

DIRECTOR:

Juan Avilés Farré

REDACTORES:

Luisa Barón Hernández José Villena Romera Marcos Rubio García

DIRECCIÓN:

Instituto Universitario de Investigación sobre Seguridad Interior Francos Rodríguez, 77 - 28039 MADRID Teléf. 91 398 85 56 - Fax 91 398 85 55 E-mail: <u>iuisi@iuisi.uned.es</u>

EDITA:

Instituto Universitario de Investigación sobre Seguridad Interior

IMPRIME:

Asociación Pro Huérfanos de la Guardia Civil (Imprenta-Escuela) Príncipe de Vergara, 248 - 28016 MADRID

Depósito legal: M-25050-2006

El Instituto Universitario de Investigación sobre Seguridad Interior no se responsabiliza de las opiniones contenidas en los artículos firmados.

EMPLEO DE SISTEMAS BIOMÉTRICOS PARA EL RECONOCIMIENTO DE PERSONAS EN AEROPUERTOS

	Página
Resumen y objetivos del proyecto	7
Reconocimiento de caras humanas	9
Descripción del proyecto	30
Conclusiones	46

Autores: Enrique Cabello Pardos Profesor Titular E.S. de Ciencias Experimentales y Tecnología Universidad Rey Juan Carlos (Investigador Principal)

> Cristina Conde Vilda Ángel Serrano Sánchez de León Licesio Jesús Rodríguez Aragón David Vázquez Bermúdez

> > Mayo, 2006



EMPLEO DE SISTEMAS BIOMÉTRICOS PARA EL RECONOCIMIENTO DE PERSONAS EN AEROPUERTOS

1.- RESUMEN Y OBJETIVOS DEL PROYECTO.

1.1.- Resumen.

El presente proyecto se desarrolló a lo largo del año 2005, probando un prototipo de un sistema de verificación facial con imágenes extraídas de las cámaras de video vigilancia del aeropuerto de Barajas. Se diseñaron varios experimentos, agrupados en dos clases. En el primer tipo, el sistema es entrenado con imágenes obtenidas en condiciones de laboratorio y luego probado con imágenes extraídas de las cámaras de video vigilancia del aeropuerto de Barajas. En el segundo caso, tanto las imágenes de entrenamiento como las de prueba corresponden a imágenes extraídas de Barajas.

Se ha desarrollado un sistema completo, que incluye adquisición y digitalización de las imágenes, localización y recorte de las caras en escena, verificación de sujetos y obtención de resultados. Los resultados muestran, que, en general, un sistema de verificación facial basado en imágenes puede ser una ayuda a un operario que deba estar vigilando amplias zonas.

1.2.- Objetivos.

Los distintos objetivos que se pretendían alcanzar se detallan a continuación.

El primer objetivo era evaluar un sistema biométrico basado en rasgos faciales en un entorno real, como puede ser el aeropuerto de Barajas. La mayoría de los sistemas biométricos utilizados en zonas de seguridad suelen estar basados en huellas dactilares, pero este es un método muy intrusivo, es decir, obliga a que el usuario sea muy colaborativo. No es posible tomar la huella de una persona sin que se percate de ello y por lo tanto, dé su consentimiento. Sin embargo un sistema biométrico basado en rasgos faciales permite el procesamiento de imágenes sin que los sujetos estudiados se percaten de ello. Además, los sistemas basados en caras son muy útiles a los operadores humanos, ya utilizan imágenes que luego pueden ser mostradas para su cotejo por parte de dicho operador.

Un objetivo adicional que se pretendía conseguir era el estudio de viabilidad del sistema de cámaras de Barajas. El uso de un sistema biométrico puede suponer el tener que realizar la instalación de ordenadores y de cámaras de video vigilancia de alta calidad, pero sólo una vez estudiados los sistemas disponibles.

Por ultimo, se deseaba probar un sistema biométrico en un entorno en el cual la iluminación no esté controlada, pero requiera, al menos potencialmente, unas condiciones de seguridad bastante altas. Estas circunstancias se dan en el aeropuerto de Barajas, sumadas además al hecho de que mantiene un flujo de sujetos muy alto.



1.3.- Antecedentes del equipo investigador.

Las primeras pruebas realizadas a sistemas de verificación facial por parte del grupo de investigación FRAV, fueron realizados mediante una pequeña y simple base de datos adquirida en condiciones no muy estrictas. Estas bases de datos preliminares fueron adquiridas en escala de grises con un número de sujetos pequeño y escasa variación y control en las condiciones de adquisición.

Un punto relevante en el desarrollo del grupo de investigación fue la realización del proyecto europeo: "Video-sensor Object Request Broker Open Architecture for Distributed Services" (VISOR BASE, IST-1999-10808), dentro del V Programa Marco de la Unión Europea. Desde enero del año 2000 hasta septiembre del 2002 se diseñó, implementó y probó un sistema de verificación facial en este proyecto. Una de las tareas que incluía este proyecto era la adquisición de una base de datos de imágenes faciales en color.

A partir de estas experiencias se procedió a adquirir la base de datos FRAV2D, con 109 personas diferentes (75 hombres y 34 mujeres) y 32 imágenes por sujeto. La finalidad principal de esta base de datos de imágenes en color era servir de conjunto de prueba de los algoritmos verificadores frente a imágenes en diferentes condiciones de adquisición. Las 32 imágenes fueron clasificadas en seis grupos según la pose y las condiciones de iluminación:

- 12 imágenes frontales: el sujeto está mirando a la cámara sin mostrar ningún gesto. Se usó luz difusa procedente de dos focos.
- 4 imágenes con un giro de 15° con respecto al eje de la cámara.
- 4 imágenes con un giro de 30° con respecto al eje de la cámara.
- 4 imágenes con expresiones faciales.
- 4 imágenes con algunos rasgos faciales ocultos: el sujeto está mirando a la cámara, por ejemplo, tapándose la mitad izquierda de su cara con la mano.
- 4 imágenes con cambios de iluminación: en este caso no se usó luz difusa, sino un fluorescente situado en el techo de la habitación.

La resolución de las imágenes de la base de datos era de 320 x 240 píxeles. En la figura 1.1 se muestran varias de las imágenes correspondientes a un individuo de la base de datos FRAV2D.

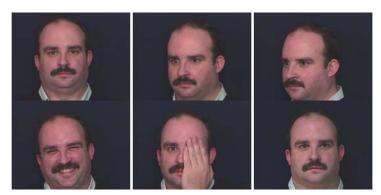


Fig. 1.1. Imágenes correspondientes a un individuo de la base de datos FRAV2D. Se muestra una imagen por condición de adquisición. De izquierda a derecha, y de arriba a abajo: frontal, giro 15°, giro 30°, gesto, oclusión e iluminación.

Actualmente se dispone de una base de datos multimodal (2D y 3D), denominada FRAV3D.



Los resultados obtenidos con estas bases de datos han sido publicados en los siguientes congresos internacionales y capítulos de libros (sólo aparecen desde el año 2004).

"Spatial approach to pose variations in face verification". Autores: Rodríguez-Aragón, L.J.; Serrano, Á.; Conde, C.; Cabello, E. Lecture Notes on Computer Science. Volumen: 3646. Springer Verlag. Editores: A. Fazel Famili, Joost N. Kok, José M. Peña, Arno Siebes, Ad Feelders. Páginas: 351-361. Fecha: septiembre 2005.

"Face verification advances using spatial dimension reduction methods: 2DPCA & SVM". Autores: Rodriguez-Aragón, L.J.; Conde, C.; Serrano, Á.; Cabello, E. Lecture Notes in Computer Science. Springer Verlag. Volumen: 3617. Editores: Roli, Fabio; Vitulano, Sergio . Páginas: 351-361. Fecha: septiembre 2005.

"3D Facial Normalization with Spin Images and Influence of Range Data Calculation over Face Verification". Autores: Conde, C.; Serrano, Á.; Rodriguez-Aragón, L.J.; Cabello, E. Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (San Diego, EEUU). IEEE Computer Society. Fecha: junio 2005.

"Influence of location over several classifiers in 2D and 3D face verification". Autores: Mata, S.; Conde, C.; Sánchez, Á.; Cabello, E. Fecha: mayo 2005. Advances Studies in Biometrics (Escuela de Verano, Alguer, Italia, 2-6 junio 2003). Revised Selected Lectures and Papers. Serie: Lecture Notes in Computer Science, Vol. 3161. Editores: Tistarelli, Massimo; Bigun, Josef; Grosso, Enrico (Eds.). Editorial: Springer-Verlag. Páginas: 153-158. ISSN: 0302-9743. ISBN: 3-540-26204-0.

"3D Facial Feature Location With Spin Images". Autores: Conde, C.; Cipolla, R.; Rodríguez-Aragón, L.J.; Serrano, Á.; Cabello, E.The International Association for Pattern Recognition (IAPR). Conference on Machine Vision Applications, Japón. Fecha: mayo 2005.

"Optimization Of A Face Verification System Using Bayesian Screening Techniques". Autores: Montes Diez, R.; Conde, C.; Serrano, Á.; Rodríguez-Aragón, L.J.; Cabello, E. IASTED International Conference on Artificial Intelligence And Applications. Innsbruck, Austria. Páginas: 427-432. Editor: M.H. Hamza. Editorial: Acta Press. Lugar de publicación: EEUU. Fecha: febrero 2005.

"Automatic Detection of the Optimal Acceptance Threshold in a Face Verification System". Autores: Montes Diez, R.; Conde, C.; Cabello, E. Lecture Notes in Computer Science. Biometric Authentication: ECCV 2004 International Workshop, BioAW 2004, Prague, Czech Republic, May 15th, 2004. Proceedings. Volumen: LNCS 3087 / 2004. Páginas: 70-79. Fecha: 2004. Editores: Maltoni, D.; Jain, A.K. Editorial: Springer-Verlag. Lugar de publicación: Alemania. ISBN: 3-540-22499-8.

"Face Verification using SVM: Influence of Illumination". Autores: Conde, C.; Ruiz, A.; Cabello, E. Libro: "Applications and Science in Soft Computing". Serie: Advances in Soft Computing Series. Páginas: 301-308. Fecha: 2004. Editores: Lotfi, A.; Garibaldi, J.M. Editorial: Springer-Verlag. Lugar de publicación: Alemania. ISBN: 3-540-40856-8.

2.- RECONOCIMIENTO DE CARAS HUMANAS.

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.). 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.

2.1.1.- 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, ya que en el tiempo de duración de este proyecto no se puede resolver de forma automática y con resultados satisfactorios. Teniendo presente que hay programas disponibles que realizan esta tarea, se han probado dos y se ha seleccionado el que mejores resultados ha proporcionado.

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. Celebrado hace unos años, 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).

2.1.2.- 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.

2.1.3.- 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 buscan 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.- Estado del arte de la investigación 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].

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 2.1 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).

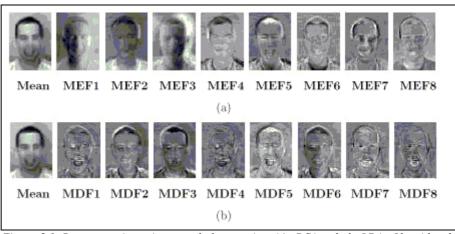


Figura 2.1. Representaciones internas de la aproximación PCA y de la LDA. Obtenidas de [Wen99]. La imagen media se muestra en la primera columna. La secuencia (a) corresponde a las características más expresivas. La secuencia (b) corresponde a las más discriminantes.

El algoritmo EBGM (véase la Figura 2.2) 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.



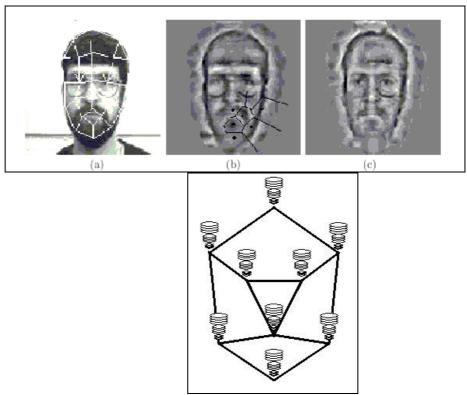


Figura 2.2. Representación interna de la aproximación EBGM (obtenidos de Wis97). (a) representa un grafo sobre una cara. (b) es la reconstrucción de la cara a partir del grafo. (c) representa la reconstrucción de la cara a partir de un grafo conjunto en el que se ha considerado en cada nodo el jet que 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 2.3. En este caso la elección de funciones kernel o base para rasgos faciales locales (por ejemplo ojos, boca, nariz, etc.) y la selección de la localización de estos rasgos es todavía una cuestión abierta.

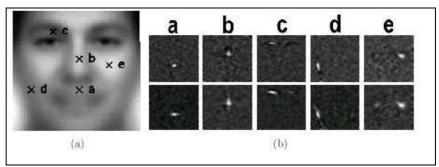


Figura 2.3. Representación interna del método LFA (obtenida de [Pen96]). (a) representa la cara media con la localización de cinco características. (b) muestra los kernels asociados a cada característica (fila superior) y la correspondiente correlación (fila inferior).

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].
- Híbridos (combinan ambos tipos de características) [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 menor medida.

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.

2.2.1.- 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.

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.



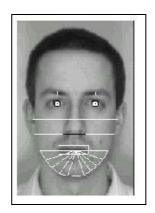


Figura 2.4. Características geométricas consideradas en [Bru93].

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.

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

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 2.5 muestra los seis primeros autovectores de un conjunto de entrenamiento.



Figura 2.5. 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 este proceso; su principal aplicación consiste en normalizar el conjunto de entrenamiento. 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 2.6.





Figura 2. 6. Cara media para la misma base de datos.

Resultados obtenidos con PCA.

Turk y Pentland [Tur91] 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 2.500 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:

- 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.

2.2.3.- 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 tambien contengan información útil para el reconocimiento.

2.2.4.- Correspondencia de plantillas.

Una serie de autores como Brunelli y Poggio [Bru93] (véase la Figura 2.4) 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 2.7. 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.

2.2.5.- 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 esta 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.

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 repuesta 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.

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 4.096 unidades de entrada, 80 unidades en la capa oculta de tipo sigmoidal y 4.096 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.

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. Con una 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 una ecuación de similaridad, 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. Sin embargo, se sale fuera del objetivo de este proyecto, ya que únicamente se han considerado imágenes provenientes de cámaras de vídeo.

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 2.8. 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.

2.3.- Sistemas y pruebas comerciales de reconocimiento y verificación facial.

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.

2.3.1.- 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].

2.3.2.- 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 Intensitive 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 Intensitive 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 1.600 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 2.9 muestra los resultados de verificación en la prueba HCInt para los ocho sistemas, con un porcentaje de falsa aceptación de un 1%.

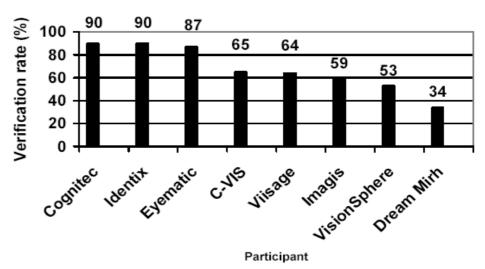


Figura 2.9. Resultados de la prueba HCInt.

Para los tres sistemas con mejores resultados en el caso anterior, la Figura 2.10 muestra la variación en el porcentaje de reconocimiento correcto en función del porcentaje de falsa aceptación.



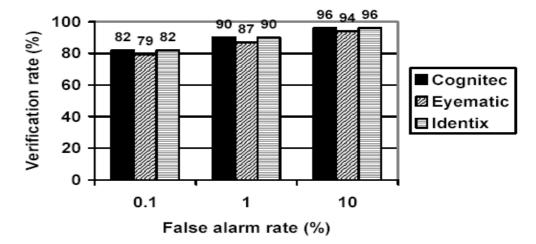


Figura 2.10. Porcentaje de reconocimiento en función del porcentaje de falsas alarmas.

La Figura 2.11 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%).

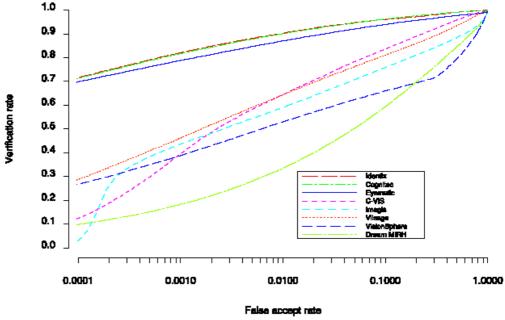


Figura 2.11. Reconocimiento correcto en función del porcentaje de falsa aceptación.

Los resultados de la prueba MCInt se muestran en las Figuras 2.12 y 2.13. La Figura 2.12 se obtuvo manteniendo un porcentaje de falsa aceptación del 0.01%. La Figura 2.13 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.

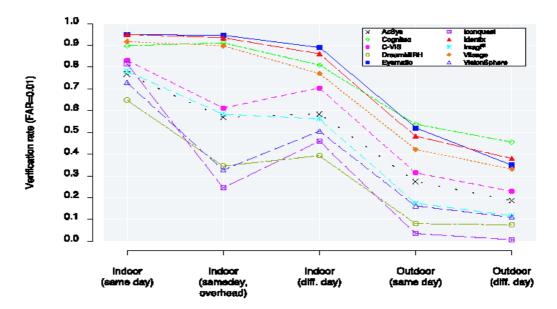


Fig 2.12. Resultados MCInt con FAR=0.01.

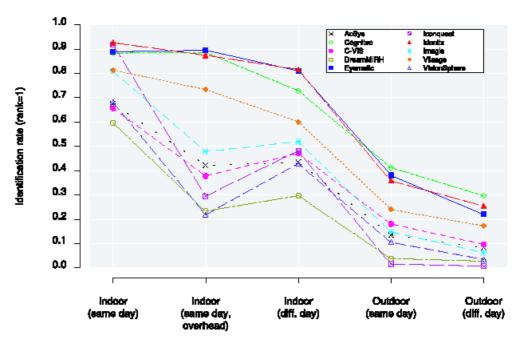


Fig 2.13. Resultados MCInt, independientemente del FAR.

3.- DESCRIPCIÓN DEL PROYECTO.

3.1. Introducción.

En una serie de visitas iniciales al aeropuerto de Barajas (6, 13 y 20 de abril, 13 y 26 de mayo y 1 y 2 de junio de 2005) se observó el entorno en el cual se iba a desarrollar el proyecto, a la vez que se adquirían las imágenes de prueba de los diferentes algoritmos.

En estas visitas iniciales se comprueba que las cámaras del aeropuerto graban en formato analógico (en cinta de video) desde el centro de cámaras en las instalaciones de AENA. Se consideró que esta es la mejor forma de adquirir las imágenes ya que se utiliza el sistema sin interferir o añadir ningún elemento. Además se observó que hay cámaras



con una calidad de imagen muy buena, con un potente zoom. Hay otras cámaras, más antiguas, en las que la calidad es bastante peor. Debido a la naturaleza del aeropuerto, la iluminación no está controlada, apareciendo iluminación natural, reflejos y brillos en la imagen. En los arcos detectores de metales, la iluminación es más uniforme, lo cual a priori nos permite suponer que los resultados de verificación facial en estas condiciones serán mejores.

En el presente proyecto se han utilizado las cámaras ya instaladas, por lo tanto las imágenes obtenidas son en el rango del espectro visible. Se decidió que la cámara se mantuviera estática (sin seguir a ningún sujeto) pero aumentando el zoom para que la cara ocupara la mayor parte de la imagen. Las cámaras no enfocan a los sujetos de frente y la posición de las mismas no es uniforme, se encuentran situadas a alturas muy variables. La implantación de un sistema de verificación o reconocimiento facial por una parte supondría la colocación de cámaras dedicadas a la verificación (por ejemplo en los controles de los arcos) y también un coste computacional (en equipos informáticos dedicados exclusivamente a este proceso). Las cámaras tienen ganancia automática.

En las visitas iniciales se comprueba que hay varias cámaras enfocando a un mismo punto. Se decidió realizar una prueba en la cual se grabe el paso de un sujeto simultáneamente con dos cámaras. Se realiza con las cámaras 60 y 2022 (también puede emplearse la cámara 61) en la terminal de facturación de T2 (ver Figura 3.1). Sin embargo los resultados no son satisfactorios. La colocación de las cámaras es tal que permite realizar video vigilancia con ellas, pero la verificación facial con dos cámaras simultáneamente no es posible. Una de las cámaras mantiene una visión prácticamente lateral de la escena (mostrando los perfiles de los sujetos), mientras que la otra cámara permite una visión frontal.



Figura 3.1.-Dos cámaras enfocando al mismo punto

En las visitas iniciales se comprueba que es posible seguir a una persona a través de varias cámaras. En concreto se realizaron varias pruebas de seguimiento por el pasillo T1 → T2 con los sujetos avanzando tanto en la cinta transportadora como andando por el pasillo. Las cámaras utilizadas son 112 (el sujeto aparece de espaldas), 2144, 87, 2085. Para realizar el seguimiento los observadores no sólo se fijan en la cara del sujeto, sino que también (o, mejor dicho, sobre todo) lo hacen por la ropa del sujeto. El interés de un sistema de verificación facial en este caso creemos que es muy limitado.

3.2.- Experimentos a realizar y cámaras seleccionadas.

Una vez realizadas la primera serie de visitas y conociendo la disposición de las cámaras, se plantearon una serie de experimentos a realizar. Para los cuales se decidió adquirir imágenes en dos tipos de entornos distintos, aunque inicialmente se consideraron



tres situaciones iniciales, descartando una de ellas. Las tres situaciones consideradas se describen a continuación.

Escenario 1: Pasillo.

En este experimento se pretende aplicar un sistema de verificación en un sujeto cuando es observado secuencialmente por dos cámaras distintas. Para ello se pretende que los sujetos deban pasar primero por una cámara y luego por la otra. Las cámaras que se consideraron son las que se encuentra en zonas como puede ser un pasillo.

Para realizar este experimento se estudiaron las siguientes secuencias de cámaras: 112 (el sujeto aparece de espaldas) – 2144 – 87 – 2085; 2144 – 2085; y 2127 – 79.

Debido a las condiciones de iluminación, posición y calidad de las imágenes, la selección final es la secuencia con las cámaras 2127 – 79 (ver Figuras 3.2 y 3.3). De estas dos cámaras la 79 es de bastante peor calidad que la 2127. Estos vídeos son los que se designan como Cinta 1 (cámara 79) y Cinta 2 (cámara 2127). En las siguientes imágenes se muestran escenas de los diferentes vídeos seleccionados.



Figura 3.2.- Fragmento del vídeo Cintal



Figura 3.3.- Fragmento del vídeo Cinta2.

Escenario 2: Arco detector.

En este experimento se pretende verificar la identidad de un sujeto. Se decidió seleccionar una cámara en un arco detector de metales. Se comparan las imágenes de laboratorio (en este caso, de la Base de datos FRAV2D) con las imágenes aquí obtenidas para el mismo sujeto.



Las cámaras estudiadas han sido: 131, 76, 129 y 215, aunque esta última puede enfocar a dos arcos. Debido a la calidad de la imagen y a otras consideraciones (el fondo en la cámara 76 tiene mucho movimiento, el ángulo de alguna cámara es muy pronunciado) las cámaras seleccionadas para este experimento son las 131 y 2154 (el arco en el cual la imagen es más frontal). Este vídeo es el que se designa como Arcos. La siguiente imagen (Figura 3.4) muestra un ejemplo de uno de los vídeos grabados en los arcos detectores de metales.



Figura 3.4.- Fragmento del vídeo Arcos

Escenario 3: Multicámara.

Este experimento se pensó como complementario al 1. Se buscaba realizar el experimento 1 pero con dos cámaras enfocando al mismo sujeto simultáneamente (en el experimento 1 el sujeto aparece secuencialmente primero en una cámara y luego en la siguiente). Este experimento fue descartado ya que no es un proceso que se realice por parte del servicio de vigilancia y además la disposición de las cámaras no esta diseñada para que sea utilizada de esta forma. Las cámaras probadas en este experimento fueron la 60 y la 2022.

Cámaras no seleccionadas.

Una de las tareas iniciales que condiciona los resultados finales, es la adecuada elección de las cámaras con las que se va a trabajar. En esta etapa fue muy importante contar con la ayuda del personal de la Guardia Civil destinado en Barajas. Con más de mil cámaras instaladas, el conocimiento experto nos permitió centrarnos desde el principio en unas pocas, candidatas a ser utilizadas. A continuación, en las Figuras 3.5, 3.6, 3.7 y 3.8, se muestran algunas imágenes de cámaras no seleccionadas, indicando las causas.



Figura 3.5.- Cámara 2144: Imágenes muy baja calidad.





Figura 3.6.- Cámara 2085: Imágenes de baja calidad.



Figura 3.7.- Cámara 76: Fondo con mucho movimiento, que dificulta la localización de la cara.



Figura 3.8.- Cámara 129: Ángulo pronunciado y poca iluminación.

Las pruebas realizadas y las cámaras seleccionadas pueden verse en la Figura 3.9, que representa el plano del aeropuerto de Barajas, con las distintas zonas en que se realizaron grabaciones.

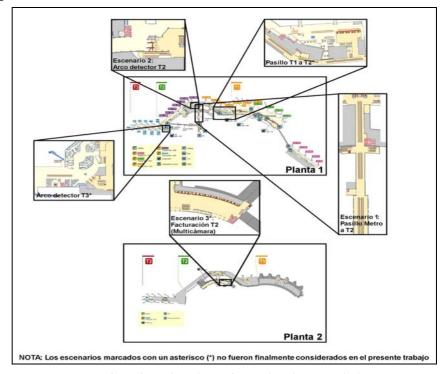


Figura 3.9.- Plano de Barajas con las cámaras probadas.



3.3.- Localización de la cara.

La primera etapa que se debe abordar en el proyecto es la localización y recorte de la cara. Esta etapa es uno de los elementos críticos del proyecto, si una cara no es detectada no puede ser verificada o reconocida. Las condiciones ambientales hacen este proceso complejo. Dos métodos se han probado en esta fase: Verilook 2.0 y Torch.

3.3.1.- Localización de la cara por VeriLook 2.0

El primer programa que se utilizó para detectar la cara es el programa VeriLook 2.0 (http://www.neurotechnologija.com/index.html). Se probó con imágenes de la base de datos obtenida en condiciones de laboratorio y con imágenes obtenidas en los arcos de seguridad, que son las de mejor calidad. Los resultados con imágenes del aeropuerto se muestran a continuación en las Figuras 3.10, 3.11, 3.12, 3.13 y 3.14; también se muestra un resultado con una imagen de otro individuo similar a los mostrados.

Individuo1: Foto_123.jpg Frame processed (0.018649 sec.). Face detection failed.



Figura 3.10.- Imagen Foto_123.jpg procesada por VeriLook v2.0.

Individuo 2: Foto_56.jpg Frame processed (0.019289 sec.).Face detection failed.



Figura 3.11.- Imagen Foto_56.jpg procesada por VeriLook v2.0.

Individuo 3: Foto_66.jpg Frame processed (0.018868 sec.). Face detection failed.



Figura 3.12.- Imagen Foto_66.jpg procesada por VeriLook v2.0.



Individuo 4: Foto_69.jpg Frame processed (0.016938 sec.).Face detection failed.



Figura 3.13.- Imagen Foto_69.jpg procesada por VeriLook v2.0.

Individuo 5: Foto_114.jpg Frame processed (0.019639 sec.). Face detection failed.



Figura 3.14.- Imagen Foto_114.jpg procesada por VeriLook v2.0.

El programa responde de manera eficiente con fotografías frontales obtenidas en condiciones de laboratorio, aunque las pruebas se han realizado sobre un número muy reducido de individuos y sería conveniente realizar pruebas más exhaustivas sobre un número mucho mayor de personas. Sin embargo el sistema se muestra muy poco fiable ante giros de la cara y elementos de ocultación, siendo incapaz de obtener las características correspondientes a la cara de una persona dentro de la fotografía (incluso con imágenes de laboratorio). Las pruebas realizadas en un entorno real como es el video de un arco de seguridad, el programa no ha podido realizar ninguna detección en las fotografías seleccionadas, siendo estas además seleccionadas a mano, tomando las imágenes más favorables (personas con el rostro descubierto y situadas de manera frontal). Este sistema se desestimo debido a los pobres resultados obtenidos con él.

3.3.2.- Localización de la cara por Torch 3.0

El segundo programa utilizado se basa en el metodo de Viola-Jones (P. Viola and M. J. Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. In Computer Vision and Pattern Recognition, 2001a.) de deteccion automática de cara, en nuestro caso hemos utilizado la implementación Torch (R. Collobert, S. Bengio, and J. Mariéthoz. Torch: a modular machine learning software library. Technical Report IDIAP-RR 02-46, IDIAP, 2002). Se trata de un algoritmo robusto que proporciona muy buenos resultados incluso con imágenes de baja calidad.

A continuación, en las Figuras 3.15, 3.16 y 3.17, se muestran ejemplos de recorte válido con el método localización de la cara aplicado, en los distintos videos probados.





Figura 3.15: BD Arcos. Situaciones de localización normal



Figura 3.16: BD Cinta 1. Situaciones de localización normal



Figura 3.17.- BD Cinta 2. Situaciones de localización normal

INSTITUTO UNIVERSITARIO DE INVESTIGACIÓN SOBRE SEGURIDAD INTERIOR

La cámara enfocada en la cinta 1 ofrecía imágenes de peor calidad que las de cinta 2 y los arcos; además la iluminación no es adecuada para un proceso de verificación o reconocimiento. Sin embargo, el algoritmo probado es capaz de localizar correctamente la cara en un porcentaje grande de casos, incluso en algunas situaciones tan adversas como las mostradas en la Figura 3.18.

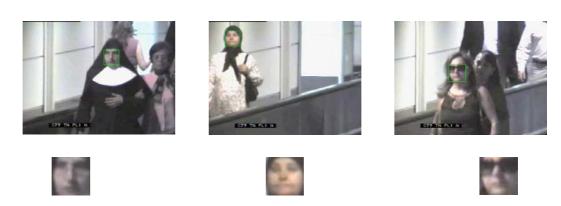


Figura 3.18.- BD Cinta1. Situaciones de localización adversas

Este método de recorte ha sido seleccionado para localizar y recortar las áreas en los diferentes vídeos. Sin embargo, en algunos casos se han observado detecciones incorrectas, como puede verse en la Figura 3.19. Hay un grupo de ellas que corresponden a aquellas zonas de una imagen que se asemejan a una cara pero que en realidad son sombras, arrugas de una tela o a que debido a la perspectiva de la cámara se funden varias caras (normalmente dos) en una sola.



Figura 3.19.- BD Cinta2 Ejemplos de fallos razonables

Otras veces el algoritmo reconoce como caras zonas que no tienen ninguna relación con el rostro humano, se muestran algunos ejemplos en la Figura 3.20. Esto se produce casi de forma anecdótica no llegando al 1% de los casos. Hemos de tener presente que, al estar trabajando con secuencias de vídeo, los fallos de localización en una imagen pueden recuperarse con localizaciones correctas en las imágenes siguientes.















Figura 3.20.- BD Cinta2 Ejemplos de fallos exagerados.

Empíricamente se ha observado que el algoritmo suele fallar en las siguientes situaciones:

- 1. El sujeto se encuentra de espaldas o de lado.
- 2. Aparece oclusión de parte de la cara: lleva gafas de sol o la cabeza cubierta, la cara aparece tapada por la mano, va hablando por el móvil, o bien sencillamente aparece detrás de otra persona.
- 3. Mantiene un giro muy grande de la cabeza (mirando al suelo o hacia los lados).
- 4. Cuando se encuentra lejos de la cámara.

3.3.2.1.- Porcentaje de aciertos

Para conocer cuál es el porcentaje caras detectados por el nuestro algoritmo, se contaron el número de personas que pasan por delante de la cámara y se les ve la cara con un tamaño mínimo para poder ser verificadas. Después se realizo una clasificación manual de los rostros detectados.

BD	Duración estudiada	Nº individuos contados	Nº de individuos detectados	Porcentaje de localización correcta
Arcos	20 min	34	34	100%
Cinta1	26 min	250	137	54%
Cinta2	20 min	238	158	66%
Frav2D				100%

Tabla 3.1.- Resultados de localización.

Como podemos observar en la Tabla 3.1, el porcentaje de caras detectadas depende directamente de la calidad de las imágenes. La cara debe de tener un tamaño mínimo de 38x38, las caras con un tamaño menor no se han tenido en cuenta en el proceso de localización. En las cintas 1 y 2 se ve a la gente desde lejos, pero sólo cuando la persona se encuentra muy próxima a la cámara ha sido tenida en cuenta. Esto provoca que la probabilidad de detección sea menor. En cambio en la cinta de Arcos, localizamos el 100% de las caras, pues éstas se encuentran en primer plano.

INSTITUTO UNIVERSITARIO DE INVESTIGACIÓN SOBRE SEGURIDAD INTERIOR

Otro parámetro de la localización de la cara que nos interesaba medir era el número de imágenes que se puede obtener de cada sujeto. En las Tablas 3.2 y 3.3 se muestra el número de individuos a los que se les localizaba un número de imágenes dado.

N° de Img. Detect.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	+ 20	+ 30	+ 40	+ 50	+ 60	+ 70	+ 80	+ 90	+ 100
Nº de Ind.	2	2	1	1	0	2	0	1	2	2	0	0	2	4	5	5	1	2	0	1	0	0	0	0
Tabla 3.2 BD Arcos. Número de imágenes obtenidas de cada individuo (Analizando los frames pares únicamente) Nº de Img. 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 + + + + + + + + +																								
Detect.																								
Nº de Ind.	33	1 8	12	5	5	9	3	4	6	5	3	4	4	6	8	12	8	5	2	1	0	0	3	2
Tabla 3.3 BD Cinta2. Número de imágenes obtenidas de cada individuo (Analizando los frames pares únicamente)																								

En las tablas puede observarse que se dispone de un número suficiente de sujetos con varias imágenes, lo cual permite verificar varias veces a un mismo sujeto. En promedio se analizaron 20 minutos de vídeo.

3.4.- Bases de datos obtenidas.

La Tabla 3.4 resume las bases de datos que se han obtenido. La primera corresponde a la grabación en los arcos de seguridad, la segunda y la tercera a la cinta 1 y a la 2, la cuarta es una base de datos de imágenes de laboratorio obtenida por el equipo de investigación (http://frav.escet.urjc.es). Puede observarse a continuación la secuencia de imágenes para un sujeto, observándose la distinta calidad de las imágenes.

Nombre de la base de datos	Nº individuos	Nº de fotos / individuo	Ejemplos
Arcos	46	10	Figura 3.21
Cinta1	94	10	Figura 3.22
Cinta2	183	10	Figura 3.23
FRAV2D	109	32	Figura 3.24

Tabla 3.4.- Características de las Bases de Datos Obtenidas.

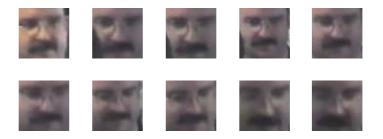


Figura 3.21: Imágenes de un sujeto en la BD Arcos.





Figura 3.22: Imágenes del mismo sujeto en BD Cinta1.

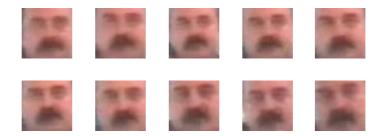


Figura 3.23: El mismo sujeto en BD Cinta2.



Figura 3.24.- El mismo individuo en la BD FRAV2D.

Puede observarse en las Figuras 3.21, 3.22, 3.23 y 3.24 que la calidad de las imágenes es muy variada, desde una alta calidad obtenida en el laboratorio a una calidad bastante baja. Otros factores que se observan es la variabilidad en la iluminación y en la posición de la cara. En el laboratorio la cara es casi frontal y en las imágenes obtenidas en el aeropuerto el sujeto puede girar libremente su cara.

3.5.- Pruebas de verificación realizadas.

3.5.1.- Evaluación de un sistema biométrico.

Habitualmente el término reconocimiento facial es usado de manera general para hacer referencia a dos aplicaciones diferentes: reconocimiento y verificación. Hablaremos de reconocimiento o identificación en el caso de que la identidad del sujeto sea deducida únicamente a partir de sus características faciales. La entrada del sistema sería la imagen del individuo, y como salida se indicaría de qué persona se trata. Los sistemas de Verificación o autenticación son aquellas en que además de la imagen, se indica qué persona reclama ser la correspondiente a esa cara. El sistema autentifica esta relación, indicando mediante una salida bivaluada si esta aseveración es correcta o es falsa.

En ambos tipos de reconocimiento, se dispone de una base de datos donde se recoge la información facial de los diferentes individuos. Esta base de datos es introducida en el sistema, para que aprenda alguna de sus características. A este conjunto se le llama conjunto de entrenamiento o de galería. Imágenes de otras personas que



pueden estar o no incluidas en el conjunto de entrenamiento, son utilizadas como conjunto de prueba o validación. En el caso de un sistema de reconocimiento o identificación, se trata de un sistema de relación uno-a-muchos, en el sentido de que cada imagen de prueba es comparada con todos los individuos del conjunto de entrenamiento, para encontrar con cuál de ellos el resultado del sistema es mejor. Un sistema de verificación o autenticación se trata de un sistema de relación uno-a-uno, ya que la imagen de prueba sólo es comparada con la persona que se "reclama" ser.

Actualmente existen muchas más aproximaciones al problema de la verificación facial que a la identificación. Esto es debido a que por un lado, la identificación es una tarea de mayor complejidad, y por otro lado a que la gran mayoría de las técnicas de verificación podrían adaptarse a identificación, aunque con un coste computacional mayor.

La evaluación de la calidad de un sistema de verificación requiere de un detallado análisis de los posibles aciertos y fallos del sistema. Como hemos explicado anteriormente, un sistema de verificación tiene como salida dos valores: aceptación o rechazo. Una salida con valor "acepto" indicará que efectivamente la imagen introducida y la persona que se reclama ser corresponden. En caso contrario indicaría que se trata de un impostor, y la salida del sistema sería "rechazo". Si se repitiera esta prueba para un conjunto de individuos, se obtendrían unas distribuciones de impostores y de auténticos semejante a la mostrada en la Figura 3.25.

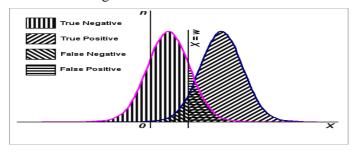


Figura 3.25. Ejemplo de distribución de casos auténticos e impostores, en función de un valor de salida X de un clasificador. El valor X=w es el umbral óptimo que separa los casos auténticos de los impostores.

Existen por lo tanto cuatro posibilidades, como se indica en la Figura 3.26:

- una persona se identifica correctamente ante el sistema y éste le acepta: verdadero positivo (TP, de "*True Positive*").
- una persona se identifica correctamente ante el sistema, pero este le rechaza: falso positivo (FN, de "False Negative").
- un impostor se identifica con la identidad de otro y el sistema acepta: Falso Positivo (FP, de "*False Positive*").
- un impostor se identifica con la identidad de otro y el sistema le rechaza: Falso Negativo (TN, de "*True Negative*").

Es decir, las salidas del sistema pueden ser aciertos (a los que llamaremos como Verdaderos positivos o negativos) o pueden ser fallos (serán los Falsos positivos y negativos). En todo sistema de verificación obviamente se pretenden minimizar los Falsos Positivos y los Falsos Negativos. En la Figura pueden verse esta clasificación de las salidas del sistema en función de las distribuciones de densidad de las salidas del clasificador y del valor del umbral de aceptación (X=w).



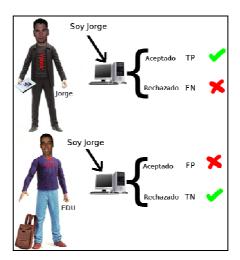


Figura 3.26. Aciertos y Fallos posibles en un sistema de Verificación Facial.

Si se tienen en cuenta todas las salidas del sistema para la confrontación de todos los sujetos del conjunto de validación, incluidos la confrontación con ellos mismos, se obtendrá una serie de valores para estos parámetros. La representación de estos fallos y aciertos en porcentajes, se realizará mediante las siguientes expresiones:

$$\%TP = \frac{TP}{TP + FN}$$

$$\%TN = \frac{TN}{TN + FP}$$

$$\%FN \circ FRR = \frac{FN}{FN + TP}$$

$$\%FP \circ FAR = \frac{FP}{FP + TN}$$

Mediante estas expresiones se cumple que:

$$\% FP + \% TN = 100\%$$
 de impostores $\% FN + \% Tp = 100\%$ de auténticos

Por lo tanto los errores del sistema serán de dos tipos: el %FN o FRR (*False Rejection Rate*) y el %FP o FAR (*False Aceptance Rate*). La decisión de aceptar o rechazar del sistema verificador viene impuesta por la medida de la similitud entre la imagen introducida y la imagen almacenada en la base de datos de la persona que se "reclama" ser. Si el valor de similitud supera un determinado umbral, se considerará una aceptación. En caso contrario, se rechazará al individuo. El valor del umbral utilizado definirá si se trata de un sistema de alta seguridad, o si se trata de un sistema más "amigable" pero de más baja seguridad. En función del valor de este umbral, el sistema generará un número concreto de FN, FP, TP y TN. Una manera global de representar la evolución del sistema en función del valor del umbral es la llamada curva ROC ("*Receiver Operator Characteristic*"). Esta curva, que puede verse en la Figura 3.27 representa el porcentaje de falsos positivos vs. el porcentaje de falsos negativos.

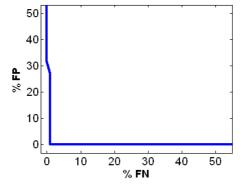


Figura 3.27. Ejemplo de curva ROC.



Para caracterizar la calidad del sistema verificador mediante un único valor, habitualmente se elige el punto de la curva ROC en que el %FN es igual al %FP. Este punto es el corte de la curva ROC con la sección y=x del primer cuadrante, como puede verse en la Figura 3.28, y se denomina "Equal Error Rate" (EER). Cuanto más bajo sea el EER nos indicará que el sistema tiene menos fallos.

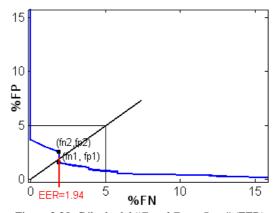


Figura 3.28. Cálculo del "Equal Error Rate" (EER).

3.5.2.- Resultados obtenidos.

Una de las características del tratamiento de imagen es la gran información de la que se dispone. Para manipular y procesar toda esta información, es necesario utilizar alguna técnica estadística de síntesis de la información, o reducción de la dimensión (número de variables). Es decir, ante un banco de datos con muchas variables, el objetivo será reducirlas a un menor número perdiendo la menor cantidad de información posible. En el transcurso de este proyecto se han utilizado diferentes técnicas de reducción de la dimensión que se describen someramente a continuación:

- PCA (Análisis de Componentes Principales): trata de hallar componentes (factores) que sucesivamente expliquen la mayor parte de la varianza total. Para ello se considera una imagen de MxN píxeles, como un vector de MxN componentes. Cada componente viene dado por la intensidad de cada uno de los píxeles.
- 2DPCA (Análisis de Componentes Principales Bimensional): se trata de una técnica semejante a PCA, pero donde se tiene en cuenta la vecindad de cada píxel dentro de la imagen, tratando una imagen como una matriz, y realizando el análisis en dos dimensiones.
- 2DLDA (Análisis Lineal Discriminante Bidimensional): técnica estadística en la que se pretende encontrar las variables que mejor discriminen o distingan a elementos de clases diferentes. Para ello se tiene en cuenta tambien la información de vecindad de los píxeles, con una imagen como matriz, realizando un tratamiento en dos dimensiones.
- CSA (Coupled Subspace Analysis): generalización de las técnicas anteriores. Mediante este método se infieren dos subespacios acoplados de baja dimensión que reconstruyen de manera óptima las matrices imagen en las direcciones fila y columna.

Se realizaron cinco experimentos. En los tres primeros se entrenó el sistema con imágenes de laboratorio y se probó con imágenes de Barajas. En los dos siguientes se entrenó y se probó con imágenes obtenidas en Barajas. El tamaño de las imágenes es de 50x50 píxeles. Las pruebas realizadas se resumen en la tabla 3.5, junto con el EER (Equal Error Rate).

	Entren	amiento	Pru	eba	N° individuos	EER				
	BD	Nº imágenes	BD	N° imágenes	maiviauos	1dpca	2dpca	2dlda	csa	
Gráfica 1	Frav2D	5	Arcos	5	6	39.44	39.44	40	42.22	
Gráfica 2	Frav2D	5	Cinta1	5	4	45.45	26.70	45.37	25	
Gráfica 3	Frav2D	5	Cinta2	5	5	27.00	30.64	47.50	33.33	
Gráfica 4	Arcos	5	Cinta 1	5	3	32.50	33.75	33.33	34.00	
Gráfica 5	Cinta 1	5	Cinta 2	5	4	28	28.66	36	29.33	

Tabla 3. 5.- Resultados obtenidos con los diferentes métodos.

Los resultados mostrados en la Tabla 3.5 corresponden al EER, es decir, al punto de la gráfica en el cual el porcentaje de falsos positivos coincide con el de falsos negativos. Aumentando el nivel de seguridad, el porcentaje de falsos positivos disminuye (un impostor tiene menos posibilidades de poder pasar) aunque el coste es que aumenta el porcentaje de falsos negativos. Con esto se pueden obtener porcentajes de falsos positivos del orden del 1%. El coste que supone el aumento del porcentaje de falsos negativos indica que a veces el sujeto tratara de entrar y el sistema no se lo permitirá. Pero si en lugar de usar una única imagen se utiliza una secuencia de video, en la cual hay varias imágenes que se procesan y el resultado final es el obtenido en la mayor parte de la simganes, el porcentaje de falsos positivos y de falsos negativos disminuye grandemente. Por otra parte se puede conseguir también un sistema más amistoso, en el cual el porcentaje de falsos negativos sea pequeño, aunque esto supone que muchos impostores podrán acceder al sistema (estos parámetros casi nunca se van a usar). Por eso el EER puede suponer una "cota superior" de los errores que comete un sistema.

3.6. Análisis de los resultados.

Puede observarse que en la localización de la cara en un entorno real, el método Torch 3.0 proporciona unos resultados mucho mejores que el sistema comercial de VeriLook 2.0, siendo este último descartado para la realización de las diferentes pruebas.

En cuanto a los experimentos de verificación, los resultados se han realizado con una base de datos bastante pequeña, debido sobre todo a que el emparejamiento en el entrenamiento entre las imágenes del mismo sujeto en las distintas cintas ha de ser hecho manualmente. Una vez recortadas las imágenes y reducida su dimensión, este emparejamiento es muy complejo.

Antes de proceder al análisis de los resultados de verificación, conviene indicar que el grupo de investigación utilizando las imágenes FRAV2D (obtenidas en condiciones de laboratorio) obtiene unos porcentajes de EER como los mostrados en la Tabla 3.6. Para un sistema comercial, como es Face-intellect, disponible en http://www.itvsec.com, la empresa afirma que obtiene un 22,6 % de identificaciones falsas. Para el análisis también se han utilizado los resultados mostrados en el apartado 2.3 del presente documento.

EER		
Conjunto de prueba	PCA	2DPCA
Imágenes frontales	1.8	1.0
Iluminación cenital	6.6	2.1
Rotación 15 grados	27.3	18.3
Imágenes con gesto	15.8	12.9
Oclusión en la cara	34.7	29.5

Tabla 3.6.- EER con FRAV2D.

La tasa de error obtenida en Barajas es bastante alta frente a los experimentos de laboratorio, pero hemos de tener en cuenta que las condiciones de trabajo son muy diferentes. En los experimentos realizados para el presente proyecto se han utilizado dos cámaras que ofrecen una calidad de imagen aceptable y una que ofrece una calidad muy pobre. Los resultados muestran que, en general, cuando se incluyen las imágenes de baja calidad provenientes de cinta 1 los resultados empeoran.

Los métodos considerados de verificación facial, desarrollados por el equipo de investigación ofrecen resultados muy esperanzadores al compararlos con los del sistema comercial Face-intellect. El sistema Face-intellect sólo funciona en interiores, con cámaras dedicadas y de alta calidad, así mismo se insiste en que el sistema Face intellect sólo funciona en condiciones de iluminación adecuadas. En nuestros experimentos nada de esto ha sido posible, ya que las cámaras no son de alta calidad y las condiciones de iluminación son las habituales de un aeropuerto, muy diferentes a las ideales necesarias incluso para los sistemas comerciales.

4.- CONCLUSIONES.

El presente estudio ha permitido probar un sistema de verificación facial en el aeropuerto de Barajas. Se han abordado todas las fases del mismo: selección de las cámaras, adquisición de vídeos, digitalización de los mismos, localización y recorte de las caras, creación de las bases de datos y pruebas de verificación.

Los resultados obtenidos son prometedores, comparables con los obtenidos por sistemas comerciales, teniendo en cuenta que las condiciones de adquisición no son las más favorables. El sistema utilizado ha buscado utilizar las cámaras instaladas en Barajas, sin añadir ningún tipo de elemento adicional o interferir con el funcionamiento habitual. Esto ha limitado tanto la posición y orientación de las cámaras como la calidad de la imagen.

Aunque las bases de datos utilizadas son de un número de sujetos reducido, permiten mostrar la validez de la aproximación considerada; mostrando que un sistema de verificación facial puede ser una ayuda a la decisión. Las pruebas realizadas indican que es posible el procesamiento de imágenes en tiempo real, a pesar de no ser uno de los objetivos del proyecto.

Uno de los objetivos alcanzados ha sido un estudio de las diferentes cámaras disponibles en Barajas y su posible utilización para un sistema de biometría facial. Las cámaras más modernas ofrecen una calidad aceptable, las más antiguas tienen una calidad muy pobre para su uso por un sistema de verificación facial. Además, una de las ventajas de un sistema basado en cámaras de video-vigilancia es que en ningún momento los sujetos se han sentido estudiados.

La apertura de la Terminal T4 de Barajas, donde los sistemas de visión son ya digitales permitirá mejorar las condiciones de adquisición de imágenes y obtener mejores resultados.



AGRADECIMIENTOS.

Queremos agradecer la concesión del proyecto realizada por el IUISI, las facilidades ofrecidas por AENA y a los voluntarios que accedieron a transitar por Barajas para ser estudiados. También queremos agradecer muy sinceramente, la inestimable y valiosa ayuda del personal de la Guardia Civil destinado en Barajas.

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.

[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.

[Bak00] S. Baker, T. Kanade. Hallucinating faces.Proc fourth IEEE International Conference on Face and Gesture Recognition. P 83-88. 2000.

[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, N° 7, p 711-720. Jul. 1997.

[Bla01] D. M. Blackburn, J. M. Bone, P. J. Phillips. FRVT 2000 Report. Technical Report. 2001. http://www.frvt.org.

[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.

[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.

[Cot90] G. W. Cottrell, M. K. Fleming. Face recognition using unsupervised feature extraction. Proc. Int. Conf. Neural Network. P 322-325. 1990.

[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.

[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.

[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.

[Hje01] E. Hjelm, B. K. Low. Face detection: a survey. Computer vision and image understanding. V 83, P 236-274. Sept 2001.

[Hwa00] W. S. Hwang J. Weng Hierarchical discriminant regression. IEEE Transactions in Pattern Analysis and Machine Intelligence, Vol 22. P 1277-1293. Nov. 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.

[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

INSTITUTO UNIVERSITARIO DE INVESTIGACIÓN SOBRE SEGURIDAD INTERIOR

[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. Lehtiö. 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. Venkatesth. 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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[Pur] Base de datos facial. http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html.

[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.



INSTITUTO UNIVERSITARIO DE INVESTIGACIÓN SOBRE SEGURIDAD INTERIOR

[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.

[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. Grosso. Active vision-based face authentication. Image and Vision Computing. Num 18. P 299-314. 2000.

[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.

[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.

[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, N° 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.

[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.

[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.

[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.

IuIsI

INSTITUTO UNIVERSITARIO DE INVESTIGACIÓN SOBRE SEGURIDAD INTERIOR



Creado por Convenio de colaboración entre el Ministerio del Interior, Dirección General de la Guardia Civil y la Universidad Nacional de Educación a Distancia.



FORMACIÓN:

Cursos:

Derechos Humanos. Cooperación Internacional. Seguridad Vial. Medio Ambiente. Técnicas de Comunicación.

Jornadas:

La colaboración de la Guardia Civil con las Comunidades Autónomas. Diciembre, 2006

INVESTIGACIÓN:

Ayudas para la realización de proyectos de investigación en materia de seguridad pública.

PUBLICACIONES:

Desde el año 2004 se editan análisis sobre materias relacionadas con la Seguridad, que pueden ser consultados en la página web del IUISI: http://www.uned.es/investigacion/IUISI_publicaciones.htm

"Cátedra de Seguridad Aeroportuaria"

Cursos:

Básico: Dirigidos a miembros de la Escala Básica de Cabos y Guardias de la Guardia Civil, a realizar en la Escuela de Especialización.

Avanzado: Dirigidos a las Escalas de Suboficiales, Oficiales y Superior de Oficiales de la Guardia Civil, a realizar en la Escuela de Especialización.

Investigación: Ayuda de Proyecto de Investigación en materia de Seguridad Aeroportuaria.

Con la colaboración de:

undación Aer

"Cátedra de Delitos Económicos y Fiscales"

Curso: sobre "Investigación Financiera y Blanqueo de Capitales", dirigido a miembros de la Guardia Civil.

Investigación: Ayudas a Proyectos de Investigación en materia de Delincuencia Económica e Investigación Patrimonial.

Fundación

 \mathbf{I}

Con la colaboración de:

Información:

Instituto Universitario de Investigación sobre Seguridad Interior Centro "Andrés Manjón" de la UNED Francos Rodríguez, 77 - 28039 MADRID Teléfs. 91 398 85 56/57 Fax 91 398 85 55 Correo: iuisi@iuisi.uned.es





INSTITUTO UNIVERSITARIO DE INVESTIGACION SOBRE SEGURIDAD INTERIOR

iuisi@iuisi.uned.es / 91 398 85 56