlled), Ref. (Controlled), Test (Controlled) cp. 5.1			
Método	Número de componentes max. Rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	44	100	100% (55)
DCT	29	99,9182	99.8365% (80)
LPP	23	99,8365	99.6729% (40)
Entrenamiento (Degraded), Ref. (Degraded), Test (Degraded) cp. 5.3.1			
Método	Número de componentes max. rendimiento	Tasa de reconocimiento (%)	Tasa de reconocimiento (Número componentes)
PCA	65	96,3731	96.0276% (55)
DCT	100	96,7185	96.5458% (80)
LPP	73	95,6822	95.1641% (40)

Tabla 8. Detalle de resultados Capítulo 5.6 y resultados de los capítulos 5.1 y 5.3.1. En la primera columna: número de componentes máxima tasa de identificación, segunda columna: máxima tasa de identificación, tercera columna: tasa de identificación para el número de componentes elegido para el caso general.

5.7 Eficiencia computacional

Uno de los principales inconvenientes si se quiere montar un sistema en tiempo real, es el tiempo necesario para su ejecución. Para ver si los métodos que se han analizado son apropiados o no para un sistema con estas características, se va a evaluar la eficiencia computacional del sistema.

Para la realización de este experimento no se tendrá en cuenta la utilización de una u otra base de datos ya que el tiempo necesario para la ejecución del sistema no dependerá del conjunto de imágenes utilizado sino del número de componentes y los diferentes procedimientos que se realizan sobre las imágenes en cada uno de los casos. En concreto para este experimento se ha optado por utilizar las imágenes ORL.

Los parámetros del sistema utilizados para este experimento son los siguientes:

- Base de datos Olivetti, formada por 400 imágenes (10 imágenes por persona con 40 individuos).
- Número de componentes variable de 10 a 100 componentes.
- Los sistemas de reconocimiento PCA, LPP y DCT.

No se tendrá en cuenta la preparación previa que requiera cada sistema, en caso de que la necesite. Esto es debido a que una vez instalado el sistema estos procedimientos no se deberán realizar de nuevo y para el usuario esa preparación previa será transparente pues no se verá reflejada en su funcionamiento.

Se va a ver el tiempo necesario para realizar un reconocimiento individual. Como se puede ver en la figura 53, todos los métodos tienen unos requisitos de tiempo parecidos ya que básicamente se trata de comparar 2 vectores de un tamaño determinado. La diferencia de tiempos entre un método y otro radica en el preprocesado que se debe realizar a las imágenes antes de calcular la distancia. Mientras que en el caso de DCT lo único que se debe hacer es

calcular la transformada en cuestión, los otros dos métodos además deben normalizar y preparar las imágenes previamente para ser proyectadas en los nuevos subespacios. Aun así el tiempo necesario no se incrementa de una manera destacable.

Otra cosa que se puede observar es que, como es lógico el tiempo necesario para hacer el reconocimiento aumenta con el número de componentes utilizadas. En este aspecto podemos ver como tanto PCA como LPP incrementan el tiempo necesario de una manera parecida, casi lineal ($t = \alpha \cdot t_i$) mientras que en DCT este incremento es poco notable.

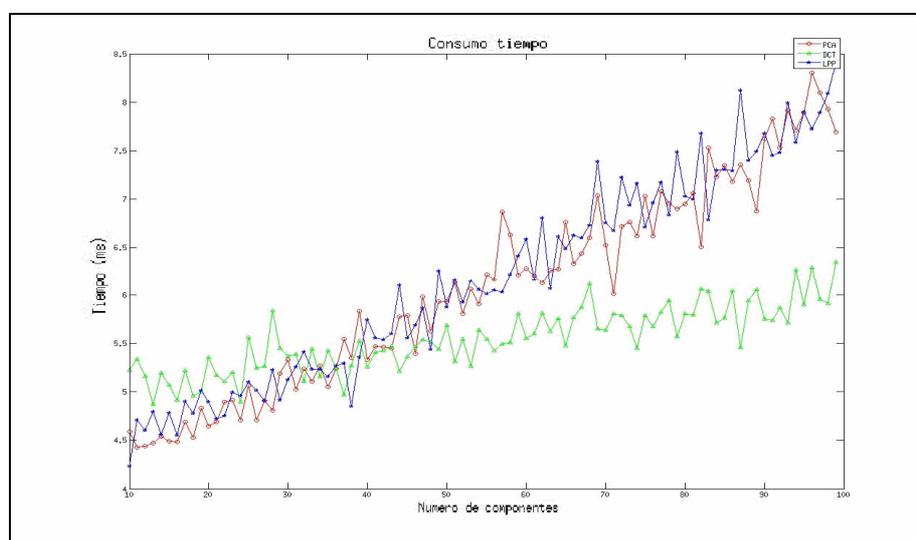


Figura 53. Gráfico de la eficiencia computacional para un reconocimiento individual en función del número de componentes utilizado.

Como conclusión, se puede decir que el método DCT es el más rápido para el reconocimiento de una imagen. Una de las razones principales por las cuales es el método más rápido es que no requiere de una matriz de proyección sino que lo que se hace es una transformación directa de la imagen. Aun así, todos los métodos serían adecuados para un sistema de reconocimiento en tiempo real ya que todos ellos requieren de tiempos parecidos y reducidos.

6 Aplicación I. Query By Example

Una vez analizado el funcionamiento de cada uno de los métodos para diferentes situaciones, se ha implementado una aplicación '*Query By Example*', que es muy parecida a la que ofrecen los programas *picasa* de Google e *iPhoto* de Apple. Dada una imagen de test, esta aplicación, realiza el reconocimiento y devuelve las 5 imágenes que tienen una menor distancia, en el nuevo subespacio, con la imagen de entrada una vez proyectada. En la ejecución de la aplicación se puede seleccionar cuales serán los parámetros utilizados para el reconocimiento (imágenes de test, imágenes de referencia, método a utilizar, número de componentes, matriz de proyección utilizada...).

Para poder evaluar el funcionamiento de esta aplicación frente a una situación lo más '*real*' posible, se va a utilizar con las condiciones reproducidas en el apartado 5.6.1 y en concreto para el caso más favorable en que se utilizan las imágenes tomadas bajo condiciones controladas como referencia. En cuanto a la elección del número de componentes utilizado, se han escogido los valores que daban un máximo rendimiento en ese experimento en particular, aunque estos valores puedan no ser los más adecuados en otras condiciones. Los parámetros utilizados son:

- Base de datos BANCA *Controlled*, con 1223 imágenes (52 individuos distintos y con aproximadamente 23 imágenes para cada individuo), como imágenes de referencia.
- Base de datos BANCA *Degraded*, con 1158 imágenes (52 individuos distintos y con aproximadamente 23 imágenes para cada individuo), como imágenes de test.
- Imágenes de la base de referencia para el entrenamiento (BANCA *Controlled* 832).
- Número de componentes variable según el método (95 PCA, 99 DCT, 31 LPP).
- Los sistemas de reconocimiento PCA, DCT y LPP.

- Distancia Euclídea.

Los resultados obtenidos para la imagen de entrada de la figura 54, son los que se pueden observar en la figura 55. Como es posible observar en la imagen, se verifican los resultados obtenidos en el apartado 5.6.1, el sistema funciona correctamente para el caso de LPP obteniendo 5 resultados correctos de entre los primeros 5, mientras que en los otros casos el sistema no tiene un correcto funcionamiento. De los métodos PCA y DCT, en ninguna ocasión, tenemos un reconocimiento correcto de entre los 5 primeros.



Figura 54. Imagen que hemos utilizado para la entrada del sistema.



Figura 55. Resultados para cada uno de los métodos (primera fila PCA, segunda DCT y tercera LPP). De izquierda a derecha resultados de la primera a la quinta posición.

7 Aplicación II. Sistema de reconocimiento en tiempo real

Consiste en una aplicación en tiempo real para ver el rendimiento del sistema en un entorno de funcionamiento práctico. Para la realización de esta aplicación se ha utilizado partes de código ya existentes en la plataforma *ImaGepus* del Grupo de Procesado de Imagen y Video de la UPC las cuales han sido modificadas para que funcionen con LPP. La aplicación se ha utilizado en el entorno de la *smart room* de la UPC y consiste en detectar y reconocer los individuos que entran en la sala.

Esta aplicación se ha ejecutado haciendo uso de un video ya obtenido de dicha sala y se ha ejecutado toda en un hilo, sin hacer uso de diferentes procesadores ni ejecuciones en paralelo, de manera que para la creación del video final, que se adjunta con el CD del proyecto, no se ejecuta en tiempo real propiamente dicha aunque se podría configurar para tal caso.

Esta aplicación realiza reconocimiento facial basado en video con fusión de los resultados individuales [33], es decir, no se muestran los reconocimientos instantáneos sino que se hace uso del resultado acumulado de las últimas N imágenes (N=10). De este modo, el resultado mostrado en el video es más estable y se evita que varíe constantemente en caso de reconocimiento incorrecto. Para realizar este reconocimiento, aunque se hace uso de modelos para identificar a los individuos, seguimos empleando el criterio de distancia mínima aunque en este caso se utiliza la distancia *Manhattan*.

Las imágenes utilizadas para esta aplicación, son sin recortar y presentan caras en diferentes posiciones y expresiones. Sobre estas imágenes se detectarán las caras frontales y laterales haciendo uso del detector facial Viola-Jones. Esta detección puede no ser siempre correcta o puede no poderse realizar, en este caso no se muestra ningún resultado.

Los métodos utilizados para la ejecución de esta aplicación han sido DCT y LPP. El número de componentes utilizado para el reconocimiento en cada caso ha sido el que se ha determinado como más adecuado para un sistema general.

Los parámetros utilizados en esta aplicación son los siguientes:

- Imágenes obtenidas del video de la sala como set de test (se trata de 2706 imágenes no recortadas con 12 individuos diferentes).
- Modelos creados a partir de las imágenes de entrenamiento usados como set de referencia (aproximadamente disponemos de unas 200 imágenes recortadas para cada individuo).
- Entrenamiento realizado con las imágenes utilizadas para la creación de modelos (un total de 3252 imágenes).
- Número de componentes variable según el método (80 DCT, 40 LPP).
- Los sistemas de reconocimiento DCT y LPP.
- Distancia Manhattan

Los resultados obtenidos del uso de esta aplicación son los que se pueden ver en los videos adjuntos en el CD final del proyecto, y para ver de manera exacta el funcionamiento se muestran en la tabla 10 las tasas de reconocimiento instantáneo para cada cara detectada y el acumulado de cada 10 imágenes.

Método	Resultados instantáneos	Resultados acumulados
DCT	78.744%	81.7%
LPP	77.456%	85.4%

Tabla 10. Resultados obtenidos haciendo uso de la aplicación en tiempo real.

Como se puede observar en la tabla los resultados instantáneos, los referentes al reconocimiento en cada imagen donde se detecta una cara, son muy parecidos en los dos métodos. En este punto debe tenerse en cuenta que se reconocen imágenes de muy baja calidad (justo cuando la persona entra por la puerta) provenientes de un video entrelazado que inducen fácilmente a error.

También se ha de tener en cuenta que se han utilizado los parámetros escogidos para el caso general y aunque para DCT a partir de cierto número de componentes su funcionamiento es bastante estable, para LPP en situaciones muy extremas, véase apartados 5.3 y 5.4, tiene su mejor funcionamiento para un número de componentes muy bajo, y en este caso las imágenes son con orientaciones muy variables.

En la segunda columna de la tabla se puede ver la tasa de reconocimiento acumulada, la cual es el indicador más importante en esta aplicación ya que se asigna una etiqueta a un individuo en concreto mediante este parámetro. Se puede observar que LPP supera con claridad a DCT en este parámetro aún teniendo peor tasa de reconocimiento instantánea. Esto indica que LPP es capaz de reconocer de forma constante a prácticamente todos los individuos mientras que DCT para algunos individuos no es capaz de realizar un buen reconocimiento.

8 Conclusiones y trabajo futuro

Una vez se ha evaluado el funcionamiento de cada uno de los tres métodos de estudio frente a diferentes situaciones, se ha podido obtener información sobre el tipo de entrenamiento que requieren los sistemas o el número de componentes necesarios para obtener unos valores de reconocimiento aceptables.

Como se ha explicado durante este documento, los métodos PCA y LPP requieren de un entrenamiento previo del sistema. La manera de realizar este entrenamiento, como se ha podido ver en el capítulo 5.5, afecta de manera distinta en cada caso. Mientras que para el caso de PCA es muy importante el número de imágenes utilizado, aún siendo estas ligeramente diferentes a las que se usarán para el reconocimiento o con mucha variedad de individuos, mientras que en el caso de LPP se debe tener en cuenta de utilizar conjuntos de imágenes que contengan un elevado número de imágenes por individuo ya que de esta manera el sistema ofrece unos mejores resultados.

En cuanto al número de componentes utilizado para el reconocimiento, se ha podido ver que cada uno de los métodos requiere un número de componentes diferente para obtener su mejor rendimiento. Al mismo tiempo se ha observado que el tipo de situación en el que se va a utilizar el sistema afecta también a este número de componentes. A la hora de hacer una elección de este parámetro, se debe buscar un compromiso entre respuesta del sistema, entorno de uso y el tiempo necesario para el reconocimiento. A la finalización de este proyecto se han escogido cuales son los valores óptimos a usar en el caso de una aplicación general, es decir, una aplicación que puede enfrentarse a todas las situaciones contempladas en el capítulo 5 con un funcionamiento correcto. Los valores han sido 55 componentes para PCA, 80 para DCT y 40 para LPP (los resultados obtenidos para estos valores se muestran en las tablas resumen de cada uno de los experimentos realizados) aunque estos pueden modificarse en caso de utilizar el sistema bajo unas condiciones muy

concretas y conocidas con anterioridad. De los resultados obtenidos se ha llegado a la conclusión de que el sistema en un entorno controlado, con un correcto entrenamiento y con una buena base de imágenes de referencia, puede funcionar hasta con 20 componentes menos que las seleccionadas para caso general, mientras que para condiciones no controladas como cambios de orientación de los rostros y cambios de iluminación uniformes, el sistema tendrá un rendimiento algo superior utilizando entre 10 y 20 componentes más que las seleccionadas. En el caso de tener imágenes en condiciones extremas (iluminación no uniforme y cambios de orientación y expresión considerables) el sistema ofrece sus mejores resultados, utilizando LPP, para valores del número de componentes muy reducidos (10-30 componentes), como se puede ver en los capítulos 5.3 y 5.4, ya que las primeras componentes son las que se ven menos afectadas por los cambios de iluminación en este método. Los casos de entrenamientos incorrectos o de bases de referencia que no cumplan un buen ratio entre el número total de imágenes y el número de individuos no se tienen en cuenta para la elección de estos valores ya que se presupone que al preparar un sistema de este tipo estos pasos se realizarán correctamente.

Después de analizar el efecto de los cambios de iluminación, los cambios en la posición de los rostros y el tipo de imágenes utilizado como referencia, se puede decir que el método más consistente y que ofrece un mayor índice de identificación correcta es LPP, el cual además es el más eficaz frente a situaciones adversas para el sistema. Estas características mostradas por LPP serían muy apropiadas para sistemas de reconocimiento facial no intrusivos, como podrían ser sistemas de vídeo vigilancia, ya que estos sistemas funcionan con imágenes que pueden presentar iluminación variable y las imágenes no presentan una posición frontal de las caras, situaciones en las que LPP ofrece mejor respuesta que los otros métodos.

Como ya se dijo en la introducción, este proyecto se ha centrado en la comparativa de diferentes métodos de reconocimiento con el fin de escoger cual sería el más apropiado para futuros trabajos o para futuras aplicaciones. La decisión que se tomaría, a la vista de los resultados, sería la de utilizar el

algoritmo LPP para el reconocimiento, aunque en futuros trabajos este sistema puede ser mejorado.

Algunas de las proposiciones que se realizan para trabajos futuros son utilizar pequeñas variaciones del método LPP original y hacer uso de otras técnicas en la etapa de decisión. Es por ello que para futuros trabajos se recomienda el estudio de los métodos derivados de LPP como son OLPP (Orthogonal laplacianfaces) [30] [32], ONPP (Orthogonal Neighborhood Preserving Projections) [31] o DLPP Discriminant Locality Preserving Projections [29], a la vez que se implementen sistemas de decisión más complejos que permitan hacer uso de modelos y hagan uso de imágenes en diferentes posiciones, teniendo en cuenta imágenes frontales y de perfil, para realizar el reconocimiento.

Por otro lado, ya que LPP preserva la estructura local de los datos y presenta un buen rendimiento frente a situaciones adversas, como cambios en la orientación y expresiones faciales, se podría estudiar la posibilidad de utilizar este método para el reconocimiento de expresiones faciales o estados de ánimo (sonrisa, felicidad, tristeza...). También puede ser interesante ver si utilizar métodos no supervisados para el entrenamiento de LPP puede ayudar a que el sistema tenga un mejor funcionamiento al ser entrenado con imágenes de individuos distintos a los que se tengan en el sistema.

Por último y no solamente para el método LPP se puede estudiar las ventajas de utilizar imágenes en color para mejorar las tasas de identificación y otros métodos de compensación de los cambios de iluminación.

9 Bibliografía

- [1] Cognitec Systems GmbH,
<<http://www.cognitec-systems.de/contact/contact.html>>.
- [2] Eyematic Interfaces Inc., <<http://www.eyematic.com>>.
- [3] Viisage, Littleton, MA, <<http://www.viisage.com>>.
- [4] Identix, Minnetonka, MN, <<http://www.identix.com>>.
- [5] International Biometric Group, <<http://www.biometricgroup.com>>.
- [6] R.Hietmeyer, "Biometric identification promises fast and secure processing airline passengers", *The Int'l Civil Aviation Organisation Journal*, vol. 55, no. 9, pp. 10-11, 2000.
- [7] Yale University face database, <<http://cvc.yale.edu/projects/yalefaces.html>>, Athinodoros Georghiades, Peter Belhumeur, and David Kriegman's paper, "*From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose*", *PAMI*, 2001.
- [8] Xiaoguang Lu, "*Image analysis for FACE Recognition*", Department of Computer Science & Engineering, Michigan State University, May 2003.
- [9] Xiaoguang Lu, Dirk Colbry, and Anil K. Jain, "*Three-Dimensional Model Based Face Recognition*", *ICPR* August 2004.
- [10] M. Kirby and L. Sirovich, "*Application of the Karhunen-Loeve procedure for the characterization of human faces*" *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103–108, Jan. 1990.
- [11] Abdallah S. Abdallah, A. Lynn Abbott, and Mohamad Abou El-Nasr. "*A New Face Detection Technique using 2D DCT and Self Organizing Feature Map*". *Proceedings of World Academy of Science, Engineering and technology* Vol. 21 May 2007 ISSN 1307-6884
- [12] Ziad M. Hafed and Martin D. Levine. "*Face Recognition Using the Discrete Cosine Transform*". *International Journal of Computer Vision* 43(3), 167–188, 2001.
- [13] Yong Xu and David Zhang. "*A New Solution Scheme of Unsupervised Locality Preserving Projection Method for the SSS Problem*", *SSPR'08*

(International Workshop on Structural, Syntactic, and Statistical Pattern Recognition),775-781.

[14] Xiaofei He & Partha Niyogi. “*Locality Preserving Projections*”, NIPS’03 volume 16.

[15] Duda R., Hart P., Stork D.. ”*Pattern Classification*” Capítulo 4.

[16] Jieping Ye, Ravi Janardan, Qi Li. “*Two-Dimensional Linear Discriminant Analysis*”, NIPS’04

[17] Naotoshi Seo. “*Eigenfaces and Fisherfaces*”. University of Maryland

[18] *ORL face database* <<http://www.uk.research.att.com/facedatabase.html>>, Ferdinando Samaria, Andy Harter, “*Parameterisation of a Stochastic Model for Human Face Identification*”, Proceedings of 2nd IEEE Workshop on Applications of Computer Vision, Sarasota FL, December 1994.

[19] Hazim Kemal Ekenel, Rainer Stiefelhagen “*Local appearance based face recognition using discrete cosinus transform*”, 13th European Signal Processing Conference (EUSIPCO 2005).

[20] Yong Xu and David Zhang. “*A New Solution Scheme of Unsupervised Locality Preserving Projection Method for the SSS Problem*”, SSPR’08 (International Workshop on Structural, Syntactic, and Statistical Pattern Recognition),775-781.

[21] Cheng J., Liu Q., Lu H. “*Supervised kernel locality preserving projections for face recognition*”, Neurocomputing 67, 443–449 (2005).

[22] Feng G., Hu D., Zhang D., Zhou Z., “*An alternative formulation of kernel LPP with application to image recognition*”, Neurocomputing 69(13-15), 1733–1738 (2006).

[23] Hu D., Feng, G., Zhou Z. “*Two-dimensional locality preserving projections (2DLPP) with its application to palmprint recognition*”, Pattern Recognition 40(1), 339–342 (2007)

[24] Cai, D., He, X. “*Orthogonal locality preserving indexing*”, SIGIR 2005, pp. 3–10 (2005)

[25] XM2VTS face database <<http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/>>, K. Messer, J. Matas, J. Kittler, “*XM2VTSDB: The extended M2VTS Database*”, AVBPA’99.

[26] Biometric Source information website, <<http://www.biometrics.gov/>>.

[27] Face Recognition Homepage, <<http://www.face-rec.org/>>.

-
- [28] BANCA face database, <<http://www.ee.surrey.ac.uk/CVSSP/banca/>>, K. Messer, J Kittler, M. Sadeghi, “*Face Authentifacacion Competition on th BANCA Database*”, ICPR’04.
- [29] Weiwei Yu, Xiaolong Teng, and Chongqing Liu, “*Discriminant Locality Preserving Projections: A New Method to Face Representation and Recognition*”, Proceedings 2nd Joint IEEE International Workshop on VS-PETS, Beijing, October 15-16, 2005.
- [30] Zhonglong Zheng, Jianmin Zhao, “*Locality Preserving Projection in Orthogonal Domain*”, 2008 Congress on Image and Signal Processing.
- [31] E. Kokiopoulou and Y. Saad, “*Orthogonal Neighborhood Preserving Projections*”, IEEE Int. Conf. on Data Mining (ICDM) 2005.
- [32] D. Cai, X. He, J. Han, and H. Zhang, “*Orthogonal Laplacianfaces for Face Recognition*”, IEEE Transactions on Image Processing, vol. 15, no. 11, pp. 3608-3614, November, 2006.
- [33] PFC Edison Cristófani “*Video-based face recognition using multiple face orientations*”, Department of Signal Theory and communications UPC.
- [34] PFC Claudi Martínez “*FACE recognition in the context of Smart Rooms*”, Department of Signal Theory and communications UPC.
- [35] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, “*The FERET evaluation methodology for face-recognition algorithms*”, IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 10, pp. 1090–1104, Oct. 2000.
- [36] J. Wang, K. N. Plataniotis, J. Lu, and A. N. Venetsanopoulos, “*On solving the face recognition problem with one training sample per subject*”, Pattern Recognit., vol. 39, no. 6, pp. 1746–1762, Sep. 2006.

