Verificación biométrica facial mediante información 2D y 3D

Biometric facial verification using 2D and 3D information

Cristina Conde[(S,*)] y Enrique Cabello[(S)]

Face Recognition and Artificial Vision Group. University Rey Juan Carlos. C/ Tulipán, s/n, Móstoles 28933, Spain

[(*)] Email: cristina.conde@urjc.es S: miembro de SEDOPTICA / SEDOPTICA member

Recibido / Received: 07/10/2008. Versión revisada / Revised version: 28/05/2009. Aceptado / Accepted: 01/06/2009

RESUMEN:

El objetivo de este artículo es resumir el trabajo llevado a cabo en una reciente tesis doctoral, centrado en la realización del diseño, implementación y prueba de un sistema de verificación facial multimodal que combine información bidimensional y tridimensional. Se realizó un estudio de la capacidad de discriminación facial de la información 2D y 3D. Se han diseñado tres sistemas verificadores diferentes, basados en distintos tipos de información: imagen de textura (2D), imagen de profundidad (2.5D) y mallado tridimensional (3D). Los verificadores 2D y 2.5D se basan en la técnica de análisis de componentes principales y máquinas de vectores soporte. El verificador 3D se basa en el algoritmo *iterative closest point*. Se ha diseñado un método de localización de rasgos faciales con la técnica de *spin images*.

Palabras clave: Biometría, Verificación Facial, Localización de Rasgos Faciales, Reconocimiento de Objetos 3D.

ABSTRACT:

The objective of this paper is to review the work in a recent thesis devoted to the design, development and test of a multimodal face verification system than combines bidimensional and threedimensional information. A deeply study of the discrimination capacity of 2D or 3D information was made. It has been designed and developed three different face verification systems, based on different data: texture image (2D), range data (2.5D) and three dimensional mesh (3D). The 2D and 2.5D verifiers are based on the principal component analysis and support vector machine. The 3D verifier it is based on the measure of the quality of surfaces adjustment using the algorithm iterative closest point. An automatic 3D facial feature location method was designed too based on the spin images technique.

Key words: Biometrics, Face Verification, Facial Feature Location, 3D Object Recognition.

REFERENCIAS Y ENLACES

[1] R.M. Bolle, J. H. Connell, S. Pankanti, N. K. Ratha and A.W. Senior, Guide to Biometrics, Springer Professional Computing (2004).

[2] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, "Overview of the face recognition grand challenge", *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005* I, art. no. 1467368, pp. 947-954 (2005).

[3] K. W. Bowyer, K. Chang, P. J. Flynn, "A survey of approaches to three-dimensional face recognition", *Proceedings - International Conference on Pattern Recognition* **1**, pp. 358-361 (2004).

[4] J. Phillips, P. Grother, R. Micheals, D. Blackburn, E. Tabassi, M. Bone, "Facial recognition vendor test", (2002). Evaluation report. http://www.frvt.org/FRVT2002/

[5] E. Bailly-Baillière, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariéthoz, J. Matas, K. Messer, V. Popovici, F. Porée, B. Ruiz, J.-P. Thiran, "The BANCA database and evaluation protocol", *4th International*

 © Sociedad Española de Óptica

Conference on Audio- and Video-Based Biometric Person Authentication, AVBPA, pp. 625-638, Springer-Verlag (2003).

[5] W. Zhao, R. Chellappa, P. J. Phillips, A. Rosenfeld, "Face recognition: a literature survey", *ACM Comput. Surv.* **35**, 399-458 (2003).

[7] A. Scheenstra, A. Ruifrok, R. C. Veltkamp, "A survey of 3D face recognition methods", *Lect. Notes Computer Sci.* **3546**, 891–899 (2005).

[8] J. Kittler, A. Hilton, M. Hamouz, J. Illingworth, "3D assisted face recognition: a survey of 3D imaging", *Modelling 3D Imaging for Safety and Security A3DISS*, San Diego, CA (2005).

[9] G. Gordon, "Face recognition based on depth maps and surface curvature", *Proceedings of SPIE* **1570**, *Geometric Methods in Computer Vision*, 234-247 (1991).

[10] T. Nagamine, T. Uemura, I. Masuda, "3D facial image analysis for human identification", *International Conference on Pattern Recognition*, 324-327 (1992).

[11] H. Tanaka, M. Ikeda, H. Chiaki, "Curvature-based face surface recognition using spherical correlation - principal directions for curved object recognition", *3rd International Conference on Automatic Face and Gesture Recognition*, pp. 372-377 (1998).

[12] M. A. Turk, A. Pentland, "Eigenfaces for recognition", *J. Cognitive Neurosci.* **3**, 71-86 (1991).

[13] C. Hesher, A. Srivastava, G. Erlebacher, "A novel technique for face recognition using range imaging", *7th International Symposium on Signal Processing and its Applications*, Vol 2, pp. 201-204 (2003).

[14] G. Pan, Z. Wu, Y. Pan, "Automated 3D face verification from range data", *International Conference on Acoustics, Speech and Signal Processing*, pp. 192-196 (2003).

[15] N. Mavridis, F. Tsalakanidou, D. Pantazis, S. Malasiotis, M. Strintzis, "The HISCORE face recognition application: Affordable desktop face recognition based on a novel 3D camera", *Proceedings of the International Conference on Augmented Virtual Environments and 3D Images* (2001), http://uranus.ee.auth.gr/hiscore

[16] K. I. Chang, K. W. Bowyer, P. J. Flynn, "Multi-modal 2D and 3D biometrics for face recognition", *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 187 (2003).

[17] X. Lu, D. Colbry, A. Jain, "Three-dimensional model based face recognition", *Proceedings - International Conference on Pattern Recognition* **1**, 362-366 (2004).

[18] V. Blanz, T. Vetter, "Face recognition based on fitting a 3D morphable model", *IEEE T. Pattern Anal.* **25**, 1063-1074 (2003).

[19] A. Ansari, M. Abdel-Mottaleb, "3D face modeling using two views and a generic face model with application to 3D face recognition", *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance*, pp. 37–44 (2003).

[20] http://www.frav.es

[21] T. Heseltine, N. Pears, J. Austin, "Three dimensional face recognition using surface space combinations", *Proceedings of the British Machine Vision Conference* (2004).

[22] P. J. Flynn, K. W. Bowyer, P. J. Phillips, "Assessment of time dependency in face recognition: an initial study", *Audio-and Video-Based Biometric Person Authentication*, pp. 44–51 (2003).

[23] C. Conde, A. Serrano, L. J. Rodríguez-Aragón, E. Cabello, "3D facial normalization with spin images and influence of range data calculation over face verification", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, AA3DISS (2005).

[24] A. E. Johnson, *Spin-Images: A Representation for 3-D Surface Matching*, PhD Thesis, Robotics Institute, Carnegie Mellon University (1997).

[25] C. Cortes, V. Vapnik, "Support vector network", *Machine Learning* **20**, 273-297 (1995).

[26] P. J. Besl, N. D. McKay, "A method for registration of 3-D shapes", *IEEE T. Pattern Anal.***14**, 239-256 (1992).

[27] Z. Zhang, "Iterative point matching for registration of free-form curves and surfaces", *Int. J. Comput. Vision* **13**, 119-152 (1994).

[28] C. Conde, A. Ruiz, E. Cabello, "PCA vs low resolution images in face verification", *Proceedings of the 12th International Conference on Image Analysis and Processing*, IEEE Computer Society (2003).

[29] C. Conde, A. Ruiz, L. Pastor, E. Cabello, "Face verification using SVM: influence of illumination", en *Applications and Science in Soft Computing*, Springer Verlag Advances in Soft Computing Series (2003).

 © Sociedad Española de Óptica

[30] C. Conde , R. Cipolla, L. J. Rodríguez-Aragón, A. Serrano, E. Cabello, "3D facial feature location with spin images", *Conference on Machine Vision Applications*, The International Association for Pattern Recognition (IAPR), pp. 418-427 (2005).

[31] C. Conde, A. Serrano, E. Cabello, "Multimodal 2D, 2.5D & 3D face verification", *IEEE International Conference on Image Processing*, pp. 2061–2064 (2006).

[32] C Conde, *Verificación Facial Multimodal 2D y 3D*, Tesis Doctoral, Universidad Rey Juan Carlos (2006).

1. Introducción

Existen muchas características fisiológicas que pueden ser cuantificadas para identificar a un individuo (conocidas como características biométricas), como pueden ser las huellas dactilares, el patrón del iris, la voz, etc. Una de las principales ventajas del reconocimiento facial, es que se trata de un método no intrusivo, es decir, los datos pueden ser adquiridos incluso sin que el sujeto se percate de ello. En la Fig. 1 se presentan distintos métodos de identificación biométrica en función de la relación entre el grado de intrusismo en la adquisición de información fisiológica y el poder de discriminación [1].

En la actualidad pueden adquirirse dos tipos de datos faciales (en referencia a su aspecto externo): las imágenes de intensidad, donde se representa la textura de la cara, y los datos tridimensionales, que recogen la estructura geométrica facial. A su vez, la información tridimensional puede representarse de dos modos diferentes: mediante una imagen de rango: se trata de una imagen en niveles de gris, donde la intensidad de cada píxel representa la profundidad del objeto en ese punto; o mediante una nube de puntos en el espacio, habitualmente aproximados a una superficie mediante un conjunto de polígonos. La representación mediante imágenes de rango se trata de una imagen 2D, donde se han proyectado los puntos 3D sobre un plano. Como se trata de una representación bidimensional de información tridimensional, se denomina habitualmente imagen 2.5D.

En la Fig. 2 pueden observarse distintos ejemplos de estos tipos de representación facial de un mismo individuo. Todos estos tipos de datos son complementarios, ya que proporcionan información diferente.

La principal limitación de los sistemas basados en una representación facial de textura es su dependencia de las condiciones de iluminación y de la posición de la cara. Por el contrario, la representación 3D, por su propia definición, no depende de la iluminación, y además permite la normalización de la cara en posición.

Existen actualmente en la comunidad biométrica dos corrientes diferentes [2]:

- Una en la que se afirma que la información tridimensional es más poderosa a la hora de realizar el reconocimiento facial (uno de los defensores de esta idea es el Prof. K. Bowyer [3]).
- Otra en la que se defiende que las imágenes de intensidad en alta resolución tienen una mayor capacidad discriminante (uno de los más destacados representantes es el Prof. J. Phillips [4]).

En nuestra opinión, ambos tipos de información son muy importantes y hacen referencia a dos aspectos diferentes de una misma realidad física compleja. Las imágenes de textura aportan información determinante de áreas de la cara donde no existe una gran variación en la estructura geométrica, como puede ser la frente, las cejas y las áreas con vello facial. En el caso de los datos 3D, la información aportada es más relevante en las áreas donde no existe una gran diferencia entre el aspecto de la textura, pero sí en la forma facial, como puede ser la mandíbula, la barbilla o las mejillas.

Nuestro enfoque del problema ha sido realizado en dos frentes: realizar un estudio exhaustivo de la capacidad verificadora de cada uno de los tipos de datos y de las condiciones de adquisición que ofrecen los mejores resultados. Para ello se han diseñado dos sistemas verificadores: uno con un motor verificador 2D y otro 3D. Cada uno de ellos se basa en características intrínsecas de la información 2D (intensidad de píxeles) y 3D (distancias en el espacio entre puntos de diferentes superficies) respectivamente. El motor verificador 2D ha sido aplicado también a las imágenes de rango, constituyendo el sistema verificador 2.5D.

 © Sociedad Española de Óptica

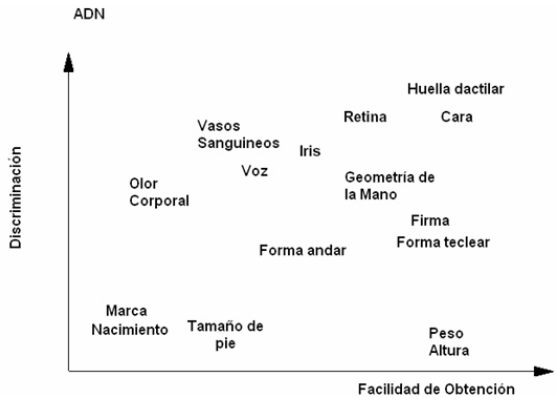

Fig. 1. Poder de discriminación vs facilidad de obtención de distintas técnicas biométricas [1].

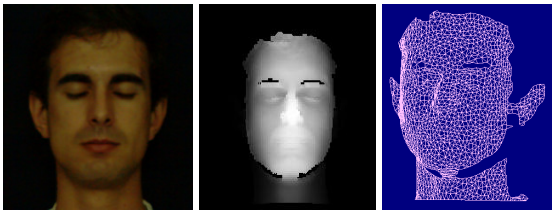

Fig. 2. Los tres tipos de datos faciales estudiados. De izquierda a derecha: imagen color (verificador 2D), imagen de rango (verificador 2.5D) y mallado 3D (verificador 3D).

Diferentes evaluaciones y competiciones realizadas en los últimos años, como el "Face Recognition Vendor Test 2002" [4] o el BANCA [5], evidencian que el estado actual de las técnicas de reconocimiento facial no es lo suficientemente maduro como para dar respuesta a la gran demanda social y comercial existente.

Varios métodos han sido aplicados al reconocimiento facial tridimensional [5-8]. Por un lado, los métodos basados en características locales utilizan la medida de distintas características geométricas de la superficie tridimensional, como pueden ser curvaturas, puntos valle o puntos silla. [9]. Por otro lado, también han sido presentados métodos basados en características globales de la superficie 3D, como la intersección de distintos planos y el estudio de los perfiles [10] o la representación mediante *imágenes gaussianas* [11]. Otros trabajos comparan la aplicación de métodos basados en *análisis de componentes principales* o PCA [12] con otros métodos como *análisis de componentes independientes* (ICA) [13] o distancia Haussdorff entre superficies [14]. Mavridis *et al* [15] calculaban autocaras tridimensionales, mientras Chang *et al* [16] comparaban autocaras aplicadas a imágenes de rango con imágenes de textura.

Otros métodos para reconocimiento facial 3D están basados en la comparación con plantillas. Por ejemplo, en [17] se presenta la utilización del algoritmo *iterative closest point* o ICP, también utilizado en el presente artículo. Otros trabajos proponen modelos 3D deformables [18] o información de textura con *active shape models* o ASM [19].

El artículo se estructura de la siguiente manera: en el apartado 2 se presenta el proceso de creación de la base de datos utilizada en esta tesis, la FRAV3D (Base de datos facial en 3D creada por el grupo de investigación FRAV) [20]. El apartado 3 se centra en la normalización facial tridimensional realizada, presentando los distintos métodos utilizados. El apartado 4 muestra los sistemas de verificación 2D y 3D diseñados y desarrollados, recogiendo los resultados y características de cada uno de ellos. En el apartado 5 se ha presentado la fusión de datos 2D y 3D, y en el último apartado se recogen las aportaciones originales de este trabajo, así como las conclusiones más importantes y las posibles líneas de trabajo futuro.

2. Base de datos FRAV3D

La creación de una base de datos destinada a reconocimiento facial es una tarea compleja, ya que la gran cantidad de factores que influyen en las condiciones de adquisición de los datos requiere de una estricta aplicación de un protocolo para conseguir que todas las imágenes de los individuos sean comparables.

Prueba de esta complejidad es la gran variedad de bases de datos faciales bidimensionales existentes [2,21,22], y la no existencia de una homogeneidad entre ellas.

En el caso de las bases de datos tridimensionales, existen muy pocas disponibles actualmente. La gran mayoría de ellas representan la información tridimensional como mapas de profundidad, y no como mallados, con la consiguiente limitación en la información disponible. Otra limitación de las bases de datos disponibles es la cantidad de imágenes por individuo. Habitualmente, se opta por crear bases de datos con muchos individuos, pero el número de imágenes de cada individuo es bajo. Por estas razones, se decidió adquirir una base de datos

 © Sociedad Española de Óptica

propia que cubriera todas estas necesidades: la base de datos FRAV3D.

La FRAV3D es una base de datos multimodal, ya que tiene información bidimensional o de textura, e información tridimensional. Fue adquirida mediante un escáner láser de Minolta modelo VIVID-700. Como uno de los objetivos de esta base de datos es que fueran datos reales, se optó por no manipular los datos durante la fase de adquisición, dejando esta tarea para una posible etapa de preprocesado. Este escáner proporciona también información de textura registrada con información 3D.

La base de datos consta de imágenes de 105 individuos (81 mujeres y 24 hombres), con 16 capturas por individuo: cuatro imágenes frontales, ocho giros en diferente sentido y grado, dos gestos y dos iluminaciones diferentes. Los voluntarios para posar en la base de datos se obtuvieron entre alumnos, profesores y personal que trabaja en la Universidad, de ahí las diferencias en edad y género que aparecen en la base de datos. En la Fig. 3 pueden verse algunos ejemplos de imágenes, modelos y mapas para un individuo.

La base de datos FRAV3D está actualmente disponible para la comunidad científica (sólo para fines de investigación, no comerciales), bajo demanda a través de la página Web del grupo de investigación [20]. Consideramos que ésta es una importante contribución, ya que puede permitir la comparación de diferentes algoritmos mediante una misma base de datos, estandarizando los resultados de diferentes sistemas de reconocimiento facial.

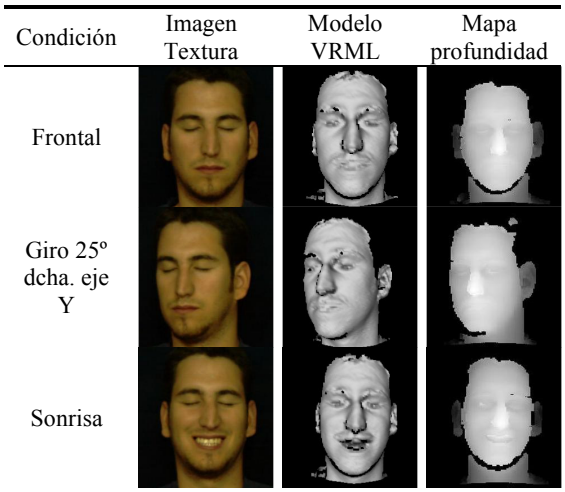

Fig. 3. Ejemplos de la FRAV3D

3. Normalización facial automática

La normalización de los datos previa a una etapa de verificación es una tarea primordial, ya que de ello depende en gran medida la eficacia del sistema. Por normalización entendemos no sólo la corrección en posición, sino también la eliminación de ruido, agujeros, etc. presentes en los datos adquiridos con el escáner. Habitualmente esta tarea se realiza de manera manual, pero consideramos que es imposible trasladar un método manual a un entorno real, por lo que se procedió a diseñar un nuevo método automático.

Se ha diseñado un método de localización facial automática basado en rasgos locales [23], mediante la técnica de registro global *spin images,* presentada por Johnson y Hebert [24]. Este método realiza una caracterización local de una superficie en torno a un punto mediante un histograma espacial re-presentado en forma de imagen, la *spin image*. Se decidió localizar tres puntos característicos: la punta de la nariz y los lagrimales. Con estos tres puntos es posible normalizar la cara tanto en posición como en tamaño.

3.1. Introducción a las *spin image*

Una Spin Image asociada con un punto (origen) en la superficie, es un histograma 2D construido a partir de la posición de los puntos vecinos al origen, representando la geometría local de la superficie en torno al punto origen.

Como se muestra en la Fig. 4, un punto orientado O (p: punto y n: vector normal perpendicular a la superficie) define una sistema de coordenadas local ($O(p,n)$) de cinco grados de libertad. Este sistema de coordenadas es definido a partir de:

- P: plano tangente que contiene al punto p, perpendicular al vector unitario de orientación n.
- L: recta que pasa por el punto p y es paralela al vector n.

Las dos coordenadas del nuevo sistema son:

- α: distancia perpendicular a la recta L.
- β: distancia con signo perpendicular al plano P.

Se trata de un sistema de coordenadas cilíndrico donde se ha perdido la coordenada del ángulo polar (ésta no puede ser determinada únicamente a partir de un punto y la normal).

 © Sociedad Española de Óptica

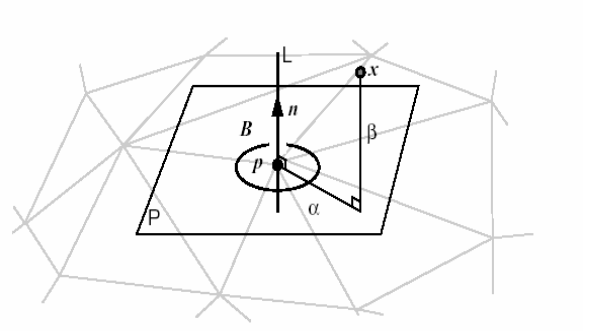

Fig. 4. Parámetros de la *spin image*.

A partir del sistema de coordenadas $O(p,n)$ podemos definir el Spin-map S_o, como la función que proyecta un punto 3D, x, al sistema de coordenadas 2D $O(p,n)$, correspondiente al punto orientado según indica la siguiente ecuación:

$$S_0 : \Re^3 \rightarrow \Re^2$$

$$S_0(x) \rightarrow (\alpha,\beta) = \left(\sqrt{\|x-p\|^2 - (n\cdot(x-p))^2}, n\cdot(x-p) \right) \quad (1)$$

La aplicación de la técnica de las *spin images* a la localización de rasgos faciales en una cara 3D es inmediata. En una cara, los puntos correspondientes a rasgos faciales tienen una geometría muy diferenciada, por lo que sus spin images serán muy características. En la Fig. 5 pueden verse Spin Images correspondientes a distintos puntos faciales. Comparando las *spin images*, se podrá decidir si un punto corresponde a un rasgo facial o no.

El método de comparación de imágenes elegido ha sido un clasificador de máquinas de vectores soporte SVM [25]. Se trata de un clasificador binario muy robusto, con gran capacidad de generalización, por lo que se consideró ideal para realizar esta tarea. Se han entrenado tres clasificadores SVM diferentes, cada uno para un punto característico diferente. El proceso de entrenamiento del SVM es supervisado. En este caso se han introducido como imágenes de entrenamiento *spin images* correspondientes a los puntos buscados (como ejemplos positivos) y *spin images* correspondientes a ruido, otras posiciones, puntos con posibilidad de crear confusión, etc. (como ejemplos negativos).

Tras un estudio inicial, se observó que la técnica de las *spin images* era efectiva, pero muy costosa computacionalmente, lo cual hacía necesario la selección inteligente de los puntos sobre los que aplicarla. Por lo tanto, se decidió introducir en el proceso una etapa de preprocesado, donde las áreas candidatas a contener los puntos característicos fueron seleccionadas. De este modo, sólo se calcularon las *spin images* de los puntos candidatos a ser puntos característicos, ahorrando una gran cantidad de tiempo de procesado.

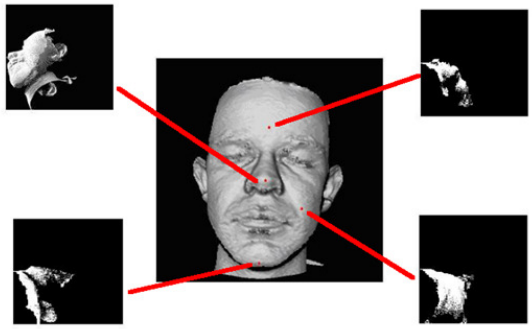

Fig. 5. Diferentes *spin images* correspondientes a distintos puntos faciales

3.2. Localización de la nariz y de los ojos

Para localizar la nariz, los puntos candidatos considerados fueron los más salientes de la cara sucesivamente. Una vez seleccionado un candidato, a continuación se procedió a calcular la *spin image* correspondiente a este punto, y a clasificarla para comprobar si se trataba de la punta de la nariz o no. El clasificador SVM discriminó entre las Spin Images correspondientes a puntos característicos y las que no lo son.

La elección de los puntos candidatos a ser lagrimales requiere de un análisis geométrico más complejo que en el caso de la nariz. Se ha aplicado una etapa de preprocesado, donde mediante curvaturas y técnicas de *clustering* se han obtenido las áreas que contienen los puntos buscados. Sólo dentro de estas áreas se han buscado candidatos, y se ha aplicado la técnica de las *spin images* y SVM.

Esta etapa puede dividirse en dos partes: en primer lugar las áreas con un mayor curvatura media discreta son seleccionadas; posteriormente, son separadas en tres grupos mediante técnicas *clustering* basadas en distancia euclidea. En la Fig. 6 puede verse el resultado de cada una de estas etapas.

Una vez localizadas las áreas de la cara que contienen los ojos, se aplicó la técnica de las *spin images* únicamente a estos puntos. Los puntos candidatos para los lagrimales se van eligiendo como aquéllos con mayor profundidad. Una vez

 © Sociedad Española de Óptica

seleccionado el candidato, la *spin image* se calcula teniendo en cuenta los puntos de la cara completa, no sólo los del cluster correspondiente. Esta imagen se introduce en el clasificador SVM (entrenado para localizar lagrimales) y es aceptada o rechazada. Se entrenaron dos SVM, uno especializado en localizar el lagrimal derecho y otro el izquierdo. Cada SVM se entrenó con 90 imágenes correspondientes a ejemplos positivos, y 90 imágenes correspondientes a ejemplos negativos. En la Fig. 7 se representan los puntos considerados como candidatos, los rechazados y los aceptados en el proceso de localización de nariz y ojos. En verde aparecen los puntos localizados correctamente. En la Fig. 8 podemos ver algunos ejemplos de puntos correctamente localizados para varios sujetos.

Tras la localización automática de los puntos faciales característicos, se procedió a la realizar las rotaciones y traslaciones que permiten dejar todos los modelos en un formato normalizado (mirada de frente y los ojos a la misma altura).

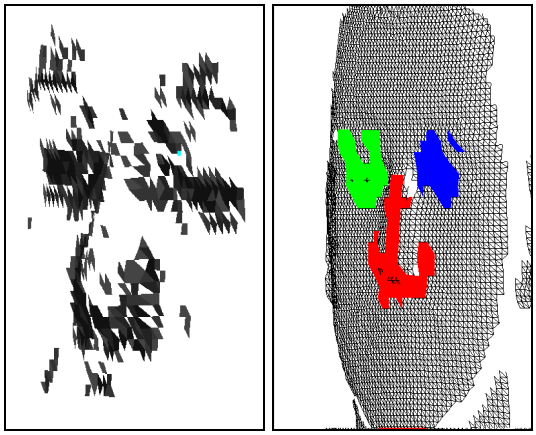

Fig. 6. Valores máximos de la curvatura media discreta (izquierda) y áreas que contienen los puntos característicos buscados, halladas mediante *clustering* (derecha).

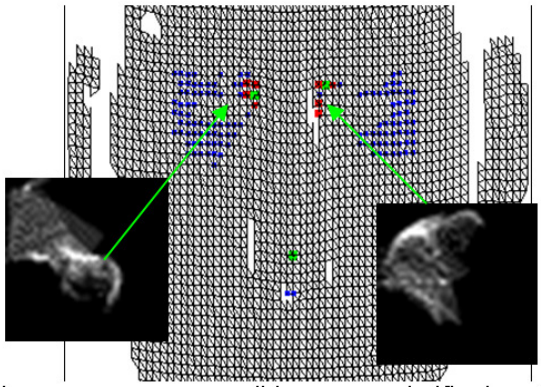

Fig. 7. Puntos candidatos no clasificados (azul), rechazados (rojo) y aceptados (verde).

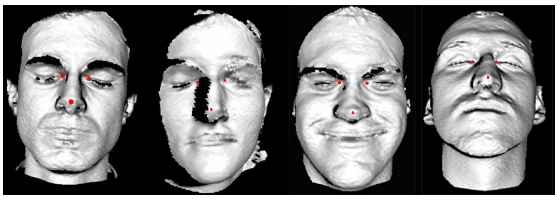

Fig. 8. Resultado final en la localización de puntos característicos (señalados en rojo)

4. Verificación

En este apartado se describen las técnicas desarrolladas para la implementación de dos tipos de verificadores faciales: uno diseñado para el tratamiento de imágenes de intensidad, y otro que toma como datos de entrada mallados de puntos tridimensionales. El primero de ellos utiliza, como información para realizar la verificación, el nivel de intensidad de cada píxel de la imagen, por lo que se trata de un sistema con un núcleo o motor 2D. El segundo de ellos se basa en la comparación de superficies 3D mediante el algoritmo de registro de superficies 3D *iterative closest point* (ICP) [26-27].

Como se ha explicado anteriormente, existen diferentes modos de representar la información tridimensional. En este trabajo se han utilizado dos diferentes: las imágenes de rango y los mallados triangulares 3D. La primera representación se trata de una imagen 2D, donde se han proyectado los puntos 3D sobre un plano. Es posible entonces realizar la verificación facial tridimensional a partir de estas imágenes de rango con el mismo sistema verificador 2D, usado también para procesar las imágenes de textura. La segunda representación es propiamente una representación tridimensional, ya que se tienen en cuenta las tres coordenadas (x,y,z) de cada punto en el espacio. En la Fig. 2 se muestran los distintos tipos de información utilizada por cada verificador.

4.1. Verificación facial 2D y 2.5D: textura e imagen de rango.

El sistema implementado se trata de un sistema de verificación facial basado en características globales [28-31]. La información utilizada para realizar la verificación es el nivel de intensidad de cada uno de los píxeles de la imagen. En la Fig. 9 se muestra un esquema de la arquitectura del sistema. El proceso de verificación facial se realiza en las siguientes etapas:

 © Sociedad Española de Óptica

- **Localización de la cara**: a partir del fondo, y mediante la convolución de la imagen con una plantilla genérica, se realiza la localización de la cara. De este modo se elimina toda la zona de la imagen que no contiene información relevante, evitando el pelo, los hombros, etc. Además, la posición facial de todos los individuos queda normalizada. En el esquema de la Fig. 9 esta fase incluye también la normalización facial automática.

- **Extracción de características**: una vez es localizada la cara, se procede a realizar una reducción en la dimensión del espacio de representación de las caras. Se pretende trabajar en un espacio de dimensión menor, pero que contenga la información más relevante para el proceso de verificación. Por ello se utilizó el método de autocaras de Turk y Pentland [12,29], basado en un análisis de características principales o PCA. Mediante este método, a partir de un análisis de autovalores y autovectores del conjunto de datos, se realiza un cambio de base del espacio de representación. El espacio final de representación, que llamaremos espacio de características, es mucho más apropiado para distinguir las diferentes caras humanas. De este modo cada cara es representada en el espacio de características como un vector de 150 componentes, que explican más de un 90% de la varianza total del conjunto de datos. Los autovectores obtenidos tienen la misma dimensión que las imágenes iniciales, y por lo tanto pueden ser vistos como una imagen en el espacio inicial: son las denominadas autocaras. Los detalles sobre las autocaras, así como ejemplos de ellas se encuentran en [12] y [28].

- **Clasificación**: la medida de la similitud entre las caras es realizada mediante un clasificador de máquinas de vectores soporte SVM [25] Se trata de un clasificador binario. Se basa en la transformación del espacio de los datos iniciales (donde los datos no son separables) a un espacio de dimensión igual o mayor (donde sí son separables). La función que realiza esta transformación, kernel, puede ser de diversos tipos. Para minimizar el error empírico cometido en la clasificación, se busca el hiperplano óptimo de separación de las dos clases en el espacio de dimensión mayor.

- **Módulo de decisión**: a partir del valor de salida del clasificador SVM, es necesario tomar una decisión para aceptar o rechazar al sujeto verificado. Se ha de fijar por lo tanto un umbral que será la frontera entre los valores del clasificador que implican aceptación o rechazo. Como parte de esta tesis se ha elaborado un método para calcular de manera automática el umbral óptimo, en función del nivel de seguridad que se le quiere otorgar al sistema.

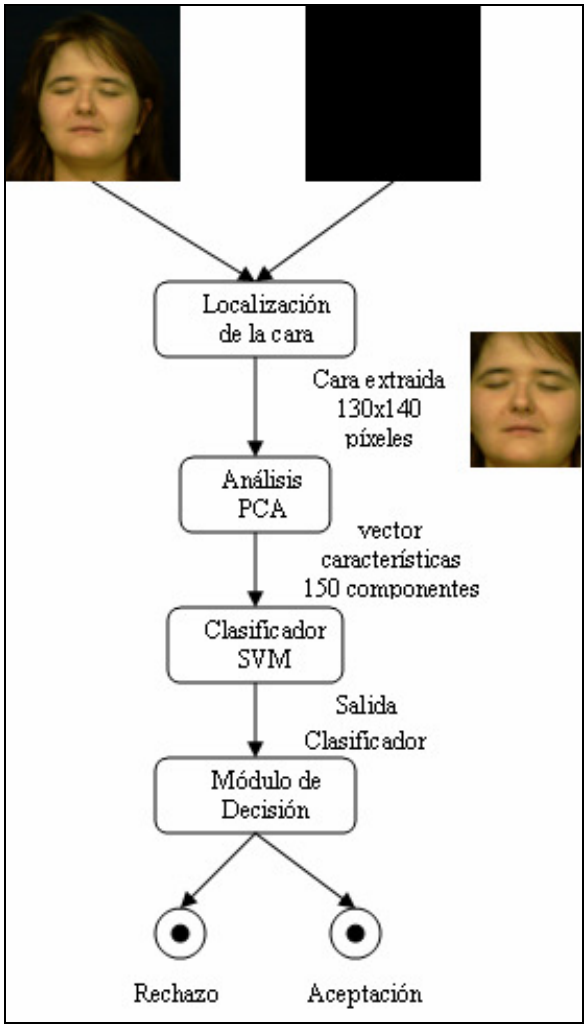

Fig. 9. Esquema Verificador 2D y 2.5D.

Tanto la técnica PCA como el clasificador SVM requieren de una etapa de entrenamiento, por lo que la base de datos ha sido dividida en dos grupos: entrenamiento y test. Estos conjuntos son siempre disjuntos. En la Tabla I se presentan los diferentes conjuntos considerados en las pruebas realizadas. A partir del análisis de los datos del conjunto de entrenamiento, se crea la matriz PCA mediante la cual se proyectará cualquier nueva imagen adquirida sobre el espacio de características. Ese mismo conjunto de entrenamiento será utilizado para que el SVM realice el aprendizaje y genere un modelo por persona. Cualquier imagen nueva de un sujeto que reclame ser una persona, será comparada con ese modelo por el SVM.

 © Sociedad Española de Óptica

TABLA I.
Pruebas realizadas. Número y tipo de imágenes que forman en cada caso el conjunto de entrenamiento y de test.

PRUEBA	Conjunto entrenamiento	Conjunto de test
1	3 frontal	1 frente
2	4 frontal	1 gestos (sonrisa)
3	4 frontal	1 gestos (boca abierta)
4	4 frontal	2 iluminación
5	4 frontal	2 giro Y 5º
6	4 frontal	1 giro Z (leve)
7	4 frontal	2 giro X
8	4 frontal	2 giro Y 25º
9	4 frontal	1 giro Z (severo)
10	4 frontal 2 iluminación	2 frontal
11	3 frontal 1 Giro Y 5º	1 frontal 1 giro Y 5º
12	3 frontal 1 Giro Y 5º 1 iluminación	1 frontal 1 giro Y 5º 1 iluminación
13	3 frontal 1 Giro Y 5º 1 iluminación	1 frontal 1 giro Y 5º 1 iluminación 1 gestos (sonrisa)
14	4 frontal	2 gestos
15	2 frontal	2 frontal

4.2. Verificación facial 3D: Método "Iterative Closest Point" (ICP)

Como verificador 3D se desarrolló un sistema basado en alguna característica intrínseca de la propia superficie tridimensional. Desde un principio se decidió utilizar como magnitud para medir la similitud entre dos caras 3D, la medida de la distancia entre dos superficies en el espacio tridimensional. Cuanta mayor diferencia hubiera entre dos caras, mayor distancia existiría entre las superficies. Se realizaron estudios previos donde se analizó la capacidad de utilizar la distancia Haussdorf, con resultados poco prometedores. Finalmente, se optó por utilizar el algoritmo de registro de superficies *iterative closest point* (ICP), introducido por Besl y Mckay en 1992 [26]. Es un método utilizado para realizar el alineamiento de superficies tridimensionales cuando se conoce una estimación inicial de la posición relativa entre ambas. Se trata de un algoritmo que, a partir de dos superficies (una llamada modelo y otra escena), "encaja" una superficie sobre la otra, mediante transformaciones rígidas sucesivas. La superficie modelo es la que permanece fija y la escena se va variando hasta aproximarse lo más posible al modelo. Para ello, se basa en el emparejamiento de puntos entre las superficies modelo y la escena, eligiendo siempre el par de puntos más cercano.

Este método fue elegido ya que se consideró apropiado para el problema propuesto de la verificación facial en 3D:

- Se basa en las características geométricas de cada una de las superficies, luego cumple el requisito impuesto de que se tratara de un método basado en algún rasgo puramente tridimensional.
- Debido a la normalización previa con las *spin images*, se dispone de una estimación inicial de la posición de las superficies, condición necesaria para llevar a cabo del registro con ICP.

La medida del error final entre las superficies alineadas puede ser utilizada como parámetro clasificador para realizar la verificación facial

El ICP es un algoritmo iterativo que trabaja en tres fases:

1. Establece correspondencias entre pares de puntos de las dos superficies, creando parejas de puntos más cercanos.
2. Estima la transformación rígida que ajusta mejor los puntos de la escena al modelo.
3. Aplica esta transformación a todos los puntos de la escena, y se calcula el error cuadrático medio entre los puntos de ambas superficies.

El proceso es repetido hasta que el error cuadrático medio entre las superficies converge a un valor mínimo. En ese momento se detiene el proceso.

En el verificador 3D basado en ICP no existe etapa de entrenamiento, por lo que no tiene sentido hablar de conjunto de entrenamiento, únicamente conjunto de test. Las capturas del conjunto de test han sido verificadas contra el modelo de cada persona (captura frontal), y el valor medio de los resultados es presentado como resultado final de cada una de las pruebas. Las pruebas realizadas coinciden con las presentadas en la Tabla I.

5. Resultados experimentales y discusión

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

 © Sociedad Española de Óptica

explicado anteriormente, un sistema de verificación tiene como salida dos valores: aceptación o rechazo. Una salida con valor "aceptación" indicará que efectivamente la imagen introducida y la persona que se reclama ser corresponden. En caso contrarío indicaría que se trata de un impostor, y la salida del sistema sería "rechazo".

Existen por lo tanto cuatro posibilidades:

- 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 negativo (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: Verdadero 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.

Para caracterizar la calidad del sistema verificador mediante un único valor, habitualmente se elige el punto en que el porcentaje de FN es igual al porcentaje de FP. Este punto se denomina *Equal Error Rate* (EER). Cuanto más bajo sea el EER nos indicará que el sistema tiene menos fallos. Los resultados experimentales van a ser mostrados como es habitual en la literatura actual del valor del parámetro EER. En la Tabla II se presentan los resultados.

En el caso de imágenes frontales, el mejor resultado es obtenido en el caso del Verificador 2.5D, con imágenes de rango, obteniéndose un 99.91% de acierto. En el caso del Verificador 2D, con imágenes en color, el porcentaje total de aciertos desciende al 97.1%, y en el caso del Verificador 3D basado en ICP, al 95.1%.

Respecto a las imágenes gestuales, en las imágenes con gesto sonrisa los resultados son mejores que los obtenidos con la boca abierta, ya que en el primer caso la geometría facial se ve menos afectada. Otra vez el mejor resultado es el obtenido por el Verificador 2.5D.

TABLA II.

Comparación de la verificación facial 2D, 2.5D y 3D. Se muestra el EER(%) obtenido en cada una de las pruebas.

Nº de test	2D	3D	2.5D
1	2.9	4.9	0.1
2	5.1	10.3	4.6
3	12.6	12.6	12.6
4	8.7	3.4	1.0
5	14.6	4.4	1.0
6	33.9	7.4	2.2
7	13.3	5.3	1.9
8	27.0	5.8	4.9
9	41.1	7.4	3.7
10	1.9	3.7	0.5
11	4.9	5.3	0.5
12	4.2	4.2	0.6
13	4.1	6.7	1.7
14	9.2	11.8	10.3
15	1.9	3.7	0.6

Los cambios en la iluminación afectan principalmente al Verificador 2D. En el caso del Verificador 3D, el resultado no se ve afectado en absoluto. Por el contrario, las imágenes de rango sí se ven más afectadas.

Los giros afectan enormemente al Verificador 2D, ya que no se ha realizado ninguna normalización. En el caso de los verificadores 2.5D y 3D, se ven mucho menos afectados por estas condiciones.

En el caso de introducir diferentes condiciones de adquisición en el conjunto de entrenamiento (pruebas 10 a 14), los verificadores con etapa de entrenamiento (2D y 2.5D) realmente mejoran sus resultados. En el caso del Verificador 3D, no se produce esta mejora ya que no posee un clasificador con etapa de aprendizaje.

En general, el Verificador 2.5D es el que muestra mejores resultados, siendo esencial la corrección en posición y orientación. También podemos destacar que el Verificador 2D ofrece mejores resultados que el Verificador 3D (excepto en el caso de los giros y la iluminación).

Podemos concluir indicando que a la vista de los resultados, la información tridimensional parece más potente a la hora de realizar la verificación, pero el motor verificador utilizado en el caso de las imágenes en color y rango (basado en PCA y SVM) es más potente que el motor verificador 3D (basado en ICP).

El verificador 2.5D ha mostrado ser en general el más potente de los tres estudiados, pero los resultados obtenidos por los sistemas 2D y 3D en

 © Sociedad Española de Óptica

algunas situaciones los mejoran. En el siguiente capítulo se muestra el método seguido para realizar la fusión de estos tres verificadores.

6. Conclusiones

En este artículo se resume la tesis doctoral [34] que ha presentado un trabajo de investigación centrado en el campo del reconocimiento facial automático. Las contribuciones principales, presentadas en las referencias [23] y [28-32] se enumeran a continuación:

- Se han presentado un método de localización automática de tres puntos característicos en la cara (ojos y nariz) mediante el uso de la técnica de registro global *spin images* unido a un clasificador de *máquinas de vector soporte*; un método global de normalización en posición a partir del ajuste de los puntos a distintos planos y rectas de regresión.
- Se ha investigado la capacidad de reconocimiento facial a partir de diferentes tipos de representación de la cara: imagen en niveles de gris (2D), imagen de rango (2.5D) y mallado tridimensional (3D).
  - Se ha propuesto un sistema de verificación facial aplicable a imágenes en escala de grises e imágenes de rango. Se basa en la reducción del espacio de características mediante un análisis de componentes principales, y la posterior clasificación mediante el método de máquinas de vector soporte (kernel lineal).
  - Se ha presentado un sistema de verificación facial aplicable a mallados tridimensionales basado en la medida de la similitud entre dos superficies mediante el algoritmo de registro *iterative closest point*.
- Para la evaluación de los métodos propuestos, se ha adquirido una nueva base de datos multimodal (2D y 3D), llamada FRAV3D. Consta de 105 individuos, con 16 capturas por individuo en diferentes condiciones de adquisición (giros, iluminación, y expresiones). Esta base de datos se encuentra disponible para la comunidad científica, para fines de investigación.

Se han probado los métodos presentados tanto de normalización como verificación facial en la base de datos FRAV3D, mediante pruebas exhaustivas, manteniendo las mismas quince pruebas en todos los casos.

Los resultados finales obtenidos han sido muy satisfactorios, alcanzando un 99.93% de aciertos en el caso de imágenes frontales, y más del 95% en el resto de condiciones (excepto en gestos con la boca abierta).

En el apartado 1 se presentaron las dos corrientes actuales existentes en la biometría facial [2]: superioridad de la capacidad de verificación de la información tridimensional o de las imágenes en alta resolución. Respecto a esta discusión, podemos afirmar que los resultados obtenidos en esta tesis muestran una mayor capacidad de reconocimiento de la información tridimensional (concretamente en forma de imagen de rango). Si bien, es importante destacar que la información bidimensional es extremadamente importante y puede contribuir en gran medida a complementar a los datos tridimensionales.

Agradecimientos

Este trabajo ha sido realizado gracias a la financiación de la Universidad Rey Juan Carlos. Los autores quieren agradecer especialmente su contribución a Jorge Pérez López.

 © Sociedad Española de Óptica