

DEPARTAMENTO DE TECNOLOGÍA
FOTÓNICA

FACULTAD DE INFORMÁTICA

Técnicas de reconocimiento facial mediante redes
neuronales

Tesis Doctoral

Autor: Enrique Cabello Pardos
Licenciado en Ciencias Físicas

Director: Luis Pastor Pérez
Doctor Ingeniero Industrial

Abril de 2004

Título:
Técnicas de reconocimiento facial mediante redes
neuronales

Autor:
Enrique Cabello Pardos

Tribunal:

Presidente: Juan Antonio Hernández Tamames

Vocales:

Antonio José Albiol Colomer

Juan Carlos Crespo Zaragoza

Carlos Cerrada Somolinos

Secretario: Angel Rodríguez Martínez de Bartolomé

Suplentes:

Javier Martínez Moguerza

Alberto Muñoz García

Acuerdan otorgar la calificación de:

Madrid, de de 2004

Agradecimientos

La presente tesis se ha desarrollado en tres Universidades, la Universidad de Salamanca, la Universidad Politécnica de Madrid y la Universidad Rey Juan Carlos. Comenzó hace mucho tiempo y ahora ya está a punto de terminar. A lo largo de tanto tiempo he conocido, trabajado y colaborado con muchas personas, me gustaría agradecerles a todas ellas su participación, su interés y esfuerzo. Aunque a continuación trataré de citar a todos, si alguno no aparece es sólo porque mi memoria lo ha omitido, no mi agradecimiento.

La primera persona que quiero citar es Luis Pastor, después de tanto tiempo juntos y de desesperarme siendo tan puntilloso en las revisiones y obligarme a cambiar las cosas en una versión para en la siguiente revisión dejarlo como estaba anteriormente, es increíble que sigamos siendo amigos. Muchas gracias, Luis.

Quisiera también agradecer a todos los amigos de la Universidad de Salamanca del Departamento de Informática y Automática y de la Facultad de Ciencias, especialmente a Araceli Sánchez, César Raposo, Ángel Labajo, Pastora Vega, Fernando Atrio y Belén Pérez.

Quisiera reconocer a todos los colaboradores del Departamento de Tecnología Fotónica de la Universidad Politécnica de Madrid, sobre todo a Ángel Rodríguez, Antonio Ruiz, Gracian Triviño y Salvador Marcalain. También a los que hemos ido coincidiendo allí: Luis Sánchez y Ángel Puebla. Quisiera reconocer también a los compañeros del grupo de Arquitectura de la Universidad Rey Juan Carlos, especialmente muchísimas gracias a Cristina Conde, Susana Mata, José Luis Bosque, Beatriz Romero, Carmen Coya, Luis Rincón, Oscar Robles, compañero Juan Antonio Hernández, Nacho Martínez y Jean Pierre Deschamps, Marta Beltrán, David Nevado, Roberto de la Prieta, a los recién llegados Licesio Rodríguez y Ángel Serrano, Susana Borromeo y Ángel Luis Alvarez. No puedo olvidarme de los que he ido conociendo en los proyectos que he participado: Antonio Guzmán, Paco Martín, Belén Moreno, Antonio Sanz, Guido Castro y Ángel Sánchez. También quiero mostrar mi agradecimiento a David Ríos, Regino Criado, Alejandro García del Amo (cuya cara fue muy útil en nuestros experimentos), Javier Moguerza, Jorge Muruzabal, Luis Solá, Raquel Montes. También quiero recordar a todos los muchachos cuyos proyectos fin de carrera he ido dirigiendo a lo largo de los últimos años. Su trabajo en los distintos laboratorios ha sido siempre una motivación adicional para seguir adelante.

Quisiera escribir unas breves líneas muy personales para agradecer la vida juntos a lo largo de todos los días de todos los años, para Almudena. Asimismo deseo manifestar mi cariño para mi madre, para Juan y Loly, Ana Belén, Carmen Nieves, Fátima y Antonio, Lalín, Márían y Sonsoles. Quisiera recordar también a Petra, Charo, Leandro y Angeles, Pedro y Petri y Fernando. Acordarme en este momento de todos los tíos y primos; ligados a los veranos en Morata, el olor a fruta recién cogida, las tardes de verano en que te sientes libre, mis recuerdos más gratos de la infancia están allí.

Para terminar, tengo un recuerdo emocionado para mi abuelita Paulina y los “compañeros del alma, compañeros”, Mariano y Leandro, Santiago y Santiago.

Resumen

En esta Tesis se exploran soluciones a la verificación facial. Se han estudiado técnicas basadas en imágenes bidimensionales y se ha realizado un estudio inicial basado en un modelo tridimensional de la cara. Se han comparado tres clasificadores (k vecinos más cercanos, redes neuronales del tipo funciones de base radial y máquinas de vector soporte), mostrando los resultados obtenidos con los diferentes datos de entrada.

En dos dimensiones se han propuesto dos técnicas de reducción de información, basadas en la utilización de análisis de componentes principales (PCA) y el empleo de imágenes de baja resolución. En la verificación partiendo de imágenes se puede observar que el clasificador que mejores resultados proporciona es el basado en máquinas de vector soporte (SVM), y el mejor método de procesamiento es PCA utilizando una plantilla por sujeto y siendo esta plantilla una imagen del sujeto que se quiere verificar.

Además se ha desarrollado un experimento inicial que permite disponer de una idea intuitiva sobre el funcionamiento de un sistema de verificación facial basado en datos tridimensionales. En esta Tesis se ha desarrollado un modelo de cara en tres dimensiones y se ha comprobado que es posible su ajuste a una nube de puntos poco densa con un error pequeño. Este tipo de datos permite obtener resultados más robustos frente a actitudes no colaborativas de los sujetos.

Abstract

In this thesis some face verification-oriented solutions are explored. Bidimensional image-based techniques have been considered and an initial study based on a 3D face model has been carried out. Three classifiers have been compared (k -nearest neighbours, neural networks-radial basis functions and support vector machines), and the results obtained with different inputs are shown.

Two information reduction techniques have been proposed in two dimensions, based on the use of a principal component analysis (PCA) and low resolution images. In the verification it can be seen that the support vector machine classifiers (SVM) yield the best results, and the best processing method is PCA using one template per subject, as long as this template is a picture of the subject to verify.

We have developed as well an initial experiment that allows us to have an intuitive idea about the performance of a 3D data-based face verification system. In this thesis a 3D face model has been developed and we have checked that a low density cloud of points fits with low error. This kind of data allows us to obtain more robust results with subjects with a non-collaborative attitude.

Índice General

Capítulo 1	1
Introducción	1
1.1.- Introducción y motivación	1
1.2.- Objetivo de la tesis	2
1.3.- Planteamiento del problema y solución propuesta	2
1.3.1.- Dificultades esperadas del problema	3
1.4.- Estructura de la memoria	4
Capítulo 2	5
Estado del arte	5
2.1.- Clasificadores	5
2.1.1.- Redes Neuronales: Funciones de base radial (RBF)	6
2.1.1.1.- Perceptrón Multicapa	8
2.1.1.2.- Funciones de base radial (RBF)	9
2.1.2.- Máquinas de vector soporte (Support Vector Machines: SVM)	14
2.1.3.- K vecinos más cercanos (K nearest neighbours, KNN)	17
2.2.- Reconocimiento de caras humanas	18
2.2.1.- Introducción	18
2.2.1.1.- Detección de la cara	19
2.2.1.2.- Representación	20
2.2.1.3.- Clasificación: Identificación o Verificación	20
2.2.2.- Estado del arte en el reconocimiento y verificación facial automática	21
2.2.3.- Características geométricas	25
2.2.4.- Análisis de Componentes Principales (PCA). Autocaras	26
2.2.4.1.- Descripción del algoritmo	26
2.2.4.2.- Resultados obtenidos con PCA	31
2.2.5.- Análisis Lineal Discriminate (LDA). Fisherfaces	34
2.2.6.- Correspondencia de plantillas	35
2.2.7.- Redes Neuronales	35
2.2.7.1.- Enfoque basado en el análisis de componentes principales (PCA). Memoria autoasociativa	36
2.2.7.2.- Redes de retropropagación: compresión de imágenes. Aplicación al procesamiento de imágenes faciales	37
2.2.8.- Correspondencia entre agrupaciones de grafos elásticos (EBGM). Transformada wavelet (onditas)	39
2.2.9.- Modelos de cara en 3D	42
2.2.10.- Otras técnicas	44
2.2.11.- Programas comerciales	45

Capítulo 3	51
Entorno de la tesis.	51
3.1.- Introducción.	51
3.2.- Sistema de visión estéreo.	52
3.2.1.- Plantilla de calibración. Algoritmo de calibración.	52
3.3.- Base de datos para adquisición de información tridimensional. Universidad de Salamanca.	53
3.4.- Sistema de adquisición de imágenes bidimensionales.	55
3.5.- Base de datos para adquisición de información bidimensional. Universidad Rey Juan Carlos.	57
 Capítulo 4	 59
Reconocimiento bidimensional de caras.	59
4.1.- Introducción.	59
4.2.- Preprocesamiento de imágenes	60
4.2.1.- Localización de la cara.	60
4.2.2.- Niveles de gris utilizando análisis de componentes principales.	63
4.2.3.- Niveles de gris utilizando imágenes de baja resolución.	64
4.3.- Clasificación facial.....	65
4.4.- Resultados experimentales y discusión.	66
 Capítulo 5	 81
Reconocimiento tridimensional de caras.	81
5.1.- Introducción.	81
5.2.- Calibración de cámaras y reconstrucción tridimensional. Marcación de puntos.	82
5.3.- Características geométricas.	83
5.3.1.- Medidas en la cara.	87
5.3.1.1.- Distancias sobre la cara.	87
5.3.1.2.- Angulos sobre la cara.	87
5.3.2.- Definición de las características.	88
5.3.3.- Captura de las coordenadas de los puntos.	91
5.3.4.- Cálculo de la robustez de los puntos somatométricos y de las características geométricas.	91
5.3.5.- Estudio del poder discriminante de las características geométricas.	93
5.3.6.- Recuperación de las características geométricas con componentes principales mayores.	94
5.3.7.- Puntos y características seleccionados finalmente.	95
5.4.- Clasificación de características.	97
5.5.- Resultados experimentales y análisis.	97
5.5.1.- Resultados KNN.	98
5.5.2.- Resultados RBF.	99
5.5.3.- Resultados SVM.	101
5.5.4.- Análisis y discusión de los resultados.	103
 Capítulo 6	 105
Conclusiones. Lineas de investigación futuras.	105
6.1. – Descripción del capítulo.	105
6.2.- Conclusiones obtenidas al comparar las diferentes técnicas desarrolladas. Aportaciones realizadas.	105
6.3.- Aplicación industrial: VISOR-BASE.	107
6.4.- Proyectos futuros.	108
 BIBLIOGRAFÍA.....	 109

Anexo A	117
Modelo tridimensional considerado.....	117
A.1.- Resultados numéricos: ajuste del modelo 3D a fotografías 2D.....	119
A.2.- Agregado de texturas al modelo tridimensional.....	121
Anexo B	123
Verificación bidimensional, datos equilibrados	123
Anexo C	129
Verificación bidimensional, datos no equilibrados	129

Índice de Figuras

Figura 2.1. Ejemplo de neurona artificial	7
Figura 2.2. Ejemplo de red neuronal	7
Figura 2.3. Ejemplo de red RBF	9
Figura 2.4. Función de activación de una neurona RBF	10
Figura 2.5. Ejemplo de red neuronal del tipo RBF.....	11
Figura 2.6. Unidad RBF utilizada por el algoritmo DDA. Los umbral es se utilizan para definir una “región de conflictos” a la cual ningún otro prototipo puede pertenecer. Además, cada patrón de entrenamiento tiene que estar en el círculo interior de al menos un prototipo de la clase adecuada.....	12
Figura 2.7. Ejemplo de entrenamiento DDA-RBF. (1) se encuentra un patrón de la clase A y se crea una nueva RBF. (2) Un patrón de entrenamiento de la clase B conduce a un nuevo prototipo de la clase B y se reduce el radio de las RBF existentes de la clase A. (3) Otro patrón de la clase B se clasifica correctamente y reduce el radio del prototipo de la clase A. (4) Un nuevo patrón de la clase A crea otro prototipo de esta clase	13
Figura 2.8. Ejemplo de SVM	14
Figura 2.9. Representaciones internas de la aproximación PCA y de la LDA. Obtenidas de [Wen99]. La imagen media se muestra en la primera columna. La secuencia (a) corresponde a las características más expresivas. La secuencia (b) corresponde a las más discriminantes.	23
Figura 2.10. Representación interna de la aproximación EBGM (obtenidos de Wis97). (a) representa un grafo sobre una cara. (b) es la reconstrucción de la cara a partir del grafo. (c) representa la reconstrucción de la cara a partir de un grafo conjunto en el que se ha considerado en cada nodo el jet que mejor se adapta. (d) es una interpretación de la representación de grafo grupo (véase el apartado 2.2.8).....	23
Figura 2.11. Representación interna del método LFA (obtenida de [Pen96]). (a) representa la cara media con la localización de cinco características. (b) muestra los kernels asociados a cada característica (fila superior) y la correspondiente correlación (fila inferior).	24
Figura 2.12. Características geométricas consideradas en [Bru93].	26
Figura 2.13. Seis primeras autocaras de una base de datos.	31
Figura 2.14. Cara media para la misma base de datos.....	31
Figura 2.15. Regiones consideradas en [Bru93] para la correspondencia de plantillas.....	35
Figura 2.16. Modelo de cara utilizando medidas antropométricas (obtenido de [Dec98]). (a) Medidas antropométricas. (b) Modelo B-spline.....	42
Figura 2.17. Modelos genéricos de cara. (a) Modelo de Water. (b) seis clases de modelos genéricos para representar la geometría facial.	43
Figura 2.18. Modelo propuesto en [Hsu02].....	44
Figura 2.19. Imagen inicial e imagen transformada en log-polar (obtenida de [Esc02]).	45
Figura 2.20. Resultados de la prueba HCInt.....	48
Figura 2.21. Porcentaje de reconocimiento en función del porcentaje de falsas alarmas.....	49
Figura 2.22. Reconocimiento correcto en función del porcentaje de falsa aceptación.....	49
Fig 2.23. Resultados MCInt con FAR=0.01.....	50
Fig 2.24. Resultados MCInt, independientemente del FAR.....	50

Figura 3.1. Esquema de adquisición de imágenes en estéreo.....	52
Figura 3.2. Plantilla de calibración (vista estéreo)	53
Figura 3.3. Ocho imágenes de un sujeto, correspondientes a cuatro sesiones.....	54
Figura 3.4. Algunos de los sujetos de la base de datos.....	55
Figura 3.5. Dos sujetos en actitud no colaborativa (cambios en la expresión).....	55
Figura 3.6. Esquema de adquisición de imágenes con luz controlada.....	56
Figura 3.7. Entorno de trabajo con los dos focos de luz difusa y la cámara CCD.....	56
Figura 3.8. Sombras faciales debidas a la iluminación cenital.....	57
Figura 3.9. Imagen iluminada con luz difusa	57
Figura 3.10. Imagen iluminada con luz no controlada	57
Figura 3.11. Sujetos de la base de datos.....	52

Figura 4.1. Recorte manual de una base de datos.....	61
Figura 4.2. Recorte utilizando el color de la piel de una base de datos.....	62
Figura 4.3. Recorte automático de una base de datos.....	63
Figura 4.4. Varianza explicada en función del número de autovectores seleccionados.....	64
Figura 4.5. Imagen inicial, Gaussiana y segmentada de un sujeto.....	65
Figura 4.6. Cara localizada correctamente	68
Figura 4.7. La misma cara pero desplazada 5x5, 10x10, 15x15 y 20x20 píxeles.....	68
Figura 4.8. Sujetos de la base de datos reducida.....	68
Figura 4.9: Resultados con imágenes correctamente localizadas: baja resolución.....	69
Figura 4.10: Resultados con imágenes correctamente localizadas: análisis de componentes principales ..	70
Figura 4.11: Resultados con imágenes desplazadas 5 píxeles: baja resolución.....	72
Figura 4.12: Resultados con imágenes desplazadas 5 píxeles: análisis de componentes principales.....	73
Figura 4.13: Resultados con imágenes desplazadas 10 píxeles: baja resolución.....	74
Figura 4.14: Resultados con imágenes desplazadas 10 píxeles: análisis de componentes principales.....	75
Figura 4.15: Resultados con imágenes recortadas con una plantilla por sujeto: baja resolución.	77
Figura 4.16: Resultados con imágenes recortadas con una plantilla por sujeto: análisis de componentes principales.....	78

Figura 5.1. Esquema de la geometría estéreo considerada.....	82
Figura 5.2. Marcado de un punto en la imagen derecha y su correspondiente recta epipolar en la imagen izquierda.....	83
Figura 5.3. Puntos somatométricos.....	85
Figura 5.4. Algunas distancias del vector de características.....	90
Figura 5.5. Algunos ángulos del vector de características.....	90
Figura 5.6. Robustez de los puntos somatométricos.....	92
Figura 5.7. Puntos considerados para su reconstrucción en 3D.....	96
Figura 5.8. Resultados de la técnica KNN para sujetos colaborativos (características geométricas en 3D).	99
Figura 5.9. Resultados de la técnica KNN para sujetos no colaborativos (características geométricas en 3D)	99
Figura 5.10. Resultados de la técnica RBF para sujetos colaborativos balanceados (características geométricas en 3D)	100
Figura 5.11. Resultados de la técnica RBF para sujetos colaborativos no equilibrados (características geométricas en 3D)	100
Figura 5.12. Resultados de la técnica RBF para sujetos no colaborativos no equilibrados (características geométricas en 3D)	101
Figura 5.13. Resultados de la técnica SVM para sujetos colaborativos equilibrados (características geométricas en 3D)	102
Figura 5.14. Resultados de la técnica SVM para sujetos colaborativos no equilibrados (características geométricas en 3D)	102
Figura 5.15. Resultados de la técnica SVM para sujetos no colaborativos no equilibrados (características geométricas en 3D)	103

Figura A.1. Diversas vistas del modelo 3D de caras humanas.....	117
Figura A.2. Modelo y su ajuste a una cara.....	119
Figura A.3. Modelo ajustado a una cara.....	122
Figura A.4. Tres vistas sintéticas obtenidas de la figura anterior.....	122
Figura B.1. Resultados con imágenes desplazadas 15 píxeles: baja resolución	124
Figura B.2. Resultados con imágenes desplazadas 15 píxeles: análisis de componentes principales	125
Figura B.3. Resultados con imágenes desplazadas 20 píxeles: baja resolución	126
Figura B.4. Resultados con imágenes desplazadas 20 píxeles: análisis de componentes principales	127
Figura C.1.- Preprocesamiento utilizando baja resolución con clasificador SVM.....	129
Figura C.2.- Preprocesamiento utilizando baja resolución con clasificador RBF.....	130
Figura C.3.- Preprocesamiento utilizando análisis de componentes principales con clasificador SVM..	130
Figura C.4.- Preprocesamiento utilizando análisis de componentes principales con clasificador RBF..	131
Figura C.5.- Preprocesamiento utilizando baja resolución con clasificador SVM.....	132
Figura C.6.- Preprocesamiento utilizando baja resolución con clasificador RBF.....	133
Figura C.7.- Preprocesamiento utilizando análisis de componentes principales con clasificador SVM..	134
Figura C.8.- Preprocesamiento utilizando análisis de componentes principales con clasificador RBF..	135

Índice de tablas

Tabla 4.1. Posibles datos de salida del sistema.....	67
Tabla 5.1. Robustez de los puntos somatométricos y de las características geométricas.....	92
Tabla 5.2. Robustez de las características.....	93
Tabla 5.3. Poder discriminante de las características.....	94
Tabla 5.4. Puntos seleccionados para su reconstrucción tridimensional.....	95
Tabla 5.5. Características geométricas empleadas en el reconocimiento 3D.....	96
Tabla A.1. Diferencia entre los puntos aproximados y los reales (en píxeles).....	120
Tabla A.2. Distribución de diferencias para toda la base de datos.....	120

Capítulo 1

Introducción.

1.1.- Introducción y motivación.

El reconocimiento automático de caras humanas es uno de los problemas que ha supuesto (y sigue siendo) un desafío en informática. Los seres humanos estamos muy acostumbrados a reconocernos entre nosotros usando los rasgos faciales. Además las medidas faciales han sido usadas en medicina legal y forense durante muchos años para identificar individuos, y por lo tanto han dado origen a una gran cantidad de estudios. Por lo tanto es lógico que surgiese la preocupación de diseñar un sistema automático de reconocimiento facial, fomentada además por los obvios intereses económicos subyacentes a este tipo de sistemas.

Una de las ventajas de un sistema automático de reconocimiento de caras humanas basado en visión artificial es su carácter no intrusivo: el sujeto no siente invadida su intimidad, no tiene que realizar ninguna acción para identificarse o someterse a ningún tipo de análisis. Otra ventaja de este tipo de sistemas es que permiten eliminar la memorización de códigos y los consiguientes riesgos de pérdida, olvido o suplantación. Además un sistema de este tipo evitaría la necesidad de llevar documentación (tarjeta, pasaporte, DNI, etc); bastaría con acercarse a una cámara fotográfica.

Sin embargo, los sistemas de reconocimiento automático de caras sólo funcionan de forma razonable cuando disponemos de sujetos colaborativos, es decir, que facilitan su reconocimiento, no enmascarando la cara. La práctica totalidad de los sistemas sólo funcionan bien en condiciones de laboratorio, ya que son muy sensibles a cambios de punto de vista, posición del sujeto o de las cámaras, iluminación etc. Otro inconveniente de la mayoría de sistemas automáticos de reconocimiento de caras es que han sido probados con bases de datos que contienen un número bajo de sujetos, muy pocos en comparación de los que debería ser capaz de examinar un sistema que estuviera en recintos con gran afluencia de personas, como aeropuertos, estadios o similares. Las pruebas realizadas hasta ahora en recintos de este tipo no han sido satisfactorias.

Los sistemas de reconocimiento automático de caras humanas basados en visión artificial ofrecen una serie de ventajas frente a otros sistemas de reconocimiento, como los basados en ADN o en huellas dactilares. Una de esas ventajas es que pueden llegar a ser muy rápidos: la respuesta suele obtenerse en menos de un segundo a un coste muy económico. Además nos hemos acostumbrado a tener cámaras de video-vigilancia en nuestro entorno y no las consideramos como una agresión a nuestra intimidad.

1.2.- Objetivo de la tesis.

El objetivo de este trabajo es realizar un sistema de verificación de caras humanas. Inicialmente se presentará un sistema bidimensional, basado en niveles de gris de imágenes, para a continuación presentar un sistema basado en información geométrica tridimensional.

El objetivo general de esta tesis es estudiar varios sistemas de verificación facial, analizando sus características de robustez, invarianza y tasas de reconocimiento frente a varios factores.

Como objetivos secundarios se busca comparar distintas formas de reconocimiento en dos dimensiones, partiendo de imágenes de niveles de gris. También se ha realizado el reconocimiento facial a partir de datos tridimensionales, lo cual permite corregir cambios de perspectiva. Para ello ha sido necesario obtener un conjunto de imágenes de sujetos, lo que llamaremos base de datos, en un entorno de cámaras controlado y calibrado. Las bases de datos utilizadas han sido adquiridas en condiciones de laboratorio, y no presentan gran variabilidad frente a cambios en iluminación, localización y gestos del sujeto.

Las ventajas de este estudio se basan en el análisis y comparación de varias técnicas de obtención de datos y de verificación de la identidad del individuo. Por lo tanto, dependiendo de sus necesidades, el usuario podrá decidir entre un sistema basado en dos dimensiones o uno con información tridimensional

Los aspectos más novedosos de esta tesis son el análisis en dos dimensiones de diversas técnicas de clasificación, así como su estudio sobre una base de datos adquirida en condiciones similares a una utilización real del sistema. Utilizando una base de datos para los diversos reconocedores faciales es posible tener una idea de la potencia y características del reconocedor. Además se puede observar la importancia de las diversas técnicas que se empleen en el preprocesamiento. Por último un estudio en tres dimensiones permite comparar varios tipos de datos distintos considerando el mismo conjunto de reconocedores.

1.3.- Planteamiento del problema y solución propuesta.

El problema de reconocimiento facial ha sido abordado por gran cantidad de investigadores en los últimos años, los cuales han ido desarrollando técnicas diversas. Algunas de estas técnicas necesitan un equipo cuyo alto coste impide su comercialización a gran escala (sensores láser, cámaras de gran resolución, etc.). El enfoque que se presenta en este trabajo sólo requiere cámaras de las habituales en trabajos de visión artificial, conectadas a una tarjeta digitalizadora de gama media en un PC comercial.

Con esta infraestructura se realizó un primer estudio analizando la cara del sujeto partiendo de los niveles de gris que tenemos en la imagen. En un entorno colaborativo, el sujeto se presenta delante de la cámara, ocupando la mayor parte de la imagen. El punto de partida es un gran número de píxeles ocupados por la cara (18200 para una imagen de 130x140 píxeles). En este caso, se puede reducir la dimensión de la imagen ocupada por la cara sin perder información utilizando el Análisis de Componentes Principales (PCA). Otra aproximación estudiada consiste en reducir información partiendo de una pérdida de resolución, reduciendo la imagen a un tamaño de unos pocos píxeles (por ejemplo 221 píxeles en una imagen de 13x17 píxeles). Una imagen de baja resolución permite a un ser humano verificar la identidad de un sujeto. El reto es conseguir que el sistema también sea capaz de conseguirlo.

Estos enfoques tienen el inconveniente de utilizar proyecciones en dos dimensiones (imágenes) de un objeto que es intrínsecamente tridimensional. Esto implica una inevitable pérdida de información. Por lo tanto, elementos tales como la posición de la cámara y del sujeto van a influir en el resultado.

Para paliar alguno de estos problemas, se procedió a adquirir una nueva base de datos con un par de cámaras calibradas (par estéreo). Se dispone así de datos tridimensionales de diversos puntos característicos en el rostro del sujeto. Utilizando características geométricas de la cara (distancias y ángulos) se pueden obtener los datos de partida para verificar a cada sujeto. En el presente trabajo la marcación de estos puntos se realizó de forma manual, siendo ello una fuente de error a considerar.

Una vez fijados los datos de partida, se deben considerar también los distintos clasificadores que se van a utilizar. En el presente estudio se consideraron tres tipos: funciones de base radial, k-vecinos más cercanos y máquinas de soporte vector. La comparación entre ellos nos permite definir para cada situación el tipo de clasificador que debemos utilizar.

1.3.1.- Dificultades esperadas del problema.

Uno de los aspectos que más ha intrigado a los estudiosos de la psicología es descubrir los mecanismos que tienen los seres humanos para reconocer caras de sus semejantes. Hay estudios que prueban que características locales muy destacadas facilitan el reconocimiento, pero en otros casos parece que es la configuración global de la cara la que se considera. Así pues, el primer problema a resolver es elegir de manera adecuada el conjunto de variables a medir que permiten caracterizar una cara humana.

Una de las mayores dificultades del problema proviene del hecho de que la cara es un objeto tridimensional y en el caso de identificación basada en imágenes se reduce a sólo dos dimensiones. Además los estudios basados en niveles de gris tienen el inconveniente de que los datos de partida se ven muy influidos por cambios en la luz, pose y aspecto del sujeto.

Una dificultad adicional radica en que es necesario recortar la cara del sujeto para evitar rasgos muy variables como es el pelo. Si se realiza de forma automática, este recorte aumenta la variabilidad. Aunque no es el objetivo de esta tesis el recortar la cara, las técnicas utilizadas nos indican que ésta es una etapa crucial en el procesamiento

posterior. Si inicialmente se disponen de datos de gran calidad, el tiempo de procesamiento se ve muy reducido.

Las bases de datos que se han utilizado o las que existen actualmente presentan un número reducido de imágenes de cada sujeto, complicando por ello los procedimientos de entrenamiento, ya que en muchas ocasiones tienen que afrontar un entrenamiento con pocas muestras y que no garantizan un buen aprendizaje por parte de los clasificadores.

Añadido a las dificultades anteriores, se puede manifestar que el hecho de usar un par de cámaras calibradas para la obtención de resultados sintéticos en tres dimensiones es una fuente importante de errores, ya que es difícil conseguir una alta precisión en las medidas realizadas con técnicas de visión artificial mediante pares estéreo.

La mejor solución es utilizar un modelo tridimensional de la cara y realizar medidas en tres dimensiones tan precisas como sea posible. Para ello será requisito casi indispensable utilizar un sensor láser.

1.4.- Estructura de la memoria.

La presente memoria se divide en seis capítulos. El primero es esta introducción. En el segundo capítulo se desarrolla el estado del arte de reconocimiento facial, mostrando las diversas técnicas y los sistemas comerciales que existen hoy día en el mercado. El capítulo tres describe los elementos que se han utilizado en esta tesis; es decir, el entorno de desarrollo que ha permitido desarrollar los experimentos de esta tesis. El siguiente capítulo presenta el trabajo realizado con imágenes en dos dimensiones, con PCA y baja resolución. El capítulo cinco aborda el tratamiento de datos en tres dimensiones y los resultados obtenidos. El último capítulo muestra las líneas futuras de investigación y las conclusiones globales.

Capítulo 2

Estado del arte.

En este capítulo se realiza una recopilación de los trabajos científicos realizados hasta la fecha centrados en el reconocimiento y verificación facial. Se ha dividido en dos partes: inicialmente se analizan los clasificadores utilizados en esta tesis y a continuación se presentan los trabajos descritos en la bibliografía de reconocimiento y verificación facial. El punto de interés del presente trabajo es realizar un sistema de verificación facial de forma automática: dada una foto y un nombre; el sistema debe responder si ambos corresponden al mismo sujeto.

El esquema que se sigue en la primera parte de este capítulo, al detallar los clasificadores, es una breve presentación teórica de los fundamentos de cada uno de ellos para a continuación mostrar ejemplos de su uso en reconocimiento facial. Sin embargo, al presentar los trabajos desarrollados en reconocimiento facial se ha optado por un enfoque levemente distinto, ya que estos temas constituyen el objetivo de la presente tesis. La literatura científica de reconocimiento y verificación facial es amplísima, y ha crecido de forma espectacular en los últimos años; por lo tanto, en este apartado se ofrecerá una introducción general al problema del reconocimiento facial para a continuación desglosar y detallar los distintos algoritmos que se han aplicado para resolver este problema.

2.1.- Clasificadores.

El objetivo de un clasificador es asignar un nombre a un conjunto de datos correspondientes a un objeto o entidad concreta. Se define un conjunto de entrenamiento como un conjunto de elementos, estando cada uno de ellos formado por una secuencia de datos correspondientes a un objeto concreto, uno de los cuales es la clase a la que pertenece. En nuestro caso, la clase corresponde al nombre del sujeto y los datos corresponden a los números que se emplean para representar la cara de ese sujeto. Un clasificador es un algoritmo que permite definir un modelo para cada clase, de tal forma que la clase a la que pertenece un elemento se puede calcular a partir de los valores de los datos que definen el elemento. Por lo tanto, de forma más práctica, el objetivo de un clasificador es asignar de forma lo más precisa posible una clase a nuevos elementos previamente no estudiados. Habitualmente se considera también un

conjunto de prueba, que nos permite medir la precisión del modelo. La clase de cada elemento del conjunto de prueba es conocida y se utiliza para validar el modelo. Inicialmente deberemos obtener un conjunto de muestras de ese sujeto (en nuestro caso, de imágenes suyas). El porcentaje de acierto del conjunto de prueba es una forma de medir la precisión del clasificador, aunque como se verá en los Capítulos 5 y 6, existen diferentes medidas de validación.

2.1.1.- Redes Neuronales: Funciones de base radial (RBF).

La primera ola de interés en las redes neuronales (también llamadas modelos conexionistas o de procesamiento distribuido paralelo) surgió con la introducción del modelo simplificado de neurona propuesto por McCulloch y Pitts en el año 1943. Estas neuronas fueron presentadas como modelos de las neuronas biológicas y componentes conceptuales para circuitos que podían realizar tareas computacionales. Cuando Minsky y Papert publican su libro “Perceptrones” en 1969, mostrando las deficiencias en el modelo del perceptrón, la mayor parte de los grupos de investigación dirigen entonces sus esfuerzos hacia otros modelos. Solo unos pocos continúan investigando sobre las redes neuronales; los más notables son Teuvo Kohonen [Koh01], Stephen Grossberg [Gro92], James Anderson [And95] y Kunihiko Fukushima [Fuk83]. El interés en las redes neuronales vuelve a emerger a principios de los años ochenta con nuevos avances teóricos (principalmente el algoritmo de la retro-propagación del error) y gracias a la mayor capacidad de cómputo por parte de los procesadores. Este interés se refleja en la gran cantidad de congresos, conferencias, revistas, libros, etc. centrados en las redes neuronales.

Las redes neuronales pueden ser descritas como “modelos computacionales” con una serie de propiedades tales como la habilidad de adaptarse o aprender, generalizar o agrupar u organizar datos, y cuya operación está basada en procesamiento paralelo. Algunas de estas propiedades pueden ser reclamadas por otros modelos; sin embargo, la pregunta clave es hasta dónde la aproximación basada en redes neuronales es mejor para una aplicación que otros modelos. Hasta la fecha no hay una respuesta clara a esta pregunta.

En muchos libros, al hablar de redes neuronales se describe su paralelismo con modelos biológicos. Pero existe tan poco conocimiento sobre los sistemas biológicos que los modelos de redes neuronales artificiales son una “super-simplificación” de estos sistemas. Por lo tanto, este trabajo se centrará únicamente en las redes neuronales artificiales desde un punto de vista informático. En este capítulo no se pretende describir exhaustivamente toda la teoría de las redes neuronales, sino ofrecer una idea de sus fundamentos y características, para posteriormente mostrar la forma en la que se han aplicado. Por ello se presentan sólo los modelos de redes neuronales utilizados en este trabajo, dejando fuera algunos modelos como los mapas auto-organizativos de Kohonen que tienen una estructura algo diferente.

Una red neuronal artificial está constituida por un conjunto de células de procesamiento llamadas neuronas. Una neurona es un dispositivo sencillo (véase la Figura 2.1) formado por una serie de entradas y una única salida. Cada neurona acepta como entrada las salidas procedentes de otras neuronas, siendo la entrada efectiva a la neurona la suma ponderada de las entradas reales a dicha neurona. Cada neurona se caracteriza por su estado de activación, que es un valor que oscila entre 0 y 1. Si el estado de activación de una neurona es 0, la neurona no está activada; mientras que

cualquier valor distinto de 0 corresponde a una neurona activa. La salida de la neurona es el estado de activación. Cada neurona realiza una tarea sencilla: recibe la información de entrada de los vecinos o del exterior y la usa para calcular una señal de salida que se propaga a otras unidades.

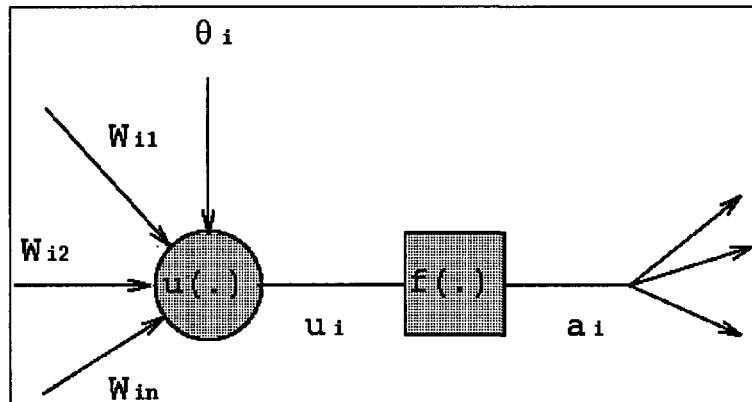


Figura 2.1. Ejemplo de neurona artificial.

En general, las neuronas se organizan en capas (véase la Figura 2.2). Dependiendo de su función en la red, se distinguen tres tipos de unidades: las unidades cuya activación son los datos de entrada del problema (unidades de entrada); las unidades cuya salida es la salida del problema (unidades de salida); y el resto de unidades, llamadas unidades ocultas (ya que no son “visibles” desde el exterior).

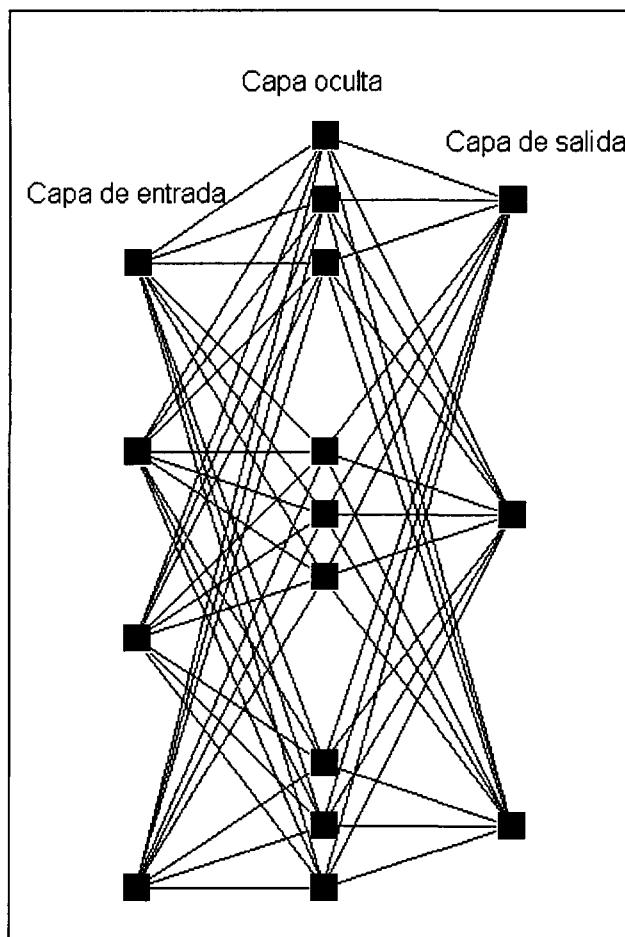


Figura 2.2. Ejemplo de red neuronal.

2.1.1.1.- Perceptrón Multicapa.

Un perceptrón multicapa está constituido por un conjunto de neuronas que se comunican enviándose señales entre sí a través de un conjunto de conexiones ponderadas. Es el caso más sencillo de red neuronal y su breve descripción se incluye para presentar los elementos de una red de este tipo (véase la Figura 2.2). Dentro de una red neuronal, podemos distinguir una serie de aspectos (Figura 2.1):

- Un conjunto de unidades de procesamiento (neuronas, células).
- Un estado de activación a_i para cada unidad, que equivale a la salida de la neurona.
- Conexiones entre las neuronas. Generalmente cada conexión se define por un peso w_{jk} , el cual determina el efecto de la señal de la neurona j en la neurona k .
- Una regla de propagación, que determina la entrada efectiva u_i de una neurona a partir de las entradas externas.
- Una entrada externa (bias, offset, momentum) θ_i para cada neurona, que va a ser uno de los componentes de la entrada efectiva $u_i(t)$ de la neurona.
- Una función de activación $f(\cdot)$ que determina el nuevo nivel de activación, a partir del valor de la entrada efectiva $u_i(t)$ y de la activación actual $a_i(t)$.
- Un método para tratar la información (regla de aprendizaje).
- Un entorno que le proporciona las señales de entrada y con el cual debe operar el sistema.

Morfológicamente, una red de neuronas está formada por múltiples capas de neuronas donde las neuronas de cada capa están conectadas con todas las neuronas de la capa siguiente. Esta conexión está ponderada mediante un peso, que es el que determina, junto con la salida, la activación o no de la neurona.

La primera tarea que debe realizarse con una red neuronal es el ajuste de los pesos entre neuronas, ya que inicialmente se les da un valor aleatorio. En nuestro trabajo se consideró un sistema de entrenamiento de tipo supervisado; durante la fase de entrenamiento se le presentan a la red tanto los patrones que queremos que diferencie como los resultados que debería proporcionar cada patrón. La red va reajustando los pesos de las conexiones en función del error cometido, hasta que para los patrones iniciales, la salida de la red neuronal es la correcta. Una vez realizada esta fase, los pesos no varían y la red está lista para clasificar correctamente los nuevos datos que se le presenten.

En una red va a ser posible distinguir una capa de entrada, a la que se le proporcionan los datos, una capa de salida, que es la que produce los resultados finales y un número indeterminado de capas intermedias. El número de neuronas en la capa de entrada depende de la cantidad de información disponible. El número de neuronas de la capa de salida depende del problema que tratemos de resolver, siendo un buen criterio el disponer al menos de tantas neuronas como elementos se quiere diferenciar. Para las capas intermedias no existe un criterio para determinar ni el número de capas ni el de neuronas por capa.

Para una discusión en detalle de los algoritmos, se recomienda la consulta de Bishop [Bis95] y Ripley [Rip96]. La implementación de los algoritmos está excelentemente descrita en Masters [Mas93].

2.1.1.2.- Funciones de base radial (RBF).

Las redes neuronales de funciones de base radial construyen aproximaciones globales a funciones usando combinaciones de funciones “base” centrados en los vectores peso, mientras que el perceptrón multicapa construye una arquitectura basada en hiperplanos. Estos hiperplanos definen las fronteras entre las clases, de tal forma que cada clase queda limitada por un conjunto de hiperplanos. Al presentar un patrón nuevo a la red neuronal, dicho patrón se encontrará en una región del espacio y el perceptrón va a activar la neurona de salida correspondiente a dicha region, aunque dicho patrón se encuentre muy alejado de los que definen la clase correspondiente a dicha neurona. En el caso de una red neuronal de funciones de base radial, si el patrón que se le presenta a la red neuronal se encuentra muy lejos de todas las clases, no se va a activar ninguna salida. Una distinción adicional es que las funciones de base radial utilizan una función distancia para convertir el vector de entrada en un escalar en la capa oculta, diferente al producto escalar de vectores que realiza el perceptrón multicapa.

Además de exhibir un conjunto excelente de propiedades teóricas, es sobre todo una construcción esencialmente práctica en problemas de discriminación y otros problemas de correspondencia [Que99] [Ren89].

Funcionalmente las RBF son redes neuronales de una capa oculta alimentados hacia delante con funciones de transferencia lineales en la salida y no lineales en la capa oculta (véase la Figura 2.3).

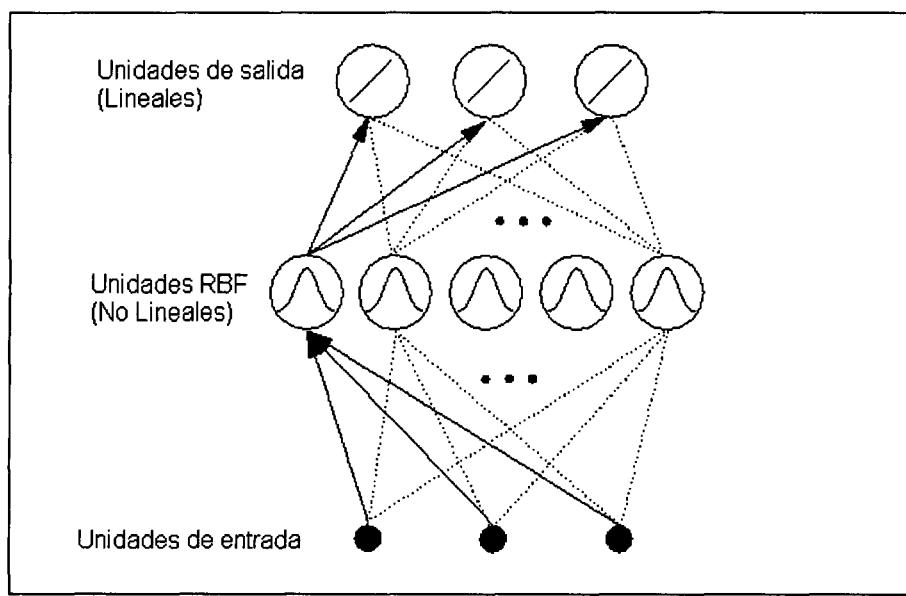


Figura 2.3. Ejemplo de red RBF.

Las redes RBF constan de dos capas de redes de aprendizaje híbrido, similares al perceptrón multicapa en términos de estructura y activación en su capa supervisada desde los nodos ocultos a los nodos de salida. Sin embargo, las capas sin supervisar, desde la capa de entrada a la oculta, difieren en que en ellas existen funciones radiales individuales del tipo campana de Gauss para cada nodo oculto simulando el efecto de los receptores humanos que están localmente solapados y cada uno de ellos está

ajustado. Se necesita un vector distancia, que puede ser la distancia euclídea, entre el vector de entrada n-dimensional y el centro de la neurona oculta. El valor de salida aumentará cuando ambos estén cerca (de forma similar a cómo varía una curva gausiana centrada en la neurona oculta).

Cada unidad oculta tiene asignado un valor σ o “anchura” (véase la Figura 2.4) que define la naturaleza y ámbito de la respuesta de la neurona, y es equivalente a la desviación estándar de la anchura de la campana de Gauss, de tal forma que valores altos indican que hay muchos puntos que van a ser considerados. Esto significa que en contra de lo que sucede en el perceptrón multicapa, las RBF poseen una función de activación que relaciona la proximidad relativa entre el conjunto de prueba y el de entrenamiento. Ello permite una medida directa de la confianza en la salida de la red para un patrón dado. Si un patrón es muy diferente de aquellos con los que la red ha sido entrenada, la respuesta será muy baja o nula.

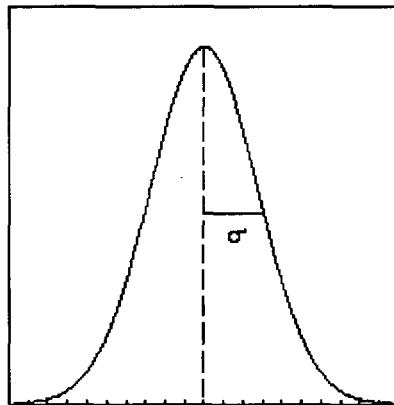


Figura 2.4. Función de activación de una neurona RBF.

Las redes neuronales del tipo RBF se han aplicado a diversos problemas. Por ejemplo, en reconocimiento facial, Howell preprocesa las imágenes faciales con un filtro de Gabor y con diferencia de Gaussianas, para a continuación aplicar una red RBF como clasificador en el reconocimiento. Los resultados que ofrecen son buenos pero muchos experimentos los realiza con una base de datos formada por 10 personas [How97].

Funciones de Base Radial – Ajuste por Descenso Dinámico (RBF-DDA).

El método seleccionado en esta tesis para implementar este tipo de redes es el RBF-DDA, que a continuación se describe. El algoritmo de ajuste por descenso dinámico (DDA: Dynamic Decay Adjustment) consiste en una extensión del algoritmo RCE [Hud92] [RCE82]. Las redes neuronales tipo RBF entrenadas con el algoritmo DDA a menudo tienen una precisión en el reconocimiento comparable a los perceptrones multicapa (MLP), pero su entrenamiento es significativamente más rápido [Ber95].

Una red RBF entrenada con el algoritmo DDA es similar en estructura al perceptrón multicapa de una capa oculta. El número de unidades en la capa de entrada representa la dimensión del conjunto de entrada. La capa oculta contiene unidades con

función de base radial (RBF). Estas unidades se van añadiendo a la capa oculta durante el entrenamiento. La capa de entrada está completamente conectada a la capa oculta. Cada unidad de la capa de salida representa una de las clases de datos existentes, codificándose en binario o en 1-de-n posibilidades. Para la clasificación, la aproximación utilizada habitualmente es del tipo “el ganador se lo lleva todo”, es decir, la salida con el valor de activación más alto determina la clase a la que se supone pertenece el patrón de entrada.

Las principales diferencias con una red tipo perceptrón multicapa son la función de activación y la regla de propagación en la capa oculta. En lugar de usar una sigmoide o cualquier otra función no lineal infinita, RBF usa funciones localizadas, Gaussianas radiales, como funciones de activación. Además, el cálculo de la distancia euclídea a un vector de referencia sustituye al producto escalar utilizado en perceptrones multicapa, según la fórmula siguiente:

$$R_i(\vec{x}) = \exp\left(-\frac{\|\vec{x} - \vec{r}_i\|^2}{\sigma_i^2}\right)$$

Esta fórmula indica que, si la red recibe como entrada un vector \vec{x} , entonces $R_i(\vec{x})$ es el valor de activación de una unidad RBF con vector de referencia \vec{r}_i y desviación típica es σ_i^2 . La capa de salida calcula la salida de clase como sigue:

$$f(\vec{x}) = \sum_{i=1}^m A_i * R_i(\vec{x})$$

donde m representa el número de funciones RBF pertenecientes a la clase correspondiente y A_i es el peso de cada RBF. Un ejemplo de red RBF de tipo DDA se muestra en la Figura 2.5. Puede observarse que no hay conexiones entre la capa de entrada y la de salida que cortocircuiten la capa oculta.

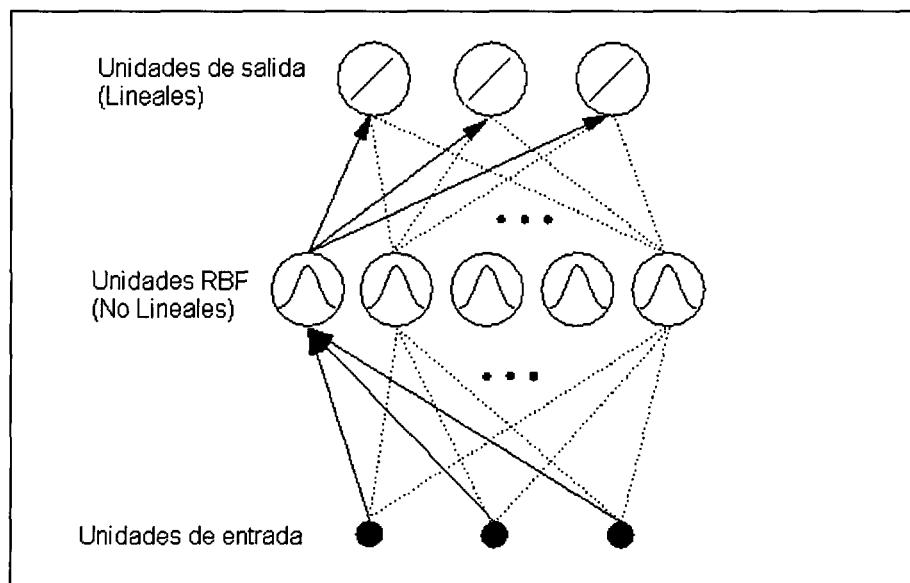


Figura 2.5. Ejemplo de red neuronal del tipo RBF.

En esta figura, el vector peso que conecta todas las unidades de entrada con una unidad oculta representa el centro de la Gaussiana. La distancia euclídea del vector de entrada con este vector referencia (o prototipo) se considera como la entrada a la función Gaussiana de la unidad correspondiente, con la cual se obtiene una respuesta local. Si el vector de entrada está cerca del prototipo, la unidad tendrá un nivel de activación alto. Por otra parte, la activación será casi nula si la distancia entre el prototipo y el vector de entrada es grande. Cada unidad de salida calcula una suma ponderada de todas las activaciones de las unidades RBF que pertenecen a la clase correspondiente.

El algoritmo DDA aporta la noción de distinguir entre emparejamiento y conflicto entre vecinos en una zona de conflicto. De acuerdo con la Figura 2.6, se consideran dos umbrales θ^+ y θ^- .

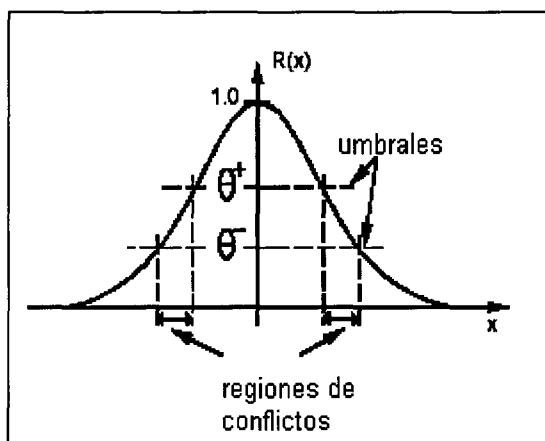


Figura 2.6. Unidad RBF utilizada por el algoritmo DDA. Los umbrales se utilizan para definir una “región de conflictos” a la cual ningún otro prototipo puede pertenecer. Además, cada patrón de entrenamiento tiene que estar en el círculo interior de al menos un prototipo de la clase adecuada.

Normalmente, θ^+ es mayor que θ^- lo cual conduce a un área de conflicto donde no está permitido que haya ningún patrón de entrenamiento. Utilizando estos umbrales, el algoritmo construye la red dinámicamente y ajusta los radios individuales. Por lo tanto, las principales propiedades del algoritmo DDA son:

- Entrenamiento constructivo: Nuevos nodos RBF son añadidos cuando se necesitan. La red se construye sólo desde un esqueleto; el número de unidades ocultas necesarias se determina durante el entrenamiento. Además, los radios individuales se ajustan durante el entrenamiento.
- Entrenamiento rápido: Normalmente cinco iteraciones bastan para completar el entrenamiento.
- Convergencia garantizada: Puede probarse que el algoritmo termina.
- Dos parámetros no críticos: Sólo los parámetros θ^+ y θ^- deben ser ajustados manualmente. Los valores de estos parámetros no son críticos.
- Propiedades garantizadas de la red: Puede demostrarse que, cuando el entrenamiento ha terminado, la red neuronal posee varias condiciones para todos los patrones de entrenamiento: los patrones incorrectamente clasificados lo están por debajo de un umbral (θ^-) y los correctamente clasificados lo están por encima de otro umbral (θ^+).

El algoritmo DDA se basa en dos pasos. Durante el entrenamiento, cuando un patrón está incorrectamente clasificado, sucede que o bien se añade una nueva unidad RBF con un peso inicial de 1 ó bien el peso de una RBF existente se incrementa (con lo cual cubre el nuevo patrón). En ambos casos, los radios de las unidades RBF conflictivas (unidades RBF de clases incorrectas) se reducen. De esta manera se garantiza que cada uno de los patrones en el conjunto de entrenamiento se cubre por una RBF de la clase correcta y ninguna de las RBF de una zona conflictiva tiene una respuesta incorrecta.

Los parámetros θ^+ y θ^- son introducidos en esta etapa. Cuando se crea una unidad nueva, ninguna de las unidades RBF existentes de la clase correcta tienen una activación por encima de θ^+ y durante la reducción del valor de los pesos de las clases conflictivas, ninguna de ellas tiene un valor de activación superior a θ^- . La Figura 2.7 muestra un ejemplo de los primeros pasos de entrenamiento del algoritmo DDA-RBF. Una vez finalizado el entrenamiento, se verifican dos condiciones para todos los pares entrada-salida $(\vec{x}; c)$ del conjunto de entrenamiento:

- Por lo menos un prototipo de la clase correcta c tiene un valor de activación mayor o igual que θ^+ .

$$\exists i : R_i^c(\vec{x}) \geq \theta^+$$

- Todos los prototipos de clases conflictivas tienen valores de activación menores o iguales que θ^- (m_k representa el número de prototipos de la clase k).

$$\forall k \neq c, 1 \leq j \leq m_k : R_j^k(\vec{x}) \leq \theta^-$$

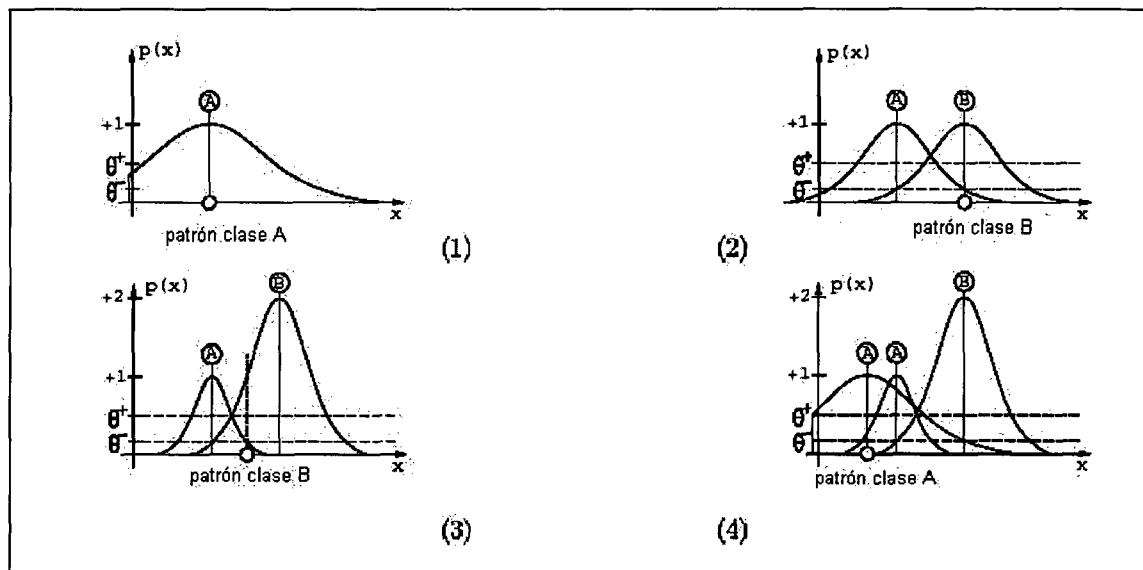


Figura 2.7. Ejemplo de entrenamiento DDA-RBF. (1) se encuentra un patrón de la clase A y se crea una nueva RBF. (2) Un patrón de entrenamiento de la clase B conduce a un nuevo prototipo de la clase B y se reduce el radio de las RBF existentes de la clase A. (3) Otro patrón de la clase B se clasifica correctamente y reduce el radio del prototipo de la clase A. (4) Un nuevo patrón de la clase A crea otro prototipo de esta clase.

En teoría, los parámetros θ^+ y θ^- son dependientes de la dimensión del espacio de características pero en la práctica los valores de los dos umbrales no son críticos.

Más importante es que los datos de entrada estén normalizados. Debido a la naturaleza radial de las RBF, cada atributo debe estar distribuido sobre un rango equivalente. Habitualmente una normalización en $[0,1]$ es suficiente.

2.1.2.- Máquinas de vector soporte (Support Vector Machines: SVM).

Las máquinas de vector soporte se han convertido rápidamente en una herramienta de uso generalizado en el campo de reconocimiento de patrones. La aplicación más sencilla de esta técnica es el problema de clasificación binaria (sólo hay definidas dos clases). La idea subyacente consiste en encontrar una hipótesis H que minimice la probabilidad de error empírico (la probabilidad de que H tenga un error en un conjunto de prueba seleccionado aleatoriamente).

En Vapnik [Vap98] se prueba que minimizar el error empírico es equivalente a encontrar el hiperplano (Figura 2.8) que se encuentra a la distancia máxima de las muestras de entrenamiento más cercanas para las dos clases. La clave del procedimiento consiste en establecer una correspondencia entre las muestras en el espacio de entrada y otro conjunto de vectores transformados en un espacio de dimensión mayor o igual, el llamado espacio de características. Para realizar esta transformación se utiliza una correspondencia previamente definida, llamada kernel. Es en este espacio de características en el que se construye el hiperplano óptimo que separa las dos clases. Por lo tanto, el procedimiento termina con una función de decisión lineal en el espacio de características, donde los pesos se calculan como la solución de un problema cuadrático con restricciones (ver por ejemplo [Gil81]).

En la figura 2.8 se muestra un ejemplo gráfico de aplicación de SVM. Se han considerado dos clases, etiquetadas como A+ y A-. Las clases definen dos grupos que por simplicidad se han considerado separables. Los Vectores Soporte (marcados como a, b y c) son los que se utilizan para calcular el hiperplano que separa las dos clases. En el dibujo se indica también el margen, distancia del hiperplano a los Vectores Soporte.

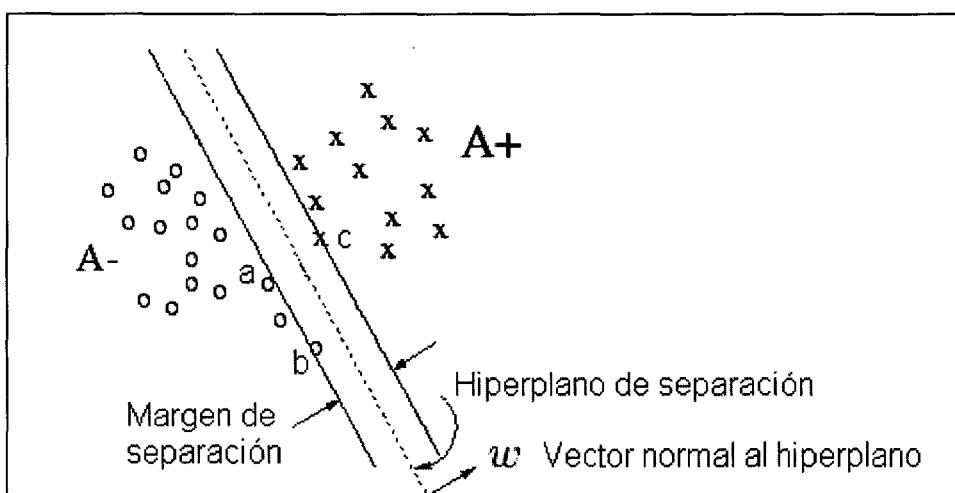


Figura 2.8. Ejemplo de SVM.

El propósito de esta sección consiste en introducir las ideas básicas que fundamentan las máquinas de vector soporte utilizadas en la presente tesis (Support Vector Machines – SVM). Estos algoritmos son en la actualidad el centro de atención dentro de este campo de investigación, y se han aplicado en multitud de casos: reconocimiento de escritura a mano [Cor95]; reconocimiento de objetos [Bla96] o categorización de texto [Joa97]. Se pueden encontrar descripciones más detalladas en [Vap95] [Vap98].

Clasificación mediante SVM

Supóngase el caso más sencillo de reconocimiento de patrones: dos clases linealmente separables. Sea un conjunto de un número m de ejemplos etiquetados $D = \{(x_i, y_i)\}_{i=1}^m$ donde se puede suponer que la etiqueta va a tomar uno de dos valores posibles (sólo disponemos de dos clases) que, por simplificar, designaremos -1 y 1 , es decir, $y_i \in \{-1, 1\}$. Para este conjunto se quiere determinar, entre todos los clasificadores lineales que separan los datos, aquel que tiene el menor error de generalización. El error de generalización mide la capacidad de la red de dar una respuesta correcta ante patrones que no han sido empleados en su entrenamiento. Intuitivamente, una buena elección consiste en el hiperplano que obtiene el mayor margen entre las dos clases, donde el margen se define como la suma de los cuadrados de las distancias del hiperplano a los puntos más cercanos de las dos clases.

Si las dos clases no son separables linealmente todavía se puede encontrar un hiperplano que maximice el margen y que minimice una cantidad proporcional al número de errores de clasificación. La compensación entre el margen y el error del clasificación se controla por una constante C positiva, seleccionada previamente. En este caso se puede mostrar que la solución a este problema es un clasificador lineal

$f(x) = signo(\sum_{i=1}^m \lambda_i y_i x^T x_i + b)$ cuyos coeficientes λ son la solución del siguiente problema, formulado en notación matricial:

$$\underset{\Lambda}{\text{Minimizar}} \quad W(\Lambda) = -\Lambda^T 1 + \frac{1}{2} \Lambda^T D \Lambda$$

$$\begin{aligned} \text{Sujeto a} \quad & \Lambda^T y = 0 \\ & \Lambda - C 1 \leq 0 \\ & -\Lambda \leq 0 \end{aligned}$$

Donde $(\Lambda)_i = \lambda_i$, $(1)_i = 1$ y $D_{ij} = y_i y_j x_i^T x_j$. Se observa que un número pequeño de coeficientes λ_i son no nulos y como cada coeficiente se corresponde con uno de los ejemplos, significa que la solución está determinada por los ejemplos asociados con los coeficientes no nulos. Estos ejemplos se llaman Vectores Soporte, y son los únicos relevantes para la solución del problema: el resto de datos ejemplo pueden ser eliminados del conjunto de entrenamiento, no alterándose por ello la solución. Intuitivamente, los Vectores Soporte corresponden a los datos del conjunto de ejemplos que están en la frontera entre las dos clases. Su número es habitualmente bajo y Vapnik [Vap98] demostró que es proporcional al error de generalización del clasificador.

Como no es muy frecuente que los problemas reales sean resolubles mediante clasificadores lineales, la técnica debe extenderse para permitir superficies no lineales. Esto se realiza fácilmente al proyectar el conjunto inicial de variables x en un espacio de dimensión mayor (llamado espacio de características) $x \in \mathbb{R}^n \Rightarrow z(x) \equiv (\phi_1(x), \dots, \phi_n(x)) \in \mathbb{R}^m$ y reformular el problema del clasificador lineal en el espacio de características. La solución tendrá la forma

$$f(x) = \operatorname{signo}\left(\sum_{i=1}^m \lambda_i y_i z^T(x) z(x_i) + b\right)$$

y por lo tanto será no lineal en el espacio de las variables iniciales. Llegados a este punto surgen dos problemas: 1) la elección de las características $\phi_i(x)$, que debe realizarse de forma tal que lleve a una clase de superficies de decisión "rica" y 2) el cálculo del producto escalar $z^T(x)z(x_i)$ lo cual puede ser computacionalmente prohibitivo si el número n de características es muy grande. Una posible solución a este problema consiste en permitir que n crezca hasta el infinito y realizar la siguiente selección de las funciones $\phi_i(x)$:

$z(x) \equiv (\sqrt{\alpha_1} \psi_1(x), \dots, \sqrt{\alpha_m} \psi_m(x), \dots)$. Donde α_i y ψ_i son los autovalores y las autofunciones de un operador cuyo núcleo o kernel $K(x,y)$ es una función simétrica y definida positiva. Con esta elección el producto escalar en el espacio de características se presenta de forma sencilla porque se obtiene:

$$z^T(x)z(y) = \sum_{i=1}^{\infty} \alpha_i \psi_i(x) \psi_i(y) = K(x, y)$$

donde la última igualdad se obtiene del teorema Mercer-Hilbert-Schmidt [Rie55] para funciones definidas positivas. El problema que a continuación se debe resolver es el mismo que el planteado en el caso anterior, pero ahora la matriz D tiene por elementos $D_{ij} = y_i y_j K(x_i, x_j)$. Como consecuencia de esta elección, el clasificador SVM tendrá la forma $f(x) = \operatorname{signo}\left(\sum_{i=1}^m \lambda_i y_i K(x, x_i) + b\right)$. Las elecciones más habituales son:

- Kernel lineal:

$$k(x, y) = x^T y$$

- Kernel polinomial:

$$k(x, y) = (x^T y + c)^q,$$

donde c es una constante y q es un entero positivo.

- Kernel Gaussiano:

$$k(x, y) = \exp\left(-\|x - y\|^2 / 2\sigma^2\right)$$

Los resultados teóricos sugieren que la eficacia de las máquinas de vector soporte se debe a su capacidad de encontrar reglas que clasifiquen objetos con un grado de confianza alto y a su capacidad para evitar el sobre-entrenamiento. En efecto, resulta frecuente que algunos algoritmos encuentren reglas complejas que clasifican perfectamente los datos del conjunto de entrenamiento; pero estas reglas pueden ser completamente inútiles para clasificar correctamente nuevas muestras, ya que están tan relacionadas con el conjunto de entrenamiento que no tienen capacidad de

generalización (este efecto se conoce como “sobre-entrenamiento”). Las redes neuronales son capaces de clasificar correctamente casi cualquier conjunto de entrenamiento, pero pueden caer en un sobre-entrenamiento en el que no puedan generalizar correctamente la clasificación.

Las máquinas de vector soporte han sido utilizadas para la clasificación de varios tipos de patrones; también en verificación facial. Por ejemplo, G. Guo et al. [Guo00] utilizan los SVM para clasificar los sujetos de la base de datos de Olivetti-Cambridge [Oli]. Procesan las imágenes con un Análisis de Componentes Principales y se obtiene un porcentaje de error del 3%, considerando una estructura en forma de búsqueda binaria para clasificar los 40 individuos de la base de datos. En [Jon00] se utiliza una base de datos de 295 sujetos, con un resultado de reconocimiento correcto del 98% (en este caso, los cambios de iluminación del conjunto de prueba también están presentes en el conjunto de entrenamiento).

2.1.3.- K vecinos más cercanos (K nearest neighbours, KNN).

Las técnicas de clasificación basadas en los vecinos más cercanos (KNN) se encuentran entre los planteamientos más clásicos de reconocimiento de patrones. Son simples y rápidos, ofreciendo valiosa información sobre el problema de clasificación que se quiere abordar.

Su principio de operación es muy sencillo: Dado un conjunto de muestras de entrenamiento t_{ij} pertenecientes a una de las J clases diferentes C_1, \dots, C_J , un nuevo patrón desconocido x se asignará a la clase C_m que tenga un $t_{im} \in C_m$ tal que sea la muestra del conjunto de entrenamiento más cercana a x de acuerdo con alguna métrica. Esta técnica se conoce como 1-NN. En general en los métodos K-NN se consideran los k vecinos más cercanos de x dentro del conjunto de entrenamiento y se considera C_m como la clase que tiene más muestras entre los k vecinos considerados.

Se han propuesto algunas modificaciones para mejorar la robustez de estas técnicas frente a elementos aislados que están fuera del rango de la clase. Estas modificaciones se basan en la edición del conjunto de entrenamiento y la eliminación de estos elementos con una desviación muy grande. También se han propuesto técnicas de condensación para acelerar el proceso de clasificación, en la cual las muestras de una clase que se encuentran completamente rodeadas de muestras de la misma clase son eliminadas, ya que no influyen en la clasificación [Web02].

Los clasificadores basados en los vecinos más cercanos han sido aplicados al reconocimiento facial, por ejemplo en Lucas [Luc97], que obtuvo un rendimiento muy alto usando $k=1$ y distancia no euclídea, tipo ciudad (la distancia entre dos puntos es la suma de la diferencia de coordenadas en cada uno de los ejes, en valor absoluto). El éxito de este clasificador tan sencillo se basa en que la base de datos considerada está muy controlada en luz y pose, ya que esta técnica no es invariante frente a estos factores. [Bel97] utiliza el clasificador del vecino más cercano comparando los resultados cuando las imágenes son procesadas con PCA y Fisherfaces. Ambos métodos obtienen buenos resultados cuando las imágenes de prueba son muy parecidas a las imágenes de entrenamiento; sin embargo, las variaciones de iluminación o de gesto en el conjunto de prueba hace que el rendimiento sea menor.

2.2.- Reconocimiento de caras humanas.

2.2.1.- Introducción.

El problema del reconocimiento de caras humanas se puede sintetizar como la identificación de un rostro utilizando únicamente los rasgos faciales, sin utilizar otros elementos (voz, huellas dactilares, firma, ADN, etc.). Además se han definido otros temas relacionados, como son la identificación del sexo y la edad, la extracción de información sobre la expresión facial, y la generación de modelos de texturas para su uso en gráficos [Che95] [FG00] [Oto91].

Es interesante notar que los trabajos pioneros en el reconocimiento automático de caras se pueden remontar al s. XIX y concretamente, a los trabajos precursores de Francis Galton. En la Bibliografía se mencionan diferentes planteamientos para el reconocimiento facial, especialmente los surgidos en los últimos años. Sin embargo la mayoría utiliza una serie de supuestos, más o menos explícitos, como son:

- a.- La imagen suele ser frontal o perfil, por lo tanto va a ser posible encontrar todos los rasgos en la imagen.
- b.- La cara está derecha, con poca o ninguna inclinación o giro.
- c.- No hay ocultamiento de trozos de imagen.
- d.- El número de casos de prueba es relativamente pequeño.
- e.- No aparece vello facial, gafas, sombreros, etc.
- f.- La mayor parte de los casos son hombres de raza blanca.

En general, todos los sistemas utilizan la misma secuencia de etapas para la identificación o verificación:

- a.- Determinar un conjunto de características independientes para representar una cara.
- b.- Representar las caras de entrenamiento en función de los valores que toman en ellas el conjunto de características seleccionado.
- c.- Determinar los valores de una cara nueva (desconocida).
- d.- Utilizar un criterio y un esquema de correspondencia para encontrar el mejor emparejamiento con las caras conocidas.

Además sólo se aborda un problema, con poca o nula incidencia de los otros: dado que se busca identificar, se supone que la localización ha sido ya hecha y las características ya extraídas. Los distintos trabajos difieren tanto en la forma de representar las caras como en el esquema de correspondencia aplicado. Podemos agrupar los cuatro puntos anteriores en tres problemas que deben ser resueltos y que son detallados a continuación.

- a) Detección de la cara en escena.
- b) Representación de la cara.
- c) Clasificación de la cara, indicando el nombre del sujeto al que pertenece. Si el sistema debe averiguar el nombre del sujeto partiendo únicamente de una imagen de la cara, se dice que el sistema identifica; si se le proporciona una

imagen de la cara junto con un nombre y el sistema determina si el nombre corresponde a esa cara, se dice que es un sistema verificador.

2.2.1.1.- Detección de la cara.

En algunos casos las condiciones bajo las que se obtiene la imagen son controladas, por ejemplo las fotografías obtenidas por la policía o las obtenidas en el laboratorio. Por lo tanto la localización de la cara en la escena puede ser fácilmente determinada. En otros casos la localización de la cara en la imagen no es conocida a priori. El primer paso, por lo tanto es determinar si en la escena hay caras y si una cara está presente, localizarla en la imagen.

Varios factores tornan este problema complejo. Uno de ellos es el problema del vello facial (bigote, barba, etc.), maquillaje, etc. que enmascaran las características faciales. Otro es la variación en la escala y orientación de la cara en la imagen. Por último, existen otros dos factores que van a dificultar la detección: la iluminación que tenga la escena y la calidad de las imágenes.

La mayoría de los métodos de detección de caras se basan en plantillas flexibles, autocaras, redes neuronales o en el color de la cara.

Nuestra investigación no aborda este problema. Las bases de datos con las que se desarrolla nuestro trabajo se han adquirido en condiciones controladas de laboratorio. En cada imagen aparece una cara (y solo una) que ocupa la mayor parte de la imagen y, en general, está situada en el centro. El fondo de la imagen es el mismo para cada base de datos.

A continuación se mencionan algunos de los trabajos más recientes que tratan el tema. Se remite al lector a la bibliografía, en particular a las revisiones [Hje01] y [Yan01], para una discusión en detalle de los diversos algoritmos propuestos.

Partiendo de imágenes foveadas (imágenes a las que se les ha aplicado la transformada log-polar), F. Jurie [Jur99] utiliza autocaras y un análisis multidimensional del histograma de color para detectar la cara y hacer su seguimiento. Afirma que el sistema es capaz de funcionar durante horas sin ningún error, pero sólo muestra resultados gráficos en un caso y no describe la base de datos ni cómo se comporta el sistema frente a cambios de luz, fondo, aparición de varias caras, etc. Las mismas técnicas las usa Koh et al. [Koh99] para detectar la cara y para el posterior reconocimiento utilizando una red neuronal.

Empleando el mismo tipo de imágenes, Smeraldi et al. [Sme00] utilizan la transformada de Gabor para seguir el movimiento de la cabeza y detectar los ojos de una cara. Otros estudios [Zho99] buscan la detección de rasgos en la cara, por ejemplo ojos para, a partir de esos rasgos, localizar la cara completamente. Otras veces será el contorno de la cara [Sam95].

En la Bibliografía se ofrecen soluciones creadas para resolver alguno de los problemas que presenta la detección de caras: por ejemplo rotación en la cara [Mia99] o iluminación no uniforme [Kon99]. En general los resultados son muy buenos en el problema que se pretende resolver, pero apenas se menciona cómo influyen los otros fenómenos a tener en cuenta.

En [Vie98] aparece la aplicación de una mezcla de redes neuronales (y su mejora con SVM) para detectar caras en entornos cerrados. El proceso se realiza variando la resolución de la imagen con el fin de compensar los cambios en la escala de las caras. Recientemente celebrado, en [FG00] aparecen más de una veintena de artículos relativos a la detección de la cara. Especial interés merecen los dedicados a la detección de una cara partiendo del color de la piel (el cual es invariante en condiciones de iluminación controladas, sólo dependiendo de la raza del sujeto).

2.2.1.2.- Representación.

El problema de la representación y el de la identificación están muy unidos. Según las características que se quieran extraer o identificar, así será la representación.

Las representaciones más utilizadas son:

- Imágenes como matrices bidimensionales de niveles de gris.
- Vectores de características.

El primer caso necesita una compactación de la información cuando el número de imágenes es elevado. Se ha puesto en evidencia que imágenes de 32 x 32 píxeles con 4 bits por píxel son suficientes para identificar a un sujeto. Ejemplos de esta representación son los métodos basados en compactación usando PCA, Fisherfaces y redes neuronales.

En el segundo caso se han usado características bidimensionales o tridimensionales, en este caso se ha basado en modelos de la cara. Se engloban en esta categoría los trabajos centrados en emparejamiento de plantillas flexibles, jets de Gabor, y demás técnicas que van a ser descritas en los apartados siguientes.

2.2.1.3.- Clasificación: Identificación o Verificación.

Identificar consiste en el proceso de asociar un nombre a un rostro. Verificar se define como la tarea de comprobar si un nombre coincide con un rostro. El primer paso es seleccionar cómo se va a realizar la identificación, para ello en la bibliografía se han empleado distintas técnicas; las más usadas son: redes neuronales, correspondencia de plantillas y autocaras, que junto con otras técnicas son descritas a partir del capítulo siguiente.

Todos estos métodos toman como hipótesis de partida el hecho de que para un rostro, los valores de las características que lo definen no varían mucho en diferentes imágenes. Es más, si un conjunto de características es muy diferente en dos imágenes los rostros correspondientes serán también diferentes.

Otros hechos de la identificación que realizan los seres humanos que se intentan simular son:

- La importancia de los rasgos faciales decrece de arriba hacia abajo (pelo > ojo > nariz > boca > barbilla). Pero caras con rasgos no habituales son identificadas rápidamente.
- Caras familiares son identificadas más rápidamente que otras menos familiares.

- Habitualmente, una persona reconoce a unos 700 casos, pero puede llegarse a algunos miles.
- Es difícil describir un rostro humano, ya que está lleno de numerosos estímulos visuales.
- El reconocimiento de una cara es el resultado de un análisis de rasgos globales junto con otros locales.

Hay que tener en cuenta que no existe una forma de calcular el número de características necesarias. Se ha estimado que depende logarítmicamente del número de caras que hace falta reconocer, de tal forma que si hubiera unas 1000 caras se necesitarían sólo 10 (lo cual es un número muy bajo).

Las características seleccionadas se busca que reúnan una serie de condiciones:

- Fáciles de estimar.
- Independientes de la iluminación.
- Independientes de cambios en la expresión facial.
- Altamente discriminantes.

Habitualmente es necesario incluir una etapa de normalización porque las medidas han de ser independientes de la posición, escala y orientación de la cara en escena.

2.2.2.- Estado del arte en el reconocimiento y verificación facial automática.

El problema del reconocimiento automático de caras ha despertado un gran interés especialmente en los últimos años debido a su importancia en temas como seguridad, comunicación hombre-máquina, etc.

La representación de una cara humana puede realizarse mediante una imagen en 2D o mediante una superficie en 3D. En función del tipo de representación utilizada, el planteamiento seguido para enfrentarse al problema del reconocimiento automático de caras es diferente. En los últimos años se han realizado amplias revisiones bibliográficas del problema del reconocimiento facial, publicados en 1995 [Che95], 1999 [Wen99] y 2000 [Zha00].

En 1997 Uwechue et al. [Uwe97] realizó un estudio de reconocimiento facial mediante redes neuronales de alto orden usando patrones de caras 2D. En 1998, [Wec98] se desarrolló una conferencia sobre reconocimiento facial usando sobre todo patrones 2D y presentando desarrollos teóricos y aplicaciones prácticas. En 1999 Hallinan et al. [Hal99] usaron tanto patrones en 2D como superficies de caras en 3D. En 1999 Gong et al. [Gon99] indicaron los excelentes resultados obtenidos usando técnicas estadísticas de reconocimiento de patrones utilizando imágenes globales. Estos estudios muestran que las técnicas de reconocimiento facial, especialmente los métodos globales basados en la teoría estadística de reconocimiento de patrones, han experimentado un gran avance.

En conclusión, los estudios anteriores muestran que las técnicas de reconocimiento facial, especialmente los métodos globales basados en teorías

estadísticas, se han desarrollado ampliamente en los últimos diez años. Excepto para las aplicaciones de reconocimiento basadas en imágenes frontales adquiridas en entornos controlados, el principal reto consiste en hacer frente al alto grado de variabilidad en las imágenes de caras. Las fuentes de variabilidad incluyen variaciones inter-sujetos (diferentes apariencias de una misma persona) y variaciones intra-sujetos (pose, expresión facial, iluminación, edad, etc.). Algunas de estas variaciones pueden ser compensadas antes del reconocimiento, pero otras no pueden ser eliminadas. Personas que tienen una semejanza facial apreciable son difíciles de distinguir unas de otras. Las variaciones debidas a diferentes poses, iluminación y expresión facial son “relativamente sencillas” de compensar. El cambio de posición es una de las principales causas de variabilidad, por lo que cada vez están tomando más auge sistemas basados en geometría 3D, pues son invariantes frente a cambios de posición.

A partir de los resultados expuestos en la bibliografía puede comprobarse que el rendimiento de un sistema varía dependiendo de la base de datos en que se ha probado. Algunas bases de datos públicas muy difundidas son: MIT [Mit], Yale [Yal], Purdue [Pur] y Olivetti [Oli]. La base de datos FERET [Phi00] no es completamente pública y la XM2VTS [Xm2] es multimodal (imágenes, vídeo y audio). Pese a ello, existen muchos artículos en que se emplean bases de datos creadas “ad hoc” localmente.

De acuerdo con Phillips [Phi98][Phi00], el sistema de evaluación de algoritmos de reconocimiento facial FERET identifica en el estado del arte actual tres técnicas:

- Métodos basados en PCA (análisis de componentes principales: Principal Components Analysis) [Tur91][Mog94][Mog97].
- Métodos basados en EBGM (Correspondencia entre agrupaciones de grafos elásticos: Elastic bunch graph matching) [Wis97].
- Métodos basados en LFA (análisis de componentes locales: Local Feature Analysis) [Pen96][Pen00].

El análisis basado en PCA hace uso de un conjunto de bases ortonormales respecto al cual se representan las características globales de las imágenes, de tal manera que se extraen los rasgos característicos de la cara en cuestión, reduciendo enormemente la cantidad de información necesaria a manipular en la fase de reconocimiento. La Figura 2.9 muestra la representación interna de la aproximación basada en PCA, las conocidas como “autocaras” así como las “Fisherfaces”, que se basan en una modificación de la PCA conocida como análisis discriminante lineal de Fisher (LDA: linear discriminant analysis).

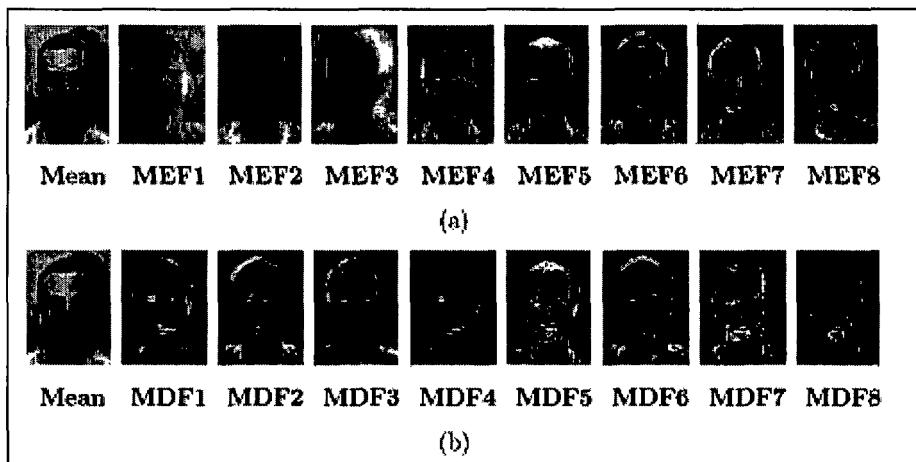


Figura 2.9. Representaciones internas de la aproximación PCA y de la LDA. Obtenidas de [Wen99]. La imagen media se muestra en la primera columna. La secuencia (a) corresponde a las características más expresivas. La secuencia (b) corresponde a las más discriminantes.

El algoritmo EBGM (véase la Figura 2.10) construye rasgos locales (considerando sólo la región próxima a un punto dado), extraídos a partir del uso de transformada wavelet de Gabor y una figura global de la cara representada como un grafo.

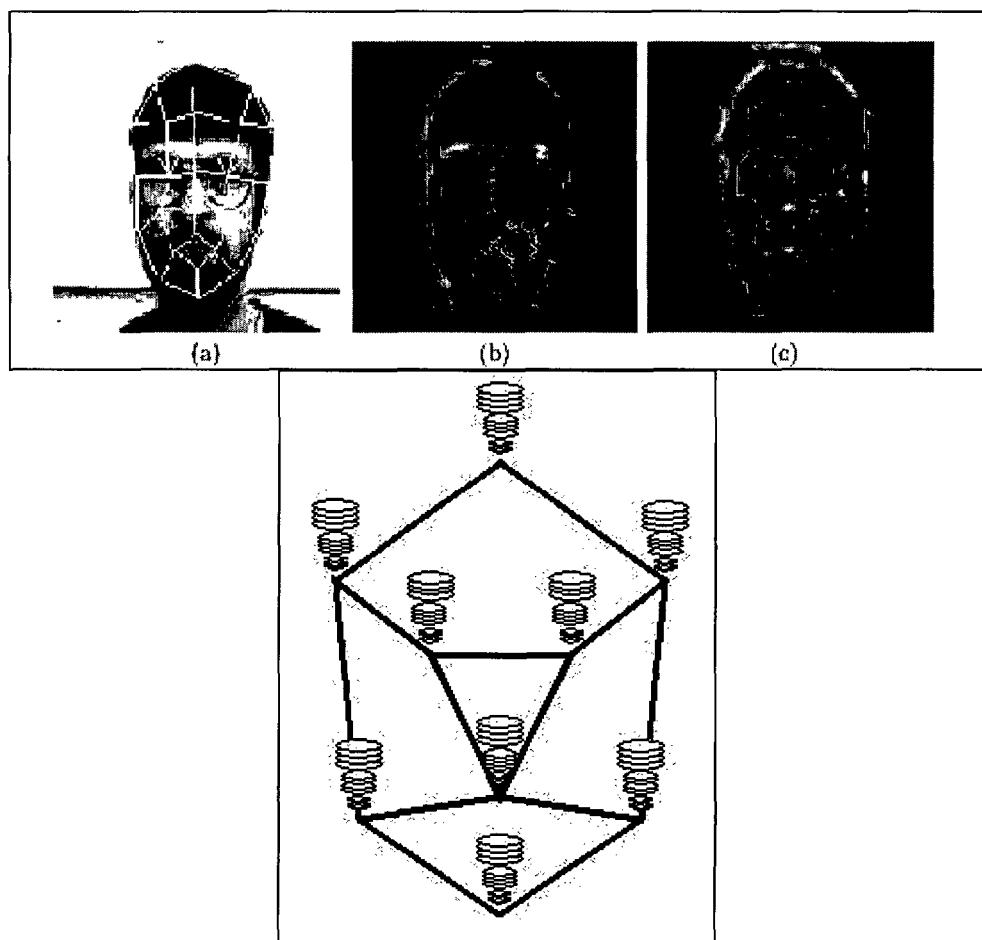


Figura 2.10. Representación interna de la aproximación EBGM (obtenidos de Wis97). (a) representa un grafo sobre una cara. (b) es la reconstrucción de la cara a partir del grafo. (c) representa la reconstrucción de la cara a partir de un grafo conjunto en el que se ha considerado en cada nodo el jet que mejor se adapta. (d) es una interpretación de la representación de grafo grupo (véase el apartado 2.2.8).

El LFA es derivado de PCA, véase la Figura 2.11. En este caso la elección de funciones kernel o base para rasgos faciales locales (por ejemplo ojos, boca, nariz, ...) y la selección de la localización de estos rasgos es todavía una cuestión abierta.

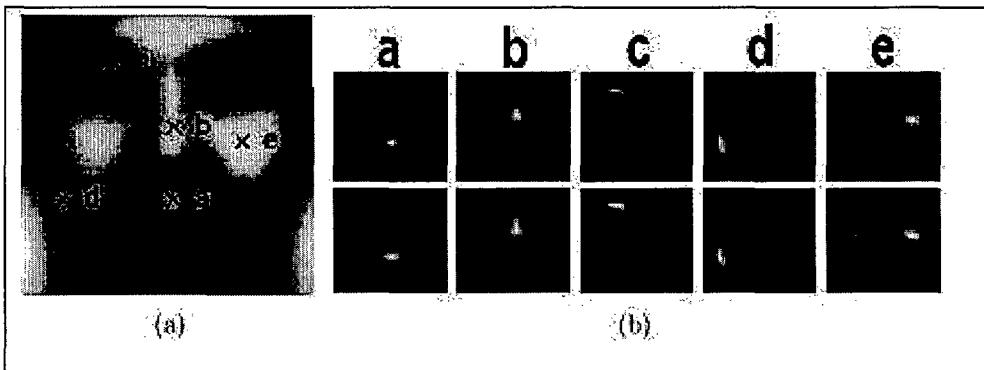


Figura 2.11. Representación interna del método LFA (obtenida de [Pen96]). (a) representa la cara media con la localización de cinco características. (b) muestra los kernels asociados a cada característica (fila superior) y la correspondiente correlación (fila inferior).

Existe aún otra clasificación de los algoritmos en función del tipo de representación facial que utilicen, es decir, si son dependientes o independientes de la posición.

En el caso de algoritmos dependientes de la posición, se almacenan un número de imágenes en 2D (apariencias) como un conjunto representativo de la cara. A su vez, pueden dividirse en tres categorías:

- Basados en características globales [Tur91][Hwa00].
- Basados en características locales distintivas [Cox96][Kam93].
- Híbridos (combinan ambos tipos de características) [Lam98] [Mog94] [Wis97] [Pen96].

Los métodos globales son muy dependientes de las variaciones intra-sujeto, ya que están basados en la información global de una imagen. Los métodos locales tienen la dificultad de detectar los puntos que guarden la información local deseada. Los métodos híbridos fueron propuestos para unir los dos tipos de información. Por ejemplo, respecto a los métodos nombrados anteriormente, PCA es un método global, y LFA y EBGM son ambos métodos híbridos, pero más cercanos a los locales.

En el caso de algoritmos invariantes con la posición, la cara es representada mediante un modelo 3D. De esta manera la variación debida a cambios de posición e iluminación afecta en mucha menor medida.

Gordon et al. [Gor95] propusieron un modelo facial representado por una serie de puntos 3D asociados con un correspondiente grado de textura. Este método requiere una estimación fiable de la posición de la cara. Lengagne et al. [Len96] propusieron un esquema de reconstrucción de una cara 3D a partir de un par estéreo de imágenes, aunque no llegaron a implementar el módulo de reconocimiento. Atick et al. [Ati96] proponen un método de reconstrucción de una cara 3D basado en auto-superficies (de manera equivalente a autovectores en PCA) y plantean su posible uso para el reconocimiento. Yang et al. [Yan98] proponen un método de reconstrucción 3D que realiza el reconocimiento facial usando el método de reconstrucción invariante bajo

rotaciones de Atick et al. Por último Zhao et al. [Zha00a] proponen un método para adaptar un modelo 3D genérico a la silueta individual para realizar el reconocimiento en diferentes condiciones de iluminación.

A continuación se detallarán los métodos de reconocimiento facial que más relevancia han alcanzado. Estos son:

- Características geométricas.
- Análisis de Componentes Principales (PCA). Autocaras.
- Análisis Lineal Discriminate (LDA). Fisherfaces.
- Correspondencia de plantillas.
- Redes Neuronales.
- EBGM. Transformada wavelet (onditas).
- Modelos de cara en 3D.
- Otras técnicas.

2.2.3.- Características geométricas.

El cálculo de características geométricas como ángulos, relaciones o distancias en una cara permite la aplicación inmediata de técnicas estadísticas de reconocimiento de patrones. El principal inconveniente de este modelo radica en que los algoritmos de localización automática de puntos en la cara no son precisos.

Un rostro puede identificarse incluso cuando los detalles de características individuales (tales como ojos, boca y nariz) no han sido discriminados. La información remanente es puramente geométrica y representa lo que queda en una resolución poco detallada. Es decir la configuración geométrica global de la cara es suficiente para discriminar dos rostros. Esta configuración puede describirse por un vector de números representando los principales rasgos.

Harmon [Har81] define nueve puntos en el perfil: nariz, la punta de la barbilla, frente, caballete de la nariz, parte inferior de la nariz, garganta, labio superior, boca y labio inferior. Con estas nueve marcas define un vector de dimensión 11 considerando distancias, áreas y ángulos. Se suelen emplear desde 13 hasta 35 números o relaciones entre datos. Un problema es la introducción de datos, ya que si es manual depende del operador. En este caso, los resultados suelen oscilar entre un 85 y un 95% de efectividad.

Uno de los trabajos pioneros fue Bledsoe [Ble64]. En su sistema un operador seleccionaba los puntos y los introducía en el ordenador. Después de calcular un conjunto de distancias, el vecino más cercano o cualquier otro clasificador se empleaba para identificar al sujeto. Posteriormente Kanade [Kan77] considera cuatro regiones: ojos, boca y nariz, junto con un conjunto de 16 parámetros (relaciones entre distancias, áreas y ángulos), obteniendo unos resultados del 45 al 75% de efectividad y utilizando una base de datos compuesta por 20 sujetos.

Brunelli y Poggio [Bru93] usaron un conjunto de características geométricas (véase la Figura 2.12), tales como la anchura y longitud de la nariz, posición de la boca y forma de la barbilla. Con una base de datos formada por 47 individuos obtuvieron un

porcentaje de aciertos del 90%. Sin embargo, comprobaron que mediante la utilización de un sistema sencillo de correspondencia de plantillas se obtenía un 100% de aciertos para la misma base de datos. Cox et al [Cox96] proponen una técnica basada en una mezcla de distancias la cual alcanza un 95% de aciertos con una base de datos de 685 sujetos (una foto por sujeto) y un conjunto de prueba de sólo 95 imágenes. En su caso, cada cara se representa por 30 distancias extraídas manualmente.

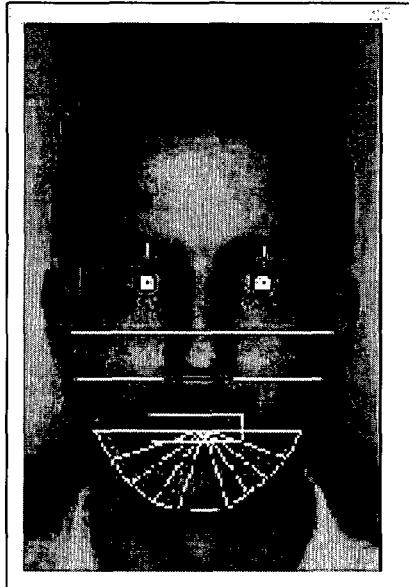


Figura 2.12. Características geométricas consideradas en [Bru93].

Los sistemas que emplean distancias medidas con precisión pueden ser muy útiles para encontrar posibles coincidencias en grandes bases de datos como las policiales. Para otras aplicaciones, es necesario que los puntos sean identificados automáticamente, y el sistema será dependiente de la precisión del algoritmo en la localización de los rasgos del individuo. Los actuales algoritmos para localización de puntos característicos no ofrecen un alto grado de precisión de forma consistente.

2.2.4.- Análisis de Componentes Principales (PCA). Autocaras.

2.2.4.1.- Descripción del algoritmo.

Una imagen digital puede ser expresada en forma de vector de números. Si el ancho y el alto de la imagen son w y h píxeles, el número de los componentes del vector es $w \cdot h$, donde cada componente es el valor en escala de grises de un pixel. Este vector se forma por la concatenación de las filas de píxeles de la imagen.

El espacio-imagen se puede generar a partir de una base de vectores. Cada vector de esta base representa la activación de un único píxel en la imagen. Este espacio vectorial nos permite representar todas las imágenes digitales (en nuestro caso, de tamaño $w \cdot h$). Imágenes que sean parecidas estarán situadas en posiciones muy cercanas. Una medida muy intuitiva de la similitud entre dos imágenes será, por tanto, la distancia euclídea.

En este espacio-imagen todas las imágenes de caras están localizadas en una zona reducida. Esto es debido a que todas las imágenes de caras son parecidas entre sí

(en general, todos tenemos dos ojos, una nariz, una boca, etc. situadas en posiciones similares). También se puede esperar que las imágenes de la cara de un sujeto formen un subregión, dentro de la región más grande de todas las caras. La importancia de que exista esta subregión es que se podría modelar con un conjunto reducido de los vectores base originales. Por ejemplo, en el espacio-imagen de 130 * 140 píxeles (18200 ejes), 150 vectores base pueden ser suficientes para explicar casi toda la varianza de la base de datos inicial de imágenes faciales y para realizar las operaciones habituales de reconocimiento, verificación, etc. Se puede constatar que esta reducción de dimensiones es fundamental para un procesado eficiente de las imágenes de caras humanas. El objetivo del Análisis de Componentes Principales (PCA) consiste en encontrar la base que nos permite representar de forma más eficiente las imágenes de caras humanas. El uso de PCA en el procesado de imágenes de caras humanas es muy habitual y puede encontrarse en numerosas referencias bibliográficas [Tur91].

El Análisis de Componentes Principales fue inicialmente desarrollado dentro del campo de la estadística. Más tarde fue reformulado en el campo de las redes neuronales artificiales. Por lo tanto, disponemos de dos formas de explicar sus principios. A continuación se detalla el paradigma basado en modelos estadísticos.

El espacio-imagen es altamente redundante cuando se utiliza para describir caras. Por lo tanto, como ha sido indicado anteriormente, es posible construir un subespacio de imágenes de caras (llamado en la bibliografía “espacio cara”) que sea menos redundante para representar imágenes faciales. Esta redundancia se debe al hecho de que cada píxel en una cara está muy correlacionado con los otros píxeles.

Sean w y h la anchura y altura de las imágenes del conjunto de entrenamiento, sea K el número de imágenes del conjunto de entrenamiento y X el conjunto de imágenes de entrenamiento, siendo x_i ($i = 1..K$) la imagen I -ésima del conjunto de entrenamiento (es el vector que corresponde a la columna i -ésima de X). En la siguiente explicación, las operaciones con vectores corresponden a vectores escritos en forma de columna.

La matriz de covarianza de X , Σ_X , es una matriz no-diagonal. Su formulación matemática es la siguiente:

$$\sum_X = X^* X^T = \begin{bmatrix} \sigma_{11}^X & \sigma_{12}^X & \dots & \sigma_{1,w^*h}^X \\ \sigma_{21}^X & \sigma_{22}^X & \dots & \sigma_{22}^X \\ \dots & \dots & \dots & \dots \\ \sigma_{w^*h,1}^X & \sigma_{w^*h,2}^X & \dots & \sigma_{w^*h,w^*h}^X \end{bmatrix}$$

Donde σ_{ij} representa la covarianza entre el píxel i y el píxel j . Hay que tener en cuenta que X debe estar previamente normalizada para que su media sea cero (en general, esto se logra restando la media a cada valor de entrada). Se puede definir la siguiente relación entre los coeficientes de covarianza y los coeficientes de correlación:

$$r_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii} \cdot \sigma_{jj}}}$$

Los coeficientes de correlación son los coeficientes de covarianza normalizados.

El objetivo consiste en realizar un cambio de coordenadas para traducir el conjunto de entrenamiento X al espacio cara. Sea Y el conjunto de entrenamiento en el espacio cara, donde y_i es la transformación de x_i .

En el espacio de las caras, cada componente no debe estar correlacionado con ningún otro, esto significa que la matriz de covarianza de los nuevos componentes es diagonal:

$$\Sigma_Y = \mathbf{Y}^* \mathbf{Y}^T = \begin{bmatrix} \sigma_{11}^Y & 0 & \dots & 0 \\ 0 & \sigma_{22}^Y & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_{w^*h, w^*h}^Y \end{bmatrix}$$

Esta forma diagonal de la matriz de covarianza implica que la varianza de una variable consigo misma es máxima, mientras que la covarianza de una variable con cualquier otra va a ser despreciable. En consecuencia, estas variables no van a estar correlacionadas y dicha construcción encuentra las direcciones que maximizan la varianza. Además, se deduce que Σ_Y es la matriz diagonal que contiene los autovalores de Σ_X . Como los elementos de la diagonal de Σ_Y son la varianza de los componentes de las caras de entrenamiento en el espacio de las caras, los autovalores de Σ_X constituyen esas varianzas.

Los componentes principales se calculan linealmente. Sea P la matriz de transformación; entonces se verifica que $\mathbf{Y} = \mathbf{P}^T * \mathbf{X}$ y $\mathbf{X} = \mathbf{P} * \mathbf{Y}$. De hecho, $\mathbf{P}^T = \mathbf{P}^{-1}$, ya que las columnas de P son ortonormales entre sí: $\mathbf{P}^T * \mathbf{P} = \mathbf{I}$, donde I es la matriz identidad. En resumen, una vez obtenida la matriz de transformación P, es posible transformar y reconstruir (ida y vuelta al espacio de cara) un vector imagen x_i (sin necesidad de normalización previa) de la forma siguiente:

$$y_i = \mathbf{P}^T * x_i \quad (\text{transformación del espacio imagen al espacio cara})$$

$$x'_i = \mathbf{P} * y_i \quad (\text{reconstrucción del espacio cara al espacio imagen})$$

En [Tur91] se describe una forma más eficiente de realizar el cálculo de la matriz P: Habitualmente el número de imágenes de caras en el entrenamiento (K) es mucho menor que la dimensión del espacio imagen (w^*h). Por lo tanto, si se calcularan los wh autovalores, sólo ($k-1$) serían distintos de cero. Para prevenir este cálculo innecesario, se deberá calcular sólo los autovalores de una matriz de tamaño ($K * K$).

Sean v_i los autovectores de la matriz de covarianza $\mathbf{X}^T \mathbf{X}$. Como se ha mencionado anteriormente, supóngase que X está normalizado en una región de media cero, en este caso se verifica que: $\mathbf{X}^T \mathbf{X} v_i = k_i v_i$ donde k_i es el correspondiente autovalor. Multiplicando ambos miembros por X, $\mathbf{X} \mathbf{X}^T (\mathbf{X} v_i) = k_i (\mathbf{X} v_i)$ y por lo tanto se puede deducir que $\mathbf{X} v_i$ son los autovectores de $\mathbf{X} \mathbf{X}^T$ y k_i son los correspondientes autovalores.

Se sigue que es posible acelerar el proceso de cálculo de la matriz P, mediante la obtención primero de los autovectores v_i de la matriz $\mathbf{X}^T \mathbf{X}$ (de tamaño $K*K$), y a

continuación mediante su conversión a los deseados autovectores, multiplicándolos previamente por X.

El objetivo del PCA es la reducción de la dimensión del espacio de trabajo. El número máximo de componentes principales es el número de variables del espacio inicial. Sin embargo, en la búsqueda de un espacio de menor dimensión, algunos componentes principales se pueden omitir.

Lógicamente, la dimensión del espacio cara es menor que la del espacio imagen:

$$\dim(\mathbf{Y}) = \text{col}(\mathbf{P}) \times \text{col}(\mathbf{X}) = \text{rango}(\mathbf{X} * \mathbf{X}^T) \times K,$$

$$\dim(\mathbf{y}_i) = \text{rango}(\mathbf{X} * \mathbf{X}^T) \times 1$$

Si se considera que $\text{rango}(\mathbf{X} * \mathbf{X}^T)$ es igual a K, la información almacenada en \mathbf{y}_i es la misma que estaba en \mathbf{x}_i . Todavía es posible reducir aún más la dimensión, ya que no hay ninguna evidencia que indique que la dimensión del espacio de caras ha de ser precisamente K.

El cerebro humano es una máquina muy buena para el reconocimiento facial. A pesar de que es una tarea muy compleja, es capaz de identificar una cara de forma muy rápida. Esta velocidad parece sugerir que la dimensión del espacio de caras ha de ser muy reducida (unas pocas decenas de componentes principales).

Por lo tanto, se podrá considerar un conjunto reducido de componentes principales para reducir aún más la dimensión del espacio de caras, de acuerdo con el desarrollo que sigue a continuación.

La reconstrucción de una cara \mathbf{x}_i , partiendo de los componentes \mathbf{y}_i viene dada por la fórmula $\mathbf{x}_i = \mathbf{P} * \mathbf{y}_i$. Si \mathbf{p}_i son los vectores columna de P, es decir, los vectores base, entonces la reconstrucción de una cara considerando sólo algunos componentes \mathbf{y}_{ij} es:

$$\overline{\mathbf{x}}_i = \sum_{j=1}^M y_{ij} \cdot p_j + \sum_{j=M+1}^K y_{ij} \cdot p_j = \sum_{j=1}^M y_{ij} \cdot p_j + \sum_{j=M+1}^K a_j \cdot p_j \quad \text{donde } a_j \text{ es una constante}$$

cuyo valor se calcula seguidamente.

Este paso introduce un error, el vector error se puede calcular como:

$$\mathbf{e} = \mathbf{x}_i - \overline{\mathbf{x}}_i = \sum_{j=M+1}^K (y_{ij} - a_j) \cdot p_j$$

El vector error es un vector aleatorio y su importancia se puede calcular por su valor cuadrático medio:

$$\langle e^2 \rangle = \mu(e^T * e) = \mu \left(\sum_{j=M+1}^K \sum_{l=M+1}^K (y_{il} - a_l) \cdot (y_{il} - a_l) \right) \cdot p_i^T \cdot p_i = \sum_{i=M+1}^K \mu[(y_i - a_i)^2]$$

Para que este error sea mínimo hay que resolver la siguiente ecuación:

$$\frac{\partial}{\partial a_i} \langle e^2 \rangle = 0$$

es decir, $-2 * \mu((y_i - a_i)) = 0$

de tal forma que $a_i = \mu(y_i)$ para $M+1 \leq i \leq K$

Esto significa que los valores omitidos de y_i han de ser reemplazados por su valor medio para minimizar el error en la reconstrucción. El vector error va ser igual a la suma de la varianza de los componentes no considerados:

$$\langle e^2 \rangle = \sum_{i=M+1}^K \mu[(y_i - \mu(y_i))^2] = \sum_{i=M+1}^K \lambda_i$$

Por lo tanto, si los autovalores se clasifican en orden decreciente, los últimos autovalores (y sus autovectores) pueden eliminarse. Esta es la reducción real de dimensiones que ofrece la PCA. En efecto, significa que se pueden eliminar algunos de los componentes principales porque éstos solo explican una pequeña porción de la información contenida en los datos iniciales, mientras que la mayor parte de la información se encuentra en los otros componentes principales. La cantidad de información que el componente principal i -ésimo explica viene dada por su autovalor, λ_i .

En otras palabras, los componentes principales (autovectores) se ordenan por sus autovalores de forma decreciente, eliminando los últimos. Este hecho es particularmente importante ya que en el caso de caras, los valores de los autovalores en general decrecen exponencialmente.

Como se ha comentado previamente, el número de dimensiones que debe tener el espacio de caras no está determinado. Sin embargo, puede realizarse un prueba sencilla que indique el número de autovectores que hay que considerar. Teniendo en cuenta que los autovalores decrecen de forma exponencial, los últimos autovalores van a tener un valor casi idéntico y muy pequeño: $\lambda_N = \lambda_{N+1} = \lambda_{N+2} = \dots = \lambda_K$, con $1 < N < K$.

En este caso, los componentes principales asociados con estos autovalores son arbitrarios y pueden ser descartados. Por consiguiente, N será la dimensión del espacio de caras.

De manera estadística, los valores λ_i representan la variabilidad del conjunto en la dirección del autovector asociado y se suele denominar varianza explicada. La suma de los autovalores λ_i considerados es la varianza explicada total y se suele expresar en forma de porcentaje referido a la suma de todos los λ_i . Esta medida da una idea de la bondad de nuestra elección, pero siempre referida al conjunto de entrenamiento. Es decir, una varianza explicada del 100% tiene poco significado si el conjunto de entrenamiento tiene unas pocas imágenes o las imágenes no son representativas del conjunto de caras con el que se va a trabajar.

Un autovector es un vector cuya dimensión es la misma que las imágenes iniciales y por lo tanto se puede ver como una imagen, en el llamado espacio imagen. Al hacerlo así, da la apariencia de imágenes de caras, pero muy diferentes del conjunto inicial. Por esta razón, en la literatura se conocen como autocaras (eigenfaces). La Figura 2.13 muestra los seis primeros autovectores de un conjunto de entrenamiento.



Figura 2.13. Seis primeras autocaras de una base de datos.

Una autocara puede ser interpretada de la siguiente forma: la dirección de un autovector es paralela a un eje en el espacio cara. Los primeros autovectores representan las direcciones de mayor varianza en la región de las caras. Cada valor de un autovector es una coordenada en el espacio imagen, representando la contribución de ese dato a la varianza de la región de caras en esa dirección. Un valor grande de esa cifra representa una contribución importante. Como las imágenes se muestran en blanco y negro, valores muy próximos al blanco o al negro representan contribuciones significativas de esas magnitudes.

La “cara media” es otro resultado de nuestro proceso; su principal aplicación consiste en normalizar el conjunto de entrenamiento X. Pero además, de forma análoga a las autocaras, su visualización aporta información intuitiva sobre la localización del conjunto de caras. Un ejemplo de “cara media” es el que se muestra en la Figura 2.14.



Figura 2. 14. Cara media para la misma base de datos.

2.2.4.2.- Resultados obtenidos con PCA.

Turk y Pentland [Tur91] presentaron un modelo en el cual la cara es proyectada según los componentes principales del conjunto de imágenes de entrenamiento. Las tareas de reconocimiento de alto nivel son modeladas típicamente con muchos pasos de procesamiento en el paradigma de Marr. Sin embargo Turk y Pentland proponen un modelo de reconocimiento basado en un procesamiento bidimensional de la imagen, de bajo nivel. Su argumento se basa en el rápido desarrollo y extrema rapidez del reconocimiento de caras que desarrollan los seres humanos, así como en experimentos fisiológicos en el córtex de monos, los cuales parecen sugerir que existen neuronas especializadas que responden selectivamente a las caras. Sin embargo, no resulta evidente ni está comprobado que estos experimentos excluyan la necesidad de realizar las operaciones del paradigma de Marr.

Turk y Pentland presentan sus resultados utilizando una base de datos de 16 sujetos con distintos grados de orientación de la cabeza, escala e iluminación. Sus

imágenes aparecen idénticas salvo pequeñas diferencias en la expresión facial, en pequeños detalles o en la posición, etc. Para variaciones de iluminación, orientación y escala su sistema alcanzó 96, 85 y 64% de clasificación correcta, respectivamente. La escala se halla normalizada al tamaño de las autocaras. La parte central de la cara se acentúa, reduciendo así los efectos negativos de los cambios en el pelo o en el fondo.

Pentland obtuvo buenos resultados en una base de datos grande (95% de aciertos para 200 personas en una base de datos de 3.000). Es difícil deducir conclusiones generales ya que muchas de las imágenes del mismo individuo parecen muy similares y la base de datos está muy alineada y con registros muy precisos. En Moghaddam y Pentland [Mog97] se obtuvieron muy buenos resultados con la base de datos FERET (sólo un error en la clasificación de 150 imágenes frontales). El sistema utilizaba procesamiento extensivo para localización de la cabeza, detección de características y normalización para la geometría de la cara, traslación, iluminación, contraste, rotación y escala.

En resumen, parece que las autocaras es un método rápido, simple y práctico que puede tener limitaciones debido al requisito de que tiene que existir un alto grado de correlación entre las intensidades de los píxeles en los conjuntos de entrenamiento y de prueba. Esta limitación ha sido enfocada utilizando preprocesamiento extensivo para normalizar las imágenes.

Turk y Pentland analizaron la robustez de la aproximación de autovectores en el reconocimiento de caras. Primero crearon una base de datos de 2500 imágenes de 16 sujetos obtenidas en diferentes condiciones de iluminación, escala y orientación. A continuación se seleccionaron varios conjuntos de 16 imágenes (una de cada sujeto), obtenidas en las mismas condiciones para formar el conjunto de entrenamiento. Las caras restantes se utilizaron como conjunto de prueba. Para cada conjunto de entrenamiento se creó una matriz de covarianza y se extrajeron los ocho autovectores cuyos autovalores fueran los más altos. Para reconocimiento, las caras del conjunto de prueba se proyectaron en esos ocho autovectores. Los patrones de peso resultantes se compararon con las proyecciones almacenadas de las caras de entrenamiento y se clasificaron como conocidos o desconocidos, utilizando un algoritmo del vecino más cercano basado en la distancia euclídea entre la proyección de la imagen objetivo y la proyección de cada imagen almacenada. Si la distancia al vecino más cercano resultaba ser inferior a un umbral determinado, la cara se clasificaba como conocida, en caso contrario se clasificaron como desconocidas.

Un conjunto de simulaciones con varios conjuntos de entrenamiento y de prueba mostraron que el sistema era relativamente insensible frente a cambios en las condiciones de iluminación (valor medio: 96% de aciertos) y orientación bidimensional de la cabeza (valor medio: 85% de aciertos). Estos resultados son lógicos ya que si los cambios en iluminación o en la orientación de la cabeza no son grandes, la correlación píxel a píxel permanece alta. Si los cambios son suficientemente grandes, entonces el rendimiento debe decrecer considerablemente. Por último, la capacidad de reconocimiento decrece con los cambios en el tamaño de la cabeza (valor medio: 64% de aciertos). Sin embargo, este problema puede ser solucionado detectando el contorno de la cara y realizando un cambio de escala antes de la fase de reconocimiento.

Los estudios de Sirovich y Kirby [Sir87] y Turk y Pentland [Tur91] proponen el análisis de componentes principales como forma de obtener una representación de la cara de forma más compacta, motivado desde un punto de vista de la teoría de la información. Estos investigadores estaban interesados en encontrar la forma óptima de representar un conjunto de imágenes faciales con el menor número posible de parámetros. La representación que proponen (es decir, tomar los autovectores con los mayores autovalores) es óptima en el sentido de que es mínimo el error cuadrático.

O'Toole [Oto93] demostró que esa representación de dimensión pequeña es en efecto óptima para reconstruir las caras, pero no constituye la representación más adecuada para el reconocimiento. En una simulación inicial, se examinó la importancia de los diferentes rangos de autovectores para discriminar caras conocidas de otras desconocidas (tarea de reconocimiento). Se creó una matriz autoasociativa con 100 imágenes (50 mujeres y 50 hombres) y se descompuso en sus autovectores. Varios rangos de 15 autovectores (ordenados en orden decreciente de acuerdo con sus autovalores) fueron utilizados para reconstruir las 100 caras conocidas y 59 caras nuevas. La calidad en la reconstrucción de cada cara se estimó poniendo en forma de vector la cara original y la reconstruida y calculando luego el coseno del ángulo entre los dos vectores. Para la tarea del reconocimiento, el valor promedio del coseno se utilizó como un criterio en la decisión. Caras con un valor del coseno por encima del criterio se consideró que eran caras conocidas, y desconocidas si el coseno estaba por debajo del umbral. Los resultados fueron los siguientes:

- 1) La calidad de la representación (similaridad física medida por el coseno entre la cara original y la reconstruida) decrece a medida que se seleccionan autovectores con autovalores más pequeños.
- 2) La habilidad del modelo para discriminar una cara conocida no seguía el decremento visto en la calidad de la representación. Alguno de los autovectores en el rango entre el 45 y 80 proporcionaba más información para reconocer caras que alguno de los 15 primeros autovectores.

Además de la información de la cara útil para identificar personas, el modelo también conserva información acerca del sexo, raza y edad. En otros trabajos, O'Toole et al. [Oto91] habían examinado la utilidad de los autovectores con autovalores relativamente grandes para representar categorías faciales. Por ejemplo, manifestaron que cuando entrenaban un conjunto formado por personas de raza japonesa y caucásica, el segundo autovector llevaba la mayor parte de la información necesaria para identificar la raza. Para ello calcularon la proyección de todas las caras en el segundo autovector. El valor medio se seleccionó como valor umbral; las caras con proyecciones mayores que el umbral se supusieron caucásicas y las que tenían un valor menor se supuso que eran japonesas. Empleando este mecanismo tan sencillo se obtuvo un porcentaje del 88.6%. Un detalle interesante es que el sistema no estaba entrenado para clasificar las caras, sino para reconstruirlas. La información necesaria para clasificar las caras por su raza emergió espontáneamente de la representación.

Sin embargo, estos resultados dependen de la homogeneidad sexual y racial del conjunto de entrenamiento. Desde un punto de vista psicológico, es una característica positiva ya que parece recordar ciertos efectos psicológicos como el “efecto de la otra raza” (es decir, las personas son capaces de reconocer más fácilmente caras de su raza que de otras razas). El primer autovector por sí mismo representa esencialmente las

características compartidas por todas las caras en el conjunto de entrenamiento. Es decir, representa la categoría “cara humana” y puede ser interpretada como un prototipo de lo que es una cara humana. Este autovector puede ser utilizado para separar caras humanas de otro tipo de objetos. El autovalor asociado con el primer autovector es generalmente muy alto (explica entre el 90% y el 99% de la varianza). Este valor tan alto refleja el hecho de que las caras humanas son objetos muy parecidos entre sí.

Una implicación muy importante de los estudios de O’Toole es que el conjunto óptimo de características de la representación no es absoluto, sino que depende de los requisitos de la tarea que se pretende realizar. Los autovectores con autovalores muy altos aportan información acerca de la forma general de la cara. Por lo tanto, una representación con una dimensión baja de una cara en el subespacio asociado a los autovalores más altos es adecuada para resolver problemas de categorización semántica tales como raza o sexo. En contraste, autovectores con autovalores pequeños muestran información sobre detalles pequeños, propios de cada individuo. Una representación de dimensión reducida pero en el subespacio asociado con los autovectores de autovalores pequeños es adecuada para la identificación de caras individuales o para la discriminación entre caras conocidas y desconocidas.

2.2.5.- Análisis Lineal Discriminate (LDA). Fisherfaces.

Una variación del Análisis de Componentes Principales denominada “Fisherfaces” fue propuesta en [Bel97], como una mejora de la primera. Teniendo en cuenta que el conjunto de entrenamiento está etiquetado, es posible utilizar esta información para reducir la dimensionalidad del espacio de características. Aplicando el Análisis Discriminante Lineal (Linear Discriminant Analysis) de Fisher, es posible construir una matriz de proyección en la cual la razón entre la dispersión intra-clase y la inter-clase sea máxima. Los resultados muestran que tanto PCA como LDA (a veces referida en la bibliografía como Fisherfaces) obtienen un buen rendimiento si las imágenes de prueba son similares a las imágenes de entrenamiento. Sin embargo, los experimentos desarrollados con la base de datos de Yale [Yal] muestran que el método LDA obtiene mejores resultados en caso de que haya variaciones en las condiciones de iluminación y de gesto (15.3% de reconocimiento incorrecto para PCA frente a 7.3% para Fisherfaces).

En [Bel97] se afirma que los autovectores que corresponden a los autovalores de mayor valor a menudo contienen información relativa a las variaciones de iluminación, más que a la identidad de los individuos. Por lo tanto, si estos autovectores se descartan (habitualmente los tres primeros como en [Oto93]), entonces la proyección de la imagen en los restantes autovectores debería tener unos valores de los cuales se han eliminado los cambios debidos a variaciones en la iluminación. Sin embargo, tanto [Bel93] como [Sim00] comprueban que al eliminar esos autovectores el rendimiento del sistema decae. Como apuntan ambos, tal vez estos tres autovectores no sólo codifican información de las condiciones de iluminación, sino que también contengan información útil para el reconocimiento.

2.2.6.- Correspondencia de plantillas.

Una serie de autores como Brunelli y Poggio [Bru93] (véase la Figura 2.12) y Yullie et al. [Yul89] realizaron correlaciones entre trozos de imágenes. En general, esta técnica es eficaz cuando las imágenes de prueba tiene la misma escala, orientación e iluminación que el conjunto de entrenamiento.

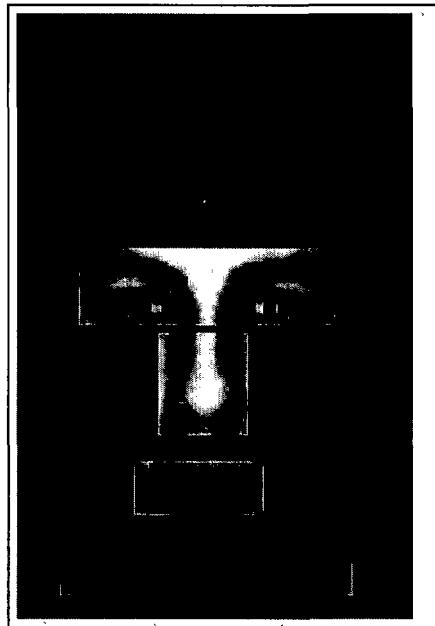


Figura 2.15. Regiones consideradas en [Bru93] para la correspondencia de plantillas.

En su versión más amplia, la imagen se compara a través de una métrica con una máscara adaptable o parametrizable que representa la cara global; en otros casos se realiza una plantilla para la boca, los ojos y/o la nariz. También en este caso la imagen ha de ser normalizada tanto en niveles de gris como en orientación. Surgen problemas tales como la iluminación o que la persona adopte un gesto que deforme la plantilla (párpado cerrado, sonrisa, etc.).

Uno de los primeros en utilizar esta técnica fue Yuille. En su trabajo emplea una máscara parametrizable para describir un ojo. Esta máscara está definida por un conjunto de parámetros que permiten un conocimiento a priori de la forma de las características para guiar el proceso de detección. La máscara interacciona con la imagen de forma dinámica, tratando de minimizar una función de energía (un coste). Todos los parámetros son experimentales.

Suelen obtenerse buenos resultados utilizando esta técnica aunque con tiempos de cómputo bastante grandes.

2.2.7.- Redes Neuronales.

Las redes neuronales pueden utilizarse dentro de un sistema de reconocimiento facial para clasificar las características de un individuo, interviniendo sólo en la parte final del proceso. En este caso dichas características han podido ser extraídas usando cualquiera de los métodos presentados en este capítulo. Sin embargo, los modelos que utilizan redes neuronales suelen considerar como entrada los niveles de gris de la imagen, de tal forma que es la propia red neuronal la que debe seleccionar las características más importantes. La representación geométrica de la cara se codifica

implícitamente pero se añaden elementos como textura y forma. Aunque este tipo de representación no crea una representación tridimensional invariante, sin embargo preserva información de la configuración y basada en características. Una de las limitaciones de este modelo es su sensibilidad frente a variaciones en la iluminación, orientación y tamaño de la cabeza. Para evitar estos inconvenientes es necesaria una etapa de preprocesamiento.

Dos características surgen cuando se aplican redes neuronales:

- La información se procesa en paralelo, en lugar de secuencialmente, por un conjunto de unidades sencillas interconectadas entre sí (las neuronas).
- La información se difunde a través de toda la red neuronal, en lugar de estar concentrada en una determinada zona.

Los modelos basados en redes neuronales se han dividido en dos grupos. En el primer grupo se presentan las redes neuronales autoasociativas (equivalente a la realización de un análisis de componentes principales) aplicadas al reconocimiento de caras humanas. El segundo grupo corresponde a la aplicación de redes de retropropagación para el reconocimiento facial.

2.2.7.1.- Enfoque basado en el análisis de componentes principales (PCA). Memoria autoasociativa

Hay una amplia clase de redes neuronales que puede ser entrenada para que realice una correspondencia de los datos de entrada a un espacio de menor dimensión [Lam98a]. Una de las formas más utilizadas es la implementación del Análisis de Componentes Principales (PCA) con una red neuronal. Inicialmente se planteó en una dimensión [Oja82], aunque se generalizó a más dimensiones [Oja89]. En [Bal89], Baldi y Hornink demostraron que entrenar una red neuronal autoasociativa de tres capas es equivalente a aplicar PCA a los datos de entrada.

Este tipo de planteamiento, generalmente denominada análisis de componentes principales (PCA) al modelo de caras, se basa esencialmente en el hecho de que las caras pueden ser representadas, bien implícitamente (a través de una memoria lineal autoasociativa) o bien explícitamente (a través del análisis de componentes principales), como una suma ponderada de autovectores extraídos de un conjunto de caras. Los autovectores se pueden considerar como un conjunto de características o de componentes básicos a partir de los cuales se construye la cara.

Kohonen [Koh81] demostró que una memoria autoasociativa podía ser usada para almacenar y recuperar imágenes faciales (es decir como una memoria direccionable de caras). Consideró una muestra de 100 caras. La eficiencia de la memoria fue puesta a prueba presentándole imágenes incompletas o con ruido y observando la imagen reconstruida por la memoria. Los resultados fueron concluyentes: si a la memoria se le presenta una imagen incompleta o parcialmente borrada, la memoria rellena las zonas perdidas de la imagen.

Usando una aproximación similar, Millward y O'Toole [Mil86] mostraron que las memorias autoasociativas podían actuar como sistemas eficientes para reconocer caras (distinguir entre caras conocidas y desconocidas). En su estudio, una memoria autoasociativa fue construida por la autoasociación de un conjunto de vectores de caras,

empleando la técnica de aprendizaje Widrow-Hoff. El reconocimiento de caras se simuló utilizando un paradigma psicológico llamado “tarea de elección forzada entre dos alternativas” (2AFC). Esto se realiza presentando a la memoria un par de vectores, donde cada par se compone de una cara previamente aprendida por la memoria y una cara no estudiada. Para cada vector la calidad de la respuesta del modelo se estima calculando el coseno entre el vector de entrada y el de salida. La cara con el coseno mayor se supone que es “reconocida” por la memoria. Los resultados muestran que este sistema es capaz de reconocer caras.

Anderson y Mozer [And81] se refieren a este tipo de características como “macrocaracterísticas” para diferenciarlas del uso más tradicional del término características. Afirman que “macrocaracterísticas es lo que la percepción y el análisis de características consideran actualmente y son entidades destacadas psicológicamente”. Aplicada al modelado facial, la aproximación PCA sugiere una definición diferente de las características que definen una cara, en lugar de la asumida en los modelos basados en características geométricas o en correspondencia de plantillas. La ventaja de representar caras en términos de macrocaracterísticas, en vez de las características tradicionales, es que las macrocaracterísticas no se definen “a priori”, sino que se generan “a posteriori” de un estudio estadístico. Reflejan por tanto la estructura estadística del conjunto de caras de las que se extraen. En otras palabras, reconstruir caras desde una memoria autoasociativa es como aplicar un filtro a las imágenes faciales (concretamente, un filtro Wiener).

Un ejemplo de la utilidad de los autovectores (tanto autoimágenes como autocaras) para la representación de caras fue aportado por Sirovich y Kirby [Sir87]. Analizaron las propiedades de una matriz de covarianza creada por 115 imágenes. Su análisis mostró lo siguiente:

- 1.- Cuando se muestran gráficamente, los autovectores de la matriz de covarianza de un conjunto de caras tiene aspecto de cara.
- 2.- Cualquier cara almacenada puede ser reconstruida exactamente como combinación lineal de los autovectores de la matriz de covarianza o aproximada utilizando sólo los autovectores de autovalores grandes.

Los resultados mostraron que las caras pueden ser reconstruidas óptimamente usando unos 40 parámetros y los correspondientes 40 primeros autovectores. Los parámetros corresponden a la proyección de la cara en los autovectores. Poco después Kirby y Sirovich [Kir90] incluyeron ojos, nariz y boca en las imágenes. Mostraron que el 82% de la varianza del conjunto de caras se encuentra en los 10 primeros autovectores y el 95% en los 50 primeros autovectores.

2.2.7.2.- Redes de retropropagación: compresión de imágenes. Aplicación al procesamiento de imágenes faciales.

En contraste con las redes asociativas lineales en las cuales las unidades de entrada están directamente conectadas a las unidades de salida, las redes de retropropagación incluyen elementos ocultos no lineales entre las unidades de entrada y de salida.

La habilidad de las redes neuronales para extraer información útil fue inicialmente aplicada a caras humanas por Cottrell y Fleming [Cot90]. Entrenaron una

red de tres capas con 4096 unidades de entrada, 80 unidades en la capa oculta de tipo sigmoidal y 4096 unidades de salida, para comprimir un conjunto de imágenes de caras y de no-caras. El conjunto de entrenamiento consistía en 64 imágenes de caras (5 ó 6 imágenes de 11 personas) y 13 imágenes de no-caras. La base de datos completa estaba formada por 231 imágenes de las cuales 204 eran caras (entre 5 y 20 imágenes de 17 personas) y 27 no contenían caras. Todas las caras en el conjunto de entrenamiento se le suministraron a la red de compresión. La representación formada en la capa oculta fue utilizada como entrada de una red de 80 unidades de entrada y 14 neuronas de salida con un rango [-1,1], entrenada para clasificar “apariencia de cara”, sexo e identidad.

La habilidad del modelo para generalizar a nuevas imágenes fue probada presentándole a la red de clasificación un conjunto de prueba formado por nuevas imágenes de las caras estudiadas, caras nuevas, elementos nuevos que no son caras y un conjunto de imágenes degradadas de caras (por ejemplo: parcialmente oscurecidas, diferente iluminación). Los resultados se muestran a continuación:

1. - El sistema identificaba y categorizaba perfectamente las imágenes con las cuales fue entrenado y clasifica casi perfectamente (3% de error) las nuevas instancias de las caras conocidas.

2. - Las nuevas caras se caracterizaban perfectamente en su “apariencia de cara” pero no de acuerdo con el sexo (37% de error: 26 de las 70 mujeres se clasificaron como hombres, mas todas las imágenes masculinas fueron correctamente clasificadas).

3. - El modelo era bastante robusto frente a imágenes parcialmente degradadas o deterioradas. Era capaz de identificar y categorizar (sólo un 3% de error) caras parcialmente oscurecidas por una barra horizontal de color gris de tamaño 1/5 de la altura de la imagen y colocada en el 4/5 inferior de la imagen. Sin embargo, cuando la banda se colocaba en el 1/5 superior de la imagen, el error aumentaba al 29% en la identificación y al 16 % en la categorización del sexo. Esto sugiere que la red estaba utilizando la información de la frente para discriminar entre caras. La importancia de la región frontal se confirmó por el hecho que la red se equivocaba más cuando la mitad superior de la cara se tapaba, que cuando se ocultaba la mitad inferior (56% frente al 0% en la categoría de “apariencia de cara”, 70% frente a un 50% en la identificación y 55% frente a un 29% en la categoría de sexo).

4. - Modificaciones en el brillo (hasta un 70%) mostraban un aumento del error inferior al 7% tanto en identificación como en clasificación.

Lawrence et al. [Law97] propusieron un sistema con un muestreo local de la imagen, un mapa auto-organizativo y una red neuronal convolutiva. En una base de datos de 40 personas y 400 imágenes (base de datos Olivetti [Oli]), el método que proponen posee un error del 3.8% frente a un 10.5% que obtienen utilizando PCA.

Lin et al. [Lin97] consideraron una red neuronal basada en decisión probabilística para detección y reconocimiento de caras. Han realizado experimentos con tres bases de datos, FERET (de esta considera sólo un subconjunto), ORL y una base de datos creada “ad hoc”. En la base de datos ORL el error que tienen es muy similar a [Law97] aunque los tiempos de entrenamiento y prueba son menores. El algoritmo propuesto por Lawrence tarda cuatro horas en entrenar frente a los veinte minutos de Lin. En el caso de FERET, afirman obtener un 99% de aciertos aunque sólo consideran un subconjunto de esta base de datos.

2.2.8.- Correspondencia entre agrupaciones de grafos elásticos (EBGM). Transformada wavelet (onditas).

En este caso, la representación de una cara toma la forma de grafos etiquetados. Los grafos están formados por vectores y nodos; los vectores se etiquetan con información geométrica (distancias) y los nodos se etiquetan con un conjunto de características locales llamados “jets”. Los jets se basan en transformaciones de Gabor, lo cual se podría tomar como un procedimiento de preprocesamiento de imágenes basado en fenómenos biológicos.

Supongamos una imagen con distribución de niveles de gris $I(x)$. Sea x un punto de la imagen, para ese punto la transformada wavelet de Gabor va a ser una onda plana que se define de acuerdo con la fórmula siguiente:

$$\psi_k(x) = \frac{k^2}{\sigma^2} e^{-\frac{k^2}{2\sigma^2}x^2} \left(e^{ikx} - e^{-\frac{\sigma^2}{2}} \right)$$

donde k es el vector de onda y σ es un parámetro que relaciona el vector de onda con el tamaño de la ventana en la cual está definida la onda plana. El segundo término del paréntesis elimina la componente continua. Una ondita, centrada en la posición x den la imagen, se usa para extraer la componente J_k de la imagen $I(x)$:

$$J_k(x) = \int I(x') \psi_k(x-x') dx'$$

La técnica EBGM (Correspondencia entre agrupaciones de grafos elásticos) lleva a cabo un muestreo en el espacio de vectores de ondas k , con cinco niveles de resolución y ocho orientaciones para cada nivel de resolución. Se obtienen por lo tanto 40 valores complejos para cada punto considerado de la imagen. Se designan las muestras en el espacio k por el índice $j = 1 \dots 40$ y se consideran todas las componentes wavelet centrados en un punto como un jet. Un jet describe las características locales del área que rodea a un punto x de la imagen. Si la densidad de puntos considerados es suficientemente alta, se podría reconstruir la imagen inicial a partir de los distintos jets.

Para poder comparar los jets y los grafos, se definen funciones de similaridad. Si dos grafos son de igual geometría, su similaridad es la suma de las similaridades de los correspondientes jets. Si los grafos tienen distorsión, se puede introducir un segundo término [Lad93] que lo considere. En el desarrollo presentado en [Oka98] y en [Wis97] se consideran dos funciones para medir la similaridad. Si los componentes de un jet J se escriben de la forma $J_j = a_j e^{i\phi_j}$, con amplitud a_j y fase ϕ_j , una forma de medir la similaridad de dos jets J y J' es el producto escalar normalizado:

$$S(J, J') = \frac{\sum_j a_j a'^*_j}{\sqrt{\sum_j a_j^2 \sum_j a'^*_j}} \quad (\text{Eq 1})$$

La otra función de similaridad tiene la forma:

$$S(J, J') = \frac{\sum_j a_j a'_j \cos(\phi_j - \phi'_{j'} - dk_j)}{\sqrt{\sum_j a_j^2 \sum_j a'^2_j}} \quad (\text{Eq 2})$$

Esta función contiene el vector de desplazamiento relativo d entre los puntos de la imagen a los que se refieren los dos jets. Cuando se comparan dos jets durante la correspondencia entre grafos, la similaridad entre ellos se maximiza respecto a d , conduciendo a una determinación precisa de la posición del jet. La primera ecuación se considera inicialmente para establecer de forma aproximada la correspondencia en el grafo, siendo la segunda ecuación la que da la posición de forma más precisa.

El proceso fundamental en el sistema consiste en la correspondencia elástica entre grafos. Mediante este planteamiento, un grafo modelo (un grafo derivado de una imagen facial con posiciones de nodos adecuadas) se compara con la imagen de prueba. Aquí los nodos del grafo modelo se colocan de forma aproximada en la imagen, se extraen los jets de estos puntos y se calcula la similaridad entre el grafo modelo y el grafo imagen así construido. Esta similaridad se optimiza variando las posiciones de los nodos en la imagen.

En una fase inicial, estas variaciones de las posiciones de los nodos toman la forma de un movimiento global de una copia rígida de las posiciones de los nodos del grafo modelo. En una segunda fase, se permite a los nodos de la imagen que se muevan individualmente, introduciendo distorsiones elásticas en el grafo; para encontrar la mejor correspondencia en esta fase se usa la Eq 1. Con esta función de similaridad, grafos y jets son atraídos a sus puntos correspondientes en la imagen utilizando el gradiente creciente de la función similaridad. Cuando se busca localizar con gran precisión un jet, se emplea la ecuación Eq 2, la cual es muy sensible frente a pequeños desplazamientos de la posición del jet.

Si se intenta encontrar una cara desconocida en una imagen y definir un grafo que la represente, se usa una estructura llamada “grafo grupo” (“bunch graph”). Es similar a los grafos descritos anteriormente, pero en lugar de considerar un jet por nodo, se une un conjunto de jets, cada uno de ellos derivado de una imagen facial distinta. Para formar un grafo grupo, se marcan en una colección de imágenes faciales (la llamada galería de imágenes del grafo grupo) las posiciones de los nodos. Se denominarán puntos característicos a estas posiciones de los nodos. Esta marcación se realiza de forma semi automática. Cuando se empareja un grafo grupo con una imagen, el jet extraído de la imagen se compara con todos los jets en el correspondiente grupo que forma el grafo grupo; el que lleve a cabo la mejor correspondencia es seleccionado. Este proceso se llama “correspondencia elástica entre grafos grupo” (“elastic bunch graph matching”). Si la selección de imágenes es adecuada, el grafo grupo es capaz de representar una gran variedad de caras con propiedades locales, en entornos reducidos cercanos a un punto dado, diferentes.

La tarea del reconocimiento se realiza en tres etapas: localización de la cara, localización de los puntos característicos y reconocimiento por comparación. Las dos primeras etapas sirven para crear un modelo invariante frente a cambios de escala de

una imagen de entrada. Ambas etapas se basan en la correspondencia elástica entre grafos grupo, aunque con un conjunto de parámetros diferente correspondiente al distinto nivel de detalle. Caras con pose diferente (en el test FERET la pose se identificaba en el nombre del fichero) se procesan de la misma forma, pero con grafos grupo adecuados a las nuevas poses. Por fin, en la última etapa, los modelos de cara se compararán para obtener el reconocimiento. Se muestran a continuación con más detalle las características más importantes de las tres etapas.

- Localización de la cara.

La primera etapa sirve para encontrar la cara en la imagen y determinar su tamaño. Esta tarea se realiza por un conjunto de emparejamientos a grafos grupo de pose adecuada y de tres tamaños diferentes. El grafo que define el mejor emparejamiento determina el tamaño y posición de la cara. A continuación se coloca un marco cuadrado alrededor de la cara, de tal forma que la cara ocupe una cuarta parte del área del marco. La imagen resultante se expande hasta 128x128 píxeles y se calcula una nueva transformada wavelet (se define así el llamado “marco de la imagen”). El marco de la imagen se pasa al siguiente módulo, el localizador de puntos característicos. La precisión de este paso al dejar la cara completamente dentro del marco de la imagen es crucial para el éxito del sistema.

- Localizador de puntos característicos.

A pesar de que en la localización de la cara se colocó sobre la misma un conjunto de nodos, el procedimiento básico se repite con un grafo grupo que contiene más nodos y una galería de imágenes del grafo grupo más grande. El propósito de este paso consiste en hallar puntos característicos faciales con gran precisión y seguridad en su posición y codificar la información de la imagen de la forma más precisa posible. Este paso es igualmente crucial, ya que un nodo desplazado de su posición correcta distorsionará los valores de similaridad en la etapa de comparación. Para una cara en posición frontal se consideran 48 nodos en esta fase para el grafo, en comparación con la etapa anterior donde se consideran 16 nodos. Los 48 nodos, junto con los correspondientes jets, definen el grafo modelo y contienen toda la información extraída de la imagen.

- Comparación de grafos.

Los grafos modelo obtenidos como resultado del localizador de puntos característicos se comparan por parejas para calcular el valor de la similaridad. Este valor se calcula como la suma de las similaridades de jets entre pares de nodos correspondientes dividido entre el número de pares (utilizando para ello la ecuación Eq 1). Como los grafos modelo para poses distintas difieren en su estructura, se emplea una tabla de conversión para identificar los nodos correspondientes al mismo punto característico. El resultado de la comparación entre grafos constituye una conclusión global de la comparación, que contiene para cada imagen la lista de imágenes ordenadas en orden descendente de similaridad.

Por otra parte, además de este procedimiento básico, en [Oka98] se describe una mejora de los algoritmos muy enfocados a la base de datos FERET (básicamente se enfocan a ecualizar el histograma, cambiar la escala de los filtros de Gabor para mejorar

el cálculo del tamaño de la cara y realizar una transformación de los jets que compense la presencia de fuertes rotaciones de la cara).

2.2.9.- Modelos de cara en 3D.

El modelado facial desempeña un papel muy importante en diversas aplicaciones tales como el seguimiento de la cabeza, animación facial, codificación y compresión de caras, reconocimiento de expresiones faciales y, más recientemente, reconocimiento facial. En las aplicaciones orientadas a reconocimiento facial, el modelado de caras humanas puede ofrecer una representación explícita de una cara que alinee a la vez la forma y la textura de la cara para realizar un emparejamiento independientemente de la pose y la iluminación de la cara.

- Modelos de cara genéricos.

Inicialmente se describen tres aproximaciones principales para modelar caras humanas y a continuación se muestra un modelo avanzado que hace uso del conocimiento a priori sobre geometría facial.

DeCarlo et al. [Dec98] usan medidas antropométricas para generar un modelo general de cara (véase la Figura 2.16). Esta aproximación empieza con superficies B-splines construidas manualmente y a continuación se aplica un mecanismo de mejora basado en optimización con restricciones y ajuste de superficies. Debido a este mecanismo de optimización el proceso es computacionalmente muy costoso.

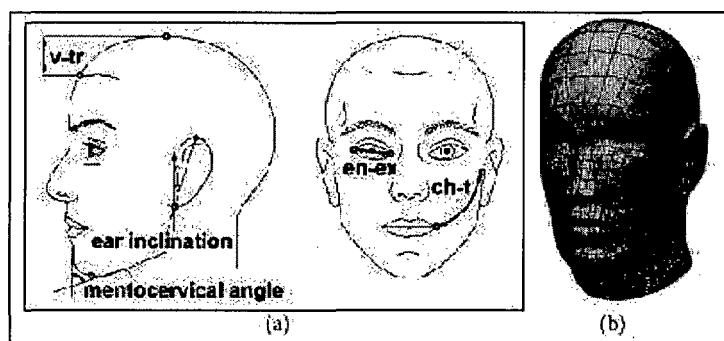


Figura 2.16. Modelo de cara utilizando medidas antropométricas (obtenido de [Dec98]). (a) Medidas antropométricas. (b) Modelo B-spline.

En la segunda aproximación, las medidas faciales se adquieren directamente partiendo de digitalizadores 3D o sensores de luz estructurada como por ejemplo en [Hal99] [Hes02]. Los modelos se obtienen después de post-procesar y triangularizar estas medidas de la forma de la cara.

La tercera aproximación se basa en calcular modelos a partir de fotografías, usando dispositivos de bajo coste. Por ejemplo, en [Len96] [Che98] partiendo de un par de cámaras estéreo se construye un modelo de cara. [Ati96] y [Yan98] reconstruyen la forma de la cara usando el sombreado y el análisis de componentes principales. En [Zha00a], se hace uso de la misma técnica de forma a partir de sombreado, mas aprovechando la forma simétrica de la cara, construyen un modelo 3D para el reconocimiento. En [Kim91], [Hou93], [Cry95], se combinan técnicas de extracción de forma a partir de estéreo y de sombreado para reconstruir caras en 3D. En [Par96] hay referencias a los distintos métodos surgidos hasta ese momento para obtener datos de la

superficie facial. Sin embargo, actualmente todavía es muy difícil extraer suficiente información sobre la geometría facial partiendo sólo de imágenes 2D. Esta dificultad es la razón por la cual Guenter et al. [Gue98] utilizan un gran número de puntos faciales para capturar la geometría 3D de la cara y obtener una animación fotorrealista. A pesar de que se pueden obtener medidas 3D muy densas partiendo de caros y modernos digitalizadores 3D, es muy costoso adquirir una base de datos con un número de sujetos grande.

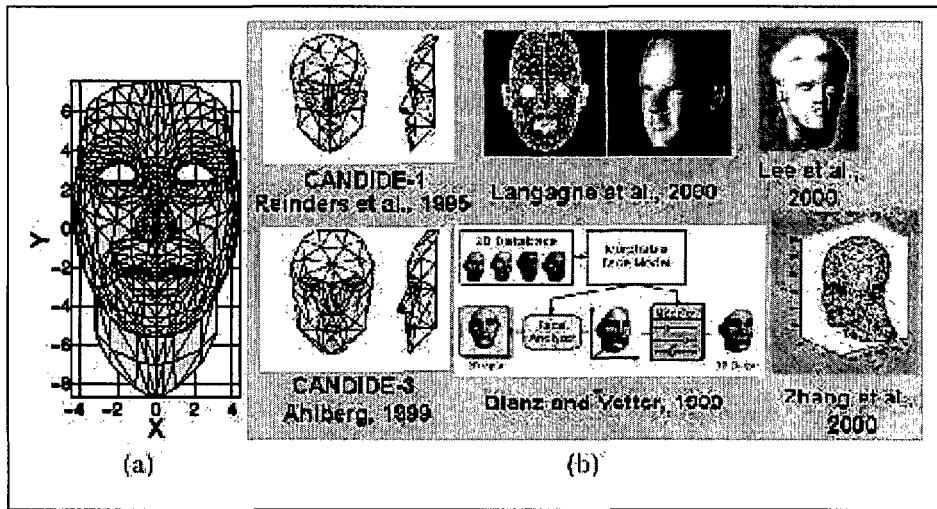


Figura 2.17. Modelos genéricos de cara. (a) Modelo de Water. (b) seis clases de modelos genéricos para representar la geometría facial.

Se analiza a continuación el trabajo de diversos autores que han incorporado información a priori sobre la geometría facial para obtener un modelo 3D de la cara. El modelo propuesto por Parke y Waters [Par96a] se basa en superficies poligonales. El modelo de Blanz y Vetter [Bla98] es un modelo estadístico que considera los componentes principales de forma y textura. Reinders et al. [Rei95] propone un modelo de alambres de la cara, orientado a su codificación, pero bastante menos elaborado que el de Waters. Yin et al. [Yin97] proponen un modelo basado en MPEG4 que usa puntos extraídos de una imagen frontal y otra de perfil. La extracción de puntos se basa en detección de bordes y en umbrales en los niveles de gris. De forma parecida, Lee et al. [Lee00] presentan un modelo para animación cuyas coordenadas iniciales son adquiridas por un escáner 3D o por medio de una imagen de frente y otra de perfil. Además, [Len00] y [Fua99] ajustan un modelo una cara humana a videos sin calibrar (utilizando un ajuste por mínimos cuadrados y por ajuste de haces (“bundle adjustment”)). Para la convergencia del método es esencial seleccionar manualmente cinco puntos característicos del rostro y estimar la posición inicial de la cámara. [Ahl01] y [Ahl01a] adapta un modelo de alambres 3D a secuencias de video (véase la Figura 2.14). [Hsu02] propone dos métodos de modelado (uno de ellos mostrado en la Figura 2.15), utilizando el modelo de Waters. El primer método usa un ajuste en 2.5D así como información acerca de la geometría de la cara, y en el segundo caso considera grafos semánticos adaptables con características orientadas al reconocimiento para obtener la geometría de un individuo.

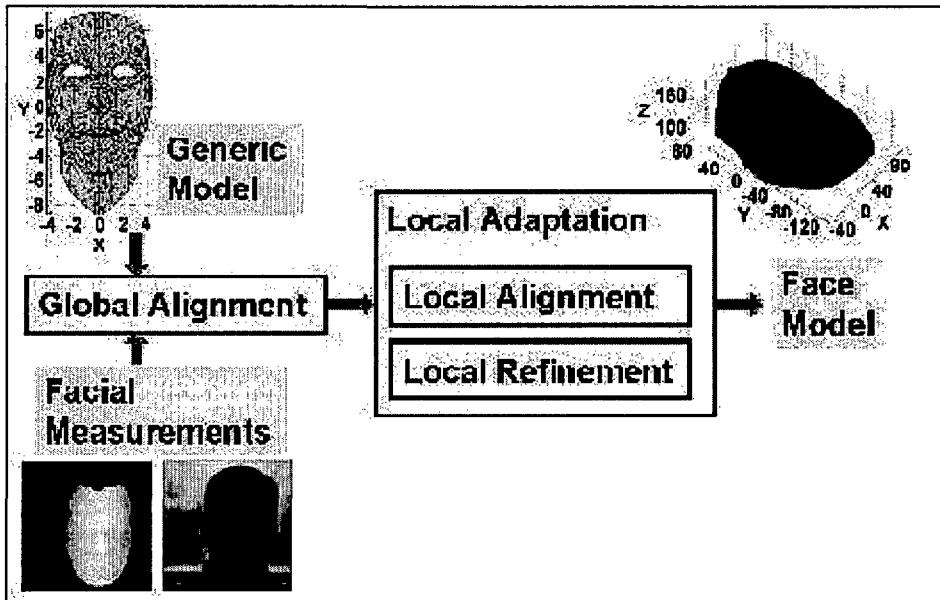


Figura 2.18. Modelo propuesto en [Hsu02].

2.2.10.- Otras técnicas.

En [Sam94] se utilizan modelos ocultos de Markov para codificar la información necesaria para reconocimiento. Un modelo oculto de Markov puede verse como una máquina de estado finito donde las transiciones entre los estados dependen de la ocurrencia de algún símbolo. Es un modelo probabilístico ya que se asocia a cada transición y a cada símbolo una probabilidad. En el caso de su aplicación a reconocimiento de caras es necesaria una selección de valores para un conjunto bastante grande de parámetros. El resultado obtenido es de un 87% de reconocimiento correcto para la base de datos ORL [Oli]. Incrementando la complejidad computacional con un modelo oculto de Markov pseudo-bidimensional, el rendimiento aumenta a un 95%. En [Eic00], se emplean modelos ocultos de Markov para reconocimiento a partir de imágenes comprimidas en formato JPEG obteniendo un 99.5% de reconocimiento correcto.

En [Tis00] se adopta un enfoque nuevo. Aquí la imagen es remuestreada en lugar de usar imágenes descritas en coordenadas cartesianas; seguidamente se pasa a coordenadas polares (situando los puntos en un plano de coordenadas polares planas con escala logarítmica, llamadas coordenadas log-pol), simulando la retina humana. Utilizando un subconjunto de la base de datos FERET, los resultados muestran un 97% de reconocimiento correcto en el mejor caso. En este trabajo se utiliza un par estéreo de cámaras, pero cambios en el tamaño, presencia/ausencia de gafas y otros factores disminuyen el porcentaje de reconocimiento corrector. En [Esc02], se comparan los resultados de EBGM y las imágenes transformadas al espacio log-polar (véase la Figura 2.19). Los resultados muestran que en la base de datos considerada (la de la Universidad de Yale), EBGM ofrece un 83.1% de reconocimiento correcto frente a un 88.93 de la transformada.

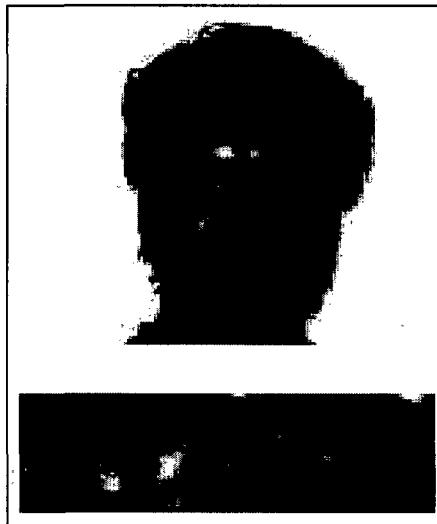


Figura 2.19. Imagen inicial e imagen transformada en log-polar (obtenida de [Esc02]).

Otras técnicas que también se han considerado en la bibliografía se basan en el uso de imágenes de baja resolución con clasificadores basados en vecindad. En [Bak00] se parte de imágenes de baja resolución para ampliarlas, interpolando los valores de los niveles de gris para obtener los nuevos puntos. No se ofrecen los resultados numéricos de reconocimiento, aunque se afirma que con un subconjunto de la base de datos FERET mejoran el porcentaje obtenido con otros sistemas (sin dar más explicaciones). T. Sim et al [Sim00] consideran imágenes de tamaño 16x16; con un subconjunto de la base de datos FERET obtienen un 96.2% de reconocimiento correcto. F. Goudail et al. [Gou96] presentan un resultado del 95% de reconocimiento correcto para una base de datos de 116 sujetos utilizando un algoritmo multiescala con correlación.

2.2.11.- Programas comerciales.

En este apartado se muestran los resultados obtenidos en las pruebas que se han realizado utilizando programas de reconocimiento o verificación de caras desarrollados por varias empresas. Se ha dividido en dos partes, una primera donde se describen algunos proyectos piloto de uso de programas comerciales, y una segunda parte en que se muestran los resultados obtenidos en el “Face Recognition Vendor Test 2002” en pruebas realizadas por varias organizaciones federales de EE.UU.

Pruebas piloto

Los resultados de las pruebas piloto son difíciles de obtener, ya que en muchos casos los resultados han sido tratados confidencialmente y enviados directamente a la empresa propietaria del programa. La mayoría de estas pruebas se desarrollaron en diversas partes de Estados Unidos a lo largo del año 2001, con resultados muy modestos.

La prueba que más difusión pública tuvo en los medios de comunicación social fue la realizada en Tampa (EE. UU.) en la Final de la Super Copa en Enero de 2001. Los únicos resultados públicos indican que se identificó correctamente a 19 delincuentes, de los que ninguno fue detenido ya que sólo tenían delitos menores. La prueba apareció como un éxito en los medios de difusión. El sistema probado fue “FaceIt” de Visionics Corporation (desde el 2002 forma parte de Identix Corporation). La policía de Tampa decidió usar el sistema en las calles de la ciudad. Según el informe

de los resultados, remitido al sindicato de derechos civiles americanos [ACLUa], se usó del 29 de Junio al 11 de Agosto del 2001. Ningún sujeto fue detenido basándose en los datos del sistema. Los resultados indican que, con un umbral de confianza muy bajo, el número de falsas alarmas era muy elevado (una media de cinco por la noche). Un umbral muy elevado disminuye el número de falsas alarmas, pero también disminuye el número de veces que el sistema reconoce a un sujeto. Además para operar el sistema, un agente debía enfocar la cara del sujeto de forma manual, ya que en caso contrario la resolución obtenida por una cámara panorámica es muy baja.

Igualmente, el Departamento de Defensa de Estados Unidos probó el mismo sistema “FaceIt” desde Julio a Octubre del 2001. [Kin02]. En esta prueba el sistema estuvo operativo durante 13 semanas, participando 270 sujetos con un total de 42.270 pruebas realizadas. El sistema funcionó correctamente en el 51% de los casos. Relajando las condiciones, se le permitió al sistema mostrar una lista de 10 sujetos, en este caso en un 81% de los casos el sistema mostró entre esos 10 sujetos a la persona correcta. El sistema “FaceIt” ha mejorado desde esta prueba, pero en esta fecha la compañía afirmaba que su porcentaje de error era 0.7 – 25 %.

Durante seis meses a lo largo del año 2002 se instaló el sistema “FaceKey” de Omron Corporation como sistema de verificación facial para acceder al reactor nuclear de la Universidad de Missouri-Rolla (EE.UU.). Los resultados fueron enviados a Omron Corporation y no se han hecho públicos [UMR] [WIR].

El sistema “Argus”, también de Visionics Corporation, fue probado en el aeropuerto de Palm Beach (Florida, EEUU) del 11 Marzo al 15 de Abril 2002. La base de datos constaba de fotos de 250 empleados, pero las pruebas se hicieron sólo con 15 de ellos. De 958 intentos, el sistema funcionó correctamente 455 veces (47.5% de aciertos). Además, cuando se dispuso el sistema en un mostrador de facturación, se obtuvo un promedio de unas 10.000 fotografías diarias con un porcentaje de falsas alarmas del 0.4%, lo cual supone unas 2-3 falsas alarmas a la hora. Además de en algunas notas de prensa (por ejemplo en [The02]), puede encontrarse el informe de los resultados en la página web del sindicato de derechos civiles americanos [ACLUb].

Face Recognition Vendor Test 2002.

El DARPA (Defense Advanced Research Project Agency: Agencia de proyectos de investigación avanzados para defensa) de Estados Unidos, ha realizado diferentes pruebas para probar la madurez de las técnicas de reconocimiento facial. Las primeras pruebas se desarrollaron con el nombre de FERET y tuvieron lugar en Agosto 1994, Marzo 1995 y Septiembre 1996 [Phi98] [Phi00]. La siguiente prueba, ya con el nombre FRVT (Face recognition vendor test: Prueba de reconocimiento facial con productos comerciales) se desarrolló en el año 2000 [Bla01], y la última prueba ha tenido lugar en el año 2002, publicada en Marzo 2003 [Phi03]. A continuación se muestran los resultados obtenidos en este último caso.

En la prueba FRVT 2002 se evaluaron diez sistemas comerciales. Se llevaron a cabo dos tipos de pruebas. En primer lugar se realizó la prueba de alta intensidad de cálculo (HCInt: High Computational Intensive test), desarrollada con una base de datos de 121.589 imágenes de 37.437 personas. De cada sujeto se tienen al menos tres imágenes, todas ellas de alta calidad y con el fondo uniforme. La segunda, denominada

prueba de media intensidad de cálculo (MCInt:: Médium Computational Intensive test) se desarrolló con una base de datos heterogénea de imágenes fijas y secuencias de vídeo con sujetos en condiciones de iluminación, actividad (gestos) o posición muy variada. Hay imágenes de interior y de exterior, existiendo imágenes de sujetos que se obtuvieron con una diferencia de tres años. En este caso, el número de imágenes y de sujetos son muy inferiores.

Las conclusiones obtenidas de este estudio muestran que:

- Para imágenes tomadas en el interior de un edificio, los cambios de iluminación no afectan significativamente el rendimiento de los sistemas evaluados (se mantiene un 90% de verificación correcta cuando el porcentaje de falsa aceptación es de 1%).
- Sin embargo con imágenes tomadas en el exterior, los mejores sistemas de reconocimiento facial comerciales ofrecen unos resultados de un 50% de acierto cuando el porcentaje de falsa aceptación es de 1%.

Una de las conclusiones alcanzadas en el ensayo FRVT 2002, fue que para los mejores sistemas de reconocimiento facial, el rendimiento cae un 5% por cada año de separación entre las imágenes de entrenamiento y las de prueba.

En FRVT 2002, ha sido la primera vez que se ha podido evaluar el efecto del tamaño de la base de datos sobre el rendimiento, cuando el número de imágenes es grande. La mejor identificación fue de un 85% con 800 sujetos, 83% con 1600 y 73% con 37.437; es decir, cada vez que se dobla el tamaño de la base de datos, el porcentaje de reconocimiento baja dos o tres puntos porcentuales.

Otro experimento consistió en mantener una lista de sujetos que se quería identificar. Si la lista es de 25 sujetos, con un porcentaje de falsa aceptación del 1%, los mejores sistemas comerciales ofrecen un porcentaje de aciertos del 77%. Si la lista aumenta a 300 sujetos, con el mismo porcentaje de falsa aceptación, el porcentaje de reconocimiento es de un 69%.

Además se observó que el porcentaje de reconocimiento en hombres es un 6-9% mayor que en mujeres. También el reconocimiento de personas de edad es mayor que en jóvenes; de media, de 18 a 22 años el porcentaje de reconocimiento es de un 62%, de 38 a 42 años se obtiene un 74%. Aproximadamente, por cada diez años de incremento en edad, el porcentaje de reconocimiento aumenta un 5%, hasta los 63 años.

Dos nuevas pruebas fueron incluidas en FRVT 2002:

- El uso de modelos 3D deformables [Bla98] para ayudar en el reconocimiento de imágenes que no son frontales. Los resultados mostraron que esta técnica mejora los resultados.
- El uso de secuencias de vídeo y no sólo de imágenes fijas. Los resultados no muestran mejoras.

Las principales conclusiones extraídas en FRVT 2002 fueron:

- Con una iluminación controlada, el estado del arte actual es de un 90% de verificación correcta, con un porcentaje de falsa aceptación de un 1%.
- El uso de modelos deformables mejora el porcentaje de verificación si el sujeto no está de frente a la cámara.
- El rendimiento del sistema mejora si se usan listas pequeñas con sujetos a reconocer.
- Características como la edad y el sexo influyen significativamente sobre el porcentaje de reconocimiento.

La Figura 2.20 muestra los resultados de verificación en la prueba HCInt para los ocho sistemas, con un porcentaje de falsa aceptación de un 1%.

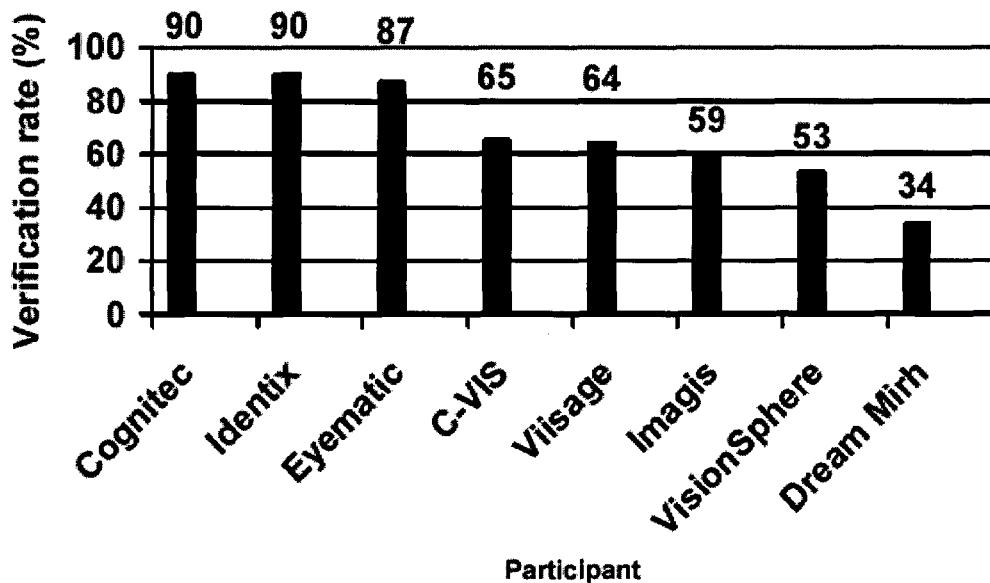


Figura 2.20. Resultados de la prueba HCInt.

Para los tres sistemas con mejores resultados en el caso anterior, la Figura 2.21 muestra la variación en el porcentaje de reconocimiento correcto en función del porcentaje de falsa aceptación.

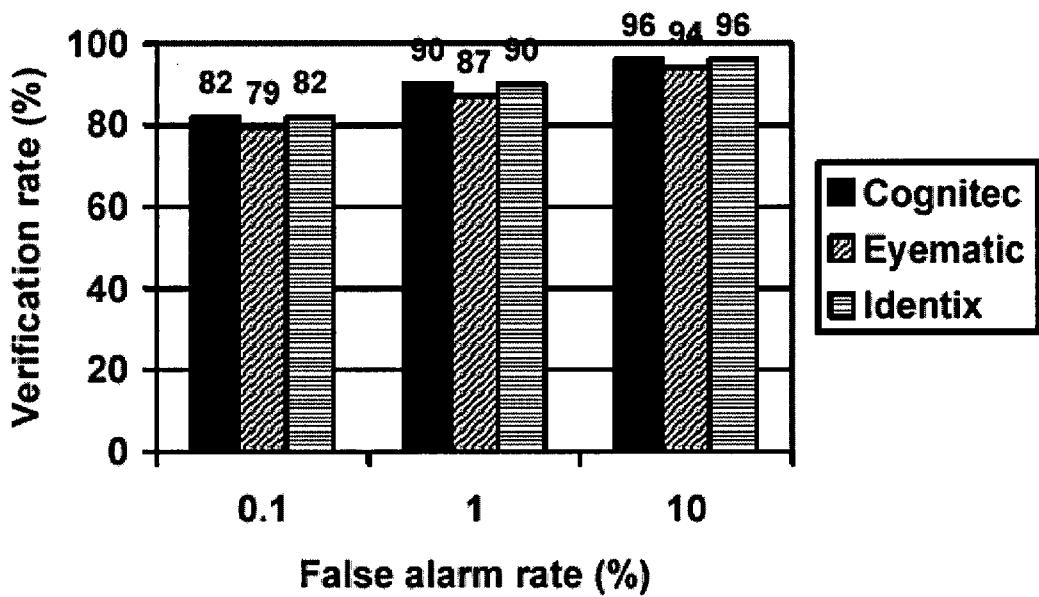


Figura 2.21. Porcentaje de reconocimiento en función del porcentaje de falsas alarmas.

La Figura 2.22 muestra los resultados de la prueba HCInt para los 8 participantes, representando el porcentaje de reconocimiento correcto en función del porcentaje de falsa aceptación. La escala en el porcentaje de falsa aceptación es logarítmica. Puede observarse que un porcentaje de falsa aceptación de 1 representa un porcentaje de 100%, es decir, el sistema deja entrar a todos los sujetos (por lo tanto también se consigue un porcentaje de reconocimiento correcto del 100%).

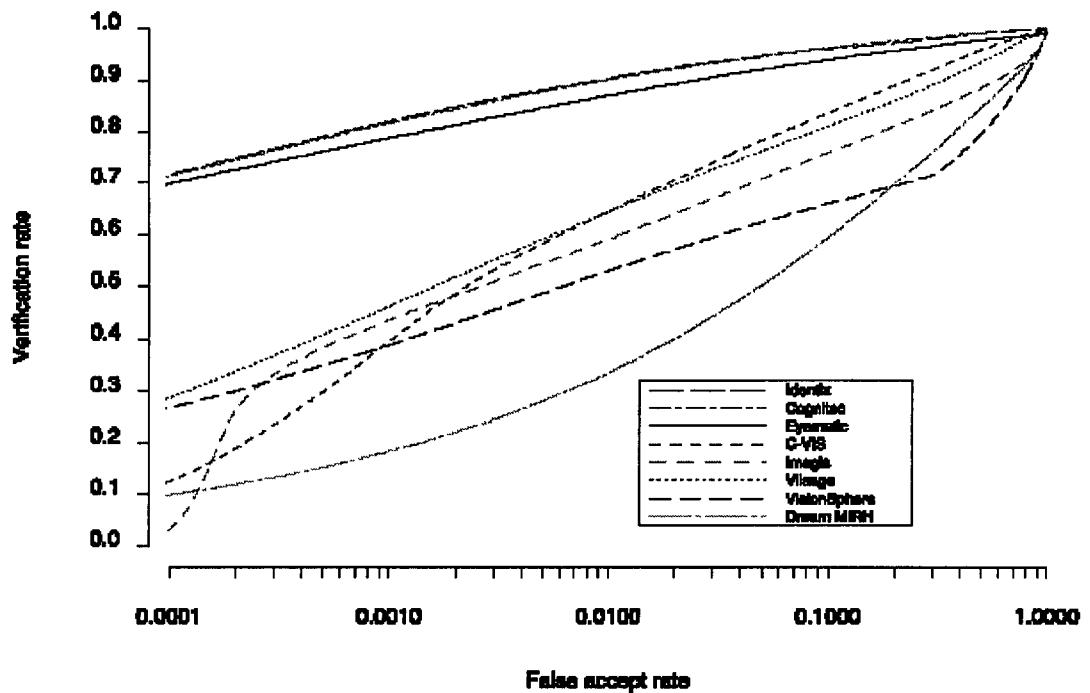


Figura 2.22. Reconocimiento correcto en función del porcentaje de falsa aceptación.

Los resultados de la prueba MCInt se muestran en las Figuras 2.23 y 2.24. La Figura 2.23 se obtuvo manteniendo un porcentaje de falsa aceptación del 0.01%. La Figura 2.24 corresponde a la misma prueba, pero sin tener en cuenta el porcentaje de falsa aceptación. Los resultados demuestran que la mayoría de los sistemas en escenas

de interior obtienen resultados aceptables independientemente del tipo de iluminación o de una pequeña diferencia de días, pero en cuanto las escenas son de exterior, los resultados empeoran mucho.

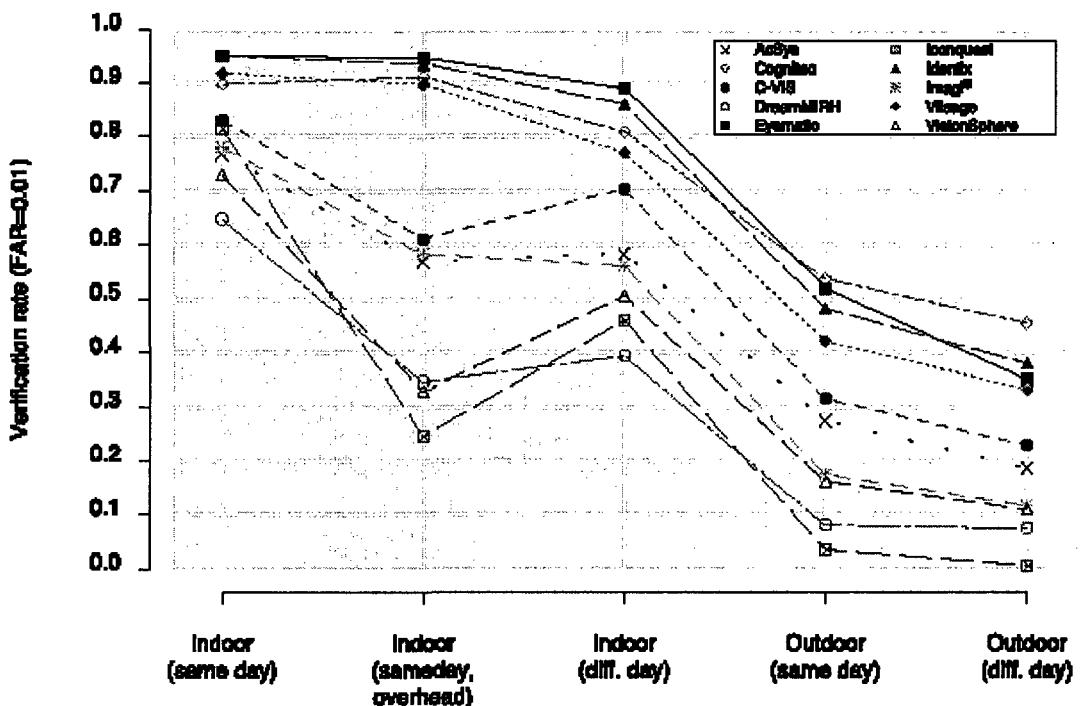


Fig 2.23. Resultados MCInt con FAR=0.01.

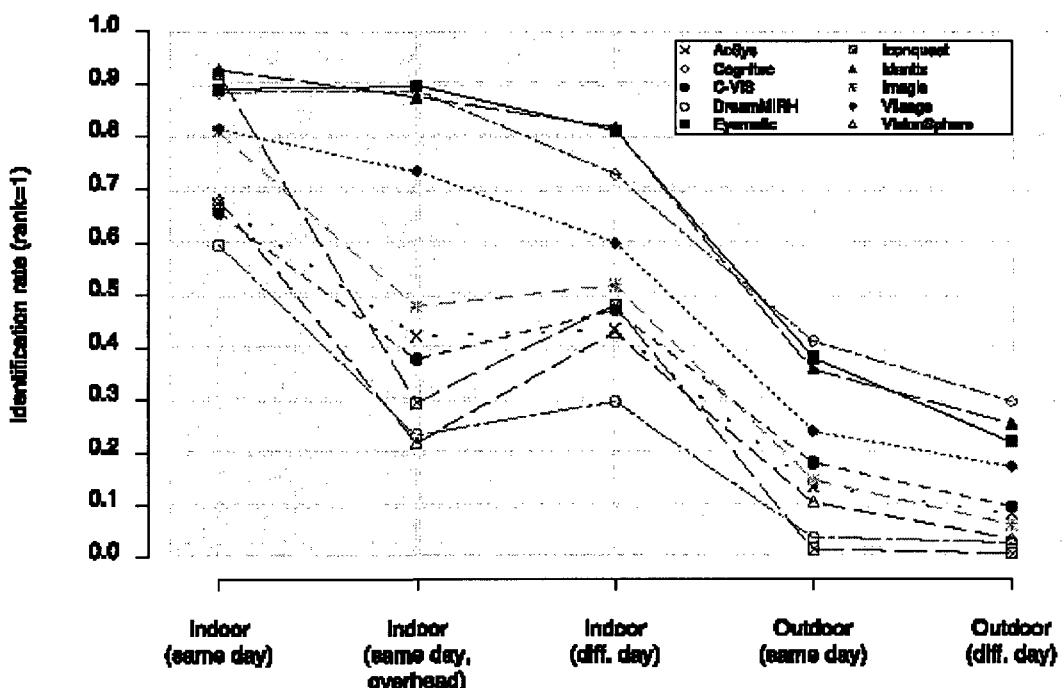


Fig 2.24. Resultados MCInt, independientemente del FAR.

Capítulo 3

Entorno de la tesis.

3.1. - Introducción.

En la realización de esta tesis se han empleado un conjunto de elementos tanto físicos como lógicos que se describen en el presente capítulo. Los elementos físicos descritos son aquellos que pueden ser de más interés: por un lado, la plantilla de calibración y el sistema de visión estéreo para el sistema de adquisición de datos tridimensionales, y por otro el equipamiento utilizado para disponer de un entorno controlado en la adquisición de datos bidimensionales. En cuanto a los elementos lógicos, se analizarán con más detalle las bases de datos utilizadas y, al describir la plantilla de calibración, se comentará el programa de calibración y reconstrucción tridimensional; el resto de elementos lógicos son los habituales en programación y visión artificial.

Se ha de tener en cuenta que ha sido necesario instalar dos equipos, cada uno de ellos orientado a la creación de una base de datos específica. Por lo tanto, a continuación se describirá cada equipo técnico, seguido de la mención y características de la base de datos con él obtenida.

El primer caso es un equipo orientado a la obtención de datos tridimensionales utilizando un par estéreo. La marcación de puntos va a ser manual, por lo que la iluminación es menos crítica que en el segundo caso. Este equipo fue instalado en el Departamento de Informática de la Universidad de Salamanca, desarrollándose también allí la adquisición de datos tridimensionales.

El segundo equipo está orientado a la obtención de datos de niveles de gris y su posterior procesamiento por medio de PCA. En este caso, la iluminación es uno de los elementos críticos. Este equipo fue instalado en el Departamento de Informática, Estadística y Telemática de la Universidad Rey Juan Carlos, desarrollándose en esta ubicación la adquisición de datos.

3.2. - Sistema de visión estéreo.

El sistema de visión constaba de dos cámaras CCD de blanco y negro junto con una tarjeta digitalizadora, trabajando en un ordenador personal de propósito general. Es un equipo muy versátil y de bajo coste, que ofrece resolución suficiente para el objetivo que se busca, así como flexibilidad para ser aplicado en otros proyectos de visión artificial.

Las cámaras CCD pertenecen a la marca Sony, con lentes comerciales de montura C. En el trabajo desarrollado se encontraban montadas sobre trípodes independientes. La separación entre trípodes era de aproximadamente un metro y el sujeto se colocaba a una distancia aproximada de unos dos metros y medio (Figura 3.1).

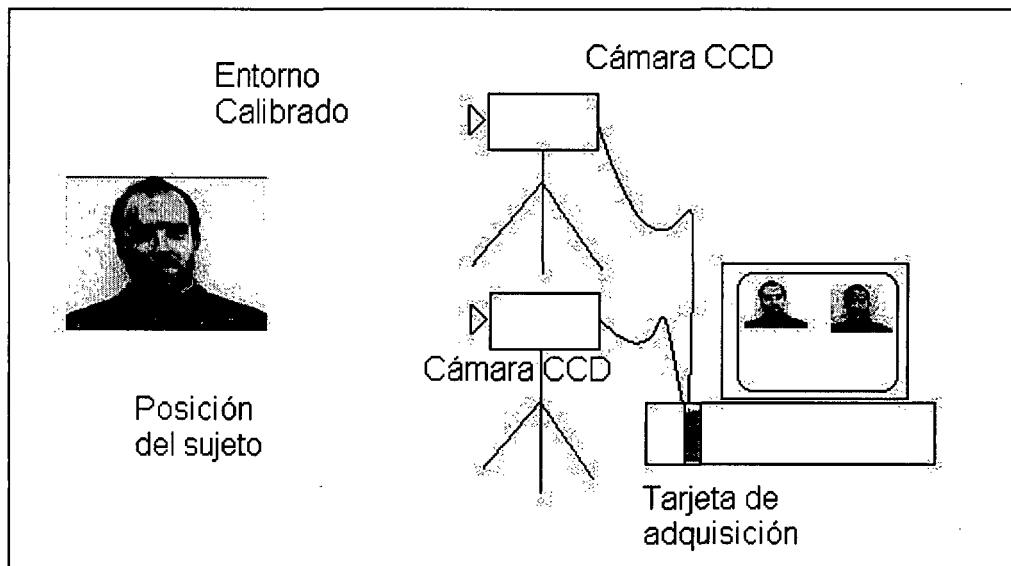


Figura 3.1. Esquema de adquisición de imágenes en estéreo.

La tarjeta digitalizadora es un modelo comercial, cuyo único requisito fue que permitiese la adquisición de imágenes en niveles de gris de dos cámaras de forma casi instantánea. La tarjeta utilizada en este trabajo permitía la adquisición de tres imágenes de blanco y negro o de una en color. La tarjeta utilizada toma la imagen de una cámara y a continuación la de la segunda cámara, necesitando unos milisegundos para hacer el cambio. Las imágenes utilizadas para obtener datos tridimensionales son todas en blanco y negro (con 256 niveles de gris).

3.2.1. - Plantilla de calibración. Algoritmo de calibración.

Para realizar una reconstrucción tridimensional de un objeto con un sistema de visión estéreo es conveniente disponer de una plantilla de calibración. Es decir, de un sistema en el cual se conozcan las coordenadas tridimensionales de una nube de puntos. Al observarlos desde el sistema estéreo es posible calcular los parámetros de cada cámara (tanto intrínsecos como extrínsecos), de tal forma que, conocidas las coordenadas tridimensionales y las coordenadas de los puntos en la imagen, sea posible realizar la calibración de las cámaras.

Para poder utilizar una plantilla de calibración de forma sencilla, ésta debe tener puntos en, al menos, dos planos distintos. Aunque es posible calibrar una cámara utilizando unos pocos puntos de coordenadas conocidas, estén o no en un único plano (por ejemplo [Zha02]), el coste computacional es mayor.

La plantilla utilizada en esta tesis (ver Figura 3.2) consta de dos bloques de acero soldados sólidamente entre sí con un conjunto de puntos marcados en una de sus caras. En concreto, dispone de 193 puntos (49 en la cara superior y 144 puntos en la cara inferior). Las caras visibles fueron pintadas en blanco y los puntos en negro. Cada punto consiste en un círculo de 6 milímetros de diámetro y se encuentra separado 20 mm de sus vecinos. La cara inferior es un cuadrado de 30 cm de lado y la superior es también cuadrada, pero de 15 cm de lado. Ambas están separadas una altura de 3 cm. Para calcular de forma precisa las coordenadas tridimensionales de los puntos, se utilizó un digitalizador tridimensional manual. Para obtener de forma precisa las coordenadas de los puntos en dos dimensiones, se procedió a calcular el centro de gravedad de cada uno de los puntos proyectados en la imagen. El diseño de esta plantilla de calibración fue realizado por el autor de este trabajo.

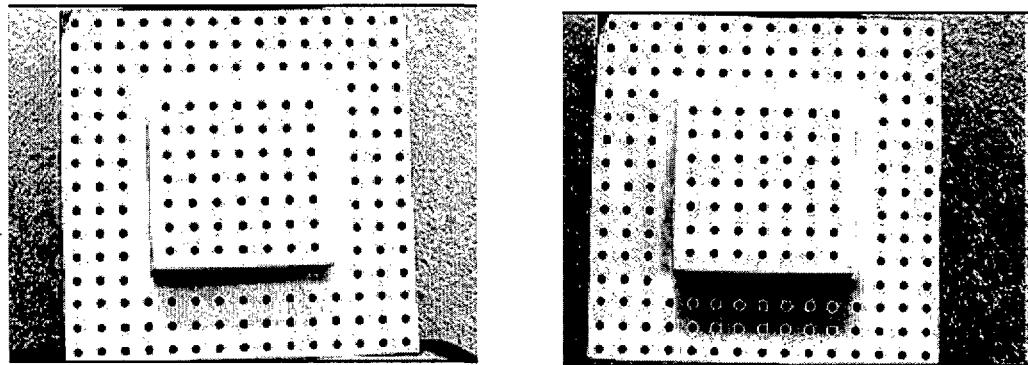


Figura 3.2. Plantilla de calibración (vista estéreo).

3.3.- Base de datos para adquisición de información tridimensional. Universidad de Salamanca.

Para alcanzar uno de los objetivos marcados en esta tesis, ha sido necesario obtener información tridimensional de los sujetos filmados por las cámaras. Se procedió a crear un sistema estéreo y a su calibrado posterior. Con ello se pudo crear una base de datos local, de la cual es posible obtener información tridimensional. Esta base de datos está formada por 20 sujetos, con 8 fotos por sujeto (que corresponden a 4 pares estéreo). Las imágenes están calibradas, es decir, es posible obtener la reconstrucción tridimensional de una nube de puntos manualmente marcados en ambas imágenes. La Figura 3.3 muestra las ocho fotos de un sujeto de la base de datos; pueden observarse las que definen un par estéreo.

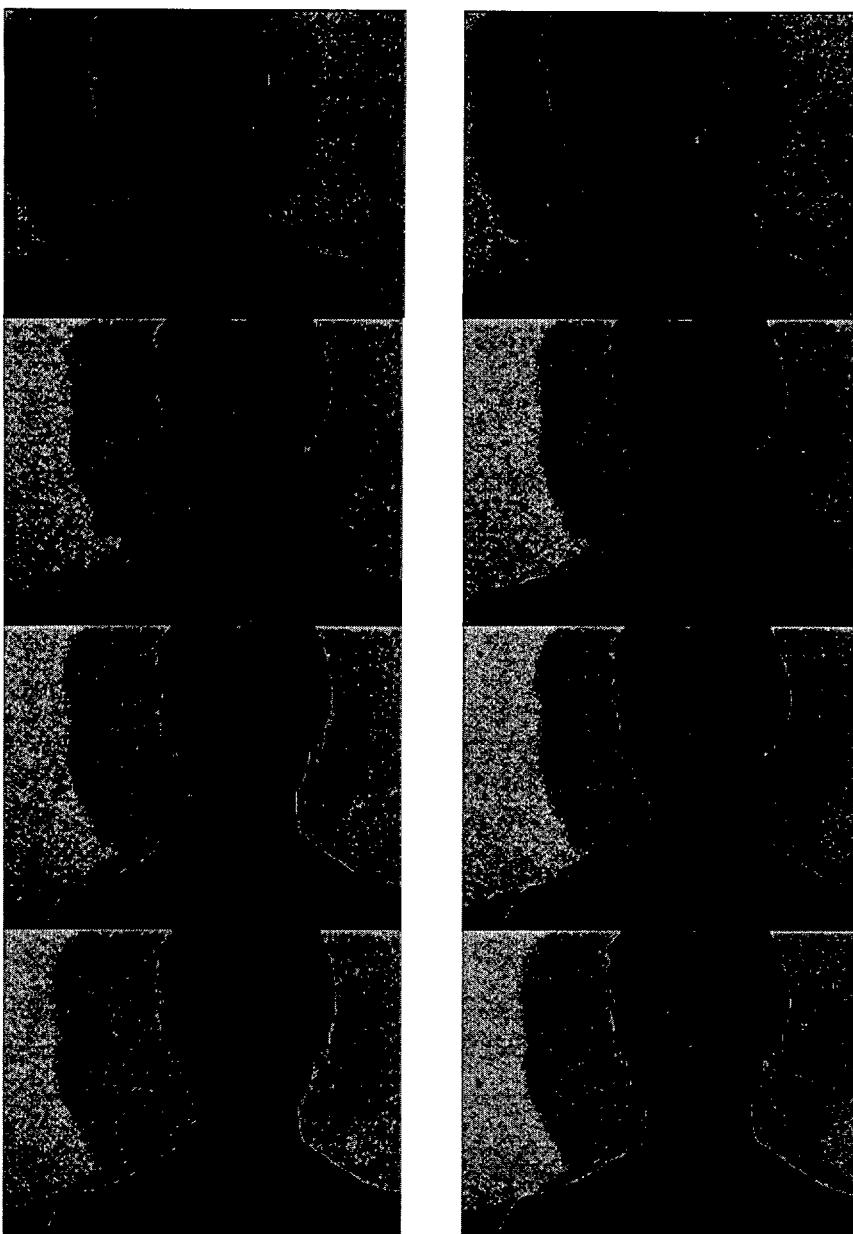


Figura 3.3. Ocho imágenes de un sujeto, correspondientes a cuatro sesiones.

De los veinte sujetos que constituyen la base de datos, la mitad son hombres y la mitad son mujeres (alguno de ellos pueden verse en la Figura 3.4). No se tuvo especial cuidado en mantener la iluminación ni la posición controlada, permitiendo variaciones que pudieran corresponder a un entorno lo más “real” posible. Hay que tener presente que esta base de datos sólo iba a ser útil para ofrecer información tridimensional, por lo tanto la iluminación no era relevante en el resultado final. Además la marcación de puntos fue realizada manualmente.

El tamaño de las imágenes es en este caso de 640 x 480 píxeles con 256 niveles de gris con un fondo de imagen de color blanco.



Figura 3.4. Algunos de los sujetos de la base de datos.

Además se completó con un pequeño conjunto de imágenes adicionales de cuatro sujetos (dos hombres y dos mujeres) que ya se encontraban en la base de datos anterior. A cada uno de ellos se les realizaron cuatro fotos con una expresión que no fuera neutra (sonrisa o ligera apertura bucal), de tal forma que el reconocimiento fuera más difícil. Este conjunto constituye una base de datos de sujetos no colaborativos y puede observarse un ejemplo en la Figura 3.5.

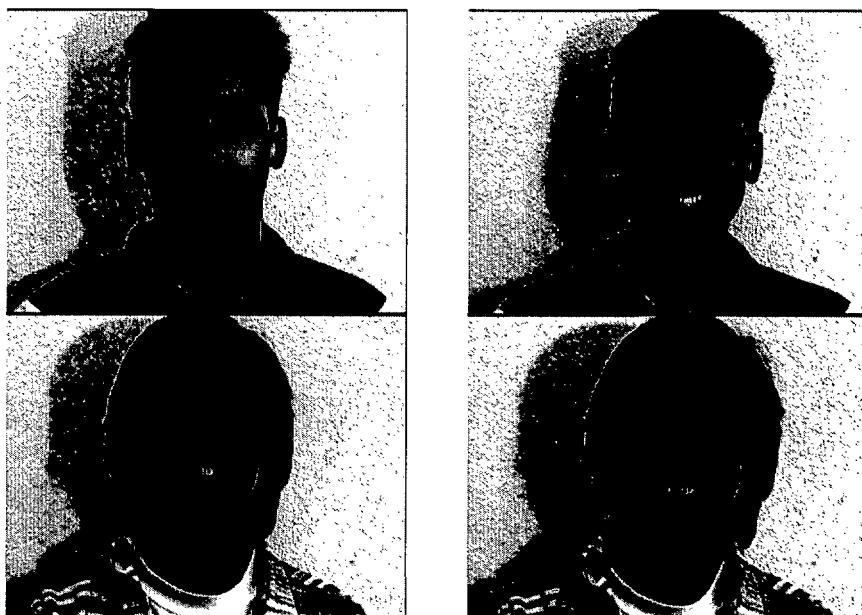


Figura 3.5. Dos sujetos en actitud no colaborativa (cambios en la expresión).

3.4.- Sistema de adquisición de imágenes bidimensionales.

En cualquier sistema de visión por ordenador, el proceso de adquisición de imágenes es de capital importancia, afectando grandemente al rendimiento global del

sistema. En un sistema de verificación facial, la necesidad de un proceso de adquisición de imágenes muy bien diseñado es incluso más imperiosa, ya que se busca que pequeños cambios de orientación de la cara afecten mínimamente a la imagen. Hay que considerar que si los algoritmos que se van a emplear para el procesamiento de las imágenes se basan en los niveles de gris, como es el caso de PCA, entonces las condiciones en las cuales se adquieran las imágenes son críticas. En esta tesis, era aconsejable el uso de luz difusa para minimizar las sombras que se proyectan en la cara, ya que éstas dependen particularmente de la expresión y de la pose. El sistema de adquisición de imágenes se muestra en las Figuras 3.6 y 3.7. Este consistía en dos fuentes de luz difusa colocadas a ambos lados de la cámara, una cámara CCD con focal de 50mm y un fondo oscuro para minimizar las posibles sombras. La tesis sólo considera la luz que proviene de los focos, en otras palabras consiste en luz difusa. Se debe tener en cuenta que una iluminación correcta del objeto permite minimizar el tiempo de procesamiento. La adquisición de imágenes de forma correcta es una etapa que influye extraordinariamente en el resultado final.

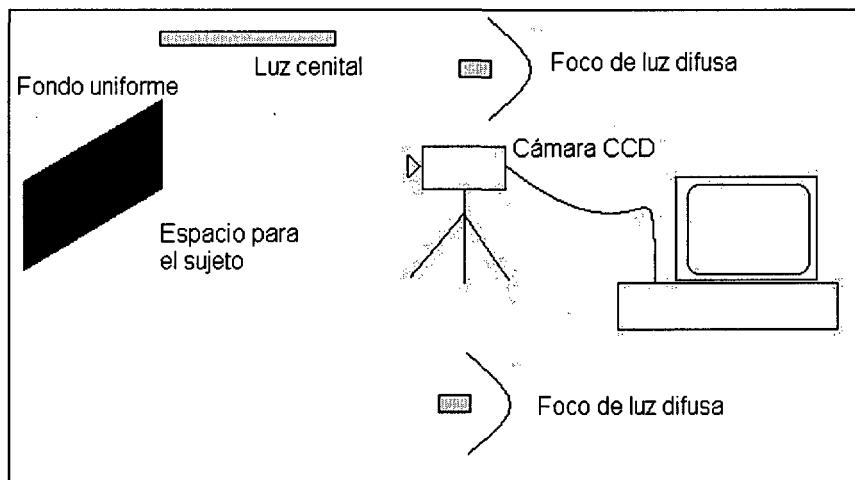


Figura 3.6. Esquema de adquisición de imágenes con luz controlada.



Figura 3.7. Entorno de trabajo con los dos focos de luz difusa y la cámara CCD.

Las Figuras 3.9 y 3.10 muestran dos imágenes del mismo sujeto enfrente de la cámara. En el segundo caso, se utilizó luz cenital (proveniente de unos tubos fluorescentes situados en el techo de la habitación) y como puede verse aparecen en el sujeto brillos y sombras dependientes de la posición y orientación de la cara respecto a las fuentes de luz y que pueden influir de manera importante en la información procesada posteriormente. En el primer caso sólo se consideró luz difusa. Fácilmente se observa el efecto de la luz difusa

evitando sombras proyectadas: la sonrisa del sujeto de la imagen en la Figura 3.8 crea importantes sombras alrededor de la boca y la nariz que pueden dificultar el proceso de verificación.



Figura 3.8. Sombras faciales debidas a la iluminación cenital.



Figura 3.9. Imagen iluminada con luz difusa

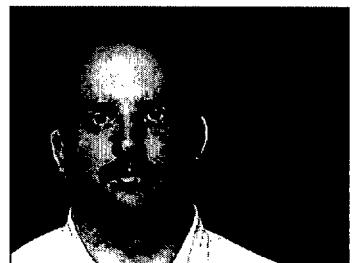


Figura 3.10. Imagen iluminada con luz no controlada

Para minimizar las distorsiones generadas por cambios en la focal de la lente o en la distancia entre la cámara y el sujeto, es aconsejable mantener ambos fijos en el entorno de trabajo. Esto requiere el uso de lentes sin zoom, así como el mantenimiento de posiciones fijas para la cámara y para la persona que está siendo identificada. Estos requisitos se cumplen fácilmente en cualquier entorno comercial en el que un sistema de verificación vaya a ser utilizado. En nuestro caso se mantuvo una disposición que cumpliera esta condición.

3.5.- Base de datos para adquisición de información bidimensional. Universidad Rey Juan Carlos.

En este caso la base de datos está formada por 29 sujetos (22 hombres y 7 mujeres), de los cuales se tomaron 12 imágenes: 8 imágenes se utilizaron para el entrenamiento y 4 para la prueba. El tamaño de la imagen consta de 320 x 240 píxeles, con la cara ocupando la mayor parte de la imagen. Las imágenes de cada sujeto se tomaron en diferentes sesiones a lo largo de varios días, de modo que aparecen pequeños cambios de pose, gesto y orientación. Las imágenes se adquirieron en color, aunque todo el procesamiento se va a realizar con imágenes en blanco y negro (con 256 niveles de gris). El fondo de todas las imágenes en este caso es azul oscuro. Los sujetos de la base de datos aparecen en la Figura 3.11.

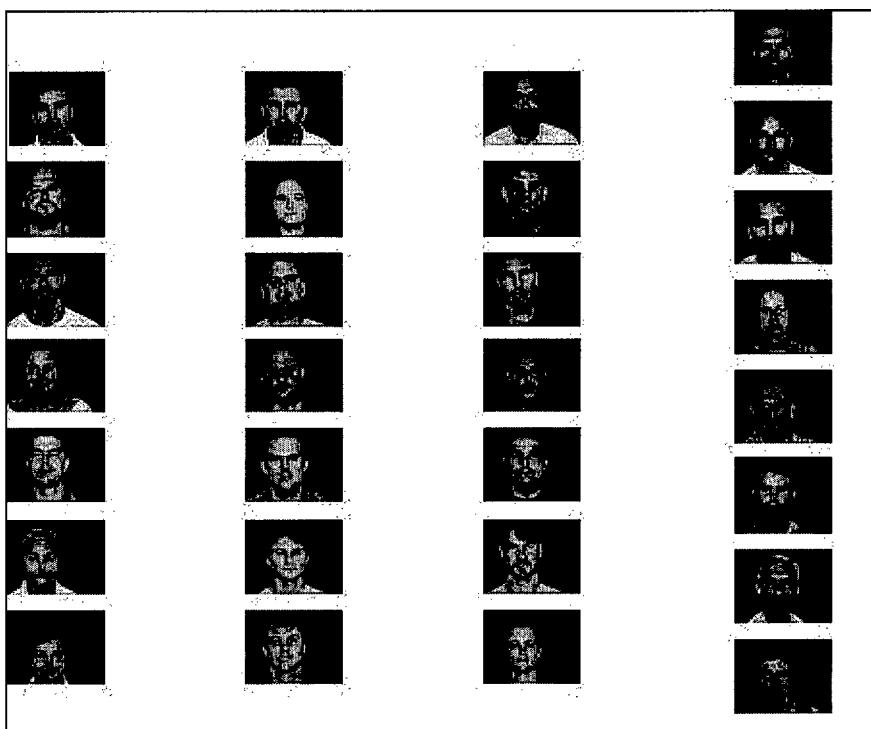


Figura 3.11. Sujetos de la base de datos.

Capítulo 4

Reconocimiento bidimensional de caras.

4.1.- Introducción.

En el presente capítulo se describen las técnicas desarrolladas y los resultados obtenidos en la verificación de caras humanas, tomando como datos de entrada una imagen de un sujeto y un número de identificación personal (PIN). El proceso de verificación se ha dividido en dos etapas: una primera etapa de extracción de información de la imagen y una segunda etapa en la que se procede a la clasificación de la información previamente extraída. La primera etapa, también llamada pre-procesamiento, constituye una fase rápida cuyos datos de salida consisten en un conjunto de números que contienen la información necesaria para identificar la cara que aparece en imagen. Se han realizado dos tipos de pre-procesamiento distintos (descritos en el apartado 4.2). Ambos tienen una etapa común inicial, denominada “método de localización de la cara en la imagen”. En el apartado 4.2.1 se describen otros estudios iniciales que se realizaron hasta adoptar la localización utilizada. Los dos preprocesamientos considerados difieren en el método de reducción de información. En el primer caso, que se describe en el apartado 4.2.2, se emplea el Análisis de Componentes Principales (PCA) como herramienta para obtener un número reducido de datos que representan una cara. En el segundo caso, descrito en el apartado 4.2.3, se considera, como método de reducción de información, la obtención de una “pirámide” de datos en que cada nivel se obtiene del anterior mediante un proceso de filtrado y submuestreo, eliminando filas y columnas entrelazadas. Posteriormente se estudia sólo el nivel de la pirámide de más baja resolución (método de baja resolución).

Una vez conseguido un conjunto de números que representan la cara humana que aparece en una imagen, la siguiente etapa es la clasificación (también llamada procesamiento). Para que un sistema sea capaz de clasificar correctamente, previamente debe haberse sometido a un proceso de entrenamiento. En esta fase se le ofrecen al sistema un conjunto de datos etiquetados, para que el sistema sea capaz de extraer las características que definen cada clase de datos a partir de ellos. Este tipo de entrenamiento se conoce como “supervisado” y es el que se ha empleado a lo largo de la presente tesis. En el caso de la verificación, surgen problemas del tipo siguiente: o bien

es el sujeto o bien no lo es. Para cada sujeto se habrá definido por lo tanto un modelo. Al introducir el PIN, se selecciona el modelo correspondiente y, al aplicarlo a los valores obtenidos del preprocesamiento, se indica si es el sujeto o no lo es. Para obtener un sistema así definido, previamente es necesario conseguir un modelo para cada sujeto de la base de datos. Dependiendo del clasificador considerado (KNN, RBF o SVM), existen tres posibles tipos de procesamiento.

A continuación se describen los experimentos realizados en el preprocesamiento de las imágenes (localización de la cara, método de análisis de componentes principales y preprocesamiento con imágenes de baja resolución). Las características de los clasificadores utilizados y los resultados obtenidos así como las conclusiones terminan este capítulo.

4.2.- Preprocesamiento de imágenes

En este apartado se muestran las distintas técnicas de preprocesamiento que se han ido utilizando a lo largo de esta tesis. Inicialmente se muestran las técnicas probadas de localización de la cara en la imagen. Una vez localizada la cara se le somete a uno de los dos tipos de preprocesamiento que se muestran a continuación: análisis de componentes principales o baja resolución. Cada uno de ellos se muestra a continuación.

4.2.1.- Localización de la cara.

Las imágenes adquiridas por el sistema de visión artificial (descrito en el capítulo 3) contienen la cara del sujeto con un campo de visión relativamente grande (ya que así es posible disponer de un entorno relativamente amplio en el que el sujeto pueda situarse). En este caso, la imagen no va a estar limitada sólo a la cara, sino que también aparecerá parte del fondo, hombros y cuello. Por otra parte, la entrada de datos al sistema de verificación facial y a los clasificadores debería ser una ventana que contenga sólo la cara, normalizada en posición y tamaño. Va a ser necesario, por lo tanto, realizar un proceso de localización facial antes de comprobar la identidad del sujeto.

En la presente tesis se desarrollaron tres sistemas de localización facial como herramienta para la verificación. Los tres fueron analizados con una base de datos inicial, para determinar el método de localización con el cual desarrollar el estudio. El primero de ellos consiste en un recorte manual de la zona en la que se encuentra la cara. El segundo de ellos se basa en la detección de los puntos con el color de piel humana. El tercero utiliza una correlación con una plantilla. El primer método sirvió para obtener una referencia frente a los otros dos. Los dos últimos métodos pueden combinarse, utilizando el color de la piel para una primera aproximación y refinando a continuación la localización utilizando una correlación con una plantilla.

Durante el proceso de localización facial se realizan otras dos operaciones: ecualización de los niveles de gris y normalización del tamaño de la ventana. Por lo tanto, los datos de salida de esta etapa son una ventana recortada a un tamaño 130x140, que contiene la cara con niveles de gris normalizados. La ecualización del histograma obtiene una distribución más uniforme de los niveles de gris a lo largo de la imagen. Este paso disminuye los cambios debidos a variaciones en las condiciones de iluminación de la escena.

Localización Manual.

El procedimiento de recorte manual sirvió para conseguir una primera aproximación del rendimiento de los algoritmos de verificación y como referencia frente a los otros métodos de localización. En este caso, una ventana rectangular se desplazaba manualmente por un operador hasta seleccionar la zona de la imagen inicial que se recortaba. Un proceso similar fue utilizado en [Rom96] como estimación de la posición de la cara.

La Figura 4.1 muestra parte de una base de datos (una imagen por sujeto) después de un recorte manual.

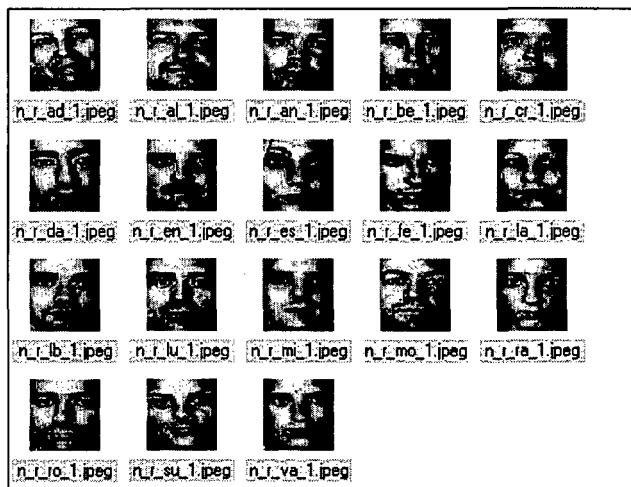


Figura 4.1. Recorte manual de una base de datos.

Localización utilizando el color de la piel.

Este detector utiliza el color de la piel como hipótesis inicial para localizar la cara. Este método es, en general, muy utilizado en detección y seguimiento de caras humanas. Estudios sobre la crominancia de la piel humana en diversas condiciones de iluminación se encuentran en [Ter00] y [Sto00]. La representación de color RGB no es adecuada para detectar características como el color de la piel, por lo tanto se realizó la transformación de la imagen en HSV. Es en este espacio en el que se realiza la detección del color de la piel. Sin embargo, diferentes imágenes de un mismo sujeto pueden presentar cantidades diversas de piel, lo cual produce imprecisiones en la detección. La Figura 4.2 muestra una parte de la base de datos después de la operación de recorte considerando la información de color en el espacio HSV. Los resultados indican que, aunque la cara está correctamente detectada y recortada, hay una gran variación en la posición.

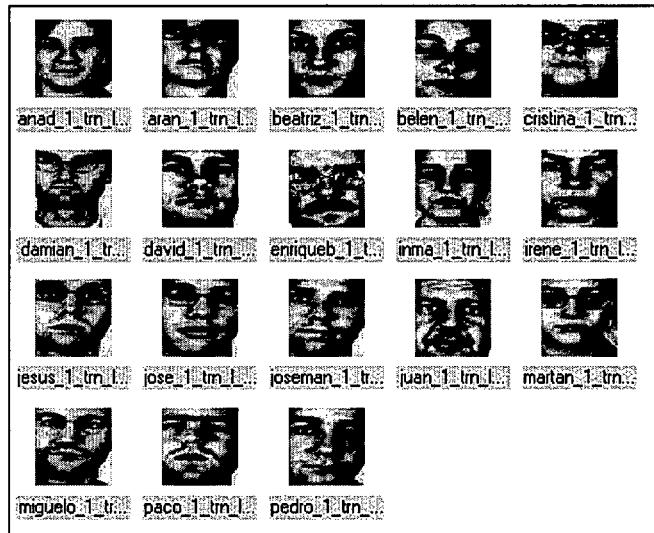


Figura 4.2. Recorte utilizando el color de la piel de una base de datos.

Localización por medio de convolución con una plantilla.

Una vez que la cara se ha detectado en escena y se conoce su posición aproximada, se puede refinar el procedimiento de localización por medio de la convolución con una plantilla. En nuestro caso, la convolución se realiza con la imagen de una cara. La convolución se lleva a cabo sólo en una zona de la imagen, con el fin de acelerar el proceso. A la imagen se le resta una foto previa que sólo contiene el fondo, computado dinámicamente durante la explotación del sistema. Se fija un umbral, y la zona en que la diferencia de niveles de gris sea superior al umbral indica la zona de la imagen en la que puede encontrar la cara y, por lo tanto, el punto inicial para realizar la convolución. Una suposición implícita en este proceso consiste en que la cara es frontal (o casi) y que está en posición derecha. Los datos de salida del proceso de localización, de nuevo componen una imagen de tamaño estándar (130x140 píxeles) conteniendo la cara del sujeto. Para maximizar la robustez del proceso, es preferible utilizar un fondo controlado en el cual los objetos queden bien definidos. En cualquier caso, este planteamiento no supone un impedimento para su posible explotación comercial. La Figura 4.3 muestra parte de la base de datos, después de su recorte por este método. La Figura muestra que la estabilidad obtenida mediante esta técnica es mayor que con las otras aproximaciones consideradas.



Figura 4.3. Recorte automático de una base de datos.

El análisis de los resultados muestra que la localización manual permite pequeñas variaciones en la localización del sujeto. La localización basada en el color de la piel es bastante robusta frente a cambios de iluminación y muy robusta frente a cambios de pose del sujeto. Pero es poco adecuada como datos de entrada de una etapa de verificación, ya que al recortar una base de datos, aparecen variaciones acusadas en la localización. El recorte utilizando una convolución con una plantilla es el que mejor resultado práctico ha obtenido. Recorta de forma muy precisa una base de datos, siendo posible realizar la verificación con los elementos recortados.

4.2.2.- Niveles de gris utilizando análisis de componentes principales.

Una vez seleccionado el preprocesamiento mediante el método de análisis de componentes principales (PCA), uno de los elementos que más influyen es lograr una localización facial muy precisa, que permita obtener imágenes de la cara con una posición definida [Tur91][Rom96]. En un laboratorio, las condiciones de iluminación y de posición del sujeto pueden ser controladas; en un entorno real, la verificación de sujetos va a realizarse con la colaboración del mismo, por lo que el rango de variación de estos elementos puede estar acotado.

Tras haber realizado el recorte de la base de datos, se calcula la matriz de proyección sobre el espacio de componentes principales por el método rápido propuesto por [Tur91]; el algoritmo de cálculo de los autovectores y autovalores procede de [Pre92]. La elección del número de autovectores que se quiere conservar es crítica. Un número muy alto supone un tiempo de procesamiento grande, sin apenas reducir información, mientras que un número de autovectores muy bajo supone que se puede perder mucha información. En nuestro caso se buscaba mantener la práctica totalidad de la varianza inicial de la base de datos con un número de autovectores reducido. La máxima dimensión de nuestro espacio en el que proyectar las imágenes iniciales viene dada por el número de sujetos en la base de datos y el número de imágenes de cada sujeto. En nuestro caso, se dispuso de 29 sujetos y de 12 imágenes por sujeto, es decir,

$29 \times 12 = 348$. Se observó (véase la Figura 4.4) que 150 autovectores son suficientes para retener casi toda la varianza del conjunto inicial. La Figura 4.4 se generó a partir de la base de datos de mayor varianza, en este caso, 150 autovectores explican el 96,43% de la varianza total. Se ha de tener presente que la varianza explicada por un autovector puede calcularse de acuerdo con la formula siguiente:

$$\text{var exp autovector}_i = \frac{\text{autovalor}_i}{\sum \text{autovalores}}.$$

Por lo tanto es posible calcular el porcentaje de varianza explicado en función del número de autovectores seleccionado.

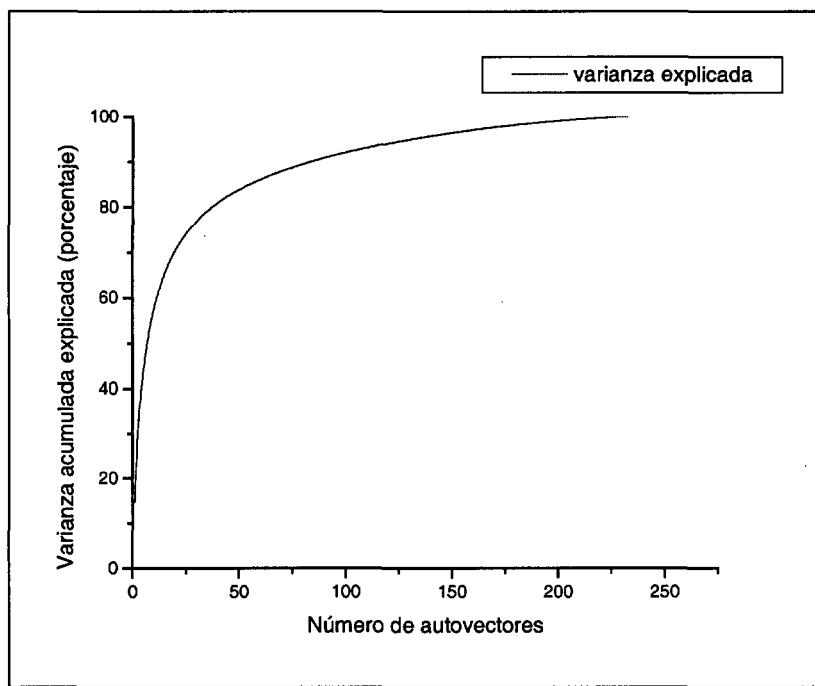


Figura 4.4. Varianza explicada en función del número de autovectores seleccionados.

A lo largo de los experimentos realizados, se observó que cada variación en las condiciones de adquisición para los conjuntos de entrenamiento y pruebas implica el cálculo de una nueva matriz PCA, ya que ésta se ve influida por las variaciones que ocurren en las características generales de las imágenes (iluminación, fondo, contraste). Estas variaciones pueden provenir de diferentes condiciones de iluminación, debido a una localización de la cara poco robusta o bien por utilizar tarjetas capturadoras con ganancias muy diferentes.

4.2.3.- Niveles de gris utilizando imágenes de baja resolución.

La etapa de preprocessamiento en este caso se compone de la localización de la cara y la reducción posterior en resolución. El uso de los niveles de gris como datos de entrada ha sido contemplado numerosas veces en visión artificial [Cab02] [Bal82] [Wec90], sobre todo si el procesamiento se realiza por medio de redes neuronales [Bis92] [Bis95].

En trabajos previos [Cab98] se consideraron dos alternativas de preprocessamiento en las que inicialmente se calculaba la pirámide Gaussiana [Jol92] de la imagen. En un caso se procesaba la información de baja resolución así obtenida directamente y en otro caso se incluía una etapa posterior de segmentación (véase la

Figura 4.5). Una pirámide Gaussiana [Der90] calcula una media ponderada de los píxeles de un nivel y posteriormente elimina filas y columnas entrelazadas para generar una imagen en el siguiente nivel. A medida que se avanza por la pirámide hacia resoluciones más bajas, los contornos se difuminan, siendo por tanto más difícil el reconocimiento. El hecho de añadir una etapa posterior de segmentación lleva a unos resultados de reconocimiento correcto muy bajos, tal como también aparece en [Sam94]. El proceso de segmentación parece que elimina información esencial en el reconocimiento, como pueden ser los detalles faciales.

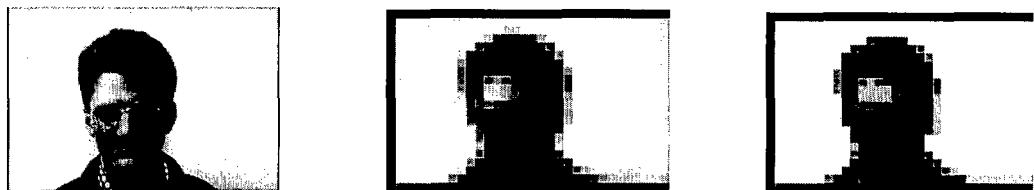


Figura 4.5. Imagen inicial, Gaussiana y segmentada de un sujeto.

El preprocessamiento seleccionado en este caso consta de una serie de pasos. El primer paso consiste en la localización de la cara, aplicando el método descrito en el apartado 4.2; se procedió a realizar una convolución con una plantilla. La adquisición de una imagen previa del fondo y su diferencia con la imagen del sujeto sirve para obtener un punto inicial y acotar una zona en la que realizar la convolución. El resultado es una imagen de tamaño 130x140 píxeles. A continuación la imagen se pasa a blanco y negro (256 niveles de gris) y se procede a realizar una ecualización del histograma. El paso siguiente consiste en calcular una pirámide en la que cada paso se construye por la eliminación de filas y columnas entrelazadas en la imagen anterior. La resolución seleccionada para realizar el presente trabajo es de $17 \times 18 = 306$ píxeles. Esta resolución permite retener información necesaria para la verificación del sujeto, así como disminuir el tiempo de cálculo en la identificación.

4.3.- Clasificación facial.

K Vecinos Más Cercanos (KNN).

El funcionamiento de este método es bastante sencillo [Cov67] [Dud73]. Dado un conjunto de vectores cuya clasificación es conocida (conjunto de entrenamiento), a un nuevo vector de prueba se le asignará la clase mayoritaria de los k vecinos más cercanos del conjunto de entrenamiento. Si se considera más de un vecino, el método es más robusto frente al ruido. En caso de realizar tareas de clasificación, k suele ser un número impar que evita empates, habitualmente $k=1, 3$ ó 5 . Si $k = 1$, el diagrama de Voronoi [Oka92] da el poliedro complejo que segmenta el espacio en regiones de puntos cercanos a un vector de entrenamiento.

Partiendo de este método básico han ido surgiendo numerosas variantes. Dependiendo del problema que se pretende resolver, se pueden considerar distancias no euclídeas. Como la búsqueda lineal no es un método de clasificación eficiente, se ha utilizado una estructura de datos llamada “árbol k-d” para indexar los vectores de entrenamiento y encontrar los vecinos más cercanos en tiempo logarítmico [Moo91]. Otras mejoras planteadas al método KNN no han sido consideradas aquí, ya que esta técnica ha sido utilizada en esta tesis por su sencillez y porque tampoco se ha considerado necesario optimizar el tiempo de ejecución de los algoritmos desarrollados.

Redes Neuronales: Funciones de base radial (RBF).

El método seleccionado en esta tesis para implementar este tipo de redes es el RBF-DDA (DDA: Dynamic Decay Adjustment), que ha sido descrito en el Capítulo 2. En dichas redes neuronales es preciso definir a priori los valores de dos parámetros (θ^+ y θ^-). En teoría, estos parámetros son dependientes de la dimensión del espacio de características pero en la práctica los valores de los dos umbrales no son críticos. En todos los experimentos de la presente tesis, la elección de $\theta^+ = 0.4$ y $\theta^- = 0.2$, ha conducido a resultados satisfactorios.

Un hecho más importante es que los datos de entrada estén normalizados. Debido a la naturaleza radial de las RBF, cada atributo debe estar distribuido sobre un rango equivalente. Habitualmente una normalización en [0,1] es suficiente. En esta tesis se ha empleado la siguiente normalización: dado un vector de datos sin normalizar x , el nuevo vector normalizado se obtiene restando a cada componente de x el valor medio de x y dividiendo el resultado entre la desviación típica de x . Esto permite construir vectores normalizados de media 0 y desviación típica 1, aunque alguna de las coordenadas de dicho vector pueda estar fuera del rango [0,1].

En esta tesis se consideró que en todos los casos, la neurona ganadora es la que decide a qué clase pertenece el patrón.

Para la implementación de las redes neuronales de tipo RBF-DDA se ha utilizado el programa SNNS [SNNS].

Máquinas de vector soporte (Support Vector Machines - SVM).

Las máquinas de vector soporte permiten resolver problemas de clasificación, pero esta clasificación se realiza sólo en dos clases (ver 2.1.2). Esta solución se adapta de forma natural al problema de verificación, ya que una clase representa al sujeto cuyo PIN ha sido introducido y la segunda clase agrupa al resto de imágenes de sujetos. Esta ha sido la aproximación considerada.

En la presente tesis se ha utilizado siempre máquinas de vector soporte con kernels lineales.

4.4.- Resultados experimentales y discusión.

Un sistema de verificación facial posee dos tipos de datos de entrada: la imagen del sujeto y un número de identificación personal (PIN) que debe introducirse. Por lo tanto, existen cuatro posibilidades en nuestro experimento: verdadero positivo, verdadero negativo, falso positivo y falso negativo. Verdadero positivo y verdadero negativo indican que el sistema ha identificado correctamente la cara de la persona, siendo o no la poseedora del PIN. Falsos positivos ocurren cuando un sujeto incorrecto es aceptado como el poseedor del PIN. Falsos negativos corresponden al caso en que la persona correcta no es identificada por el sistema. Estos cuatro casos pueden verse en la siguiente tabla.

La imagen corresponde al dueño del PIN	Respuesta del sistema: Verdadero	Respuesta del sistema: Falso
Verdadero	Verdadero Positivo (TP)	Falso Negativo (FN)
Falso	Falso Positivo (FP)	Verdadero Negativo (TN)

Tabla 4.1. Posibles datos de salida del sistema.

Para cada una de los experimentos realizados, se usará la siguiente notación :

- Porcentaje de reconocimiento correcto: $(TP+TN)/(TP+TN+FP+FN)$
- Porcentaje de falsos negativos: $FN/(FN+TN)$
- Porcentaje de falsos positivos: $FP/(FP+TP)$

En donde TP, TN, FP y FN son los valores de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos, respectivamente.

En todas las pruebas, los conjuntos de prueba y de entrenamiento carecen de elementos comunes. A la hora de presentar los resultados se ha optado por agrupar los datos de prueba de dos formas distintas. La primera forma se denomina "no equilibrada". En este caso cada valor del PIN correspondiente a un sujeto (cada prototipo o modelo correspondiente a clase) ha sido probado con cada imagen del conjunto de prueba. En el conjunto de prueba hay cuatro imágenes del sujeto y 112 imágenes del resto de sujetos. Al ir variando el parámetro sucede que, si el sistema

rechaza a todos los sujetos, el porcentaje de acierto es $\frac{112}{116} = 96.55\%$, ofreciendo por

tanto un resultado que puede ser engañoso. Esta es la forma en que se suelen mostrar los resultados en la bibliografía (por ejemplo [Tur91]). En consecuencia, se ha optado por mostrar también los resultados llevados a cabo con un conjunto de prueba, en el que el número de imágenes del sujeto sea igual al número de imágenes de otros sujetos (en nuestro caso ese número es de cuatro). Este conjunto se denomina "equilibrado". Para mostrar los resultados, se ha decidido aplicar el método de validación cruzada, considerando la media de diez pruebas para cada modelo. Cada una de estas pruebas tiene cuatro imágenes del sujeto y cuatro imágenes seleccionadas aleatoriamente entre las imágenes de los otros sujetos.

Seguidamente se muestran los resultados obtenidos en los dos experimentos realizados. El primer experimento consiste en realizar un desplazamiento en la localización de la imagen (véanse las Figuras 4.6 y 4.7). Se entrena el sistema con la cara recortada correctamente (donde la correlación con la plantilla es máxima) y se realizan varias pruebas. La primera emplea la cara correctamente recortada en las imágenes de prueba (desplazamiento de 0 píxeles en cada eje, o con la notación adoptada: 0x0); las siguientes pruebas se realizan desplazando la cara 5, 10, 15 y 20 píxeles en cada eje. El segundo experimento cambia la plantilla con la que se recorta la cara. En el primer caso, se recorta toda la base de datos con una única plantilla, en el segundo caso se recorta la base de datos con una plantilla para cada sujeto. Esta plantilla propia de cada sujeto corresponde con una imagen del sujeto cuya identidad se quiere verificar. En cada caso se muestran los resultados obtenidos tanto para imágenes de baja resolución como para imágenes procesadas con PCA. Se indican asimismo los resultados generados para cada uno de los tres clasificadores estudiados (KNN, SVM y RBF).



Figura 4.6. Cara localizada correctamente



Figura 4.7. La misma cara pero desplazada 5x5, 10x10, 15x15 y 20x20 píxeles.

En el caso de imágenes de baja resolución, se han modificado las imágenes reduciendo su resolución hasta 17x18, obteniéndose los resultados que se muestran en la Figura 4.8. Tanto las redes neuronales del tipo RBF como las máquinas de vector soporte (SVM) se han entrenado con los vectores de dimensión $17 \times 18 = 306$.

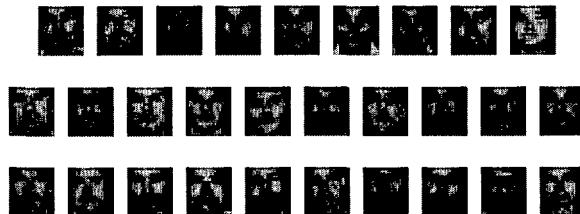


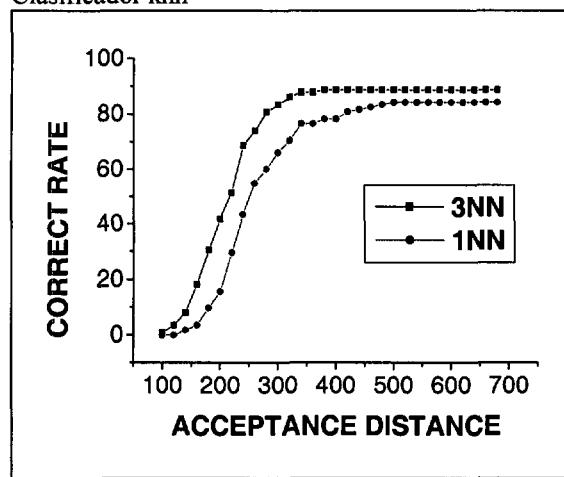
Figura 4.8. Sujetos de la base de datos reducida.

En el Anexo B se muestran las gráficas con los datos no incluidos en las páginas siguientes (valores de desplazamiento de 15 y 20 píxeles por eje) ya que cualitativamente no aportan ningún elemento nuevo en la discusión de los datos. En el Anexo C se muestran las gráficas obtenidos en el caso de datos no equilibrados.

En las tablas que se muestran a continuación en el caso del clasificador KNN el parámetro corresponde a los valores de la distancia euclídea y a continuación se muestran los valores obtenidos en la verificación correcta. En el caso de los clasificadores RBF y SVM; el valor del parámetro corresponde a los valores obtenidos por la neurona ganadora (RBF) o a la salida obtenida por el clasificador (SVM).

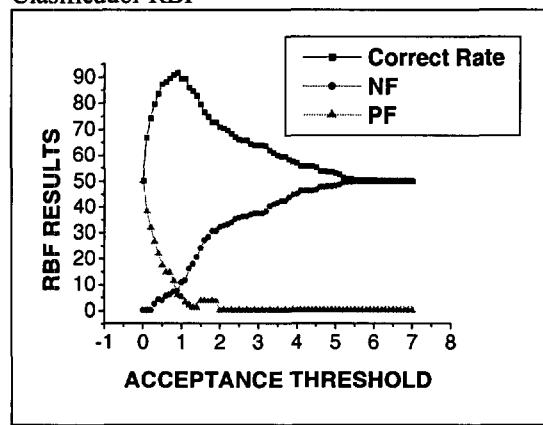
Figura 4.9: Resultados con imágenes correctamente localizadas: baja resolución.

Clasificador knn



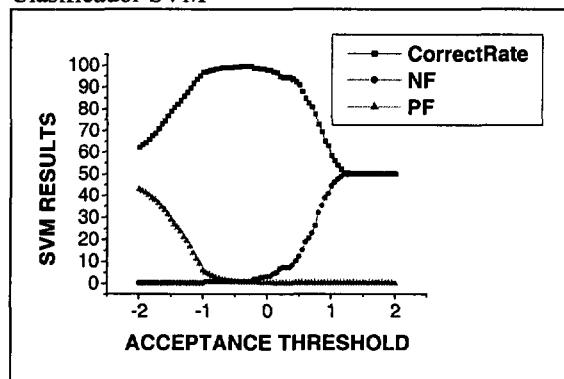
Parámetro	3NN	1NN
100	0,86	0
160	18,26	3,47
180	30,43	9,56
220	51,30	29,56
240	68,69	43,47
280	80,86	60
300	83,47	66,08
340	87,82	76,52
360	87,82	76,52
400	88,69	78,26
420	88,69	80,86
460	88,69	82,60
480	88,69	83,47
520	88,69	84,34

Clasificador RBF



Parámetro	% Correctos	%FN	% FP
0	50,03	0	49,96
0,9	91,57	7,41	5,90
1,3	84,48	17,89	1,07
1,9	72,61	30,46	3,53
2,2	69,54	33,01	0,08
2,8	64,39	36,87	0
3,3	61,73	40,11	0
4	56,89	44,92	0
4,2	55,59	46,28	0
4,9	53,43	47,89	0
5,4	50,83	49,54	0
6	50,39	49,79	0
6,3	49,96	50,03	0
6,9	49,96	50,03	0

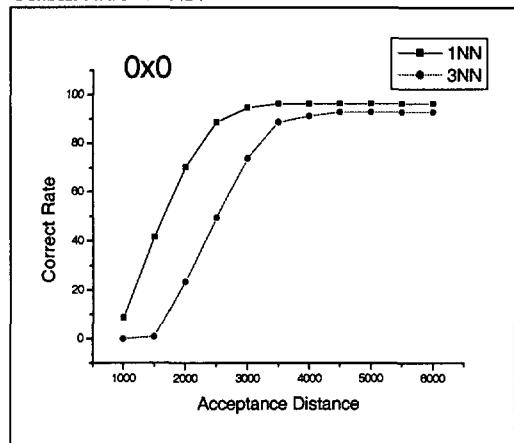
Clasificador SVM



Parámetro	%Correctos	%FN	%FP
-1,55	76,59	0	31,20
-1,35	83,79	0	23,59
-0,9	97,45	0,68	3,89
-0,6	99,01	0,68	1,05
-0,35	99,39	0,68	0,32
0	97,80	2,98	0,08
0,1	96,93	4,38	0,11
0,45	92,24	9,96	0
0,7	81,03	22,69	0
1	58,18	44,25	0
1,15	52,15	48,60	0
1,45	50	50	0

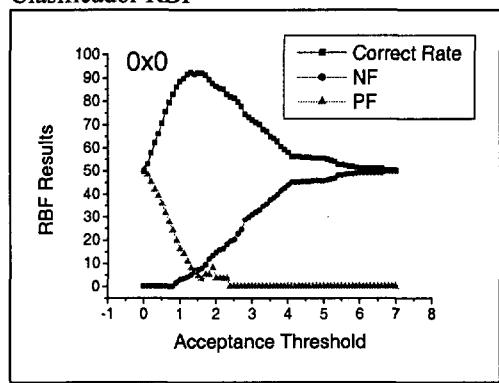
Figura 4.10: Resultados con imágenes correctamente localizadas: análisis de componentes principales.

Clasificador KNN



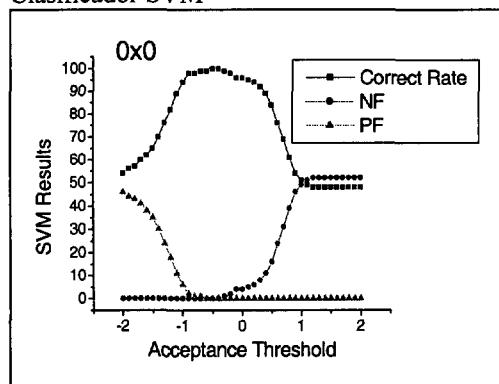
Parámetro	1-KNN	3-KNN
1000	8,7	0
1500	41,7	0,9
2000	70,4	23,5
2500	88,7	49,6
3000	94,8	73,9
3500	96,5	88,7
4000	96,5	91,3
4500	96,5	93
5000	96,5	93
5500	96,5	93
6000	96,5	93

Clasificador RBF



Parámetro	% Correctos	% FN	% FP
0	49,75	0	50,24
0,9	86,15	1,63	20,02
1,3	92,37	5,04	8,02
1,9	87,78	13,30	8,11
2,2	85,13	16,34	3,53
2,8	74,44	28,68	0
3,3	67,91	34,61	0
4	58,03	43,95	0
4,2	56,28	45,11	0
4,9	55,41	45,68	0
5,4	52,83	48,19	0
6	51,10	49,26	0
6,3	51,10	49,26	0
6,9	50,24	49,75	0

Clasificador SVM



Parámetro	% Correctos	% FN	% FP
-2	54,81	0	46,11
-1,6	62,15	0	38,55
-1,4	70,32	0	30,23
-1,1	89,47	0	11,15
-1	94,54	0	6,87
-0,7	99,69	0	1,12
-0,5	100	0	0
-0,2	98,72	2,27	0
-0,1	96,59	4,87	0
0,2	94,44	6,16	0
0,4	89,37	11,32	0
0,7	69,12	31,56	0
1,1	49,01	51,66	0
1,3	48,87	52,17	0
1,6	48,45	52,87	0

Discusión de los resultados.

Los resultados mostrados han sido obtenidos con imágenes de prueba correctamente localizadas y podemos afirmar que son muy buenos, tanto utilizando baja resolución como análisis de componentes principales, independientemente del clasificador seleccionado.

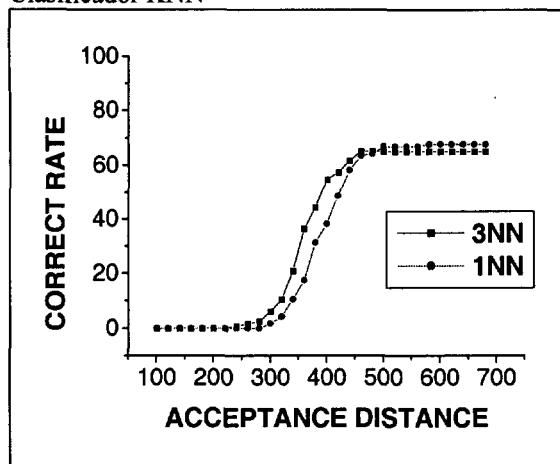
Los mejores resultados se obtienen procesando las imágenes con análisis de componentes principales frente a baja resolución, aunque las diferencias son muy pequeñas para cada clasificador. Por ejemplo SVM obtiene un 100% de aciertos en el caso de PCA frente a un 99,4 en el caso de baja resolución. RBF obtiene un porcentaje de aciertos de 92,38 procesando con PCA y 91,58% en caso de baja resolución. En el caso de KNN podemos ver que con PCA se obtiene un resultado e reconocimiento correcto de 96,5 % frente a un 88,5 % en el caso de baja resolución. Este hecho nos indica que al ser baja resolución un método global (emplea información de toda la cara) es capaz de retener información suficiente para verificar a una persona, aunque sufre una pérdida de información frente a PCA podemos considerarla como bastante pequeña.

Atendiendo a los clasificadores considerados, los mejores resultados se obtienen para SVM, seguido de RBF y los peores resultados son obtenidos por KNN. Este hecho depende de la complejidad del clasificador y era un hecho esperado. Sin embargo a pesar de que KNN es un clasificador lineal poco complejo presenta las ventajas de una fácil implementación y nos ofrece una idea intuitiva de cómo se agrupan los datos (en este caso las caras). SVM es un clasificador muy potente que aunque no ha sido optimizado demuestra su gran capacidad.

Los resultados mostrados en las tablas y gráficas precedentes ofrecen resultados comparables a los referenciados en la bibliografía. Turk y Pentland [Tur91] obtienen un 96% de clasificación correcta con una base de datos de 16 sujetos y procesando con PCA. En nuestro caso, el empleo de clasificadores con mayor poder discriminante permite obtener mejores resultados a pesar de tener una base de datos mayor. Los resultados son comparables a los obtenidos por Brunelli y Poggio [Bru93] utilizando correspondencia de plantillas para una base de datos de 47 sujetos (100% de acierto) pero con un tiempo de procesamiento mucho menor en nuestro caso. Los resultados mostrados en el “Face recognition vendor test” [Phi03] muestra que en el caso de imágenes obtenidas en el interior de un edificio los sistemas comerciales ofrecen un 90% de verificación correcta cuando el porcentaje de falso positivo es de 1%. En nuestro caso, el clasificador RBF no llega a tales porcentajes (aproximadamente un 70% en el caso de baja resolución y un 80% en caso de PCA). Sin embargo, SVM mejora dicho umbral al obtener un 99% de resultados correctos con un 1% de falsos positivos tanto en el caso de baja resolución como en el caso de PCA.

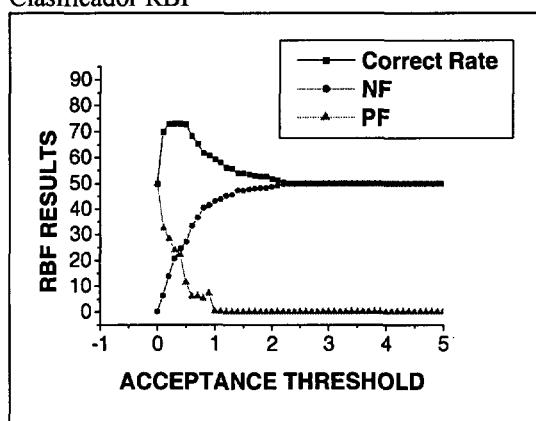
Hemos de tener en cuenta que los métodos presentados en la presente tesis se basan en el estudio de los niveles de gris de la imagen y por lo tanto, los resultados obtenidos dependen del error cometido en la localización de la cara en la imagen. Para medir la influencia de errores en la localización en el porcentaje de reconocimiento correcto se diseñó el siguiente experimento, en el cual las imágenes de prueba se desplazan un determinado número de píxeles de la localización correcta.

Figura 4.11: Resultados con imágenes desplazadas 5 píxeles: baja resolución.
Clasificador KNN



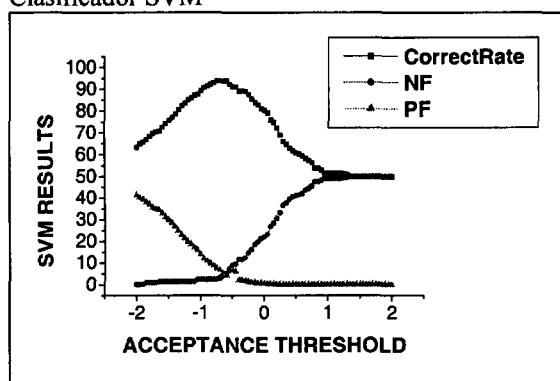
Parámetro	1NN	3NN
220	0	0
240	0	0,87
280	0	2,61
300	1,74	6,08
340	10,43	20,86
360	17,39	36,52
400	38,26	54,78
420	48,70	57,39
460	63,48	65,22
480	64,35	65,22
540	66,96	65,22
580	67,83	65,22
600	67,83	65,22

Clasificador RBF



Parámetro	% Correctos	%FN	% FP
0	49,95	0	50,04
0,1	69,75	6,21	32,32
0,3	72,98	20,57	23,86
0,4	72,99	24,40	22,21
0,6	68,26	33,15	5,91
0,7	65,33	36,53	6,09
0,9	60,68	41,24	7,06
1	59,11	42,80	0,08
1,2	56,13	45,05	0
1,5	53,96	47,26	0
1,8	52,66	48,16	0
1,9	52,66	48,16	0
2,1	51,36	49,12	0
2,2	50,49	49,71	0

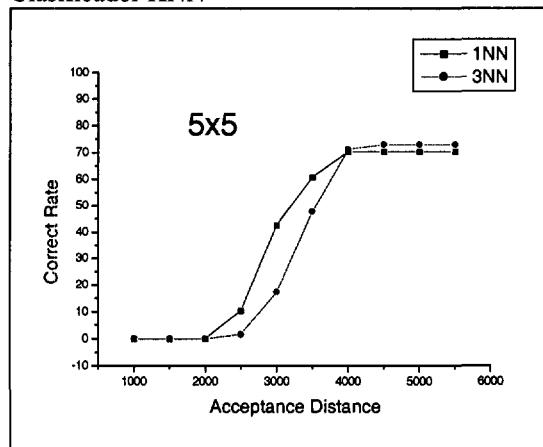
Clasificador SVM



Parámetro	%Correctos	%FN	%FP
-2	63,27	0	41,26
-1,55	74,39	1,25	31,17
-1,35	80,64	1,25	24,77
-1,05	87,71	2,34	15,98
-0,9	91,42	2,27	10,84
-0,6	94,14	4,73	4,09
-0,35	89,35	11,59	1,88
0	80,34	21,98	0,47
0,1	76,12	26,81	0,32
0,45	61,63	40,42	0
0,7	57,32	44,31	0
1	51,72	48,93	0
1,45	50	50	0
1,6	50	50	0

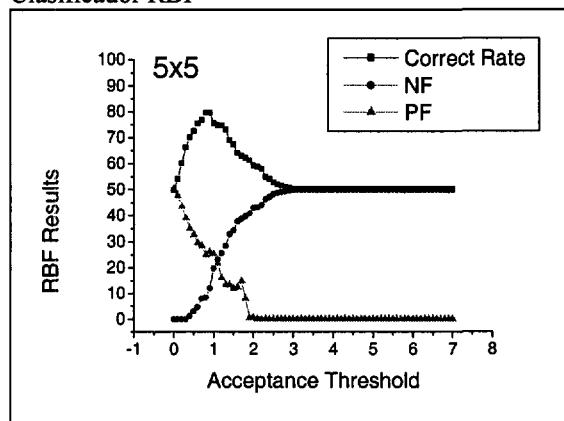
Figura 4.12: Resultados con imágenes desplazadas 5 píxeles: análisis de componentes principales.

Clasificador KNN



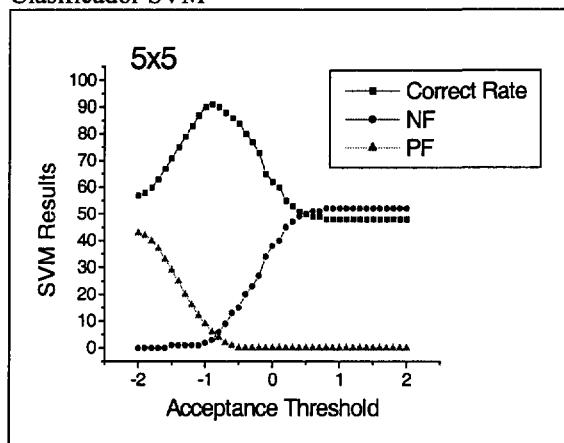
Parámetro	1-KNN	3-KNN
1000	0	0
1500	0	0
2000	0	0
2500	10,42	1,71
3000	42,63	17,42
3500	60,96	47,86
4000	70,47	71,37
4500	70,47	73,23
5000	70,47	73,23
5500	70,47	73,23

Clasificador RBF



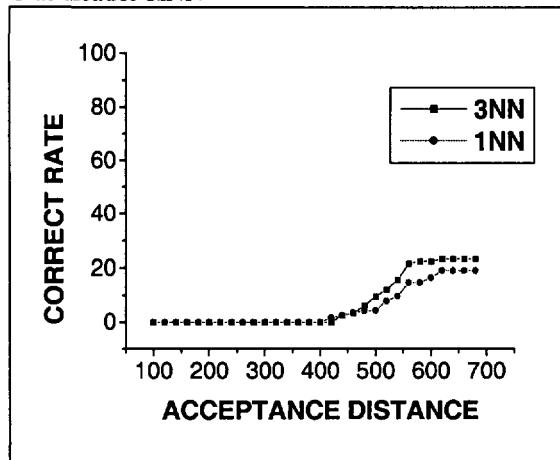
Parámetro	% Correctos	% FN	% FP
0	49,75	0	50,24
0,1	54,19	0	47,59
0,4	70,09	1,18	35,14
0,6	75,50	4,61	29,58
0,9	79,53	12,12	26,21
1	75,57	19,79	25,43
1,3	73,16	28,44	13,57
1,5	67,37	34,41	11,99
1,8	62,33	39,70	8,05
1,9	61,36	40,78	0,57
2,2	58,01	44,08	0
2,4	54,12	47,15	0
2,7	51,53	49,01	0
2,8	51,11	49,26	0
3,1	50,24	49,75	0
3,3	50,24	49,75	0

Clasificador SVM



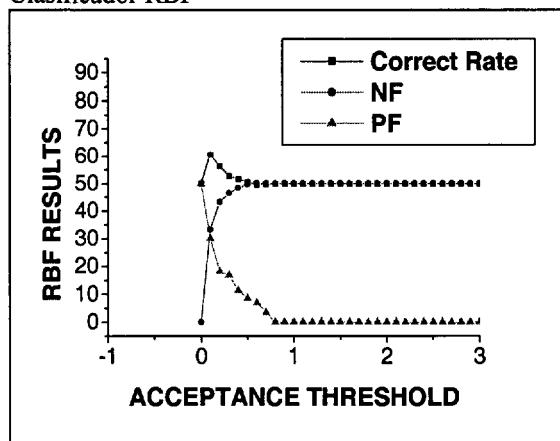
Parámetro	% Correctos	% FN	% FP
-2	57,12	0	43,89
-1,6	67,26	0	33,23
-1,4	75,39	1,24	25,56
-1,1	87,42	1,88	12,34
-1	90,56	2,12	9,89
-0,7	88,87	9,18	2,02
-0,5	84,64	15,33	0
-0,2	73,52	27,45	0
-0,1	65,39	34,61	0
0,2	55,26	45,73	0
0,4	51,10	49,91	0
0,7	49,01	51,13	0
0,8	48,87	52,24	0
1,1	48,87	52,24	0

Figura 4.13: Resultados con imágenes desplazadas 10 píxeles: baja resolución.
Clasificador KNN



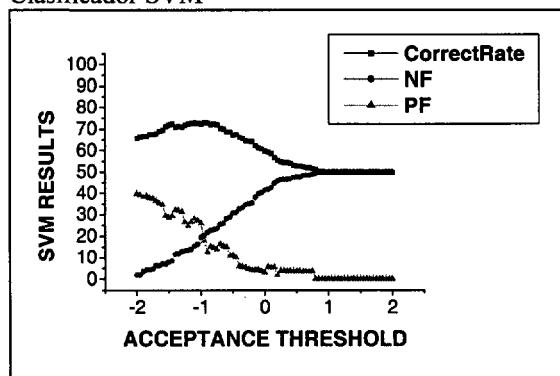
Parámetro	3NN	1NN
400	0	0
420	0	1,73
440	2,60	2,60
460	3,47	3,47
480	6,08	4,34
500	9,56	4,34
520	12,17	7,82
540	15,65	9,56
560	21,73	14,78
580	22,60	14,78
600	22,60	16,52
620	23,47	19,13

Clasificador RBF



Parámetro	% Correctos	%FN	% FP
0	50,03	0	49,96
0,1	60,58	33,27	30,05
0,2	56,29	43,17	18,13
0,3	52,79	46,41	16,79
0,4	51,71	48,36	11,38
0,5	50,61	49,57	8,52
0,6	49,56	50,25	6,89
0,7	49,82	50,10	3,44
0,8	49,96	50,04	0
0,9	49,96	50,04	0
1	49,96	50,04	0

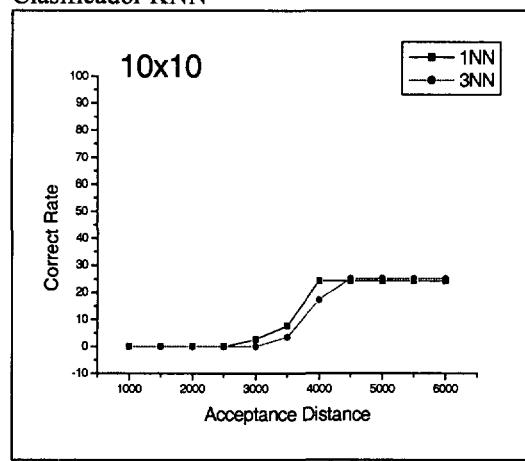
Clasificador SVM



Parámetro	%Correctos	%FN	%FP
-2	65,68	1,81	39,64
-1,95	66,33	1,81	39,14
-1,55	71,16	6,81	29,51
-1,35	71,25	11,77	31,82
-1,05	72,71	16,16	27,19
-0,9	72,80	21,50	12,58
-0,6	68,79	28,08	14,99
-0,35	65,56	33,18	5,49
0	59,69	41,33	2,91
0,1	58,53	42,50	5,25
0,45	53,14	47,34	3,53
0,65	52,06	48,43	3,44
0,85	50,43	49,75	0
0,95	50	50	0

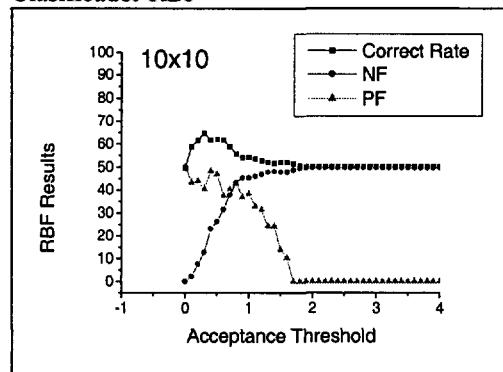
Figura 4.14: Resultados con imágenes desplazadas 10 píxeles: análisis de componentes principales.

Clasificador KNN



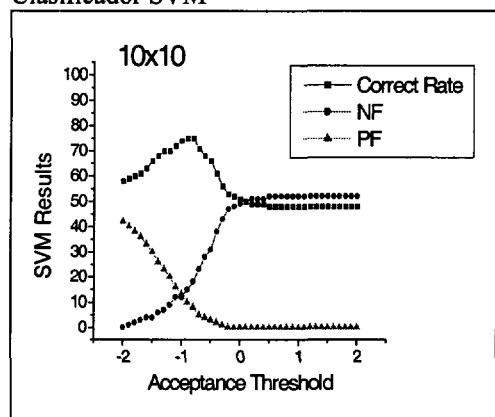
Parámetro	1-KNN	3-KNN
1000	0	0
1500	0	0
2000	0	0
2500	0	0
3000	2,62	0
3500	7,85	3,51
4000	24,36	17,44
4500	24,36	25,28
5000	24,36	25,28
5500	24,36	25,28
6000	24,36	25,28

Clasificador RBF



Parámetro	% Correctos	%FN	% FP
0	49,75	0	50,24
0,1	58,89	2,21	43,33
0,3	64,82	12,75	40,64
0,4	61,89	23,18	48,35
0,6	61,89	31,54	37,47
0,7	58,89	37,76	40,41
0,9	54,11	45,27	36,84
1	54,29	45,35	38,38
1,2	52,86	46,96	31,41
1,3	52,22	47,76	24,37
1,5	52,14	47,91	13,79
1,6	52,18	47,89	10,34
1,8	50,67	49,50	0
1,9	50,24	49,75	0

Clasificador SVM



Parámetro	% Correctos	%FN	%FP
-2	58,82	0	42,15
-1,9	59,14	1,22	40,23
-1,6	63,17	4,12	33,33
-1,4	68,44	6,34	26,89
-1,1	72,56	12,12	16,76
-1	74,78	12,01	13,56
-0,7	71,23	23,55	5,09
-0,5	66,33	31,11	3,11
-0,2	53,66	47,22	0
-0,1	52,98	48,01	0
0,2	49,01	51,12	0
0,4	49,01	51,12	0
0,7	48,23	52,01	0
0,8	48,23	52,00	0

Discusión de los resultados.

Los resultados precedentes muestran las gráficas y los valores numéricos cuando los errores en la localización corresponden a 5 y 10 píxeles por eje. Los valores gráficos para desplazamiento de 15 y 20 píxeles se muestran en el anexo B. Obviamente los valores absolutos en los porcentajes de acierto empeoran a medida que aumentamos el desplazamiento. Los resultados con un desplazamiento de 10 píxeles son peores que si se desplazan 5 píxeles por eje.

Puede observarse que el clasificador SVM es el que mejores resultados ofrece, seguido de RBF y por último KNN. SVM demuestra ser el clasificador más robusto, ya que con un desplazamiento de 5 píxeles obtiene porcentajes de verificación correcta del 94 % (baja resolución) y 90 % (PCA). RBF en las mismas condiciones obtiene porcentajes de 72,99 % y 79,54 %. Los porcentajes de KNN son más bajos. En el caso de desplazar 10 píxeles, el clasificador SVM obtiene un porcentaje de verificación correcta de 72,8 (baja resolución) y 74% (PCA). En las mismas condiciones RBF obtiene 60,58% y 64,82%.

Además puede observarse que, en general, los resultados con imágenes procesadas con PCA son mejores que con baja resolución, salvo en el caso de desplazar la imagen 5 píxeles y utilizar como clasificador SVM. En este caso, sorprendentemente, baja resolución ofrece mejores resultados que PCA. Aunque podemos considerarlo como una excepción a una regla general.

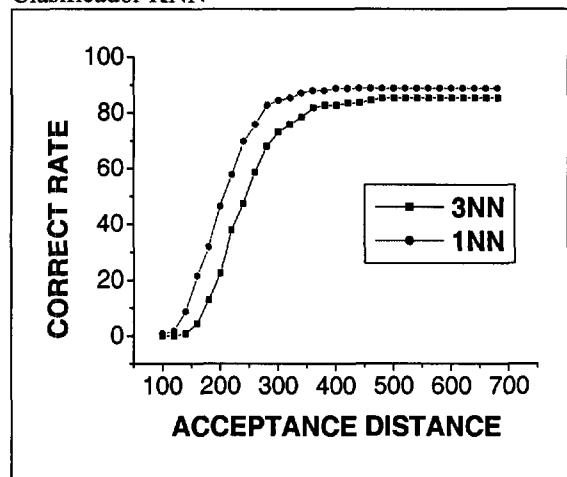
En el caso de errores de 5 píxeles, los resultados siguen siendo comparables a los mostrados en la bibliografía con localización correcta, en especial utilizando el clasificador SVM. Un error de localización de 10 píxeles (o superior) aparece como dramático en la respuesta de los clasificadores.

Sin embargo los porcentajes de acierto se consiguen a costa de valores muy altos en los porcentajes de falsos positivos y negativos. Por ejemplo utilizando PCA con desplazamiento de 10 píxeles el porcentaje de acierto de SVM (74%) implica porcentajes de falsos positivos (13%) y falsos negativos (12%) bastante altos; en el caso de RBF estos porcentajes son del 40% y del 13%.

Una vez estudiados los efectos de los errores en la localización debido a desplazamientos se decidió mejorar el proceso estudiando el efecto que tiene el usar una única plantilla en el recorte de la base de datos o el usar una plantilla por sujeto. A continuación se muestran las gráficas y resultados numéricos obtenidos al recortar a cada sujeto con una plantilla propia, ya que las gráficas y resultados numéricos obtenidos al recortar a todos los sujetos con una única plantilla común se muestran en las Figuras 4.9 y 4.10.

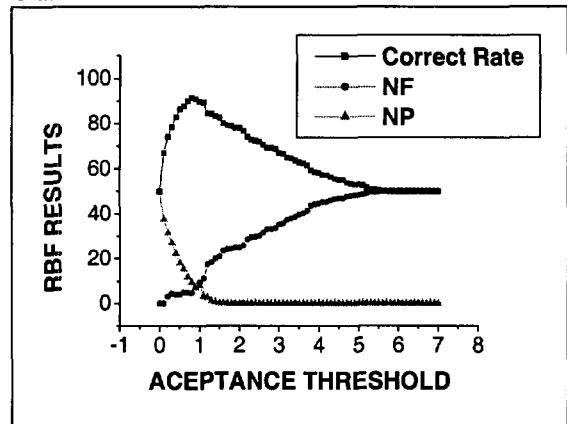
Figura 4.15: Resultados con imágenes recortadas con una plantilla por sujeto: baja resolución.

Clasificador KNN



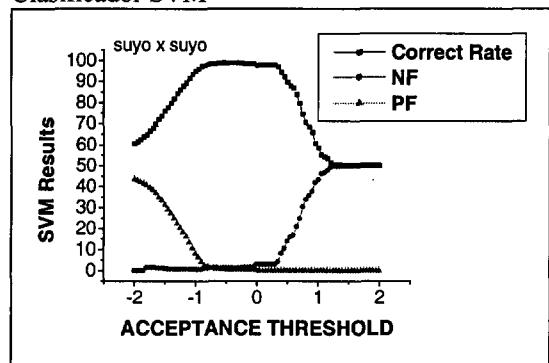
Parámetro	1NN	3NN
100	0,86	0
140	8,62	0,86
160	21,55	4,31
200	46,55	22,41
220	57,75	37,93
260	75,86	58,62
280	82,75	68,10
320	85,34	75,86
340	87,06	78,44
380	87,93	82,75
440	88,79	83,62
460	88,79	84,48
500	88,79	85,34
520	88,79	85,34

Clasificador RBF



Parámetro	% Correctos	%FN	% FP
0	50	0	50
0,8	91,59	4,67	9,43
1,2	84,67	17,23	2,53
1,8	78,87	24,30	0
2,1	76,72	25,69	0
2,7	69,39	32,97	0
3,2	65,08	37,19	0
3,9	58,62	44,20	0
4,1	57,32	45,18	0
4,8	53,44	47,94	0
5,3	50,86	49,50	0
5,9	50	50	0

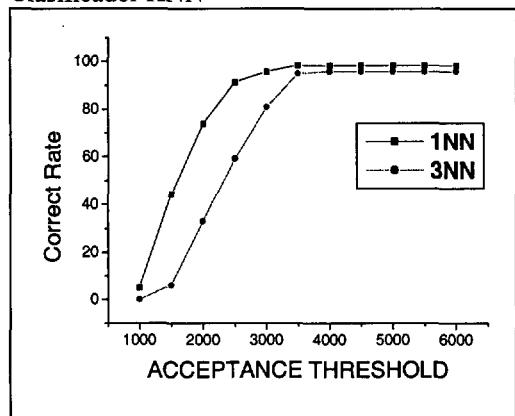
Clasificador SVM



Parámetro	%Correctos	%FN	%FP
-2	60,77	0	43,56
-1,75	66,46	1,49	39,48
-1,3	84,39	0,81	22,42
-1,1	91,50	0,71	13,35
-0,95	96,03	0,68	6,34
-0,65	98,66	1,14	0,89
-0,4	98,92	1,14	0,41
-0,25	98,96	1,14	0,32
0,05	97,84	2,857	0
0,4	94,39	8,14	0
0,7	79,74	24,87	0
1,1	54,31	46,63	0
1,4	50	50	0

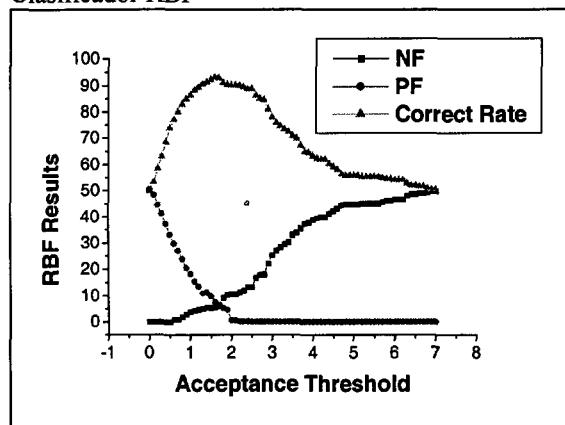
Figura 4.16: Resultados con imágenes recortadas con una plantilla por sujeto: análisis de componentes principales.

Clasificador KNN



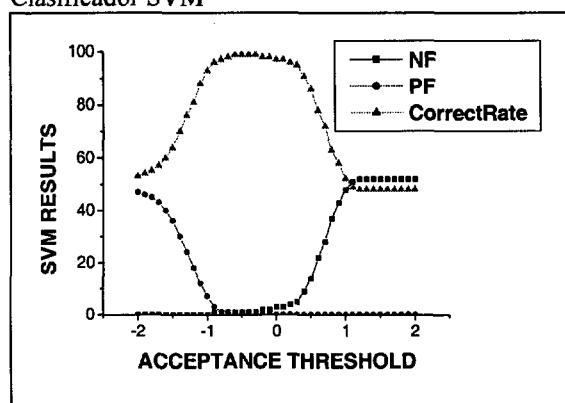
Parámetro	1-KNN	3-KNN
1000	5,21	0
1500	44,34	6,19
2000	73,94	33
2500	91,35	59,15
3000	95,79	80,91
3500	98,36	94,82
4000	98,36	95,67
4500	98,36	95,67
5000	98,36	95,67
5500	98,36	95,67
6000	98,36	95,67

Clasificador RBF



Parámetro	% Correctos	% FN	% FP
0	49,75	0	50,23
0,8	82,79	1,79	23,79
1,2	89,32	4,69	13,39
1,6	93,12	5,71	7,49
2,1	90,33	10,61	0,50
2,7	84,91	17,76	0
3,2	74,56	28,40	0
3,9	64,65	37,84	0
4,1	62,5	39,58	0
4,8	55,84	44,77	0
5,3	55,41	45,02	0
5,9	54,55	46,17	0
6,2	54,12	46,63	0
6,8	50,67	49,50	0

Clasificador SVM



Parámetro	% Correctos	% FN	% FP
-2	53,24	0	46,76
-1,5	63,88	0	36,12
-1,2	81,01	0	17,99
-0,8	97,60	0,50	1,10
-0,6	100	0,91	0
-0,2	98,10	1,92	0
0,1	96,87	2,13	0
0,6	77,18	22,81	0
0,7	72,45	27,55	0
1,2	48,33	51,67	0
1,5	48,33	51,67	0
1,9	48,33	51,67	0

Discusión de los resultados.

Inicialmente consideraremos las gráficas y los valores numéricos obtenidos utilizando imágenes de baja resolución, recortadas considerando una plantilla común (Figura 4.9) o bien una plantilla por sujeto (Figura 4.15). En este caso se observa que tanto KNN como RBF mejoran el porcentaje de aciertos cuando se considera una plantilla para cada sujeto. En el caso de 1NN el porcentaje de acierto con una plantilla común es de 84,35% y con una plantilla por sujeto es de 88,79%. En el caso de RBF con una plantilla común el porcentaje de verificación correcta es de 91,57 % y en caso de una plantilla por sujeto es de 91,59 %. Sorprendentemente en el caso de SVM con una plantilla común se obtiene un porcentaje de acierto máximo de 99,39 % y con una plantilla por sujeto se obtiene 98,96 %. Hemos de tener presente que no se ha considerado ninguna técnica de optimización de SVM para adaptarlo a los datos que estamos considerando y que el kernel empleado es el lineal, por lo tanto podemos observar pequeñas diferencias en SVM dependiendo del procesamiento que hemos realizado en las imágenes (la diferencia es de un 0.4 %). El uso de un kernel más adaptado a nuestro problema permitiría obtener mejoras en el porcentaje de verificación correcta. En cualquier caso (una plantilla común o una por sujeto) el clasificador que mejores resultados ofrece es SVM, seguido de RBF y ambos claramente superiores a KNN.

Considerando el análisis de componentes principales, las gráficas y resultados numéricos se muestran en las Figuras 4.10 (una única plantilla) y 4.16 (una plantilla por sujeto). Se observa que si se considera una plantilla por sujeto los resultados obtenidos por los distintos clasificadores son mejores que si se recorta con una única plantilla común. Considerando el clasificador 1NN el porcentaje es 98,3 % frente a 96,5%. Para RBF se obtiene un 93,13% frente a 92,37%. En caso de SVM ambos clasificadores tienen un 100% de aciertos, sin embargo el rango del parámetro en el cual se obtiene dicho porcentaje en el caso de una plantilla por sujeto es mayor que en el caso de una plantilla única. De nuevo, puede observarse que el clasificador que mejor resultado obtiene es SVM, seguido de RBF y por último, KNN.

Considerando únicamente imágenes recortadas con una plantilla por sujeto, los mejores resultados se obtienen en el caso de utilizar imágenes procesadas con análisis de componentes principales frente a las imágenes de baja resolución, independientemente del clasificador considerado.

Podemos concluir que el uso de una plantilla por sujeto nos permite tener una localización de la cara más robusta, lo cual repercute en la obtención de mejores resultados de verificación correcta por parte de los clasificadores.

Capítulo 5

Reconocimiento tridimensional de caras.

5.1.- Introducción.

En este capítulo se detalla el trabajo realizado en la verificación de personas partiendo de datos tridimensionales de la cara. Para ello, se utilizaron características geométricas en el reconocimiento. Los datos en tres dimensiones se obtuvieron a partir de dos cámaras calibradas (par estéreo). En estas condiciones es posible obtener una nube de puntos poco densa en tres dimensiones que representa a la cara. Con esta condición se decidió realizar un experimento limitado para analizar el problema de reconocimiento de caras humanas en el espacio tridimensional. Ante el surgimiento recientemente de sistemas de scanner láser a precios cada día más asequibles, se realizaron una serie de experimentos limitados para analizar la validez del enfoque tridimensional ante sujetos no colaborativos, ya que en dos dimensiones se observó que la tasa de verificación correcta empeoraba cuando el sujeto adoptaba gestos o muecas.

Una de las decisiones que se tomó en esta tesis fue la de realizar la marcación manual de los puntos en la cara. Una vez tomada esta decisión y seleccionados un conjunto amplio de puntos en la cara (obtenidos de los habitualmente considerados en medicina legal y forense), era preciso estudiar su robustez frente a variaciones en el sujeto que efectúa la marcación. Esto permitió descartar una serie de puntos con gran variabilidad en el marcado manual. Otros puntos se descartaron debido a que el pelo o una ligera rotación de la cara los oculta muy fácilmente. El resultado fue la selección de 14 puntos. El estudio se realizó partiendo de imágenes bidimensionales, ya que la marcación de puntos se hizo sobre este tipo de imágenes. Teniendo en cuenta que el objetivo de este capítulo es realizar un trabajo preliminar de la viabilidad del reconocimiento tridimensional de caras humanas no se desarrolló un estudio exhaustivo de los puntos y características mejores (que, tal vez, constituya por sí mismo otra tesis).

Un proceso análogo se desarrolló con el estudio de características geométricas. Inicialmente se consideraron 198 características. El coeficiente de variación de Pearson permitió ordenar aquellas cuya variabilidad era menor. El ratio de Fisher permitió seleccionar aquellas cuyo poder discriminante es mayor. Por último, se calculó la transformada PCA y se obtuvieron aquellas características que más influyen en el primer componente principal. Como resultado final de este proceso se ha obtenido un conjunto de características adecuadas para el reconocimiento. El estudio se llevó a cabo considerando las características en el plano de la imagen. Se seleccionaron aquellas que

incluían los 14 puntos seleccionados anteriormente y se procedió al cálculo de las características en tres dimensiones. Puede observarse que el hecho de utilizar características en tres dimensiones evita la corrección de ángulos que debería realizarse si se estudiaseen características en dos dimensiones debido a la posición relativa cámara-sujeto y a la pose que adopte el sujeto.

En la presente tesis se describe también un modelo de cara en 3D, al cual se le puede añadir textura y que permite adaptarse a la reconstrucción en 3D de los puntos marcados en una cara. Este modelo no fue considerado para la verificación facial ya que los resultados iniciales, aunque prometedores, no tienen el grado de detalle necesario para distinguir a varios individuos. Sin embargo, con una apropiada captura de datos, por ejemplo proveniente de un escáner láser, el modelo puede ser adecuado para el reconocimiento facial a partir de vistas generadas sintéticamente.

5.2.- Calibración de cámaras y reconstrucción tridimensional. Marcación de puntos.

La Figura 5.1 muestra la geometría estéreo considerada. La calibración de las cámaras permite calcular la matriz de proyección de un punto tridimensional en la imagen (sobre un plano de dos dimensiones). La plantilla de calibración se utilizó para calcular dicha matriz para cada cámara por medio de un ajuste de mínimos cuadrados, ya que se disponía de un número elevado de puntos. Para estimar los errores cometidos en la reconstrucción tridimensional, se procedió a marcar una serie de puntos sobre un objeto y se midieron las distancias reales entre esos puntos. El objeto mide aproximadamente unos 10 cm de largo y los puntos reconstruidos estaban separados entre sí unos 3 cm. Se procedió a colocar el objeto y a obtener dos imágenes (un par estéreo), se estimaron las coordenadas tridimensionales de los puntos (tras un marcado manual de los mismos), y se estimaron con los puntos reconstruidos las distancias que los separaban. Mediante el método de mínimos cuadrados, se halló que el error medio cometido es de 2.1 mm (con una desviación típica de 1.6). Se pueden encontrar más detalles en [Gui98].

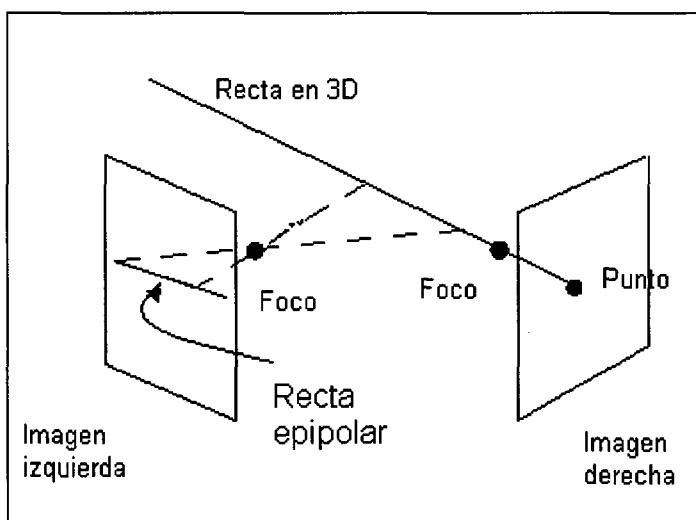


Figura 5.1. Esquema de la geometría estéreo considerada.

El marcado manual de puntos es una tarea bastante tediosa y sujeta a errores. Pero en el caso de utilizar imágenes obtenidas de un par estéreo calibrado, se puede

emplear la recta epipolar para precisar el marcado de puntos. En la Figura 5.2 se ha marcado manualmente un punto en la imagen derecha (concretamente la comisura externa del párpado derecho del sujeto) y aparece la recta epipolar en la imagen izquierda, donde se marcará el punto correspondiente.

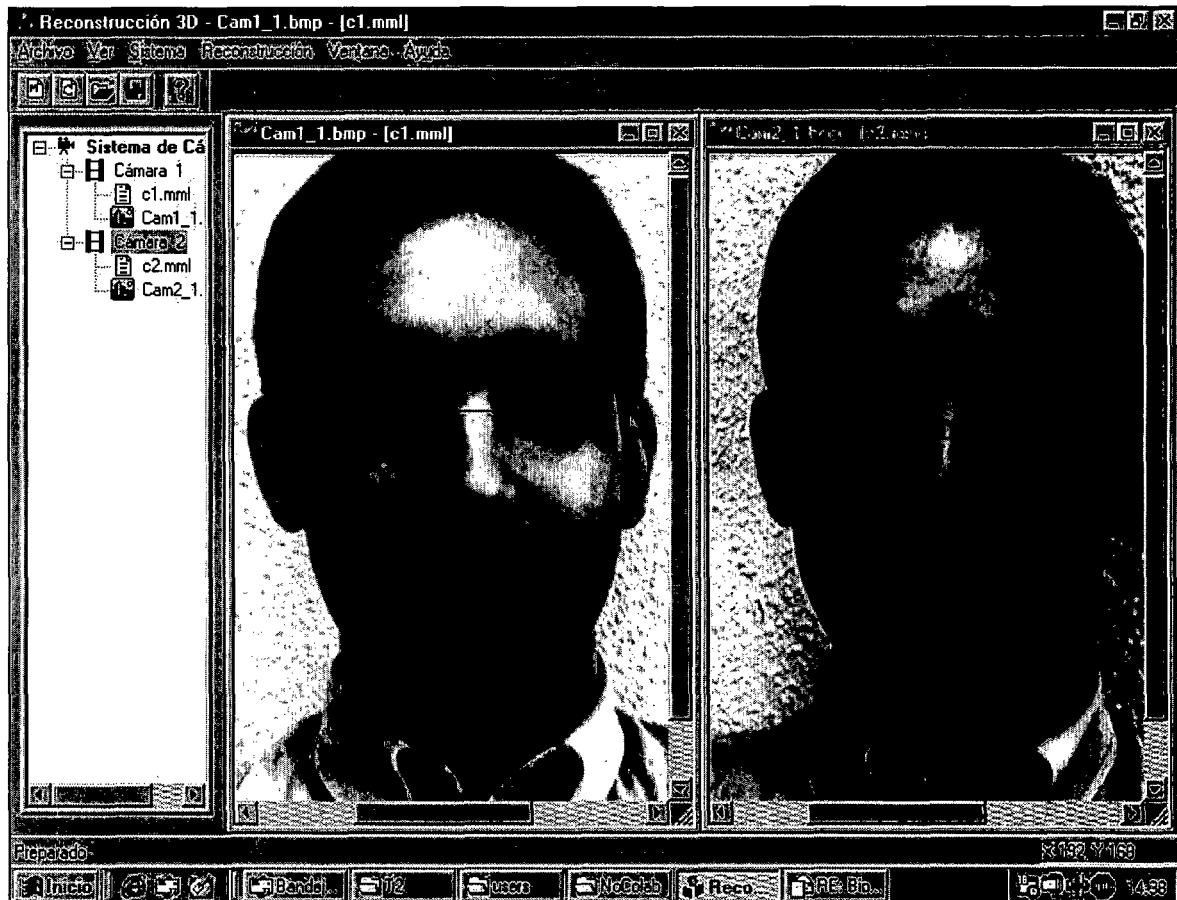


Figura 5.2. Marcado de un punto en la imagen derecha y su correspondiente recta epipolar en la imagen izquierda.

5.3.- Características geométricas.

Antes de describir los experimentos realizados en el reconocimiento tridimensional de caras, se van a definir los conceptos que se van a utilizar. La *somatometría* es la parte de la antropología que se encarga de estudiar todas las medidas cuantificables sobre el cuerpo del vivo. Los *puntos somatométricos* son puntos característicos del cuerpo humano en vivo fácilmente determinables, de forma estándar. Entre estos puntos se pueden calcular medidas características del cuerpo humano. También se puede hablar de *puntos craniométricos*, que son puntos característicos del cráneo humano. Muchos de estos puntos tienen correspondencia directa con los puntos somatométricos que se sitúan sobre el rostro humano.

Desde el punto de vista antropológico, se define la *cara* como la parte anterior de la cabeza desde el principio de la frente a la punta de la barbilla. Se puede situar sobre la cara una línea imaginaria llamada *línea media*, que se define como aquella línea

que divide el rostro en dos partes simétricas. Los puntos somatométricos no pertenecientes a la línea media aparecen siempre emparejados, debido a la simetría de la cara. A estos puntos se los llama *pares* o *laterales*. A los situados sobre la línea media se los llama *impares*.

Se denomina *expresión neutra* de la cara a aquella en la que el rostro aparece con rasgos de expresividad normales y carente de gestos. Es decir el rostro no aparece con aspecto ni sonriente ni triste, los ojos no aparecen guiñados, etc.

La presente tesis considera solamente puntos somatométricos del frontal del rostro humano; gran parte de ellos presentan una correspondencia directa con alguno de los puntos craniométricos. Estos puntos son especialmente interesantes para un sistema automático que tenga como objetivo el reconocimiento de personas a partir de su rostro, debido a que son muy robustos frente a cambios que puedan aparecer sobre fotografías frontales de un mismo individuo que hayan sido tomadas en situaciones distintas. Las fotografías tomadas con una diferencia de tiempo adecuada van a presentar diferencias debido al crecimiento del cabello, a la presencia o no de barba, o simplemente diferencias debidas al uso de gafas o no. No es imprescindible que las fotografías sean tomadas con diferencias de tiempo considerables, ya que unas fotografías tomadas de forma casi consecutiva pueden presentar cambios debidos a variaciones en la situación anímica del individuo. Así, la posición de los puntos de las comisuras labiales cambian cuando el rostro de una persona presenta un aspecto serio o cuando está sonriendo. Sin embargo algunos puntos craniométricos no tienen correspondencia directa con puntos somatométricos y son bastante difíciles de marcar sobre un rostro vivo.

En la Figura 5.3 se muestra una serie de puntos somatométricos sobre el frontal del rostro del ser humano en vivo. Aunque aquí se muestren únicamente los puntos somatométricos en una vista frontal, algunos de estos puntos se pueden marcar con mayor exactitud sobre una vista de perfil de la cabeza. En la Figura 5.3 se han indicado todos sobre el frontal ya que, aun así, pueden ser situados con relativa facilidad sobre la cara. Este es el caso de los puntos marcados sobre las orejas. En la figura se han suprimido algunos puntos, que se pueden encontrar en libros de antropometría, por la dificultad que presentaban para ser situados sobre el rostro vivo.

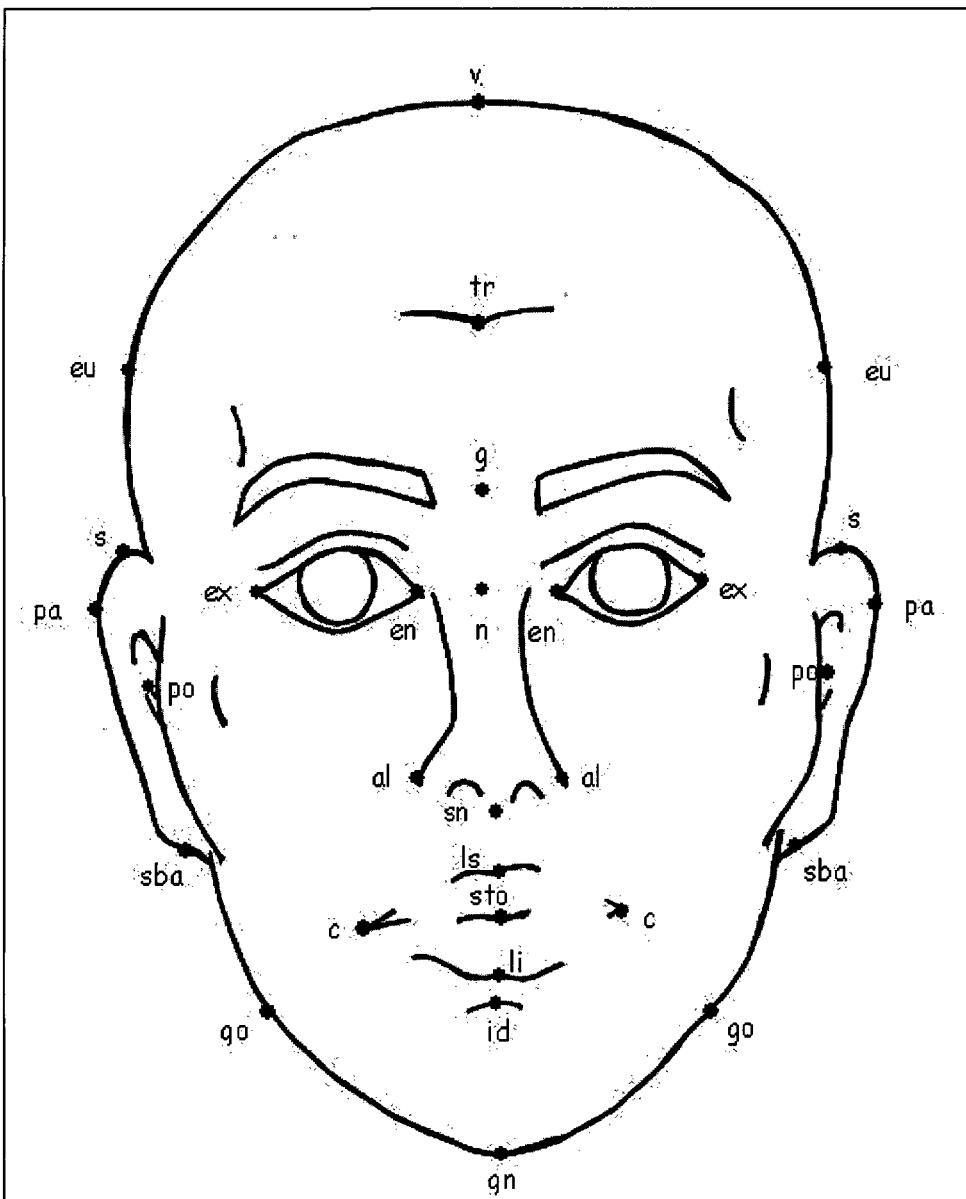


Figura 5.3. Puntos somatométricos.

Los puntos somatométricos que aparecen en la Figura 5.3 se definen a continuación, ya que, aunque algunos de ellos quedan completamente determinados en la figura otros no [Rev91]. Sólo se consideran en esta tesis aquellos puntos que han sido obtenidos a partir de la medicina legal y forense.

Puntos somatométricos en la línea media o impares:

VERTEX (*v*): Punto más alto del cráneo.

TRIQUIO o CRINION (*tr*): Punto situado en el centro de la línea que separa la frente de la cabellera. En un punto muy variable de unas personas a otras.

GLABELA (*g*): Punto medio en la eminencia de la parte inferior de la frente por encima de la raíz de la nariz en el entrecejo. En la mayor parte de los casos es el punto más saliente hacia delante, pero a veces sobresalen más los arcos supraciliares.

NASION (n): Se sitúa en la raíz de la nariz, donde se une la sutura frontal con los huesos propios de la nariz. Sobre el rostro del vivo se suele encontrar entre los lacrimales.

ACANTION o SUBNASAL (sn): Punto inferior del tabique nasal. Sobre el rostro vivo es el punto en el que la nariz alcanza su punto más bajo.

LABIAL SUPERIOR (ls): Punto medio de la línea que une la piel con la mucosa labial, por encima del labio superior.

ESTOMIO (sto): Punto medio de la abertura bucal con los labios cerrados.

LABIAL INFERIOR (li): Punto medio de la línea donde se une la piel con la mucosa del labio inferior.

INFRADENTAL (id): Punto anterior y superior entre los incisivos medios situado en el borde alveolar del maxilar inferior. Es el punto equivalente al *prostio* del maxilar superior.

GNATIO o MENTON (gn): Es el punto mas bajo de la barbilla (maxilar inferior).

Puntos somatométricos pares o laterales:

EURION (eu): Punto más saliente del cráneo hacia los lados. Este punto es bastante variable ya que puede situarse o bien en el parietal o bien en el temporal, dependiendo de cual de ellos sobresalga más.

FORNIX EXTERNO (ex): Punto situado en la comisura exterior de los párpados.

FORNIX INTERNO o LAGRIMAL (en): Punto situado en la comisura interna de los párpados.

ALA NASAL (al): Punto en el que la nariz alcanza su anchura mayor.

COMISURA LABIAL (c): Punto de unión de los labios superior e inferior.

PORION (po): Punto más elevado del contorno del agujero auditivo. Sobre el rostro del vivo se sitúa hacia la mitad de la parte anterior de la oreja.

GONION (go): Punto más bajo, más posterior y más externo del ángulo mandibular, donde se unen las ramas horizontal y ascendente del maxilar inferior.

SUPERAURAL (s): Punto más alto del borde del hélix en la oreja.

POSTAURAL (pa): Punto más saliente de la oreja a izquierda o derecha.

SUBAURAL (sba): Punto más bajo del lóbulo de la oreja.

5.3.1.- Medidas en la cara.

Las medidas más sencillas de calcular a partir de los puntos somatométricos son distancias y ángulos. Estas son precisamente las características consideradas para describir los rostros. En la presente tesis se han tomado las características que emplean sólo puntos somatométricos del frontal. Para realizar un estudio de las características más adecuadas para la descripción de rostros, se ha tenido que elegir un número elevado de ellas. A continuación se muestran, a modo de ejemplo, algunas de estas características.

5.3.1.1.- Distancias sobre la cara.

Según Reverte [Rev91] las *distancias* más usuales en la cara son su longitud, su anchura bizigomática y la longitud y anchura esquelética de la nariz. Sobre el rostro vivo, la longitud de la cara se toma desde la raíz de los cabellos al mentón.

Por otro lado se pueden obtener resultados interesantes con un número muy reducido de características. Basta con emplear la longitud y la anchura de la cara para distinguir unas razas de otras. De este modo se puede usar un índice Longitud/Anchura para la distinción entre razas. El uso de índices resulta esencial cuando se trabaja con distancias tomadas sobre fotografías de individuos que pueden ser realizadas a distancias variables. El problema de la variación de las distancias medidas en distintas fotografías puede resolverse mediante normalizaciones con distancias, también medidas sobre la cara; este sistema es muy robusto frente a la inestabilidad de los datos. Además, el uso de índices no sólo facilita las operaciones matemáticas, sino que también evita el empleo de unidades, si bien es cierto que se produce una pequeña pérdida de información. Ejemplos de estos índices son los siguientes:

INDICE FACIAL: es debido a Broca [Bro75] y compara el *lado ofrio alveolar* con el *diámetro bizigomático*. Esta definición queda mucho más clara en la siguiente fórmula:

$$IF = \frac{d(tr, sn)}{d(po, po)}$$

INDICE AURICULAR: relación entre la anchura y la altura de la oreja.

INDICE ALVEOLAR: mide el grado de prominencia de los maxilares.

5.3.1.2.- Angulos sobre la cara.

El otro tipo de medidas que se puede realizar sobre el rostro, a partir de los puntos somatométricos, son los *ángulos*. Además de proporcionar más medidas de la cara, los ángulos permiten efectuar correcciones sobre la posición del rostro en fotografías que no sean frontales.

ANGULO FACIAL: según Camper [Cam70] es el definido por las siguientes líneas: una vertical que pasa por el punto más saliente de la frente y por el punto alveolar, y otra horizontal que pasa por el conducto auditivo y también por el punto alveolar. El ángulo facial de Camper ha sido modificado en alguno de sus puntos por otros autores.

ANGULO PALATINO: es debido a Cuvier y Geoffroy Saint-Hilaire [11] y está definido por la línea horizontal de Camper con la línea media palatina que divide en dos el arco alveolar.

5.3.2.- Definición de las características.

Las características geométricas que se han empleado han sido distancias y ángulos calculados a partir de unos puntos característicos del rostro, usando distancias euclídeas y corrigiendo la distorsión introducida por las cámaras de vídeo.

Todos los puntos característicos del rostro considerados han sido del frontal propuestos por la literatura en antropometría. En la Figura 5.4 se muestran estos puntos. Sin embargo en antropometría se utilizan más puntos, algunos de los cuales pueden ser marcados sobre el frontal y no se han tenido en cuenta por su dificultad para ser situados sobre una imagen frontal del rostro vivo (el cabello dificulta la localización de algunos puntos).

Todas las distancias se han normalizado a la distancia $d(ex, ex)$, entre los dos puntos *fornix externos*. Con esto se consigue que las características sean invariantes frente a la distancia a la que hayan sido tomadas las imágenes. Por otro lado la distancia $d(ex, ex)$ es una distancia muy robusta frente a cambios en el rostro debidos a expresiones faciales, ya que los puntos no cambian de posición al cambiar la expresión de la cara. Además, incluso si hay presencia de gafas o vello facial son puntos relativamente fáciles de marcar. Con los ángulos no se realiza transformación alguna ya que estos ya son invariantes frente a la distancia de toma de la imagen.

A partir de los puntos somatométricos que se indican en la Figura 5.3 se han calculado un total de 199 características geométricas, de las cuales 143 eran distancias y 56 ángulos. La distancia de normalización sólo ha sido empleada a tales efectos, por lo tanto se calcularon 142 distancias normalizadas. Todas las características han sido escogidas de forma intuitiva sin preocupación por posibles redundancias. El objetivo buscado consistía en capturar la mayor información estructural posible. Una aplicación posterior de una transformada tipo PCA se puede encargar de eliminar redundancias. A continuación se muestran las características escogidas para la descripción de los individuos.

Distancias:

<i>Distancia de normalización:</i>	$D(ex_1, ex_2)$
<i>Contorno del rostro:</i>	$D(po_1, po_2) \quad d(eu_1, eu_2) \quad d(v, gn)$
<i>Ojos:</i>	$D(en_1, en_2) \quad d(ex_1, en_1) \quad d(ex_1, en_2) \quad d(ex_2, en_2) \quad d(ex_2, en_1)$
<i>Nariz:</i>	$D(al_1, al_2) \quad d(sn, al_1) \quad d(sn, al_2) \quad d(n, sn) \quad d(n, al_1) \quad d(n, al_2)$
<i>Boca:</i>	$d(c_1, c_2) \quad d(ls, li) \quad d(c_1, sto) \quad d(sto, c_2) \quad d(ls, sto) \quad d(sto, li) \quad d(c_1, ls)$ $d(ls, c_2) \quad d(c_2, li) \quad d(c_1, li)$
<i>Orejas:</i>	$d(s_1, s_2) \quad d(pa_1, pa_2) \quad d(pa_1, sba_1) \quad d(sba_1, sba_2) \quad d(s_1, sba_1) \quad d(s_2, sba_2)$ $d(pa_2, sba_2) \quad d(s_1, pa_1) \quad d(s_2, pa_2)$
<i>Mixtas: Distancia del gonian al resto de puntos.</i>	$d(g, v) \quad d(g, eu_1) \quad d(g, eu_2) \quad d(g, po_1) \quad d(g, po_2) \quad d(g, gn) \quad d(g, id) \quad d(g, n)$ $d(g, ex_1) \quad d(g, ex_2) \quad d(g, en_1) \quad d(g, en_2) \quad d(g, al_1) \quad d(g, al_2) \quad d(g, sn) \quad d(g, c_1)$ $d(g, c_2) \quad d(g, ls) \quad d(g, li) \quad d(g, sto) \quad d(g, s_1) \quad d(g, s_2) \quad d(g, pa_1) \quad d(g, pa_2)$ $d(g, sba_1) \quad d(g, sba_2)$
<i>Mixtas: Distancia del nasion al resto de puntos.</i>	$d(n, v) \quad d(n, eu_1) \quad d(n, eu_2) \quad d(n, po_1) \quad d(n, po_2) \quad d(n, gn) \quad d(n, id) \quad d(n, g)$ $d(n, ex_1) \quad d(n, ex_2) \quad d(n, en_1) \quad d(n, en_2) \quad d(n, c_1) \quad d(n, c_2) \quad d(n, ls) \quad d(n, li)$ $d(n, sto) \quad d(n, s_1) \quad d(n, s_2) \quad d(n, pa_1) \quad d(n, pa_2) \quad d(n, sba_1) \quad d(n, sba_2)$
<i>Mixtas: Distancia del estomio al resto de puntos.</i>	$d(sto, v) \quad d(sto, eu_1) \quad d(sto, eu_2) \quad d(sto, po_1) \quad d(sto, po_2) \quad d(sto, gn)$ $d(sto, id) \quad d(sto, n) \quad d(sto, ex_1) \quad d(sto, ex_2) \quad d(sto, en_1) \quad d(sto, en_2)$ $d(sto, al_1) \quad d(sto, al_2) \quad d(sto, sn) \quad d(sto, sto) \quad d(sto, s_1) \quad d(sto, s_2)$ $d(sto, pa_1) \quad d(sto, pa_2) \quad d(sto, sba_1) \quad d(sto, sba_2)$
<i>Mixtas: Distancia del infradental al resto de puntos.</i>	$d(id, v) \quad d(id, eu_1) \quad d(id, eu_2) \quad d(id, po_1) \quad d(id, po_2) \quad d(id, gn) \quad d(id, g)$ $d(id, n) \quad d(id, ex_1) \quad d(id, ex_2) \quad d(id, en_1) \quad d(id, en_2) \quad d(id, al_1) \quad d(id, al_2)$ $d(id, sn) \quad d(id, c_1) \quad d(id, c_2) \quad d(id, ls) \quad d(id, li) \quad d(id, sto) \quad d(id, s_1) \quad d(id, s_2)$ $d(id, pa_1) \quad d(id, pa_2) \quad d(id, sba_1) \quad d(id, sba_2)$
<i>Distancias entre puntos de la boca y ojos:</i>	$D(c_1, ex_1) \quad d(c_1, en_1) \quad d(c_2, ex_2) \quad d(c_2, en_2)$
<i>Distancias entre puntos de la nariz y los ojos:</i>	$d(sn, en_1) \quad d(sn, en_2) \quad d(sn, ex_1) \quad d(sn, ex_2) \quad d(al_1, ex_1) \quad d(al_1, en_1)$ $d(al_2, ex_2) \quad d(al_2, en_2)$

Angulos:

<i>Ángulos formados con la línea media:</i>	$ang(sto, n, v) \quad ang(sto, n, eu_1) \quad ang(sto, n, eu_2) \quad ang(sto, n, po_1)$ $ang(sto, n, po_2) \quad ang(sto, n, go_1) \quad ang(sto, n, go_2) \quad ang(sto, n, gn)$ $ang(sto, n, id) \quad ang(sto, n, tr) \quad ang(sto, n, ex_1) \quad ang(sto, n, ex_2)$ $ang(sto, n, en_1) \quad ang(sto, n, en_2) \quad ang(sto, n, al_1) \quad ang(sto, n, al_2)$ $ang(sto, n, sn) \quad ang(sto, n, c_1) \quad ang(sto, n, c_2) \quad ang(sto, n, ls)$ $ang(sto, n, li) \quad ang(sto, n, sto) \quad ang(sto, n, s_1) \quad ang(sto, n, s_2)$ $ang(sto, n, pa_1) \quad ang(sto, n, pa_2) \quad ang(sto, n, sba_1) \quad ang(sto, n, sba_2)$
<i>Ángulos de la parte derecha del rostro, formados con la línea que une los dos puntos fornix externos (ex) de los ojos.</i>	$ang(en_1, en_2, v) \quad ang(en_1, en_2, tr) \quad ang(en_1, en_2, g) \quad ang(en_1, en_2, n)$ $ang(en_1, en_2, sn) \quad ang(en_1, en_2, ls) \quad ang(en_1, en_2, sto) \quad ang(en_1, en_2, li)$ $ang(en_1, en_2, id) \quad ang(en_1, en_2, gn) \quad ang(en_1, en_2, eu_1)$ $ang(en_1, en_2, eu_2) \quad ang(en_1, en_2, po_1) \quad ang(en_1, en_2, po_2)$ $ang(en_1, en_2, go_1) \quad ang(en_1, en_2, go_2) \quad ang(en_1, en_2, en_1)$ $ang(en_1, en_2, en_2) \quad ang(en_1, en_2, al_1) \quad ang(en_1, en_2, al_2) \quad ang(en_1, en_2, c_1)$ $ang(en_1, en_2, c_2) \quad ang(en_1, en_2, s_1) \quad ang(en_1, en_2, s_2) \quad ang(en_1, en_2, pa_1)$ $ang(en_1, en_2, pa_2) \quad ang(en_1, en_2, sba_1) \quad ang(en_1, en_2, sba_2)$

En la Figura 5.4 se muestran algunas de las distancias definidas anteriormente:

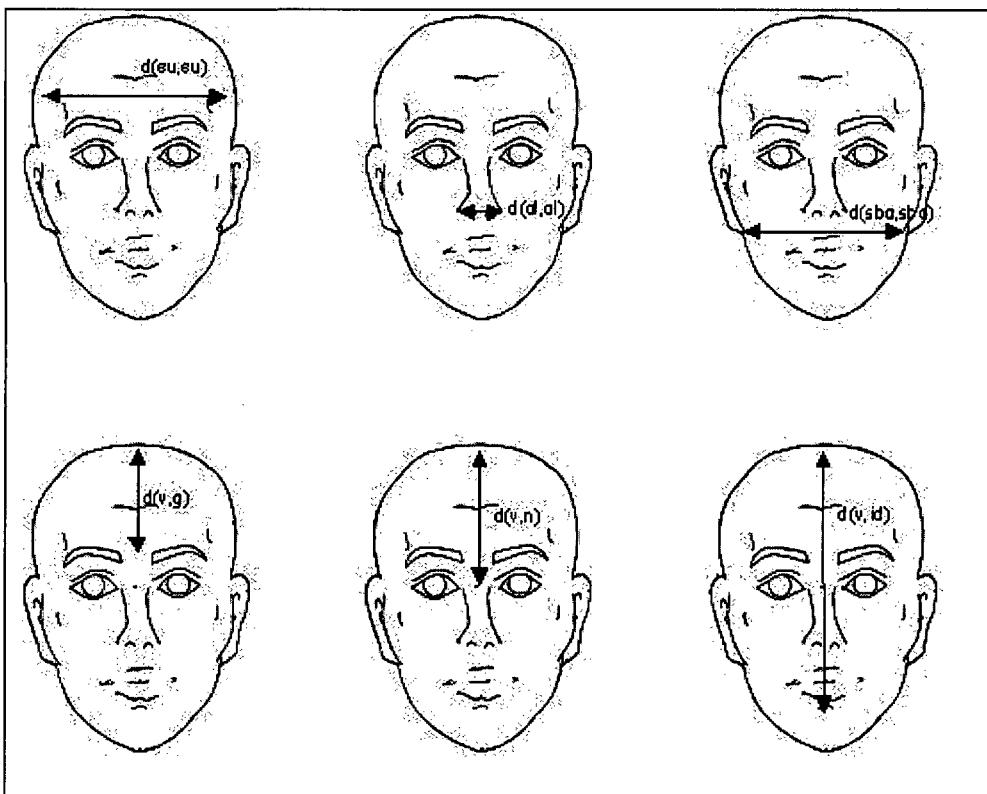


Figura 5.4. Algunas distancias del vector de características.

En la Figura 5.5 se muestran algunos de los ángulos definidos:

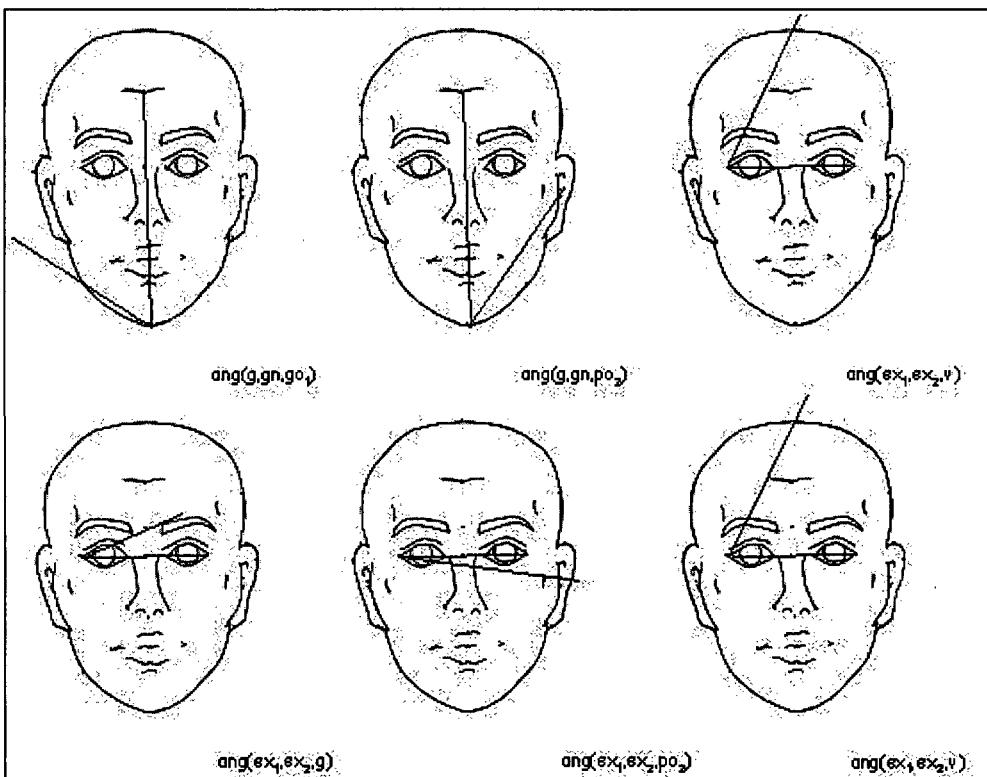


Figura 5.5. Algunos ángulos del vector de características.

Una vez efectuado el cálculo de estas características geométricas, cada clase pasa a ser representada por un vector, $\bar{x} = (x_1, \dots, x_{198})$, en el espacio de dimensión 198 constituido por las características.

5.3.3.- Captura de las coordenadas de los puntos.

La construcción de los vectores de características exige partir de un conjunto de datos que contengan las coordenadas de los puntos somatométricos marcados sobre cada imagen de la base de datos facial. El programa de captura de datos se proporcionó a varias personas distintas. El programa muestra las imágenes de la base de datos facial y un operador humano marca con un puntero las coordenadas de los puntos somatométricos sobre cada una de las imágenes. Nueve personas distintas marcaron los puntos somatométricos sobre imágenes frontales de 20 sujetos. Después del cálculo de las características geométricas, se obtuvo un vector de características por cada imagen y por persona. En total se calcularon 180 vectores de características.

5.3.4- Cálculo de la robustez de los puntos somatométricos y de las características geométricas.

Para medir la robustez de los puntos, se consideraron las 9 muestras de cada punto somatométrico marcadas en cada cara. Con la ayuda de los puntos marcados sobre las distintas caras, se calculó la media de las 9 muestras de un mismo punto somatométrico marcado sobre la misma imagen, así como la desviación típica de cada una de las dos coordenadas de los puntos y en cada una de las imágenes. Se obtuvieron dos desviaciones típicas para cada una de las dos coordenadas de los puntos: σ_x y σ_y . Para el cálculo de la desviación típica, primero se calculó la varianza de las dos coordenadas de cada uno de los puntos y de cada una de las imágenes. También se calculó la media de las desviaciones típicas de cada una de las imágenes correspondientes a los mismos puntos. Para cada punto se calcularon dos desviaciones típicas medias, correspondientes a las dos direcciones del rostro: $\bar{\sigma}_x$ y $\bar{\sigma}_y$.

Los puntos más robustos son aquellos que presentan desviaciones típicas medias menores, ya que son los que se han podido marcar con mayor precisión. Las características más robustas son aquellas calculadas a partir de los puntos somatométricos más robustos.

La robustez de los puntos se ha evaluado mediante la desviación típica calculada a partir de las coordenadas de los puntos somatométricos de todas las imágenes y todas las muestras tomadas por imagen. Los resultados se resumen en la Figura 5.6.



Figura 5.6. Robustez de los puntos somatométricos.

En la Figura 5.6 se han marcado todos los puntos somatométricos y en torno a ellos se dibujó un círculo cuyo radio representa la desviación típica de cada punto. Hay que destacar que se calcularon desviaciones típicas medias para ambas direcciones del plano y se tomó como radio del círculo dibujado el mayor de los dos para que la figura resultara más significativa. Además para marcar los puntos se tomó la media de todas las muestras tomadas de la clase de la figura. En ella se observa que los puntos menos robustos son también los que más dificultad presentan para su marcación, debido a que están tapados por el pelo y la barba. Los más robustos se sitúan en la boca, ojos y nariz y son precisamente los que se marcan con mayor facilidad. En la Tabla 5.1 se presentan las dispersiones de los puntos somatométricos.

Desviación típica	Número de puntos	Puntos somatométricos
$\sqrt{\sigma_x^2 + \sigma_y^2} > 4$	2	v, tr
$\sqrt{\sigma_x^2 + \sigma_y^2} > 3$	9	v, eu ₁ , eu ₂ , go ₁ , go ₂ , tr, sn, pa ₁ , pa ₂
$\sqrt{\sigma_x^2 + \sigma_y^2} > 2$	12	v, eu ₁ , eu ₂ , po ₁ , go ₁ , ex ₁ , tr, sn, pa ₁ , pa ₂ , sba ₁
$\sqrt{\sigma_x^2 + \sigma_y^2} < 2$	18	Po ₂ , gn, id, g, n, ex ₂ , en ₁ , en ₂ , al ₁ , al ₂ , c ₁ , c ₂ , ls, li, sto, s ₁ , s ₂
$\sqrt{\sigma_x^2 + \sigma_y^2} < 1$	2	ls, sto

Tabla 5.1. Robustez de los puntos somatométricos y de las características geométricas.

Para el cálculo de la robustez de las características se empleó el coeficiente de variación de Pearson [Alv94]:

$$C.V. = \frac{\sigma}{m}$$

En donde σ es la desviación típica de las 9 muestras del conjunto de imágenes frontales de una característica y m es la media.

Para poder estimar la robustez de las características, se calculó de nuevo una dispersión media de las mismas divididas por un valor medio de éstas. En la Tabla 5.2 se muestran las características más robustas.

Características con $\sigma/m < 0,002$			
Distancias	Total = 12	Angulos	Total = 7
Característica		Característica	
$D(po_1, po_2)$		$ang(sto, n, eu_1)$	
$D(s_1, s_2)$		$ang(sto, n, eu_2)$	
$D(pa_1, pa_2)$		$ang(en_1, en_2, g)$	
$D(sba_1, sba_2)$		$ang(en_1, en_2, id)$	
$D(g, gn)$		$ang(en_1, en_2, eu_1)$	
$D(g, id)$		$ang(en_1, en_2, g)$	
$D(sto, ex_1)$		$ang(en_1, en_2, c_1)$	
$D(sto, ex_2)$			
$D(sto, en_1)$			
$D(sto, en_2)$			
$D(sto, sn)$			
$D(c_1, en_1)$			
$D(c_2, ex_2)$			

Tabla 5.2. Robustez de las características.

5.3.5.- Estudio del poder discriminante de las características geométricas.

El poder discriminante de las características fue estimado mediante la *ratio de Fisher* [Alv94]. Para su cálculo se necesitaban las características de cada una de las muestras de todas las clases. La ratio de Fisher se define en general partiendo de un conjunto de N clases (en nuestro caso N sujetos), de cada una de ellas se tienen p muestras (p veces se han obtenido las características para cada sujeto) . En esta situación la ratio de Fisher se define como:

$$F = \frac{\frac{1}{N} \sum_{j=1}^N (m_j - \bar{m})^2}{\frac{1}{Np} \sum_{j=1}^N \left\{ \sum_{i=1}^p (X_{ij} - m_j)^2 \right\}}$$

donde m_j es la media del valor de una característica para cada sujeto ($j=1, \dots, N$), \bar{m} es la media de todas las características según la fórmula $\bar{m} = \frac{1}{N} \sum_{j=1}^N m_j$, X_{ij} es el valor de la característica correspondiente a la clase j y a la muestra i .

Para su cálculo se partió de los puntos somatométricos marcados por los 9 sujetos, para generar los valores de las características. Mediante la *ratio de Fisher* se consiguió evaluar la capacidad de las características para separar clases distintas, es decir, su poder de discriminación. Una característica será más discriminante cuanto mayor sea su *ratio de Fisher*. En la Tabla 5.3 se presenta un resumen de estos resultados, en el que se incluyen las 20 características más discriminantes.

Distancias		Angulos	
Ratio de Fisher	Característica	Ratio de Fisher	Característica
251,86	$D(sto, en_2)$	371,59	$ang(sto, n, eu_1)$
223,69	$D(sto, ex_2)$		$ang(sto, n, eu_2)$
207,52	$D(sto, en_1)$		$ang(en_1, en_2, eu_1)$
195,10	$D(id, ex_2)$	30,81	$ang(sto, n, c_1)$
191,92	$D(id, g)$	25,29	$ang(sto, n, gn)$
172,51	$D(al_1, al_2)$	22,93	$ang(sto, n, en_1)$
168,84	$D(g, gn)$	22,32	$ang(sto, n, al_2)$
160,21	$D(sto, ex_1)$	17,22	$ang(sto, n, c_2)$
159,11	$D(c_1, en_1)$	15,66	$ang(sto, n, ex_2)$
149,76	$D(n, li)$	14,34	$ang(en_1, en_2, al_1)$

Tabla 5.3. Poder discriminante de las características.

5.3.6.- Recuperación de las características geométricas con componentes principales mayores.

Una vez realizados los estudios anteriores, se pasó a comprobar aquellas características que influyen más en los componentes principales en el espacio de las características. Para ello se seleccionaron varios conjuntos de características considerando las de mayor poder discriminante junto con algunas cuya medida era muy sencilla y se realizó el cálculo de la correspondiente transformación PCA. Una vez aplicada la transformada PCA a las distintas combinaciones de las muestras de los vectores de características, se llevó a cabo una proyección en un subespacio según los autovalores mayores de la matriz de *autocorrelación*. A los patrones ya proyectados se les aplicó la transformación inversa a la PCA, recuperando así los patrones en el espacio original. Las características con coeficiente mayor son las que se corresponden con las componentes principales mayores. De esta forma, además de disponer de las características con mayor poder discriminante, se puede obtener las que explican la mayor parte de la varianza, para la base de datos considerada.

Dado que la transformada PCA es una aplicación lineal, se puede obtener mediante una ecuación matricial:

$$\vec{y}_i(t_j) = \Phi \vec{x}_i(t_j)$$

En donde los índices i, j indican la clase i y la muestra j . Φ es la matriz de cambio de base, formada por los autovalores de la matriz de autocorrelación.

Para obtener la transformación inversa basta con multiplicar, por la izquierda y en ambos miembros de la igualdad anterior, por la matriz de cambio de base inversa:

$$\begin{aligned}\Phi^{-1} \vec{y}_i(t_j) &= \Phi^{-1} \Phi \vec{x}_i(t_j) \\ \vec{x}_i(t_j) &= \Phi^{-1} \vec{y}_i(t_j)\end{aligned}$$

Finalmente se realizó un estudio estadístico de los coeficientes correspondientes a las mismas características, que consiste en el cálculo de las medias y desviaciones típicas de cada uno de los coeficientes.

Aunque sólo unas pocas características se repiten de forma apreciable entre todas las combinaciones que se probaron, cabe destacar por su relevancia, ya que son las que más porcentaje de la varianza explican, las siguientes características geométricas:

$$d(eu_1, eu_2), d(po_1, po_2), d(sto, sn), d(sto, gn), d(sto, ex_2), d(en_1, en_2), d(sn, al_2).$$

5.3.7.- Puntos y características seleccionados finalmente.

Tras haber obtenido características más discriminantes, se procedió a estudiar los puntos que las definen. Al utilizar una marcación manual de forma directa, se decidieron considerar los puntos más robustos, es decir, se eliminaron aquellos que no aparecen en imágenes frontales o que son fácilmente tapados por el pelo. El resultado final son los 14 puntos cuya definición se incluye en la Tabla 5.4. Su representación gráfica aparece en la Figura 5.7.

<i>Glabela (g):</i>	Punto más saliente del entrecejo.
<i>Nasión (n):</i>	Punto situado en la raíz de la nariz.
<i>Comisura externa derecha (ex₁):</i>	Comisura exterior del párpado del ojo derecho.
<i>Comisura externa izquierda (ex₂):</i>	Comisura exterior del párpado del ojo izquierdo.
<i>Lagrimal derecho (en₁):</i>	Comisura interna del párpado del ojo derecho.
<i>Lagrimal izquierdo (en₂):</i>	Comisura interna del párpado del ojo izquierdo.
<i>Ala nasal derecha (al₁):</i>	Punto de unión (derecha) de la nariz con la cara.
<i>Ala nasal izquierda (al₂):</i>	Punto de unión (izquierda) de la nariz con la cara.
<i>Subnasal (sn):</i>	Punto situado en la base de la nariz.
<i>Labio superior (ls):</i>	Punto medio superior de los labios.
<i>Comisura labial derecha (c₁):</i>	Punto de unión derecho de los labios.
<i>Comisura labial izquierda (c₂):</i>	Punto de unión izquierdo de los labios.
<i>Labio inferior (li):</i>	Punto medio inferior de los labios.
<i>Gnathion (gn):</i>	Punto más bajo de la mandíbula.

Tabla 5.4. Puntos seleccionados para su reconstrucción tridimensional.

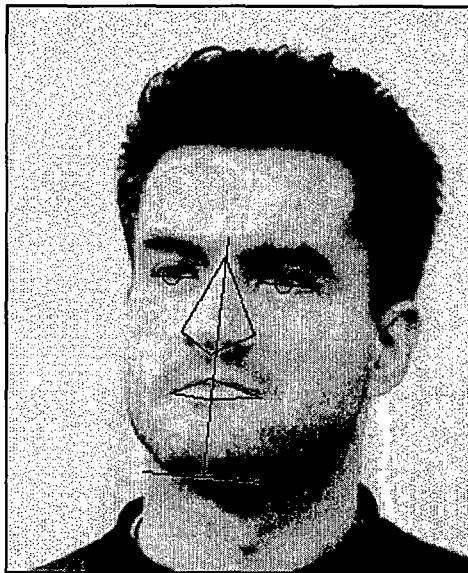


Figura 5.7. Puntos considerados para su reconstrucción en 3D.

Una vez seleccionados estos puntos se procedió a su marcado manual sobre la base de datos. Las características geométricas de nuestro estudio se muestran en la Tabla 5.5. Hay que tener presente que, al ser distancias que se miden en el espacio tridimensional, no hace falta su normalización o la corrección de ángulos debido a la pose del sujeto.

<i>Distancias</i>	$d(n,al_1)$ $d(al_1,al_2)$ $d(sn,al_1)$ $d(al_1,c_1)$ $d(al_1,ex_1)$ $d(c_1,ex_2)$ $d(gn,ex_1)$ $d(en_2,al_2)$ $d(c_2,al_2)$ $d(g,ex_2)$ $d(ex_2,en_1)$ $d(al_2,ex_2)$ $d(sn,c_1)$ $d(ls,c_1)$ $d(c_1,c_2)$ $d(c_1,n)$ $d(c_2,ex_2)$ $d(c_2,li)$ $d(g,ex_1)$ $d(g,en_1)$ $d(g,c_1)$ $d(g,c_2)$ $d(gn,ex_2)$ $d(gn,c_2)$ $d(gn,li)$ $d(c_2,li)$ $d(ls,li)$ $d(al_1,ls)$ $d(al_2,ls)$ $d(c_2,ls)$ $d(en_1,ex_1)$ $d(en_1,al_1)$ $d(en_1,c_1)$ $d(g,en_2)$ $d(n,en_2)$ $d(c_2,en_2)$ $d(n,g)$ $d(n,en_1)$ $d(n,al_2)$ $d(n,sn)$ $d(sn,al_2)$ $d(sn,ls)$ $d(sn,c_2)$
<i>Ratios</i>	$R1 = d(ex_1,en_1) / d(ex_2,en_2)$ $R2 = d(al_1,al_2) / d(n,sn)$ $R3 = d(c_1,c_2) / d(g,li)$ $R4 = d(al_1,al_2) / d(c_1,c_2)$

Tabla 5.5. Características geométricas empleadas en el reconocimiento 3D.

Llegados a este punto, se pudo ya realizar la verificación facial de los sujetos de nuestra base de datos, considerando estas características geométricas. Los resultados del reconocimiento se muestran en el apartado 5.5. Sin embargo, antes de iniciar el proceso del reconocimiento, se decidió realizar un modelo de la cara en 3D que pudiese permitir corregir giros de la cara en la imagen e interpolar los puntos que no pudieran ser marcados manualmente debido a ocultación o a giros en la cara. Este modelo parte de una malla triangular que se adapta a los puntos tridimensionales de la cara de un sujeto y cuyo paso final es dotar al modelo geométrico de textura. Los detalles del modelo se encuentran en el Anexo A.

5.4.- Clasificación de características.

De forma análoga al capítulo precedente, en este caso se consideraron tres clasificadores: K vecinos más cercanos (KNN); redes neuronales del tipo funciones de base radial (RBF); y máquinas de vector soporte (SVM). La entrada a los clasificadores es un vector, en este caso de características geométricas formado por 47 datos (43 distancias y 4 ratios).

El clasificador KNN es lineal y permite definir de forma intuitiva cómo se sitúan los puntos que representan a las diferentes clases en el espacio de características. Con ello se puede deducir si las clases están muy cerca una de otra o si los vectores que definen una clase están muy dispersos o no. Además se ha seleccionado un clasificador basado en redes neuronales del tipo funciones de base radial (RBF), ya que es no lineal y permite mejorar los resultados obtenidos con el anterior clasificador (KNN) y con un coste computacional aceptable. Por último, se decidió incluir las máquinas de vector soporte (SVM) con un kernel lineal, por su adecuación al problema considerado, ya que admite de forma inmediata distinguir entre elementos pertenecientes a dos clases diferentes. Otra razón que motivó la selección de este clasificador son las propiedades teóricas que posee.

5.5.- Resultados experimentales y análisis.

En este apartado se presentarán de forma gráfica los resultados obtenidos por cada clasificador.

De forma análoga a lo presentado en el capítulo anterior, para cada una de los experimentos realizados, se usará la siguiente notación :

- Porcentaje de reconocimiento correcto: $(TP+TN)/(TP+TN+FP+FN)$.
- Porcentaje de falsos negativos: $FN/(FN+TN)$.
- Porcentaje de falsos positivo: $FP/(FP+TP)$.

En donde TP, TN, FP y FN son los valores de verdaderos positivos (TP), verdaderos negativos (TN), falsos positivos (FP) y falsos negativos (FN).

La base de datos con la que se han realizado todas las pruebas está formada por 20 sujetos, con 8 fotos por sujeto (que corresponden a 4 pares estereoscópicos). La mitad de los sujetos son hombres y la mitad son mujeres. Además se completó con un pequeño conjunto de imágenes adicionales de cuatro sujetos (dos hombres y dos mujeres) que ya se encontraban en la base de datos anterior. A cada uno de ellos se les tomaron cuatro fotos con una expresión que no fuera neutra (sonrisa o ligera apertura bucal), de tal forma que el reconocimiento fuera más difícil. Este conjunto constituye la denominada base de datos de sujetos no colaborativos.

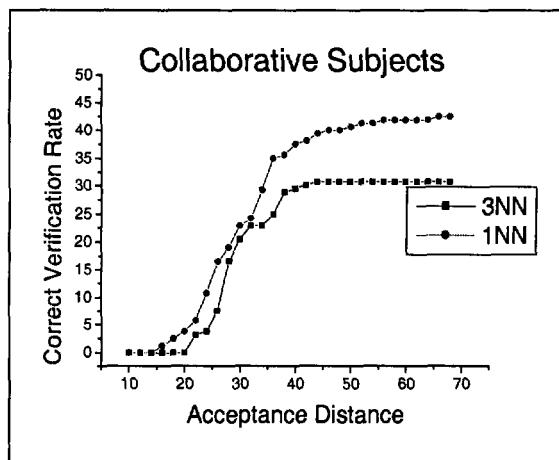
En todas las pruebas, los conjuntos de prueba y de entrenamiento carecen de elementos comunes. A la hora de presentar los resultados, se ha optado por agrupar los datos de prueba de dos formas distintas. La primera forma se ha llamado “no equilibrada”. En este caso cada modelo correspondiente a una clase (a un sujeto) ha sido probado con todos los datos 3D del conjunto de prueba. En el conjunto de prueba existe

una única secuencia de datos 3D para cada sujeto, una para el sujeto cuyo modelo se está probando y 19 secuencias 3D del resto de sujetos. Al ir variando el parámetro, se encuentra que si el sistema rechaza a todos los sujetos, el porcentaje de acierto es $\frac{18}{19} = 94.73\%$, ofreciendo un resultado que puede ser engañoso. Esta es la forma en que se suelen mostrar los resultados en la bibliografía (por ejemplo [Tur91]).

En consecuencia, se ha optado por mostrar también los resultados con un conjunto de prueba en el que el número de datos 3D del sujeto fuese igual al número de datos 3D de otros sujetos (en nuestro caso ese número es de uno). Este conjunto se denominó “equilibrado”. Para mostrar los resultados se ha utilizado una validación cruzada, considerando la media de ocho pruebas para cada modelo. Cada una de estas pruebas consta de una secuencia de datos 3D del sujeto y otra secuencia de datos 3D seleccionada aleatoriamente entre las secuencias de los otros sujetos.

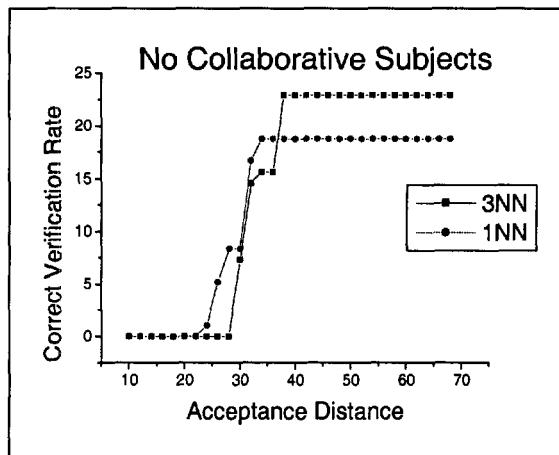
5.5.1.- Resultados KNN.

Se empleó la técnica K vecinos más cercanos (KNN); se presentan aquí los resultados obtenidos para $k = 1$ y $k = 3$. Se han utilizado dos conjuntos de prueba: uno formado por características geométricas de los sujetos en actitud colaborativa, y otro conjunto de prueba formado por sujetos en actitud no colaborativa (gestos exagerados). Hay que tener en cuenta que se dispone de pares estéreo de todos los sujetos en actitud colaborativa, pero sólo se dispone de datos 3D de unos pocos sujetos en actitud no colaborativa.



Parámetro	1NN	3NN
12	0	0
16	1,25	0
18	2,5	0
22	5,72	3,16
24	10,76	3,78
28	19,05	16,63
30	22,90	20,39
34	29,22	22,93
36	34,98	24,94
40	37,48	29,46
42	38,11	30,08
46	39,98	30,71
48	39,98	30,71
54	41,26	30,71
64	41,89	30,71
66	42,51	30,71

Figura 5.8. Resultados de la técnica KNN para sujetos colaborativos (características geométricas en 3D).



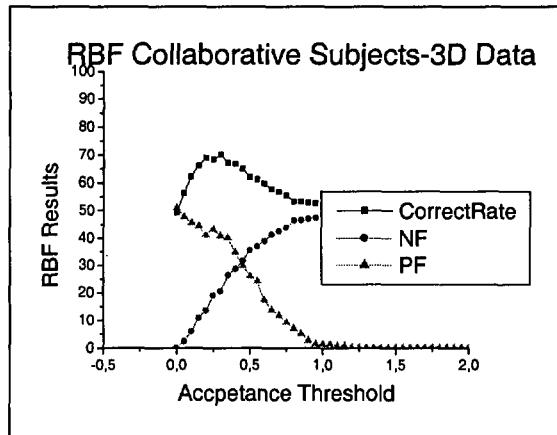
Parámetro	1NN	3NN
22	0	0
24	1,04	0
26	5,20	0
28	8,33	0
30	8,33	7,29
32	16,67	14,58
34	18,75	15,62
36	18,75	15,62
38	18,75	22,91
44	18,75	22,91
50	18,75	22,91
52	18,75	22,91
60	18,75	22,91
68	18,75	22,91

Figura 5.9. Resultados de la técnica KNN para sujetos no colaborativos (características geométricas en 3D).

5.5.2.- Resultados RBF.

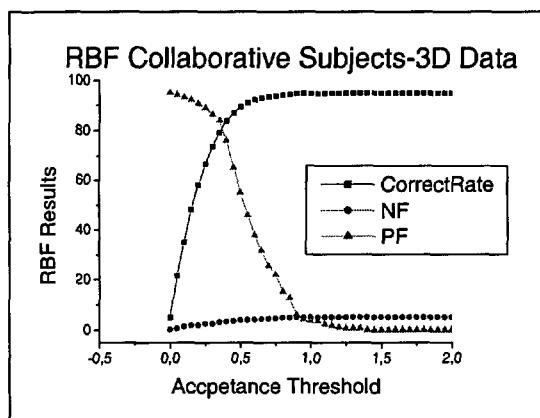
Tras utilizar la técnica de redes neuronales del tipo funciones de base radial (RBF), se presentan aquí los resultados obtenidos. Al igual que con KNN, se han utilizado dos conjuntos de prueba: uno formado por datos 3D de los sujetos en actitud colaborativa y otro conjunto de prueba formado por sujetos en actitud no colaborativa, igual que en el caso anterior.

La primera gráfica (Figura 5.10) muestra los resultados con sujetos colaborativos; en este caso el conjunto de prueba contiene el mismo número de sujetos para los que la verificación debería ser positiva y negativa (es lo que se ha llamado caso “equilibrado”). La Figura 5.11 corresponde a las mismas condiciones de la Figura 5.10, pero ahora el conjunto de prueba contiene un solo sujeto de verificación correcta y 19 de verificación incorrecta (es el caso “no equilibrado”). La Figura 5.12 corresponde a la prueba realizada con los sujetos no colaborativos, en donde sólo existe el caso de “no equilibrados”, al carecer de pares estéreo de todos los sujetos. En este caso no aparece la gráfica correspondiente a sujetos no colaborativos balanceada ya que sólo se dispone de muy pocos casos de prueba (según se describe en 3.3).



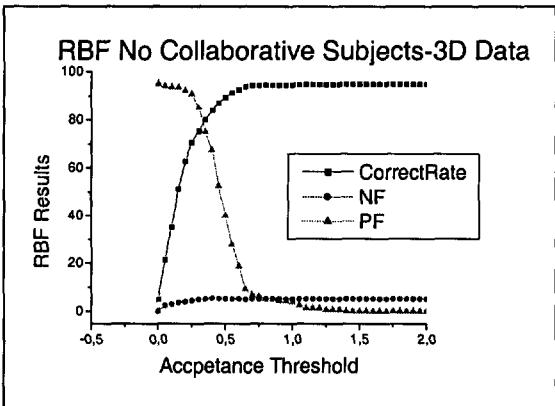
Parámetro	%Correctos	%FN	%FP
0	49,16	0	50,83
0,05	56,37	2,41	47,70
0,3	70,14	20,51	40,74
0,45	64,96	31,54	29,78
0,65	57,86	41,23	13,79
0,75	55,53	43,82	9,30
0,95	52,66	47,27	1,30
1,35	51,14	48,85	0
1,65	51,14	48,85	0
1,8	51,14	48,85	0
2,45	50,83	49,16	0
2,7	50,83	49,16	0
3	50,83	49,16	0
3,25	50,83	49,16	0
3,45	50,83	49,16	0

Figura 5.10. Resultados de la técnica RBF para sujetos colaborativos balanceados (características geométricas en 3D).



Parámetro	%Correcto	%FN	%FP
0	5	0	95
0,3	73,68	2,36	86,33
0,45	87,03	3,37	64,99
0,65	92,92	4,26	31,72
0,75	93,74	4,51	21,95
1,1	94,86	4,94	2,29
1,4	95,03	4,96	0
1,65	95,03	4,96	0
1,8	95,03	4,96	0
2	95	5	0
2,45	95	5	0
3	95	5	0
3,45	95	5	0

Figura 5.11. Resultados de la técnica RBF para sujetos colaborativos no equilibrados (características geométricas en 3D).



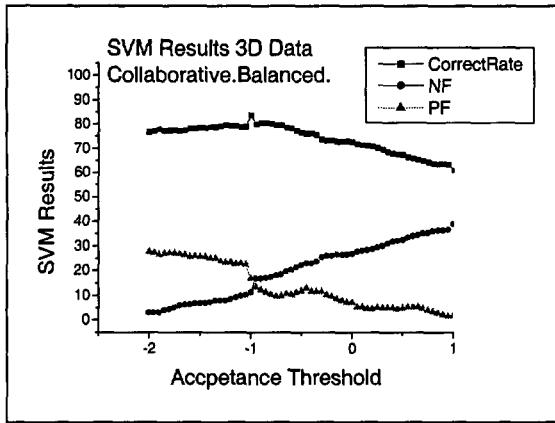
Parámetro	%Correctos	%FN	%FP
0	5	0	95
0,05	21,30	2,23	94,22
0,3	75,41	4,89	85,23
0,45	87,13	5,14	52,50
0,65	94,01	5,02	9,37
0,75	94,47	5	6,25
0,95	94,63	5	4,37
1,1	94,89	5	1,25
1,4	95	5	0
1,65	95	5	0
2	95	5	0
2,3	95	5	0
2,7	95	5	0
3	95	5	0
3,3	95	5	0
3,45	95	5	0

Figura 5.12. Resultados de la técnica RBF para sujetos no colaborativos no equilibrados (características geométricas en 3D).

5.5.3.- Resultados SVM.

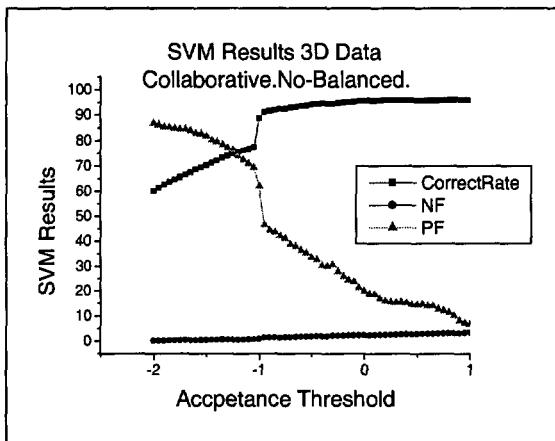
Una vez utilizada la técnica del clasificador tipo máquinas de vector soporte (SVM), con kernel lineales, se presentan los resultados obtenidos. También, como para las dos técnicas anteriormente mencionadas, se han utilizado dos conjuntos de prueba: uno formado por datos 3D de los sujetos en actitud colaborativa y otro conjunto de prueba formado por sujetos en actitud no colaborativa.

La Figura 5.13 se ha obtenido con sujetos colaborativos, en este caso el conjunto de prueba contiene el mismo número de elementos en que la verificación debería ser positiva y negativa (es lo que anteriormente ya se ha llamado caso “equilibrado”). La Figura 5.14 corresponde a las mismas condiciones de la Figura 5.13, pero ahora el conjunto de prueba contiene un solo sujeto de verificación correcta y 19 de verificación incorrecta (es el caso previamente mencionado “no equilibrado”). La Figura 5.15 corresponde a la prueba realizada con los sujetos no colaborativos, donde sólo existe el caso de “no equilibrados”, al carecer de pares estéreo de todos los sujetos.



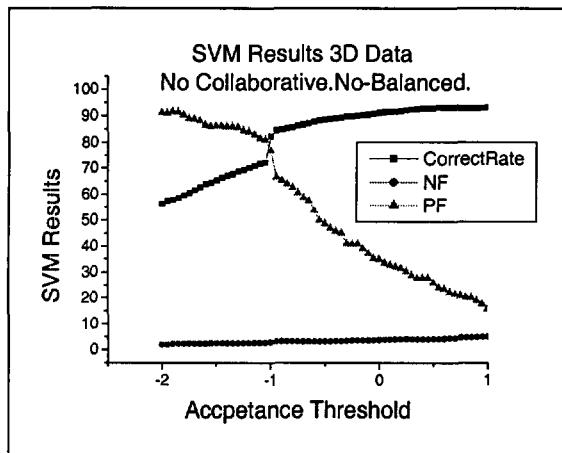
Parámetro	%Correctos	%FN	%FP
-2	76,66	2,86	27,64
-1,55	78,06	6,59	25,86
-1,35	78,61	7,85	24,83
-1,05	78,78	10,07	22,35
-0,9	80,34	16,58	12,18
-0,6	78,02	20,13	10,30
-0,35	75,43	23,53	11,42
0	72,61	26,74	7,11
0,1	71,38	28,00	5,23
0,45	67,64	31,97	4,26
0,7	64,7	34,96	4,54
1	61,05	38,78	1,42
1,1	56,88	43,02	2,58
1,3	54,65	45,26	1,93
1,45	53,11	46,85	1,25
1,7	47,60	52,38	0,62
1,75	47,28	52,69	0,62
2	46,97	53,00	0,62

Figura 5.13. Resultados de la técnica SVM para sujetos colaborativos equilibrados (características geométricas en 3D).



Parámetro	%Correctos	%FN	%FP
-2	60,20	0,30	86,76
-1,55	69,64	0,58	82,66
-1,35	73,41	0,69	78,43
-1,05	77,50	0,86	69,30
-0,9	91,65	1,50	44,31
-0,6	93,59	1,78	36,41
-0,35	94,62	2,11	30,03
0	95,61	2,42	19,87
0,45	95,88	2,91	14,59
0,7	95,91	3,15	12,69
1	96,00	3,53	6,68
1,3	95,61	4,11	4,37
1,45	95,58	4,20	3,43
1,7	95,28	4,59	1,87
2	95,31	4,65	0,62

Figura 5.14. Resultados de la técnica SVM para sujetos colaborativos no equilibrados (características geométricas en 3D).



Parámetro	%Correctos	%FN	FP
-2	56,40	1,87	91,15
-1,7	61,66	2,26	88,77
-1,35	67,81	2,40	85,88
-1,25	69,42	2,45	84,58
-0,9	85,05	3,16	65,28
-0,65	87,29	3,09	57,35
-0,35	89,32	3,25	45,10
-0,2	90,10	3,46	40,83
0,1	91,61	3,80	32,70
0,3	92,50	4,00	28,75
0,7	93,22	4,38	21,45
0,75	93,02	4,66	20,83
1,1	93,59	4,95	13,12
1,25	93,85	4,95	11,87
1,45	94,16	5	9,37
1,55	94,27	5	8,12
1,75	94,37	5	7,5
2	94,73	5	3,12

Figura 5.15. Resultados de la técnica SVM para sujetos no colaborativos no equilibrados (características geométricas en 3D).

5.5.4.- Análisis y discusión de los resultados.

A continuación se discuten los resultados más significativos obtenidos.

Técnica KNN:

Para la técnica KNN, pueden observarse los pobres resultados que se obtienen. En el caso de sujetos colaborativos la tasa de acierto no supera el 45%, cayendo a menos de un 25% para los no colaborativos.

Puede observarse también que para los sujetos colaborativos, son mejores los resultados en el caso de $k=1$ frente a $k=3$. En el caso de no colaborativos, sucede al contrario, los mejores resultados se obtienen si $k=3$.

De ambos hechos se deduce que las clases que representan a los sujetos se encuentran muy agrupadas, en una zona muy reducida del espacio de características que con un clasificador lineal a tramos es difícil de separar correctamente. Este hecho se pone de manifiesto incluso en el caso de bases de datos con pocos sujetos, como es el caso presentado en este trabajo. Las fuentes de error que se han detectado son las siguientes:

- Imperfecta calibración de las cámaras.
- Variaciones en la marcación de los puntos.
- Nube de puntos poco densa.

A pesar de haber intentado minimizar los dos primeros problemas, la conclusión es que ambos factores afectan mucho al resultado final, generando una gran dispersión de las clases. El tercer factor influye ya que se están obteniendo datos de la cara, con un número reducido de puntos. Aumentar el número y la precisión de los puntos seleccionados lleva a pensar en otra forma de adquisición de datos, en concreto utilizando un escáner láser. Además esta forma de adquisición de datos permitiría trabajar con un modelo más preciso en 3D, dotado de textura y forma.

Técnica RBF:

Para la técnica RBF, los resultados obtenidos son mejores que en el clasificador anterior (KNN). En el caso de sujetos colaborativos, la gráfica considerando conjuntos de prueba balanceados muestra que el máximo de reconocimiento (centrado en un valor del parámetro aproximado de 0.25) corresponde a un 75% de reconocimiento correcto, con un 40 % de falsos positivos y un 20 % de falsos negativos. La curva de falsos positivos y falsos negativos se cruzan cuando el parámetro vale casi 0.5, que corresponde a un porcentaje de 32.5 %. En este caso el porcentaje de reconocimiento correcto es de un 65%.

Las dos gráficas de RBF permiten ver que el reconocimiento no varía mucho si los sujetos muestran una actitud no colaborativa. Las gráficas obtenidas de tipo no balanceadas, para sujetos colaborativos y no colaborativos, son muy parecidas. Como es lógico, se reconoce mejor a los sujetos colaborativos que a los que no lo son. La elección de puntos realizada por lo tanto parece adecuada; la mayoría de los 14 puntos considerados son puntos craniométricos que no van a variar con los gestos del sujeto. Sólo 5 de los 14 puntos van a variar al mover la boca y estos influyen en 24 de las 47 características consideradas. Este apartado muestra que la elección de puntos y características fue correcta; también indica que los datos en 3D deben ser tenidos en cuenta para la verificación de sujetos.

Técnica SVM:

Para la técnica SVM, los resultados son bastante mejores que con la técnica RBF. El porcentaje de resultados correctos es un poco más pequeño que en el caso de RBF. Sin embargo, SVM consigue minimizar los falsos positivos y los falsos negativos. En este caso, se obtiene un 80% de verificación correcta en el caso de colaborativos (conjunto equilibrado) con un 16% de FN.

Como conclusión final de este capítulo se puede observar que usando información tridimensional de la cara (incluso con las limitaciones presentes en este capítulo) los valores de reconocimiento no se degradan excesivamente incluso con sujetos en actitudes no colaborativas.

Capítulo 6

Conclusiones. Líneas de investigación futuras.

6.1. – Descripción del capítulo.

La motivación general que ha guiado esta tesis ha sido contribuir al estudio y al desarrollo de técnicas que permitan realizar de forma adecuada el proceso de verificación facial. Se han estudiado técnicas basadas en imágenes bidimensionales y en un modelo tridimensional. Además se han utilizado tres clasificadores: K vecinos más cercanos (KNN), máquinas de vector soporte (SVM), y redes neuronales del tipo funciones de base radial (RBF). En este capítulo se comparan las distintas técnicas utilizadas y un resumen de las principales aportaciones realizadas en esta tesis, posteriormente se muestran las aplicaciones que han surgido y las líneas de trabajo que aparecen de forma natural como continuación de los trabajos realizados en la misma.

6.2.- Conclusiones obtenidas al comparar las diferentes técnicas desarrolladas. Aportaciones realizadas.

En la tesis se han considerado tres clasificadores para su uso en sistemas de verificación facial. Los datos de entrada de los clasificadores provienen tanto de imágenes bidimensionales de niveles de gris, como de una nube de puntos poco densa en tres dimensiones.

En la presente tesis se ha desarrollado y analizado un sistema de verificación facial basado en el preprocesamiento de características bidimensionales (análisis de componentes principales e imágenes en baja resolución) para su posterior verificación por medio de un clasificador (k vecinos más cercanos, redes neuronales del tipo funciones de base radial y máquinas de vector soporte). Las pruebas realizadas muestran que el clasificador que ofrece mejores resultados es el basado en máquinas de soporte vector. Los mejores resultados se obtienen preprocesando por medio del análisis de componentes principales.

Además se ha desarrollado un experimento inicial que permite disponer de una idea intuitiva sobre el funcionamiento de un sistema de verificación facial basado en datos tridimensionales. Este tipo de datos permite obtener resultados más robustos frente a actitudes no colaborativas de los sujetos.

A continuación se enumeran las principales contribuciones de este trabajo:

A) Utilización de imágenes bidimensionales:

- Considerando los datos de entrada, se obtienen mejores resultados con niveles de gris que con una nube de puntos en tres dimensiones. Este resultado puede parecer sorprendente, ya que podría pensarse que con una nube de puntos en tres dimensiones se puede conseguir más información que con una proyección bidimensional. Sin embargo, se ha utilizado un par estéreo de cámaras para obtener una reconstrucción tridimensional a partir de un marcado manual de puntos. Este proceso genera bastantes errores, debido a la marcación de puntos y a la calibración de las cámaras. Por lo tanto, es muy susceptible de ser mejorado.
- En dos dimensiones se han propuesto dos técnicas de reducción de información, basadas en la utilización de análisis de componentes principales (PCA) y el empleo de imágenes de baja resolución. En el primer caso (PCA), la imagen se reduce a un vector de dimensión 150, en el segundo caso (Baja Resolución), la dimensión necesaria es 306 para obtener resultados similares en el porcentaje de reconocimiento correcto. Los resultados son ligeramente mejores utilizando PCA que Baja Resolución. Hay que tener en cuenta que la base de datos se ha obtenido empleando luz controlada y a una distancia cámara – sujeto constante.
- El clasificador que mejores resultados ofrece para los casos analizados es SVM. Se ha comprobado que los resultados dependen mucho del clasificador empleado; las máquinas de vector soporte (SVM) ofrecen siempre el mayor porcentaje correcto con un número de falsos positivos y falsos negativos muy bajo, seguido del clasificador tipo redes neuronales-funciones de base radial (RBF).
- En la verificación partiendo de imágenes se puede observar que el clasificador que mejores resultados proporciona es SVM, y el mejor método de procesamiento es PCA utilizando una plantilla por sujeto y siendo esta plantilla una imagen del sujeto que se quiere verificar. De este modo se permiten errores en la localización de hasta cinco píxeles para una resolución de la imagen facial de 130x140 píxeles, obteniéndose a pesar de ello resultados muy buenos. Errores mayores de 10 píxeles (aproximadamente un 7%) traen consigo que los resultados ya no sean aceptables.

B) Utilización de datos tridimensionales:

- Respecto a la utilización de datos tridimensionales, en esta tesis se ha desarrollado un modelo de cara en tres dimensiones y se ha comprobado que es posible su ajuste a una nube de puntos poco densa con un error pequeño. Los puntos obtenidos en tres dimensiones pueden emplearse en un sistema de reconocimiento basado en distancias geométricas.
- Se ha comprobado que la medida de características geométricas en el rostro humano es una técnica adecuada que permite verificar si el sujeto es quien dice ser o no. Sin embargo, las limitaciones inherentes a las técnicas utilizadas para la obtención de datos en 3D implican que los resultados obtenidos empleando bien

la técnica de K vecinos más cercanos (KNN) o bien la técnica de redes neuronales con funciones de base radial (RBF), no sean todo lo satisfactorios que debieran.

- Se ha tratado de aislar el problema de la verificación de la etapa del procesamiento de las imágenes. Por ello se procedió a la marcación manual de los puntos. Una vez seleccionados y definidos los puntos a buscar, debería encontrarse la manera de automatizar el proceso. En efecto, el uso de un algoritmo preciso para la localización de puntos, sin duda redundaría en una disminución de la variabilidad en la marcación manual de puntos. Se podría esperar que los resultados mejorasen mediante dicha localización automática.
- En definitiva, la selección de puntos y características geométricas ha buscado maximizar la robustez del proceso. Así mismo, la selección de puntos y características parece adecuada a la vista de los resultados que con ellas se han obtenido. Por último, aunque el proceso de adquisición de datos introduce una serie de errores durante el proceso de extracción de información, estos pueden ser superados por un clasificador de tipo máquinas de soporte vector (SVM).

6.3.- Aplicación industrial: VISOR-BASE.

Parte de este trabajo ha sido financiado por la Unión Europea, con el proyecto “Video Sensor Object Request Broker Open Architecture for Distributed Services” abreviadamente VISOR-BASE (Referencia: IST-1999-10808) perteneciente al V Programa Marco. Se inició el 1 Enero 2000 y finalizó el 27 Septiembre 2002 [VISOR].

En este proyecto se desarrolló un sistema de verificación facial, tomando como datos de entrada una imagen y un número de identificación personal (PIN), asignado a cada sujeto de la base de datos. La respuesta del sistema informa si la imagen y el PIN corresponden al mismo sujeto o no, dando también la confianza que el sistema tiene en esa afirmación (con un número de 0 a 100, 0 representa baja confianza). Los datos considerados son imágenes de niveles de gris procesadas con PCA. La localización de la cara en la imagen se produce de forma automática, utilizando la diferencia con la imagen del fondo para estimar la posición de la cara. Posteriormente se procede a su localización fina realizando una convolución con una plantilla que representa una cara. Los clasificadores considerados han sido: K vecinos más cercanos (KNN), redes neuronales – funciones de base radial (RBF), y máquinas de vector soporte (SVM).

Toda esta tarea se realiza de forma muy rápida utilizando hardware comercial no específicamente diseñado para el procesamiento de imágenes.

Actualmente se ha iniciado el proceso de patente del sistema, quedando abierta así su posible comercialización por parte de alguna empresa de prestación de servicios de seguridad.

6.4.- Proyectos futuros.

A la vista de los resultados y de los experimentos realizados han surgido varias líneas de continuación de las técnicas presentadas en esta tesis:

- La primera línea de trabajos futuros que promete resultados más robustos y precisos es el empleo de datos en tres dimensiones. Este proyecto es el más importante que se plantea después del desarrollo de esta tesis. Sin embargo, para ello es preciso cambiar la fase de adquisición de datos. Actualmente se dispone de un escáner láser con el cual se pueden obtener información tanto de la textura como de la forma de la cara en tres dimensiones. Con estos datos es posible identificar un rostro a partir de la forma en tres dimensiones (no sólo de una nube poco densa de puntos) y de la textura (no sólo de una proyección de dicha textura).
- El modelo tridimensional creado en esta tesis está siendo empleado en varios trabajos de investigación. En una primera línea de trabajo, el modelo creado puede utilizarse para generar vistas sintéticas del sujeto. Estas vistas sintéticas pueden utilizarse como imágenes que se utilizan para aumentar el número de imágenes de la base de datos. La segunda línea se basa en un primer paso que consiste en adaptar la imagen de un sujeto al modelo tridimensional. Una vez ajustado, se puede corregir la rotación de la cara y obtener una nueva imagen en la cual la cara del sujeto aparezca frontal. Una vez obtenida la imagen frontal, es posible analizarla con las técnicas descritas en la presente tesis.
- En cuanto al trabajo con clasificadores, han surgido varias líneas. Una línea que no ha sido explotada en la realización de esta tesis consiste en la fusión de clasificadores. Con los resultados obtenidos de los distintos clasificadores considerados (KNN, SVM, RBF), es posible definir un nuevo clasificador cuyo comportamiento global sea mejor que cada uno de los clasificadores que lo componen. Además, pueden añadirse más clasificadores, por ejemplo utilizando la técnica de “boosting”; a partir de un clasificador “débil” como es KNN, es posible definir un clasificador más complejo.
- Otra línea de investigación posible es la fusión de varios elementos en el reconocimiento facial. Hasta ahora el trabajo se ha realizado utilizando una imagen procesada con PCA o mediante baja resolución. Sin embargo, se puede procesar la imagen de las dos formas y así poder obtener un resultado más fiable. Dentro de esta línea se pretende también añadir un nuevo tipo de preprocesamiento, basado en la extracción de características locales que permita complementar los dos métodos globales estudiados en la presente tesis.

BIBLIOGRAFÍA

- [ACLUa] American Civil Liberties Union. Drawing a blank: Report on Tampa police records reveals poor performance of Face Recognition Technology. http://archive.aclu.org/issues/privacy/drawing_blank.pdf.
- [ACLUb] American Civil Liberties Union. Palm Beach International Airport's Interim Report on Facial Recognition Test. http://archive.aclu.org/issues/privacy/FaceRec_data.pdf.
- [And95] J. A. Anderson. An Introduction to Neural Networks. MIT Press. 1995.
- [Ahl01] J. Ahlberg. An experiment on 3D face model adaptation using the active appearance algorithm. Technical Report. Dept of EE, Linköping University, Sweden. Jan. 2001.
- [Ahl01a] J. Ahlberg. Candide-3 an updated parametrized face. Technical Report. Dept of EE, Linköping University, Sweden. Jan. 2001.
- [Alv94] M. Alvarez Sainz. Estadística. Ed. Universidad de Deusto. 1994.
- [And81] J. A. Anderson, M. C. Mozer. Categorization and selective neurons. Parallel models of associative memory, Eds G. E. Hinton, J. A. Andersons. P 213-236. Erlbaum. 1981.
- [Ati96] J. J. Atick, P. A. Griffin, A. N. Redlich. Statistical approach to shape from shading: reconstruction of 3D face surfaces from single 2D images. Neural Computation. V 8. N 6. P 1321-1340. Aug. 1996.
- [Bak00] S. Baker, T. Kanade. Hallucinating faces. Proc fourth IEEE International Conference on Face and Gesture Recognition. P 83-88. 2000.
- [Bal82] D.H. Ballard, C.M. Brown. Computer Vision. Prentice Hall. 1982.
- [Bal89] P. Baldi, J. Hornink. Neural networks and principal components analysis: learning from examples without local minima. Neural networks. N 2, V 1. P 53-58. 1989.
- [Bel97] P.N. Belhumeur, J. P. Hespanha and D. J. Kriegman. Eigenfaces vs Fisherfaces: Recognition Using Class Specific Linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol 19, Nº 7, p 711-720. Jul. 1997.
- [Ber95] M. R. Berthold and Jay Diamond. Boosting the performance of rbf networks with dynamic decay adjustment. In G. Tesauro, D. S. Touretzky, and T.K. Leen, editors, Advances in Neural Information Processing Systems, volume 7, 1995.
- [Bis92] H. Bischof, A. Pinz. Neural Networks in Imagen Pyramids. In IJCNN'92, V 4, P 374-379, 1992.

- [Bis95] C. M. Bishop. Neural networks for Pattern Recognition. Oxford Univ. Press.1995.
- [Bla96] V.. Blanz, B. Schölkopf, H. Bülthoff, C. Burges, V. Vapnik, y T. Vetter. Comparison of view-based object recognition algorithms using realistic 3d models. In C. von der Malsburg, W. von Seelen, J. C. Vorbrüggen, and B. Sendhoff, editors, Artificial Neural Networks - ICANN'96, pp. 251-256, Berlin, 1996. Springer Lecture Notes in Computer Science, Vol. 1112.
- [Bla98] V. Blanz, T. Vetter. Generalization to novel views from a single face image. In Face Recognition: from theory to applications. Springer Verlag. H. Wechsler et al., Eds. P. 310-326. 1998.
- [Bla01] D. M. Blackburn, J. M. Bone, P. J. Phillips. FRVT 2000 Report. Technical Report. 2001. <http://www.frvt.org>.
- [Ble64] W. W. Bledsoe. The model method in facial recognition. Technical Report PRI:15. 1964.
- [Bro75] P. Broca. Instructions craniologiques et craniometriques de la Société d'Anthropologie de Paris. Mem. Soc. Anthropol. de Paris. 1875.
- [Bru93] R. Brunelli and T. Poggio. Face Recognition: Features versus Templates. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol 15, Nº 10, p.1042-1052. Oct. 1993.
- [Cab98] E. Cabello, M. A. Sánchez, L. Pastor. Some experiments in face recognition with neural networks. In Face Recognition: from theory to applications. Springer Verlag. H. Wechsler et al., Eds. P. 589-598. 1998.
- [Cab02] E. Cabello, M A, Sánchez, J. Delgado. A new approach to Identify big rocks with applications to the mining industry. Real Time Imaging. V 8, N 1, P 1-9, Feb. 2002.
- [Cam70] P. Camper. Dissertation sur les différences que présentent les traits du visage chez les hommes des différents pays et de différents âges. Presented to the Academie du Dessin. Amsterdam. 1770.
- [Che95] R. Chellappa, C. L. Wilson and S. Sirohey. Human and Machine Recognition of Faces: A Survey. Proceedings of the IEEE. Vol. 83. Nº 5. pp 705-740 .May 1995.
- [Che98] Q. Chen, G. Medioni. "Building human face models from two images. IEEE 2nd Workshop on Multimedia Signal Processing. P 117-122. Dec. 1998.
- [Cor95] C. Cortes and V. Vapnik. Support vector networks. Machine Learning, 20, pp. 273-297, 1995.
- [Cot90] G. W. Cottrell, M. K. Fleming. Face recognition using unsupervised feature extraction. Proc. Int. Conf. Neural Network. P 322-325. 1990.
- [Cou53] R. Courant y D. Hilbert. Methods of Mathematical Physics. Interscience, 1953.
- [Cov67] T. Cover and P. Hart. Nearest Neighbor pattern classification, In IEEE transactions in Information Theory, Vol.13, pages 21-27. 1967.
- [Cox96] I. J. Cox, J. Ghosn, P. N. Yianilos Feature-based face recognition using mixture-distance. Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition. P 209-216. Jun 1996.
- [Cry95] J. Cryer, P. S. Tsai, M. Shah. Integration of shape from shading and stereo. Pattern Recognition. V 28. N 7. P 1033-1043. July 1995.
- [Dec98] D. DeCarlo, D. Metaxas, M. Stone "An anthropometric face model using variational techniques". Proc. SIGGRAPH Conf. P 67-74. 1998.

- [Der90] R. Deriche. Fast Algorithms For Low-Level Vision. IEEE Trans. Pami, Vol 12, No 1, Jan 1990.
- [Dud73] R. Duda and P. Hart. Pattern Classification and Scene Analysis. Ed. Wiley, 1973.
- [Eic00] S. Eickeler, S. Müller, G. Rigoll. Recognition of JPEG compressed face images based on statistical methods. Image and Vision Computing. Num 18. P 279-287. 2000.
- [Esc02] M. J. Escobar, J. Ruiz-del-Solar. Biologically-based Face Recognition using Gabor Filters and Log-Polar Images. Int. Joint Conf. on Neural Networks. 2002.
- [FG00] Fourth IEEE International Conference on Automatic Face and Gesture Recognition. March 26 - 30, Grenoble, France. 2000.
- [Fua99] P. Fua. Using model-driven bundle-adjustment to model heads from raw video sequences. Proc. IEEE Int. Conf. Computer Vision. P 46-53. Sept.1999.
- [Fuk83] K. Fukushima, S. Miyake, Takayukiito. Neocognitron: A neural network model for a mechanism of visual pattern recognition. IEEE Trans. On Systems, man and cybernetics. V 13, N 5, Oct. 1983.
- [Gil81] P. E. Gill, W. Murray, M. Wright. Practical Optimization. Academic Press, New York, 1981.
- [Gon99] S. Gong, S. J. Mckenna, A. Psarrou. Dynamic Vision: from images to face recognition. Imperial College Press. 1999.
- [Gor95] G. G. Gordon, M. E. Lewis. Face recognition using video clips and mug shots. Proc. Office of National Drug Control Police. International technical Symposium. Oct. 1995.
- [Gou96] F. Goudail, E. Lange, T. Iwamoto, K. Kyuma, N. Otsu. Face recognition using local autocorrelations and multiscale integration. IEEE Transactions in Pattern Analysis and Machine Intelligence, Vol 18. N 10. P 1024-1028. Oct. 1996.
- [Gro92] S. Grossberg, G. A. Carpenter. Neural networks for vision and image processing. MIT Press, 1992
- [Gue98] B. Guenter, C. Grimm, D. Wood, H. Malvar, F. Pighin. Making faces. Proc. SIGGRAPH Conf. P55-66. July 1998.
- [Gui98] G. J. Castro, J. Nieto, L. M. Gallego, L. Pastor, E. Cabello. An effective camera calibration method. Proc. V Int. Workshop on Advanced Control Motion. P 171-174. 1998.
- [Guo00] G. Guo, S. Z. Li, K. Chan. Face recognition by support vector machines. Proc fourth IEEE International Conference on Face and Gesture Recognition. P 196-201. 2000.
- [Hal99] P. L. Hallinan, G.G. Gordon, A. L. Yuille, P. Giblin, D. Mumford. Two- and Three dimensional patterns of the face. A. K. Peters. 1999.
- [Har81] L. D. Harmon, M. K. Khan, R. Lasch, P. F. Raming. Machine identification of human faces. Pattern Recognition, V 13, P 97-110. 1981.
- [Hes02] C. Hesher, A. Srivastava, G. Erlebacher. PCA of Range Images for Facial Recognition. International Multiconference in Computer Science. 2002
- [Hje01] E. Hjelm, B. K. Low. Face detection: a survey. Computer vision and image understanding. V 83, P 236-274. Sept 2001.
- [Hou93] D. R. Hougen, N. Ahuja. Estimation of the light source distribution and its use in integrated shape recovery from stereo and shading. Proc IEEE Int. Conf. Computer Vision, P 148-155, May 1993.

- [How97] A. J. Howell. Automatic face recognition using radial basis function networks. PhD Thesis. University of Sussex. 1997.
- [Hsu02] R. L. Hsu. Face detection and modeling for recognition. PhD. Thesis. Michigan State University. Dpt. Computer Science and Engineering. 2002.
- [Hud92] M. J. Hudak. Rce classifiers: Theory and practice. *Cybernetics and Systems*, P 483-515, 1992.
- [Hwa00] W. S. Hwang J. Weng Hierarchical discriminant regression. *IEEE Transactions in Pattern Analysis and Machine Intelligence*, Vol 22. P 1277-1293. Nov. 2000.
- [Joa97] T. Joachims. Text categorization with support vector machines. Technical report, LS VIII Number 23, Universidad de Dortmund, 1997.
- [Jol92] J. M. Jolion, A. Montanvert. The adaptive pyramid, a framework for 2D image analysis. *CVGIP: Image Understanding*. N 55, V 3, P 3339-348. 1992.
- [Jon00] K. Jonsson, J. Matas, J. Kittler, Y. P. Li. Learning support vectors for face verification and recognition. Proc fourth IEEE International Conference on Face and Gesture Recognition. P 205-210. 2000.
- [Jur99] F. Jurie. A new log-polar mapping for space invariant imaging. Application to face detection and tracking. *Pattern Recognition*. Vol 32. P 865-875. 1999.
- [Kan77] T. Kanade. Computer Recognition of human faces. Ed. Birkhauser. 1977.
- [Kam93] M. S. Kamel, H. C. Shen, A. K. C. Wong R. I. Campeanu. System for the recognition of human faces. *IBM System journal*. V 32, N 2, P 307-320. 1993.
- [Kim91] B. Kim, P. Burger. Depth and shape from shading using the photometric stereo method. *Computer Vision, Graphics and Image Processing: Image Understanding*. V 54. N 3. P 416-427. 1991.
- [Koh01] T. Kohonen. *Self-Organizing Maps*. Springer Verlag. 2001.
- [Kin02] S. King. H. Harrelson. G. Tran. Testing Iris and Face Recognition in a Personnel Identification Application. Departamento de Defensa, EE. UU. Feb. 2002.
http://www.itl.nist.gov/div895/isis/bc2001/FINAL_BCFEB02/FINAL_1_Final%20Steve%20King.pdf
- [Kir90] M. Kirby, L. Sirovich. Application of the Karhunen-Loève procedure for the characterization of human faces. *IEEE Trans. Pattern Analysis Mach. Intell.* N 12, P 103-108. 1990.
- [Koh81] T. Kohonen, E. Oja, P. Lehtio. Storage and processing of information in distributed associative memory systems. *Parallel models of associative memory*, Eds G. E. Hinton, J. A. Andersons. P 49-81. Erlbaum. 1981.
- [Koh99] L. H. Koh, S. Ranganath, M. W. Lee, Y. V. Venkatesh. An integrated face detection and recognition system. 10 ICIAP. P 532-537. 1999.
- [Kon99] T. Kondo, Hong Yan. Automatic human face detection and recognition under non-uniform illumination. *Pattern Recognition*. Vol 32. P 1707-1718. 1999.
- [Lad93] M. Lades, J. C. Vorbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Würz, W. Konen. Distortion Invariant object recognition in the dynamic link architecture. *IEEE Trans. Computers*. N 42. V 3. P 300-311. 1993.

- [Lam98] K. M. Lam, H. Yan. An analytic-to-holistic approach for face recognition based on a single frontal view. *IEEE Transactions in Pattern Analysis and Machine Intelligence*, Vol 20. N 7 P 673-686. July 1998.
- [Lam98a] J. Lampinen, E. Oja. Pattern recognition. *Image Processing and Pattern recognition*. Ed. C. T. Leondes. P 1-59. Academic Press. 1998.
- [Law97] S. Lawrence, C. L. Giles, A. C. Tsoi, A. D. Back. Face recognition: a convolutional neural-network approach. *IEEE Transactions on Neural Networks*. Vol 8. N 1 P 98-113. Jan 1997.
- [Lee00] W. Lee, N. Magnenat-Thalmann. Fast head modeling for animation. *Image and vision computing*. V 18, N 4. P 355-364. Mar. 2000.
- [Len96] R. Lengagne, J. P. Tarel, O. Monga. From 2D images to 3D face geometry. Proc. IEEE International Conference Automatic face and gesture recognition. P 301-306. Oct. 1996.
- [Len00] R. Lengagne, P. Fua, O. Monga. 3D stereo reconstruction of human faces driven by differential constraints. *Image and Vision Computing*. V 18. N 4. P 337-343. Mar 2000.
- [Lin97] S. Lin, S. Kung, L. Lin. Face recognition/detection by probabilistic decision-based neural network. *IEEE Transactions on Neural Networks*. Vol 8. N 1 P 114-132. Jan 1997.
- [Luc97] S. M. Lucas. Face recognition with the continuous n-tuple classifier. Proc. British Machine Vision Conference. Eds: A. F. Clark. BMVA Press. P 222-231.1997.
- [Mas93] T. Masters. Practical neural network recipes in C++. Morgan Kaufmann Publishers. 1993.
- [Mia99] J. Miao, B. Yin, K. Wang, L. Shen, X. Chen. A hierarchical multiscale and multiangle system for human face detection in a complex background using gravity-center template. *Pattern Recognition*. Vol 32. P 1237-1248. 1999.
- [Mil86] R. Millward, A. J. O'Toole. Recognition memory transfer between spatial-frecuency analyzed faces. *Aspects of face processing*. Eds. H. D. Ellis, M. A. Jeeves, F. Newcombe, A. Young. P 34-44. Nijhoff. 1986.
- [Mit] Base de datos facial. <ftp://whitechapel.media.mit.edu/pub/images>.
- [Mog94] B. Moghaddam, A. Pentland. Face recognition using view-based and modular eigenspaces. *Automatic Systems for the Identification and Inspection of Humans*, Proc SPIE. V 2257, P 12-21. July 1994.
- [Mog97] B. Moghaddam, A. Pentland. Probabilistic visual learning for object representation. *IEEE Transactions in Pattern Analysis and Machine Intelligence*, Vol 19. N 7 P 696-710. July 1997.
- [Moo91] A. W. Moore. An introductory tutorial on kd-trees. Technical report 209. Computer laboratory. University of Cambridge.1991.
- [Oja82] E. Oja. A simplified neuron model as a principal components analyzer. *J. Math Biol.* N 15. V 3. P 267-273. 1982.
- [Oja89] E. Oja. Neural networks, principal components and subspaces. *Int. J. Neural Systems*. N 1. V 1. P 61-68. 1989.
- [Oka92] Okabe, Boots and Sugihara. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. Ed John Wiley & Sons. 1992.
- [Oka98] K. Okada, J. Steffens, T. Maurer, H. Hong, E. Elagin, H. Neven, C. von der Malsburg. The Bochum/USC face recognition system and how it fared in the FERET phase

- III Test. In Face Recognition: from theory to applications. Springer Verlag. H. Wechsler et al., Eds. P. 186-205. 1998.
- [Oli] Base de datos facial. <http://www.cam-orl.co.uk/facedtbase.html>.
- [Oto91] A. J. O'Toole, H. Abdi, K. A: Deffenbacher, J. C. Bartlett. Classifying faces by race and sex using an autoassociative memory trained for recognition. Proc.13th Annual Conf. Cognitive Sci. Soc. P 847-851. 1991.
- [Oto93] A. J. O'Toole, H. Abdi, K. A: Deffenbacher, D. Valentin. A low-dimensional representation of faces in the higher dimensions of the space. J. Opt. Soc. Amer. V 10. P 405-411. 1993.
- [Par96] F. I. Parke, K. Waters. Modeling faces. In Computer facial animation. P 55-104. A. K. Peters. 1996.
- [Par96a] F. I. Parke, K. Waters. Appendix I: Three dimensional muscle model facial animation. In Computer facial animation. P 337-338. A. K. Peters. 1996.
- [Pen96] P. S. Penev, J. J. Atick. Local feature analysis: a general statistical theory for object representation. Network: Computation in Neural Systems. V 7, N 3, P 477-500, 1996.
- [Pen00] P. S. Penev, L. Sirovich. The global dimensionality of face space. Proc. Fourth IEEE International Conference on Face and Gesture Recognition. P 264-270. 2000.
- [Phi98] P. Phillips, H. Wechsler, J. Huang, P. Rauss. The FERET database and evaluation procedure for face recognition algorithms. Image and Vision Computing, V 16, N 5, P 295-306. Mar. 1998.
- [Phi03] P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, M. Bone. FRVT 2002. Technical Report. 2003. <http://www.frvt.org>.
- [Phi00] P. J. Phillips, H. Moon, S. A. Rizvi, P. J. Rauss. The FERET evaluation methodology for face recognition algorithms. IEEE Transactions in Pattern Analysis and Machine Intelligence, Vol 22. N 10 P 1090-1104. Oct 2000.
- [Pre92] W. H. Press, W. T. Vetterling, S. A. Teukolsky. B. P. Flannery. Numerical Recipes in C. Cambridge Univ. Press. 1992.
- [Pur] Base de datos facial. http://rv11.ecn.purdue.edu/~aleix/aleix_face_DB.html.
- [Que99] R. Queiroz F., C. E. Thomaz, A. Veiga. Comparing the Performance of the Discriminant Analysis and RBF Neural Network for Face Recognition. Int Conf on Information Systems Analysis and Synthesis. Aug. 1999.
- [RCE82] D.L. Reilly, L.N. Cooper, and C. Elbaum. A neural model for category learning. Biol. Cybernet., 45, 1982.
- [Rei95] M. J. T. Reinders, P. J. L. van Beek, B. Sankur, J. C. A. van der Lubbe. Facial feature location and adaptation of a generic face model for model base coding. Signal Processing: Image Communication. V 7, N 1. P 57-74. Mar 1995.
- [Ren89] S. Renals, R. Rohwer. Phoneme classification experiments using radial basis functions, Int. Joint Conf on Neural Networks, V 1, P 461-467. 1989.
- [Rev91] J. M. Reverte Coma. Antropología forense. Ed. Ministerio de Justicia. 1991.
- [Rie55] F. Riesz, B. S. Nagy. Functional Analysis. Ed. Ungar. 1955.
- [Rip96] B. D. Ripley. Pattern recognition and neural networks. Cambridge Univ. Press. 1996.

[Rom96] S. Romdhani. Face Recognition using Principal Components Analysis. MSC thesis. University of Glasgow, 1996.

[Sam94] F. S. Samaria. Face Recognition using Hidden Markov Models. PhD. Thesis. Trinity College. University of Cambridge. Cambridge. UK. 1994.

[Sam95] A. Samal, P. A. Iyengar. Human face detection using silhouettes. Int. Journal of Pattern Recognition and Artificial Intelligence. Vol. 9. N. 6. P. 845-867. 1995.

[Sim00] T. Sim, R. Sukthankar, M. Mullim, S. Baluja. Memory-based face recognition for visitor identification. Proc fourth IEEE International Conference on Face and Gesture Recognition. P 214-220. 2000.

[Sir87] L. Sirovich, M. Kirby. Low-dimensional procedure for the characterization of human face. J. Opt. Soc. Amer. V 4. P 519-524. 1987

[Sme00] F. Smeraldi, O. Carmona, J. Bigün. Saccadic search with Gabor features applied to eye detection and real-time head tracking. Image and Vision Computing. Num 18. P 323-329. 2000.

[SNNS] Stuttgart Neural Network Simulator SNNS v4.2. <http://www-ra.informatik.uni-tuebingen.de/SNNS/>.

[Sto00] M. Störring, H. J. Andersen, E. Granum. Estimation of the illuminant colour from human skin colour. Proc fourth IEEE International Conference on Face and Gesture Recognition. P 62-67. 2000.

[Ter00] J. C. Terrillon, M. N. Shirazi, H. Fukamachi, S. Akamatsu. Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images. Proc fourth IEEE International Conference on Face and Gesture Recognition. P 54-61. 2000.

[The02] T. C. Greene. Face recognition kit fails in Fla airport. The Register. May 2002. <http://www.theregister.co.uk/content/55/25444.html>.

[Tis00] M. Tistarelli, E. Grossos. Active vision-based face authentication. Image and Vision Computing. Num 18. P 299-314. 2000.

[Tou74] J. T. Tou, R. C. González. Pattern Recognition Principles. Ed. Addison-Wesley Publishing Company .1974.

[Tur91] M. Turk, A. Pentland. Eigenfaces for Recognition. Journal of Cognitive Neuroscience. V 3, N 1, P 71-86. 1991.

[UMR] University Missouri - Rolla. Putting a new face on security. <http://web.umr.edu/~newsinfo/facekey.html>.

[Uwe97] O. A. Uwechue, A. S. Pandya, Eds. Human Face Recognition using third-order synthetic neural networks. Kluwer Academic Publishers.1997.

[Vap95] V. Vapnik. The Nature of Statistical Learning Theory. Springer-Verlag. 1995.

[Vap98] V. Vapnik. Statistical learning theory. Ed Wiley. 1998.

[Vie98] E. Viennet, F. F. Soulié. Connectionist Methods for Human Face Processing. In Face Recognition: from theory to applications. Springer Verlag. H. Wechsler et al., Eds. P. 124-156. 1998.

[VISOR] VISOR-BASE. <http://www.visual-tools.com/visorbase/index.html>.

[Web02] A: R. Webb. Statistical Pattern Recognition. Ed: John Wiley.2002.

- [Wec90] H. Wechsler. Computational Vision. Academic Press. 1990.
- [Wec98] H. Wechsler, P. Phillips, V. Bruce, F. Soulie, T. Huang, Eds. Face Recognition: from theory to applications. Springer Verlag. 1998.
- [Wen99] J. Weng, D. L. Swets. Face Recognition in Biometrics: Personal Identification in Networked Society. A. k. Jain, R. Bolle, S. Pankanti, Eds. P 67-86. Kluwer Academic. 1999.
- [WIR] J. Scheeres. Nuke Reactor: Show Me Your Face. Wired News. Aug. 2002. <http://www.wired.com/news/technology/0,1282,54423,00.html>.
- [Wis97] L. Wiskott, J-M Fellous, N. Krüger, C. von der Malsburg. Face Recognition by Elastic Bunch Graph Matching. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol 19, Nº 7. p 775-789. Jul. 1997.
- [Xm2] Base de datos facial. <http://xm2vtsdb.ee.surrey.ac.uk/home.html>.
- [Yal] Base de datos facial. <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>.
- [Yan98] Y. Yan, J. Zhang. Rotation-invariant 3D recognition for face recognition. Proc. IEEE Int. Conf. Image Processing. V 1, P 156-160. Oct. 1998.
- [Yan01] M. H. Yang, N. Ahuja, D. Kriegman. Detecting faces in images: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol 24, N 1. p 34-58. Jan. 2001.
- [Yin97] L. Yin, A. Basu. MPEG4 face modeling using fiducial points. Proc. IEEE Int. Conf. Image Processing. P 109-112. Oct 1997.
- [Yul89] A. L. Yullie, D. S. Cohen and P. W. Hallinan. Feature Extraction from Faces Using Deformable templates. Proceedings Computer Vision and Pattern Recognition. 1989.
- [Zha00] W. Zhao, R. Chellappa, A. Rosenfeld, P. J. Phillips. Face recognition: a literature survey. CVL Technical report, Center for automation research. University of Maryland at College Park. Oct. 2000.
- [Zha00a] W. Zhao, R. Chellappa. 3D model enhanced for face recognition. Proc. IEEE Int. Conf. Image Processing. Sept 2000.
- [Zha02] Z. Zhang. Camera calibration with one dimensional object. Technical report MSR-TR-2001-120. Aug. 2002.
- [Zho99] J. Zhou, C. Zhang, Y. Li. Directional symmetry transform for human face location. Opt. Eng. Vol 38. Num 12. P 2114-2117. Dec. 1999.

Anexo A

Modelo tridimensional considerado.

Como ayuda en la marcación se puntuó el modelo 3D de una cara humana. El modelo está formado por facetas triangulares (véase la Figura 5.8). De cada faceta triangular se almacenan las coordenadas tridimensionales de cada vértice (en total el modelo tiene 383 vértices) y los vértices que se unen formando una arista. El modelo permite definir en el mismo y de forma manual, las coordenadas de una serie de puntos, los llamados puntos guía. Estos puntos guía son los 14 puntos definidos en la Tabla 5.4. Este modelo no fue utilizado en el reconocimiento, ya que los resultados obtenidos, aunque prometedores, no tienen el nivel de detalle necesario para distinguir claramente entre dos individuos diferentes. Empleando una captura de datos mejorada, por ejemplo mediante un escáner láser, el modelo podría tenerse en cuenta para la verificación facial.

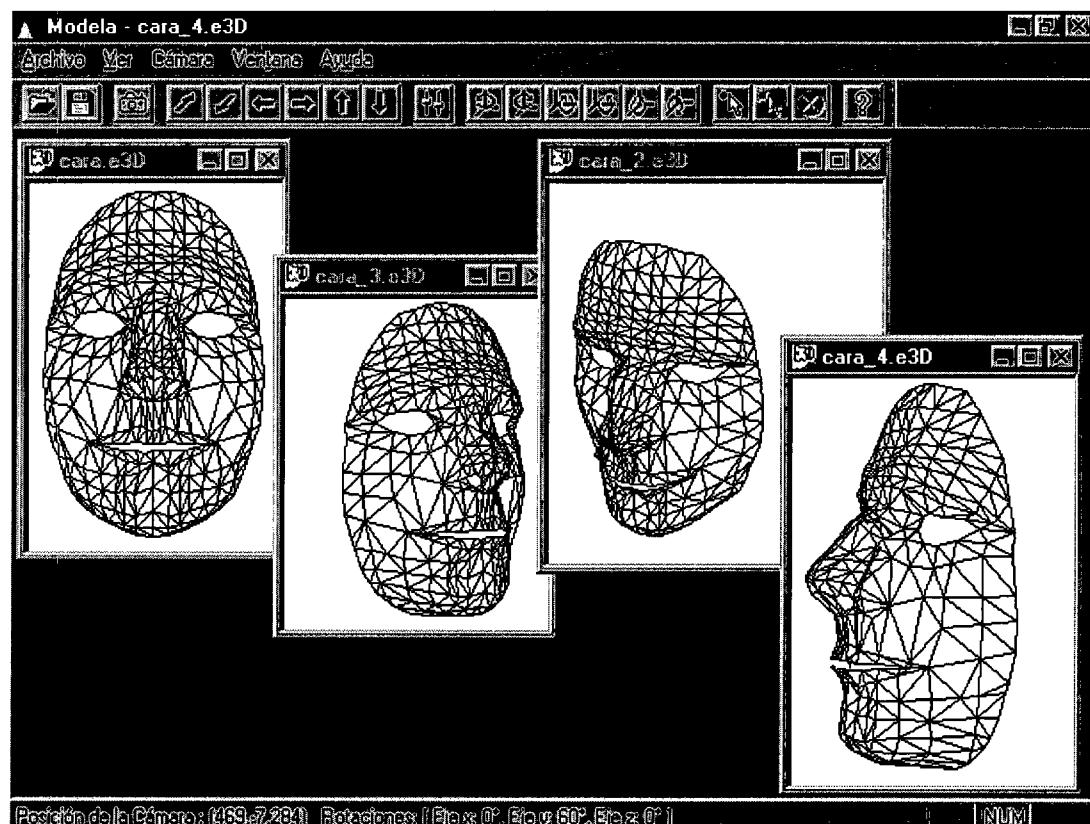


Figura A.1. Diversas vistas del modelo 3D de caras humanas.

A través de una serie de transformaciones que se detallan a continuación, utilizando los puntos guía, se consigue un modelo que puede ajustarse a una cara procedente de una captura concreta. De los 14 puntos guía, 11 se utilizan para el ajuste del modelo, y el resto, para la comprobación de errores. El resultado final es una *deformación* del modelo, como si fuese una máscara flexible, acercando los puntos guía a los correspondientes en la foto y transmitiendo esa adaptación al resto de puntos. La *deformación* se produce tras la aplicación secuencial de las siguientes transformaciones sobre el modelo.

Ajuste de orientación:

Para corregir la orientación del modelo, se crea el plano que ajusta en el modelo los 11 puntos de control, y el plano que por el mismo procedimiento ajusta los 11 puntos correspondientes marcados en el sujeto. Los vectores normales a cada plano se utilizan para calcular los ángulos que debe girar el modelo para ajustarse al sujeto.

Ajuste de escala:

Se halla el factor de escala comparando las medias de las distancias de un punto a los otros 10, en los puntos guías y en los homónimos en el modelo. Una vez obtenido este factor para cada uno de los ejes coordinados, se aplicará el cambio de escala en cada eje al modelo.

Ajuste de posición:

Se halla la media de las distancias entre cada uno de los 11 puntos de control y sus homónimos en el modelo. Con ello se obtiene el vector desplazamiento para la translación del modelo.

Ajuste de cambio de escala ponderado:

Aquí se trabajará ya sólo con el modelo ajustado. Lo primero será igualar los puntos marcados de control a los que les corresponden en el modelo, para obtener un modelo personalizado. El objetivo es "acercar" el resto de los 372 puntos alrededor de los 11 puntos guías. La función que nos da el peso de la ponderación del cambio de escala viene dado por:

$$f(d) = \begin{cases} 0.76 \cdot d & Si \quad d \leq 4\% \\ 0.04 \cdot d & Si \quad 4\% < d \leq 45\% \\ 0 & Si \quad d > 45\% \end{cases}$$

donde d es la distancia del punto del modelo al punto guía que estamos considerando, dividido entre la distancia máxima (la distancia del punto guía al más alejado del modelo).

Los puntos que están muy cerca de un punto guía (hasta el 4% de la distancia máxima) son muy atraídos hacia él. Los puntos que están en el intervalo entre el 4% y el 45% de la máxima distancia son atraídos en menor medida hacia dicho punto, y los que están a partir del 45% no son atraídos. Mediante esta transformación se consigue que cada uno de los puntos del modelo se acerque, dependiendo de la distancia a los puntos guías, a cada uno de esos puntos.

Al final del proceso, se obtiene una proyección del modelo tridimensional adaptado a una cara humana. La visualización del modelo ya ajustado al sujeto que aparece en la foto aparece en la Figura A.2. Es posible colocar la fotografía detrás de nuestro nuevo modelo, y así comprobar visualmente tanto el ajuste, como el error cometido.

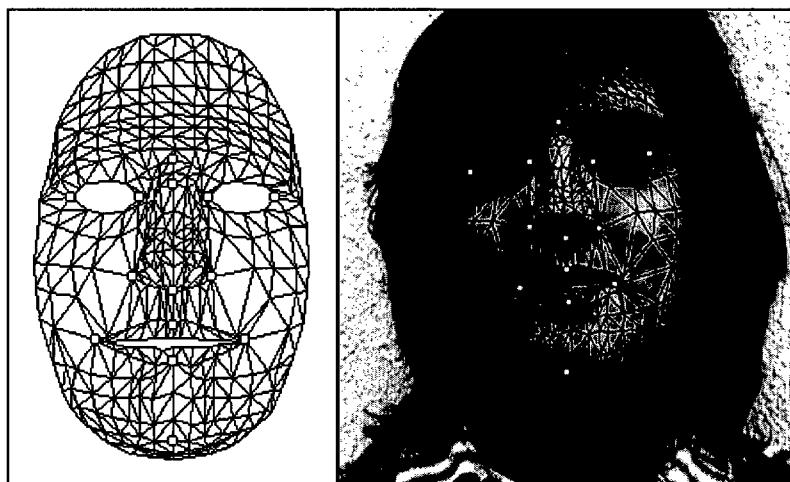


Figura A.2. Modelo y su ajuste a una cara.

A.1.- Resultados numéricos: ajuste del modelo 3D a fotografías 2D.

En este apartado se detallan los resultados obtenidos al adaptar el modelo a las imágenes que forman la base de datos de caras humanas. Para realizar el presente trabajo se consideraron 14 puntos que han de ser manualmente marcados en la imagen. 11 puntos se utilizan para el ajuste del modelo y tres son puntos de control. Los puntos de control en ningún momento intervinieron en el proceso manual de adaptación entre imagen y modelo y sirven para medir el error cometido entre el ajuste del modelo y la medición manual.

Una vez adaptado el modelo genérico a la cara concreta, se midió la diferencia en píxeles entre el ajuste (proyectado en dos dimensiones) y la marcación manual. La Tabla 5.6 muestra, para los tres puntos de control la diferencia media y la desviación típica en cada uno de los ejes (medidos en píxeles). La base de datos considerada está formada por 20 sujetos. De cada uno de ellos se obtuvieron 8 imágenes (que corresponden a cuatro adquisiciones de imágenes por un par estéreo), casi frontales, aunque aparecen cambios en la expresión, giros en la cabeza, variaciones en la iluminación.. Los puntos guía fueron manualmente marcados en las 160 imágenes dos veces por medio de un operador.

Eje X	Diferencia media	Desviación típica
Glabela	1.6002	1.3218
Nasión	2.2518	1.8753
Subnasal	1.5170	1.2421

Eje Y	Diferencia media	Desviación típica
Glabela	2.0952	1.8720
Nasión	2.8816	2.0695
Subnasal	1.6730	1.2190

Tabla A.1. Diferencia entre los puntos aproximados y los reales (en píxeles).

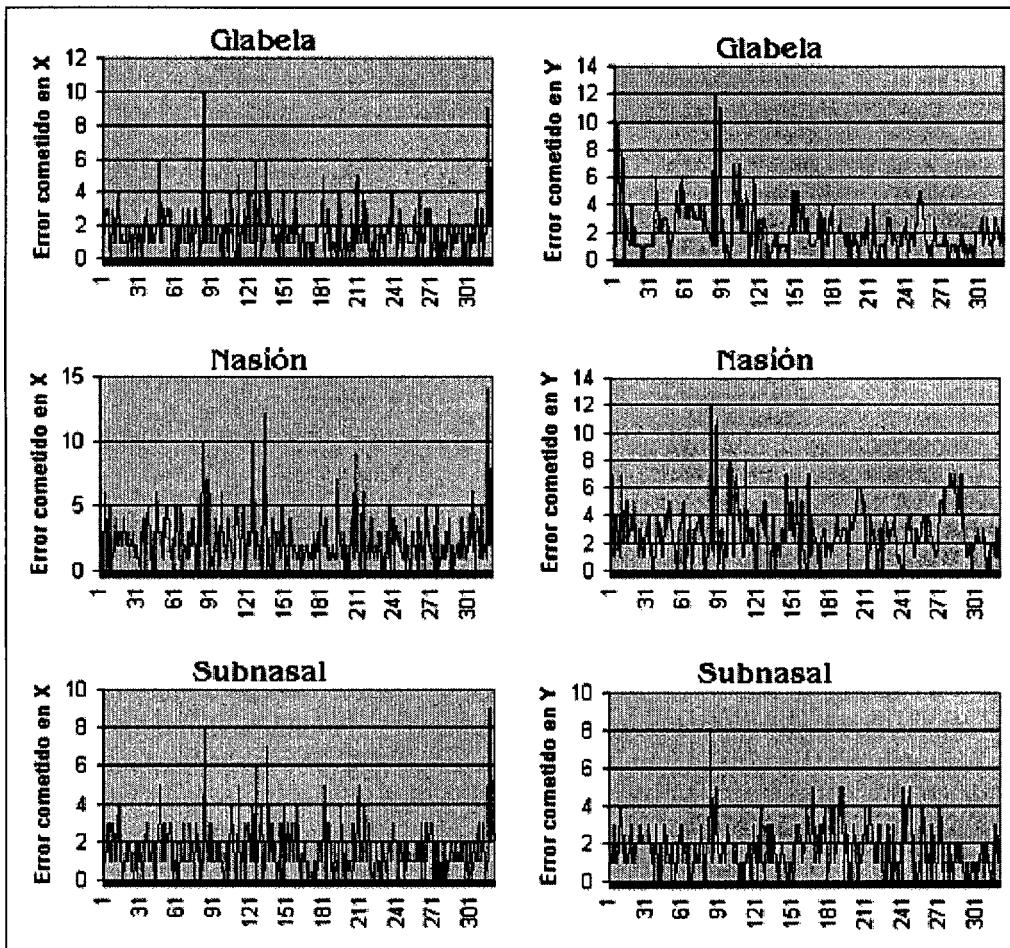


Tabla A.2. Distribución de diferencias para toda la base de datos.

En la Tabla A.2 se detalla la distribución de diferencias para toda la base de datos. Aunque la mayor separación entre los puntos marcados manualmente y los adaptados es de doce píxeles en uno de los casos, tanto los errores medios como su desviación toman valores muy pequeños (menores de tres píxeles de valor medio).

En el caso de caras humanas, las diferencias en los valores obtenidos en la adaptación de un modelo geométrico y la marcación manual son aceptables para su uso

posterior. Así mismo se ha comprobado la facilidad en la ampliación de la aplicación, con un coste muy bajo.

A.2.- Agregado de texturas al modelo tridimensional.

Una vez que se tiene perfectamente ajustado el modelo, la fase siguiente consiste en obtener un modelo con textura. Los pasos a realizar son:

- Conseguir el nivel de gris de cada uno de los 383 puntos del modelo.
- Conseguir una superficie que interpole esos puntos.
- Conseguir el nivel de gris de cada uno de los puntos interpolados.
- Conseguir la proyección en cada una de las cámaras de la superficie interpolada con su nivel de gris.

Cada uno de los pasos anteriores se detalla brevemente a continuación.

Nivel de gris de los puntos:

Para ello se proyectarán cada uno de los puntos sobre una cámara y se recogerá el nivel de gris de la foto asociada a esa cámara. Para proyectar un punto en 3 dimensiones a 2 dimensiones, se usará la matriz que representa el modelo de la cámara. Se analiza el nivel de gris que se encuentra en la foto y dicha información se introduce en el punto en 3 dimensiones. En la Figura 5.10 se aprecia el modelo superpuesto a una cara, de tal forma que se pueden obtener los niveles de gris de cada punto.

Superficie interpolatoria y nivel de gris:

En esta etapa se parte de los puntos que forman el modelo; cada punto tiene asociado un nivel de gris junto con sus tres coordenadas cartesianas. Partiendo de los niveles de gris de los puntos vértice de cada triángulo que define el modelo, es posible llenar el color correspondiente a cada punto del triángulo. En nuestro caso, para calcular el nivel de gris de un punto intermedio, se ponderó el nivel de gris de cada uno de los vértices, teniendo en cuenta la distancia del punto a dicho vértice.

Proyección de la superficie con textura en una imagen:

Con ayuda nuevamente de las matrices de calibración, se pasan los puntos en 3 dimensiones a una proyección en 2 dimensiones, con su nivel de gris respectivo. De cada ajuste se pueden conseguir ahora tantas proyecciones como se desee, variando los parámetros que definen la matriz de proyección de una cámara. La Figura 5.11 muestra un modelo ajustado a una cara y tres vistas sintéticas generadas con dicho modelo.

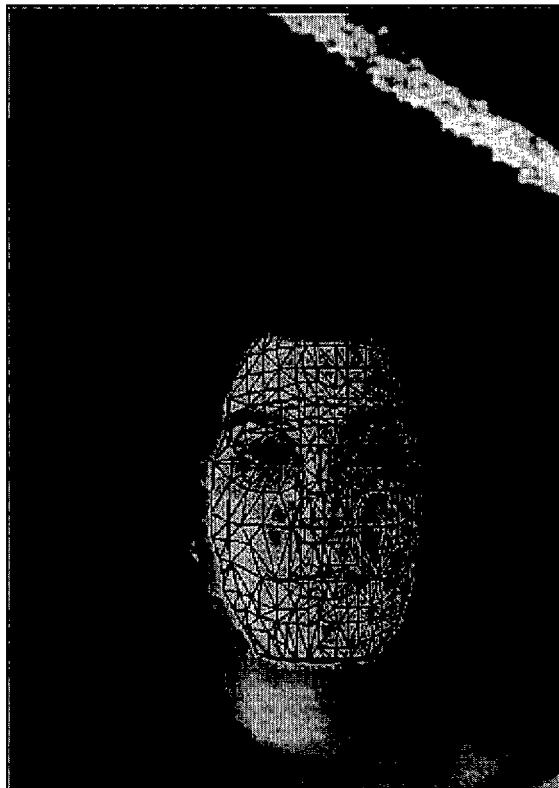


Figura A.3. Modelo ajustado a una cara.

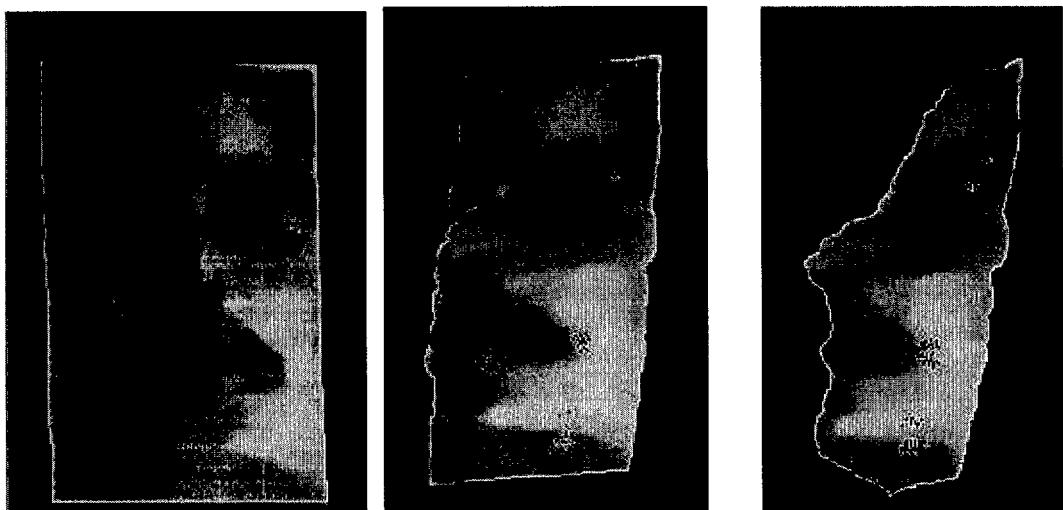


Figura A.4. Tres vistas sintéticas obtenidas de la figura anterior.

La observación de dichas vistas indica que es necesario refinar aún más el modelo si se desea utilizar para reconocimiento, haciendo una correspondencia de la textura punto a punto entre imagen y modelo, evitando la interpolación. Aunque los rasgos aparecen definidos, para reconocimiento es preciso refinar la adaptación de textura.

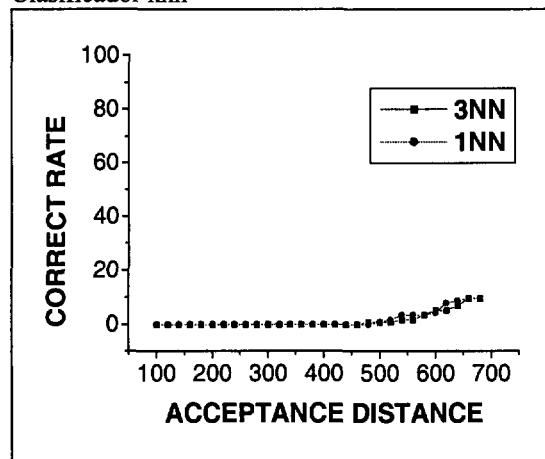
Anexo B

Verificación bidimensional, datos equilibrados.

En el presente anexo se muestran los valores no incluidos en el capítulo 4, correspondientes a valores de desplazamiento de 15 y 20 píxeles por eje para los distintos métodos de preprocesamiento y clasificadores considerados. Todas las gráficas en este anexo corresponden a datos equilibrados.

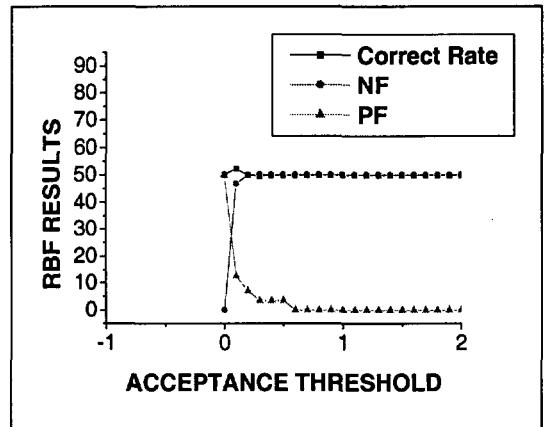
Figura B.1. Resultados con imágenes desplazadas 15 píxeles: baja resolución

Clasificador knn



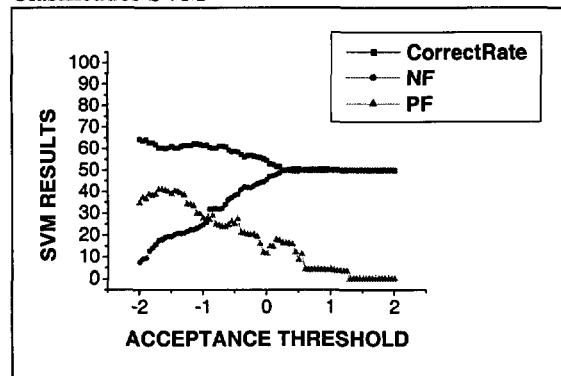
Parámetro	3NN	1NN
100	0,00	0,00
220	0,00	0,00
260	0,00	0,00
300	0,00	0,00
380	0,00	0,00
420	0,00	0,00
500	0,87	0,87
580	3,48	3,48
620	5,22	7,83
660	9,57	9,57

Clasificador RBF



Parámetro	% Correctos	% FN	% FP
0	50,06	0,00	49,94
0,4	49,90	50,08	3,45
0,8	49,94	50,06	0,00
1,6	49,94	50,06	0,00
2	49,94	50,06	0,00

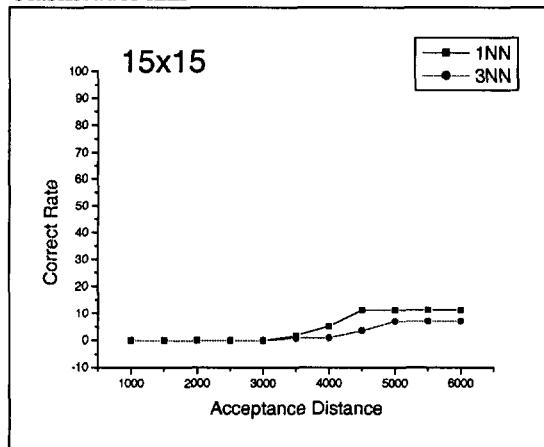
Clasificador SVM



Param	%Correctos	%FN	&FP
-2	64,27	7,37	34,65
-1,6	60,04	18,85	40,51
-1,2	61,47	22,41	33,86
-1	61,55	24,68	27,62
-0,6	59,22	36,12	24,50
0	54,44	45,40	11,40
0,8	50,39	49,67	4,02
1	50,09	49,93	4,24
1,8	50,00	50,00	0,00
2	50,00	50,00	0,00

Figura B.2. Resultados con imágenes desplazadas 15 píxeles: análisis de componentes principales

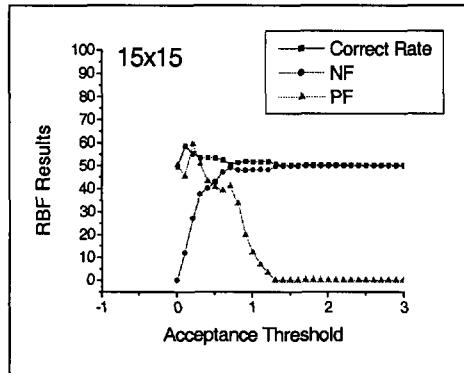
Clasificador knn



Parámetro 1-KNN 3-KNN

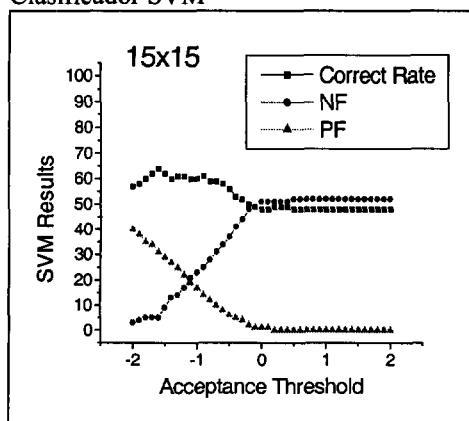
Parámetro	1-KNN	3-KNN
1000	0	0
1500	0	0
2000	0	0
2500	0	0
3000	0	0
3500	1,7	0,9
4000	5,2	0,9
4500	11,3	3,5
5000	11,3	7
5500	11,3	7
6000	11,3	7

Clasificador RBF



Parámetro	% Correctos	% FN	% FP
0,00	49,75	0,00	50,25
0,50	53,45	42,89	40,72
0,80	51,49	48,01	33,64
1,20	51,88	48,07	3,45
1,40	50,25	49,75	0,00
1,80	50,25	49,75	0,00
2,10	50,25	49,75	0,00
2,60	50,25	49,75	0,00
2,70	50,25	49,75	0,00
3,20	50,25	49,75	0,00

Clasificador SVM



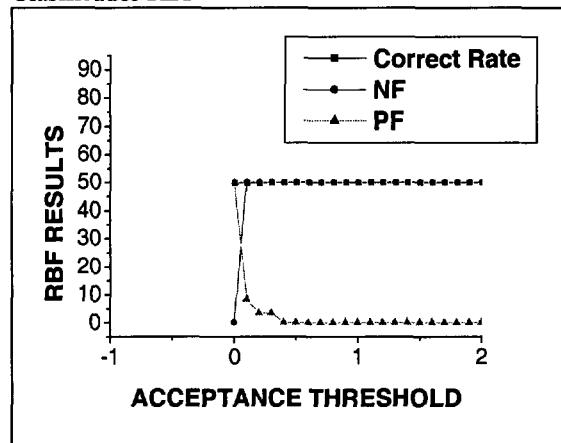
Parámetro	% Correctos	% FN	% FP
-2,00	57,12	3,00	40,00
-1,70	62,00	5,23	34,45
-1,20	61,00	17,34	22,01
-0,80	59,12	28,56	12,45
-0,60	58,34	34,00	8,21
-0,30	51,91	44,00	4,09
0,10	47,34	51,00	1,67
0,30	49,00	51,00	0,00
0,70	48,00	52,00	0,00
1,00	48,00	52,00	0,00
1,60	48,00	52,00	0,00
1,90	48,00	52,00	0,00

Figura B.3. Resultados con imágenes desplazadas 20 píxeles: baja resolución

Clasificador knn

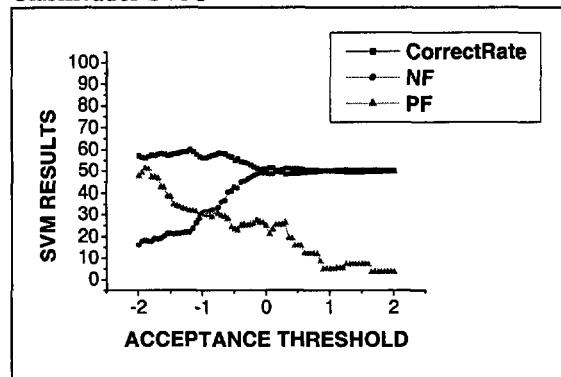
El resultado obtenido en este caso es nulo, por lo que no se muestra.

Clasificador RBF



Parámetro	% Correctos	% FN	% FP
0,00	49,95	0,00	50,05
0,30	50,01	49,97	3,45
0,80	50,05	49,95	0,00
0,90	50,05	49,95	0,00
1,40	50,05	49,95	0,00
1,70	50,05	49,95	0,00
1,80	50,05	49,95	0,00
2,10	50,05	49,95	0,00
2,60	50,05	49,95	0,00
2,70	50,05	49,95	0,00

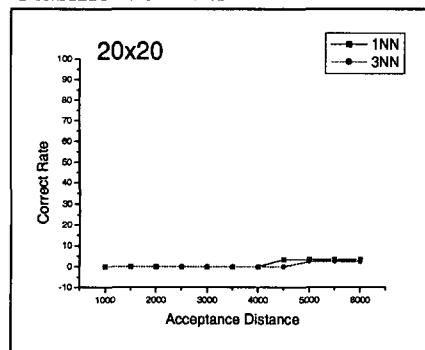
Clasificador SVM



Parámetro	% Correctos	% FN	% FP
-2,00	57,28	16,08	47,44
-1,60	57,93	20,00	42,38
-1,40	58,36	21,26	33,93
-1,10	57,80	25,89	31,28
-0,95	56,38	31,22	29,82
-0,65	58,10	36,58	28,93
-0,40	54,35	44,75	24,79
-0,05	50,86	48,72	25,95
0,05	51,51	48,45	20,97
0,40	48,66	50,85	19,12
0,65	49,48	50,26	11,98
0,95	50,17	49,80	4,67
1,10	49,78	50,12	5,25
1,40	49,53	50,26	6,90
1,60	49,70	50,16	6,90
1,85	49,83	50,09	3,45
1,90	49,83	50,09	3,45

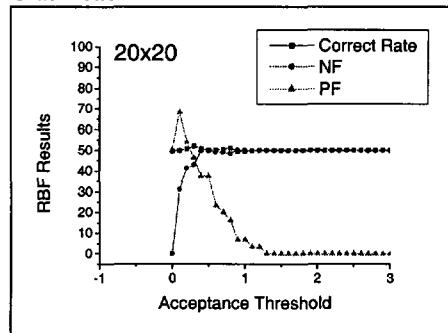
Figura B.4. Resultados con imágenes desplazadas 20 píxeles: análisis de componentes principales

Clasificador knn



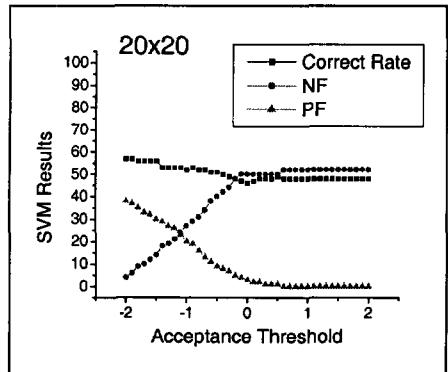
Distance	1-KNN	3-KNN
1000,00	0,00	0,00
1500,00	0,00	0,00
2000,00	0,00	0,00
2500,00	0,00	0,00
3000,00	0,00	0,00
3500,00	0,00	0,00
4000,00	0,00	0,00
4500,00	3,50	0,00
5000,00	3,50	2,60
5500,00	3,50	2,60
6000,00	3,50	2,60

Clasificador RBF



Parámetro	% Correctos	% FN	% FP
0,00	49,75	0,00	50,25
0,30	52,20	42,95	46,29
0,50	50,19	49,57	37,60
0,80	51,32	48,44	16,14
1,20	50,16	49,80	3,45
1,70	50,25	49,75	0,00
1,80	50,25	49,75	0,00
2,30	50,25	49,75	0,00
2,60	50,25	49,75	0,00
3,00	50,25	49,75	0,00

Clasificador SVM



Parámetro	% Correctos	% FN	% FP
-2,00	57,00	4,00	38,00
-1,70	56,00	10,00	33,00
-1,50	56,00	14,00	30,00
-1,20	53,00	21,00	26,00
-0,80	52,00	31,00	16,00
-0,60	51,00	38,00	11,00
-0,30	49,00	44,00	7,00
0,10	47,00	50,00	2,00
0,30	48,00	50,00	1,00
0,60	48,00	52,00	0,00
1,00	48,00	52,00	0,00
1,20	48,00	52,00	0,00
1,60	48,00	52,00	0,00
1,90	48,00	52,00	0,00

Anexo C

Verificación bidimensional, datos no equilibrados.

En el presente anexo se muestran los resultados con datos no equilibrados (el conjunto de prueba contiene cuatro fotos por cada sujeto de la base de datos, es decir en total hay menos casos de imágenes del sujeto analizado que la suma de imágenes de otros sujetos).

Sólo se muestran los clasificadores RBF y SVM, ya que los resultados del clasificador KNN se han mostrado en el capítulo 4. Sólo se muestran resultados gráficos ya que los resultados numéricos han sido comentados, analizados y estudiados en el capítulo 4.

Figura C.1.- Preprocesamiento utilizando baja resolución con clasificador SVM.

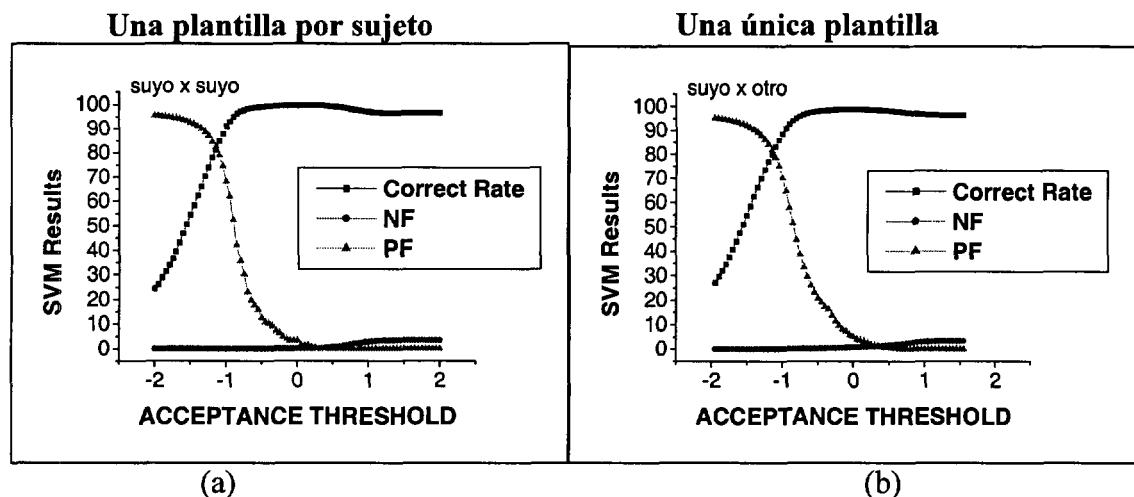


Figura C.1. La gráfica (a) muestra el resultado obtenido al usar una plantilla por sujeto y (b) con una plantilla común para toda la base de datos.

Figura C.2.- Preprocesamiento utilizando baja resolución con clasificador RBF.

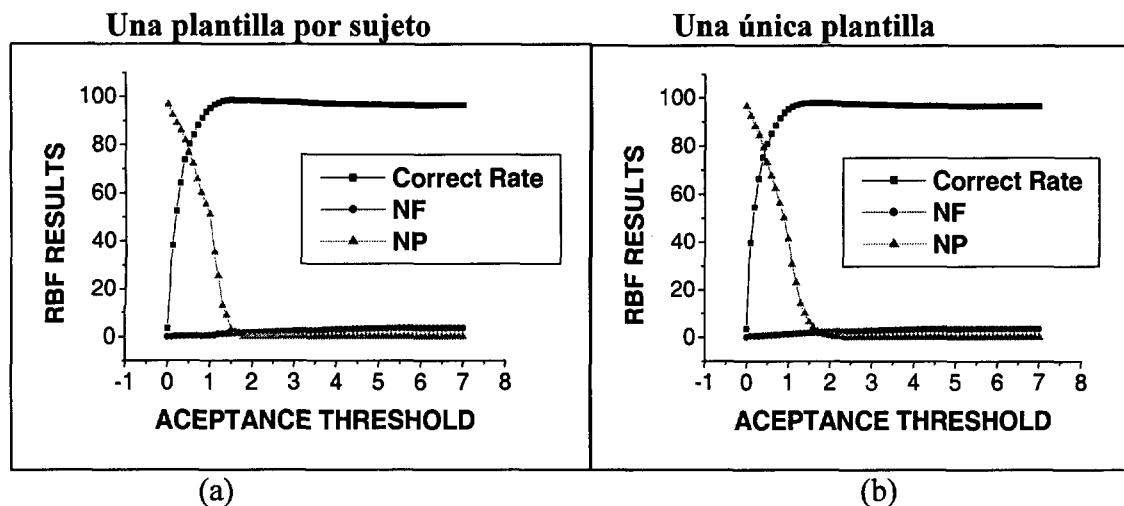


Figura C.2. La gráfica (a) muestra el resultado obtenido al usar una plantilla por sujeto y (b) con una plantilla común para toda la base de datos.

Figura C.3.- Preprocesamiento utilizando análisis de componentes principales con clasificador SVM.

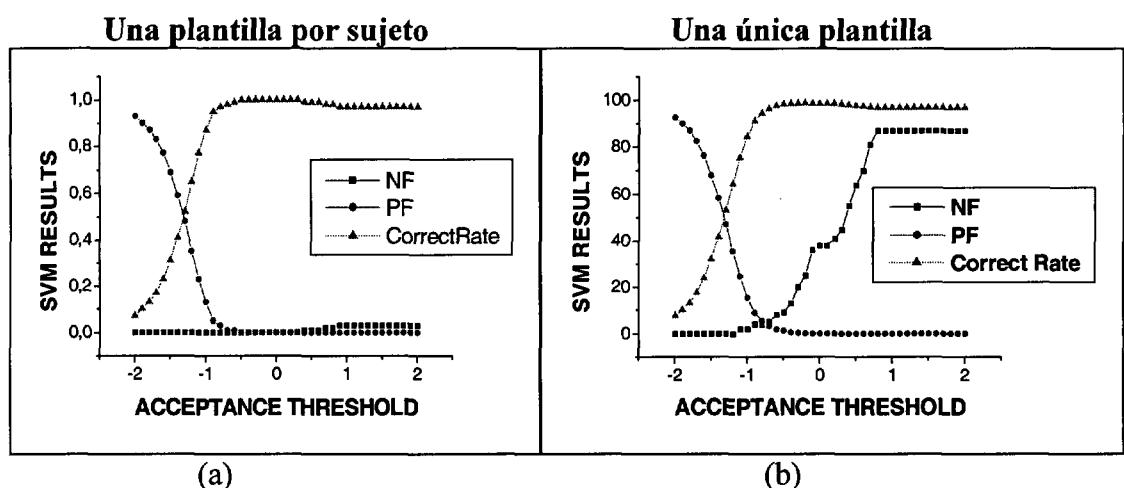


Figura C.3. La gráfica (a) muestra el resultado obtenido al usar una plantilla por sujeto y (b) con una plantilla común para toda la base de datos.

Figura C.4.- Preprocesamiento utilizando análisis de componentes principales con clasificador RBF.

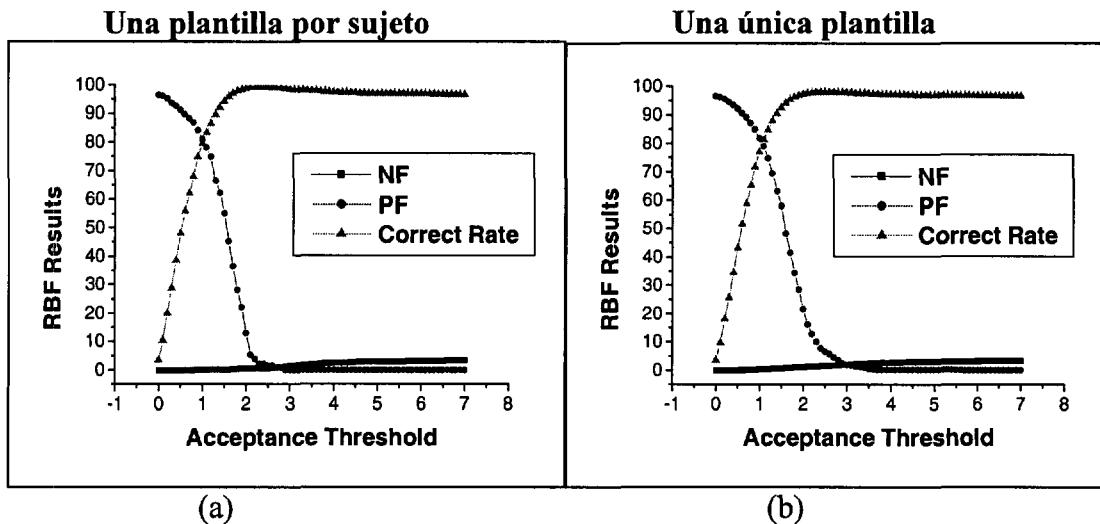


Figura C.4. La gráfica (a) muestra el resultado obtenido al usar una plantilla por sujeto y (b) con una plantilla común para toda la base de datos.

Figura C.5.- Preprocesamiento utilizando baja resolución con clasificador SVM.

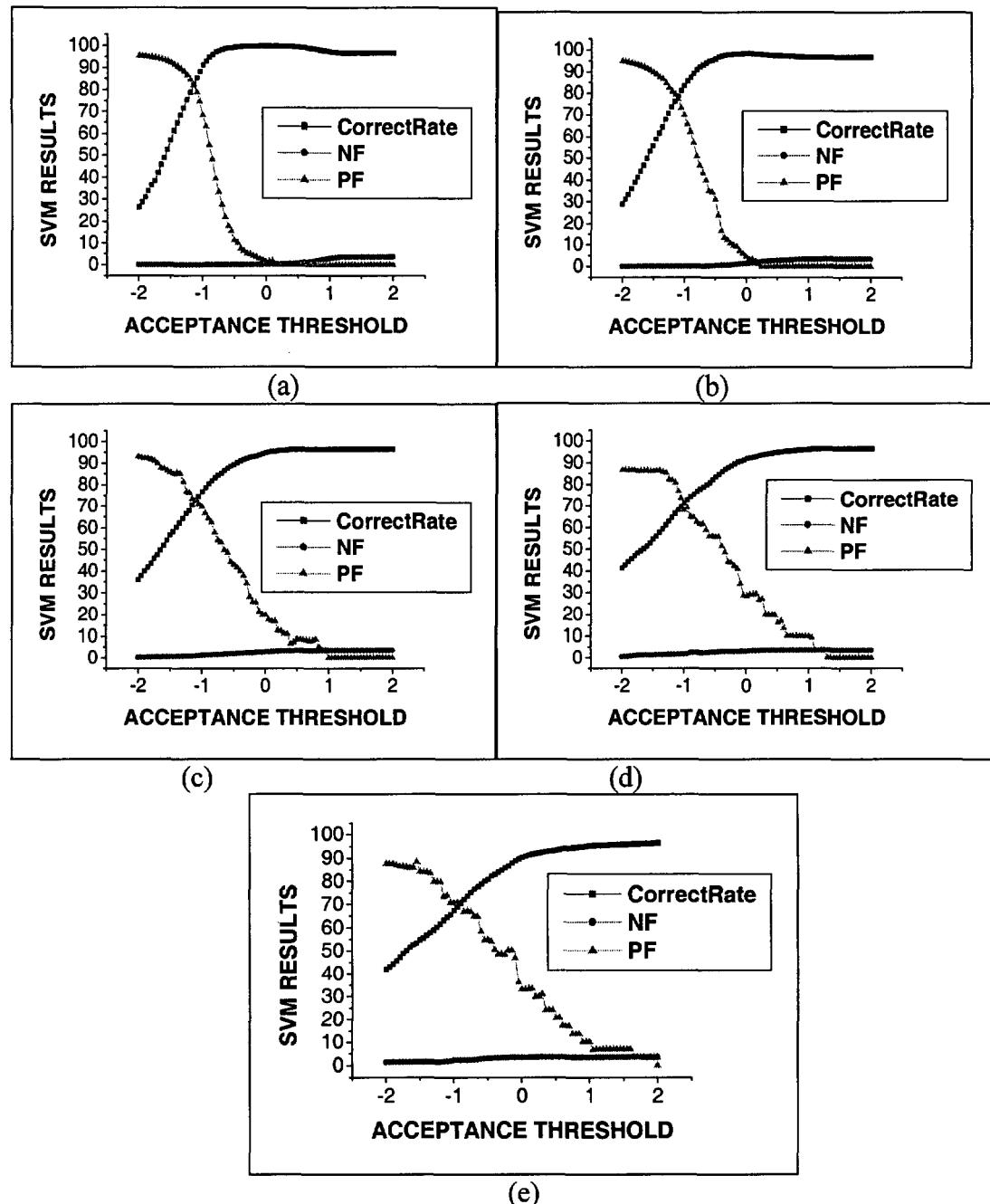


Figura C.5. Cada gráfica muestra los resultados (a) con las imágenes correctamente localizadas, (b) con un desplazamiento de 5 píxeles en cada eje, (c) desplazamiento de 10 píxeles, (d) desplazamiento de 15 píxeles, (e) desplazamiento de 20 píxeles.

Figura C.6.- Preprocesamiento utilizando baja resolución con clasificador RBF.

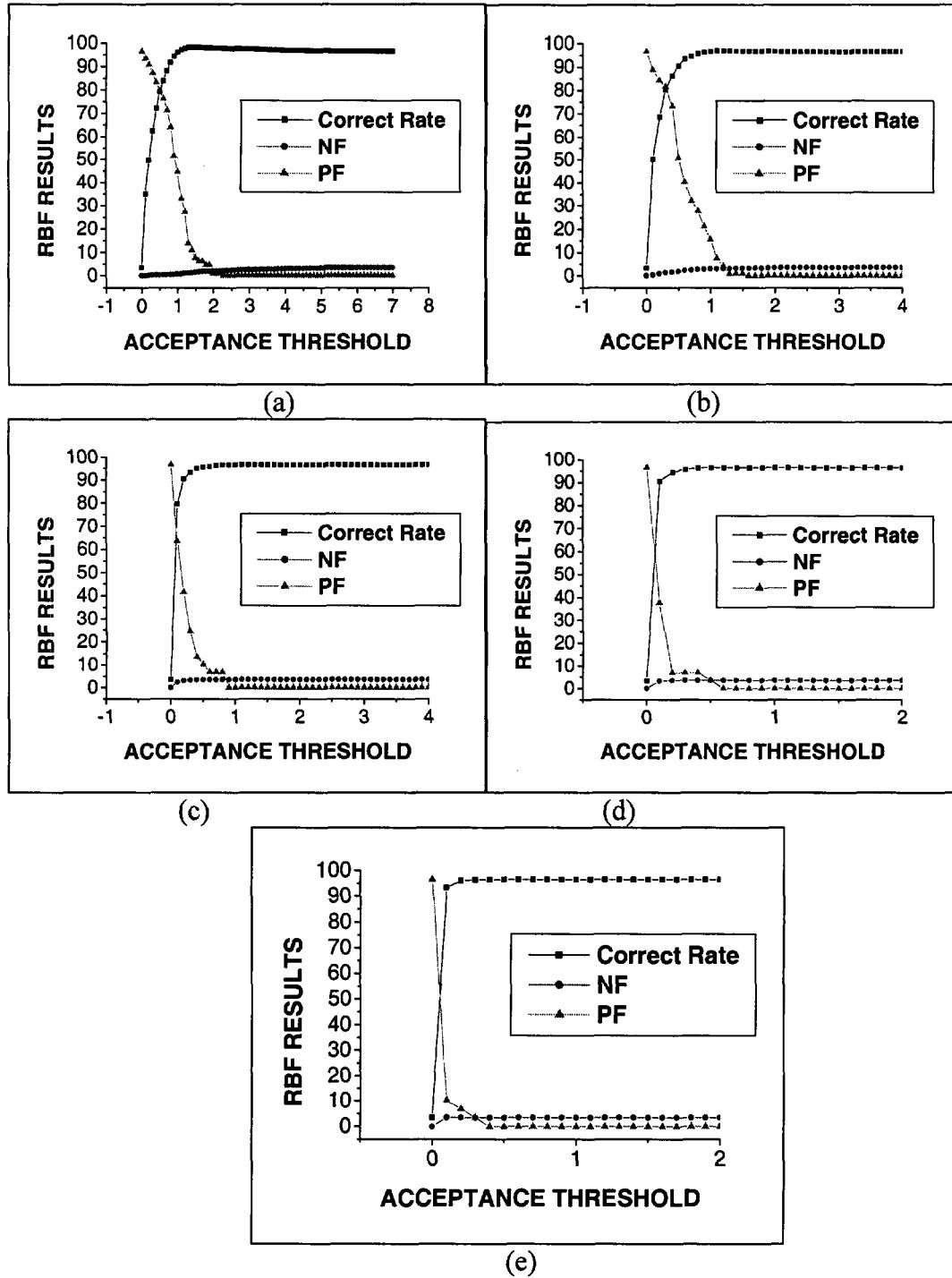


Figura C.6. Cada gráfica muestra los resultados (a) con las imágenes correctamente localizadas, (b) con un desplazamiento de 5 píxeles en cada eje, (c) desplazamiento de 10 píxeles, (d) desplazamiento de 15 píxeles, (e) desplazamiento de 20 píxeles.

Figura C.7.- Preprocesamiento utilizando análisis de componentes principales con clasificador SVM.

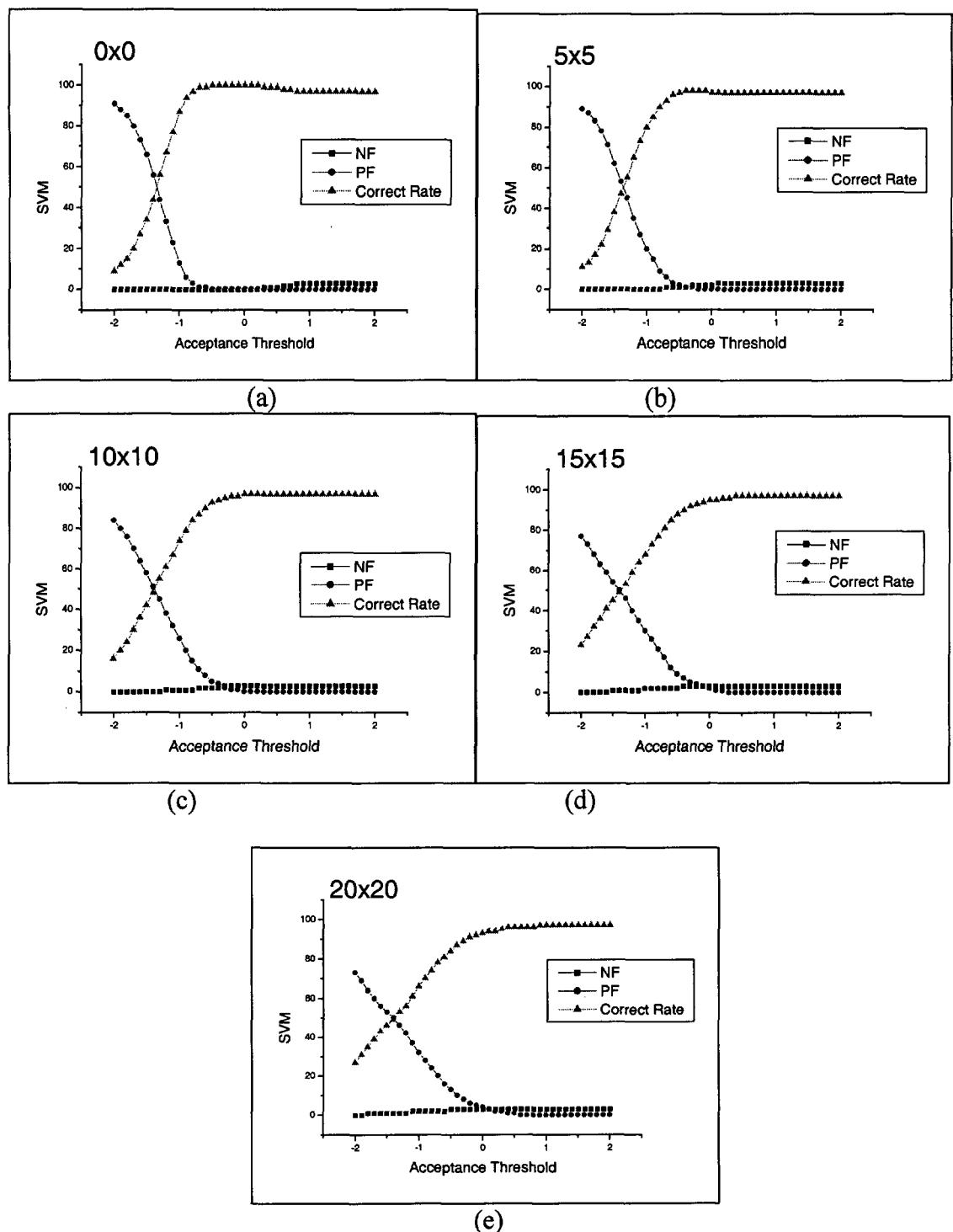


Figura C.7. Cada gráfica muestra los resultados (a) con las imágenes correctamente localizadas, (b) con un desplazamiento de 5 píxeles en cada eje, (c) desplazamiento de 10 píxeles, (d) desplazamiento de 15 píxeles, (e) desplazamiento de 20 píxeles.

Figura C.8.- Preprocesamiento utilizando análisis de componentes principales con clasificador RBF.

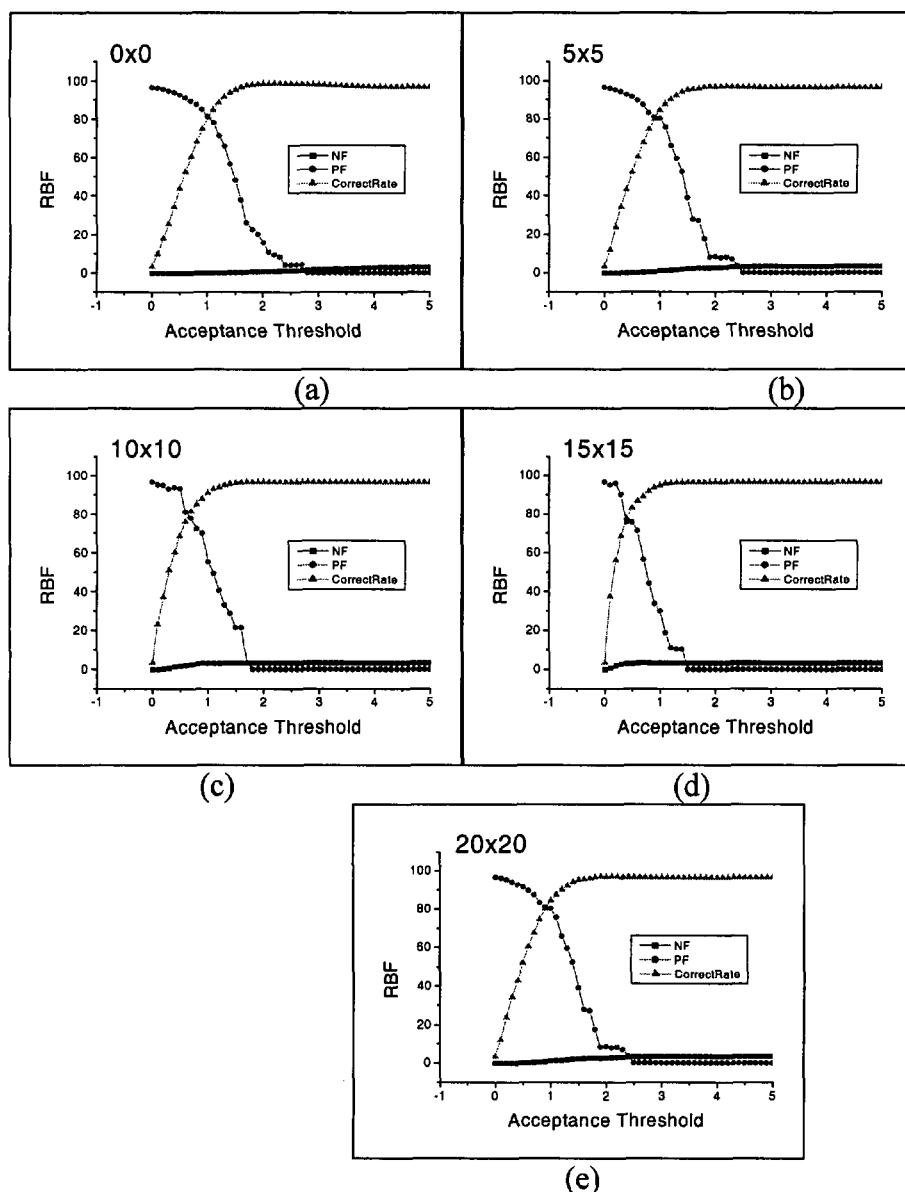


Figura C.8. Cada gráfica muestra los resultados (a) con las imágenes correctamente localizadas, (b) con un desplazamiento de 5 píxeles en cada eje, (c) desplazamiento de 10 píxeles, (d) desplazamiento de 15 píxeles, (e) desplazamiento de 20 píxeles.