

Sistema de reconocimiento facial de conductores

Sebastián Bronte, Luis M. Bergasa, Jesús Nuevo, Rafael Barea

Resumen—En este documento se presenta un sistema para el reconocimiento facial de conductores a bordo de vehículos. Con objeto de diseñar un control del encendido del vehículo en función de la identidad del conductor. Los sistemas básicos para el reconocimiento de caras funcionan bien en entornos controlados. La iluminación ha de ser uniforme o debe seguir un cierto patrón que ha de estar incluido en la base de datos. Introducir un reconocedor a bordo de un coche supone unos nuevos retos que hay que estudiar. En este artículo se presentan algunas mejoras introducidas en los sistemas de reconocimiento facial actuales para un funcionamiento sobre un vehículo. También se muestran algunos resultados experimentales y las conclusiones.

Index Terms—2DPCA, Clustering, Independencia a iluminación, Visión artificial, SSR.

I. INTRODUCCIÓN

EL diseño de un sistema de reconocimiento facial de conductores puede ser útil para aumentar la seguridad de los vehículos ante robos. En este tipo de sistemas se integran varios subsistemas: Primeramente es necesario la presencia de caras en la imagen. Una vez detectada, hay que hacer un tracking de la misma para no perderla durante el reconocimiento y tenerla centrada. Cuando se tiene una imagen estable de la cara, se evalúa la región de interés para realizar el reconocimiento. A continuación se realiza la comparación con las caras que están en una base de datos, siendo éstas realizadas calculando las distancias entre características de las imágenes de la base de datos y las que llegan desde el dispositivo de entrada al sistema. Por último, se evalúa cuál es el conductor que más se parece dentro de la base de datos. Este sistema aparentemente ideal, presenta una serie de problemas que a continuación se pasa a detallar.

Tomando exclusivamente un criterio de distancia mínima, no se es capaz de decidir si el usuario pertenece o no la base de datos. El sistema lo que hace en realidad es buscar el más parecido dentro de la base de datos del sistema aunque el parecido sea muy bajo, lo cual no es aceptable, por lo se plantea un subsistema para evitar este inconveniente.

Por otro lado, el sistema básico no es inmune a los cambios de iluminación respecto de las imágenes de la base de datos, por lo que podría dar lugar a reconocimientos válidos cuando el sujeto no estuviese en el base, y el caso contrario, esto es, que el sujeto esté en la base y debido a las no linealidades introducidas por la iluminación no sea reconocido.

Para solucionar el problema del rechazo de usuarios, se aplicará un algoritmo de clustering.

Respecto al problema de la iluminación, se presentarán los resultados de la aplicación de distintos algoritmos al sistema base:

Sebastián Bronte, Luis Miguel Bergasa, Jesús Nuevo y Rafael Barea pertenecen al Departamento de Electrónica de la Universidad de Alcalá.

Email: {sebastian.bronte; bergasa; jnuevo; barea}@depeca.uah.es

- Primero se estudiará el efecto que tiene sobre el sistema el centrarnos únicamente sobre la zona de la cara de los ojos y la nariz.
- Por otro lado se analiza el efecto de la ecualización del histograma para mejorar la iluminación de la cara.
- A continuación se probará un método basado en [1], en el que se entrena al sistema con varios patrones de iluminación.
- Por último se analizarán los efectos de SSR [2] implementado en nuestra aplicación.

II. TRABAJOS PREVIOS

Se parte de un sistema de detección de caras y reconocimiento de usuarios básico en el que se realizan funciones de detección, tracking y reconocimiento [3].

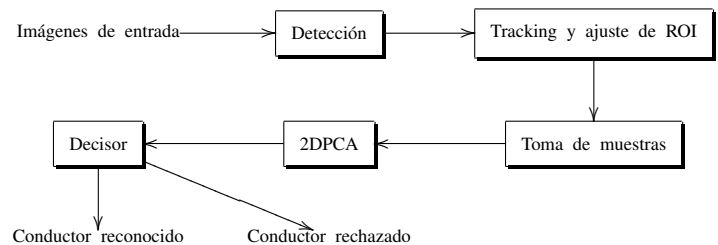


Figura 1. Esquema principal del sistema de reconocimiento

Como clasificador se usaba la distancia euclídea entre la imagen a reconocer y las imágenes de la base de datos, y como decisor se usaba un criterio de mayoría sobre una secuencia de imágenes de usuario.

II-A. 2DPCA

Sea X un vector columna a partir del que podamos proyectar una imagen A en un vector Y en el que tendremos las características básicas.

$$Y = AX \quad (1)$$

Para obtener la mejor X que pueda representar las características de A en nuestro espacio, empezamos calculando la matriz de covarianza de todos los usuarios de nuestra base de datos:

$$\bar{A} = \frac{1}{N} \sum_{i=1}^N A_i \quad (2)$$

$$C = \frac{1}{N} \sum_{i=1}^N (A_i - \bar{A})(A_i - \bar{A})^T \quad (3)$$

Siendo \bar{A} la imagen media, N el número de imágenes en nuestra base de datos, A_i la imagen i-ésima de la base de datos, y C la matriz de covarianza.

$$C = PQP^T \quad (4)$$

A partir de la matriz de covarianza, calculamos sus autovalores Q y sus autovectores P, los cuales utilizaremos para formar la base de nuestro espacio. No tomamos todos los autovectores, sino que tomamos los primeros que son los más representativos, los que nos darán más información acerca de las características de las caras. Esta última matriz de autovectores será nuestra matriz X.

Por último calculamos las proyecciones de las imágenes de los conductores permitidos en la base del espacio obtenida anteriormente, como hemos explicado en la ecuación 1.

II-B. Cálculo de distancias

En este artículo se realizará una comparativa entre 2 métodos de cálculo de distancias para comparar sus resultados. Utilizaremos distancia euclídea y distancia de Mahalanobis.

En el primer caso la distancia se calcula de la siguiente manera:

$$d_{euclidea} = \sum_{i=1}^w \sqrt{(a_i - b_i)^T (a_i - b_i)} \quad (5)$$

Siendo a_i y b_i las columnas i-ésimas de las matrices de características de la imagen de entrada y de la base de datos respectivamente, y w la anchura de la imagen.

Para el caso de la distancia de Mahalanobis, utilizamos la siguiente ecuación:

$$d_{mahalanobis} = \sum_{i=1}^w \sqrt{(a_i - b_i)^T C^{-1} (a_i - b_i)} \quad (6)$$

Donde C es la matriz de covarianza de las imágenes de características de entrada al sistema, calculada de igual forma que en la matriz de covarianza para generar la base de datos (ver ecuación 3).

Esta última distancia mejorará los resultados como veremos a continuación, pues estamos multiplicando las distintas componentes de distancia por un peso que viene dado por la matriz de covarianza con lo que estamos adaptando mejor las distancias a la base de datos y al usuario de entrada.

III. MEJORA DEL DECISOR

Para implementar el decisor, se parte de las imágenes características a las que se hacía referencia en el apartado anterior como puntos en el espacio de decisión. Utilizando los métodos de cálculo de distancias anteriores, se tiene una medida de cómo estamos de cerca o lejos de una región del espacio que define el conjunto de puntos de cada conductor. Con esto se suplirá una de las carencias que el sistema original tenía, como es el rechazo automático de usuarios que no estaban en la base de datos.

Este decisor se ha implementado basándonos en técnicas de clustering:

- Se calcula el centro de todos los puntos que tenemos en el espacio, esto es, la media de las imágenes características. Con este centro tenemos una referencia para poder realizar un rechazo sencillo de los usuarios externos a la base de datos, pues normalmente éstos generarán distancias mayores que las de la base de datos. Para tener cierta flexibilidad, se aceptarán imágenes que cumplan que su distancia calculada al centro sea menor que $d_{cuni\ max} (1 + U_1)$.
- Si pasa este primer filtrado, se comprueba que la proyección de la imagen de características del conductor actual se ha quedado en una región intermedia entre las agrupaciones formadas por los conductores permitidos por el sistema. Para ello, se ha calculado para cada conductor, el punto central que le representa y la distancia máxima que le caracteriza. Si el punto obtenido desde el sistema está entre 2 centros de conductores de la base de datos, pero está muy alejado de los 2, el sistema se dará cuenta y rechazará esta imagen para el reconocimiento. Así como se ha hecho anteriormente, para que una imagen sea aceptada, tendrá que cumplir que sea menor que al anterior $d_{cusr\ max} (1 + U_2)$.

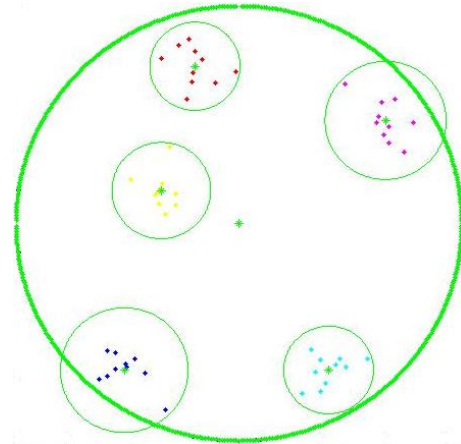


Figura 2. Representación simplificada de los 2 primeros filtrados

- Si ha pasado el filtro de las 2 etapas anteriores, si las regiones de decisión se entremezclan o son cercanas, comprobaremos a cuál de todas las imágenes de la base de datos se parece más, esto es, cuál es la que más cerca está. También calcularemos cuál es el segundo conductor en la base más cercano. Si está por encima de $d_{min} (1 + U_3)$ el reconocimiento será válido. Si no, la imagen también será rechazada.
- Repetimos todo el proceso anterior con varias imágenes tomadas como muestras y se comprueba si en sucesivos intentos el conductor está dentro de la base de datos o no. Si se confirma el mismo conductor en las suficientes repeticiones, daremos el proceso por válido y tendremos seguridad sobre la identidad del conductor. En algunos casos, para asegurar mejor el reconocimiento se realiza la media de las distancias a cada punto de nuestro universo

y se aplica el algoritmo de umbrales utilizado en el apartado anterior.

Es importante señalar que en el sistema que se presenta, pesa mucho más el hecho de que se acepte un conductor no permitido, que rechazar a un usuario permitido o confundirlo con otro usuario de la misma base de datos.

IV. MEJORAS ANTE CAMBIOS DE ILUMINACIÓN

Partimos de una serie de imágenes para cada conductor en la base de datos a partir de las cuales realizaremos el procesado para el resto de pruebas:

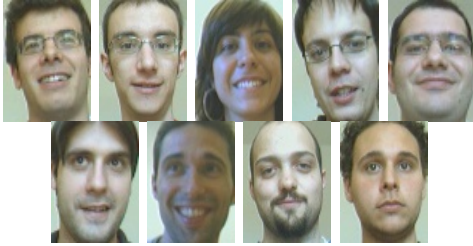


Figura 3. Ejemplos de imágenes de la base de datos sin procesar

IV-A. Identificación de la cara recortada únicamente a ojos y nariz

Se ha seleccionado para las pruebas esta zona de la cara pues es la que posee las partes más invariantes de la cara. Aunque los ojos se cierran o las pestañas se levanten, varía menos que si se toma la cara completa, pues puede haber cambios de peinado, movimientos de boca, etc. Además, con esto sería inmune a los cambios en la barba en el caso de los hombres, pues no es lo mismo ir afeitado que tener barba de unos días. El problema fundamental de este método es que no hay suficiente información con el área recortada para realizar un reconocimiento correcto, y si además el usuario utiliza gafas y estas no están incluidas en la base de datos, se producirán rechazos o errores.

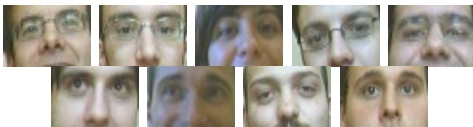


Figura 4. Ejemplos de imágenes recortadas

IV-B. Ecuilización de histograma + filtro de mediana

Con este método tradicional de compensación de los efectos de la iluminación, lo que se pretende es aplanar el histograma, regulando la iluminación en la imagen global, aumentando el contraste de la imagen. El problema de este método es que si la cara está muy iluminada por alguno de los lados, al procesarla, sigue quedando con bastante iluminación al no poder aplanar correctamente el histograma, lo cual no es bueno a la hora de

comparar las imágenes con la base de datos, aunque esta ya esté procesada. Al realizar la ecualización de histograma, se introduce en algunos pixels un ruido que puede perjudicar a la hora de realizar el reconocimiento, con lo que aplicamos un filtrado de mediana para eliminar ese ruido.

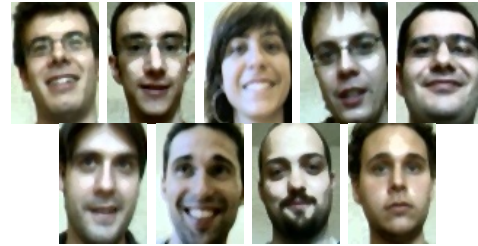


Figura 5. Ejemplos de imágenes con ecualización de histograma

IV-C. Máscaras de Iluminación

Con este método se generan una serie de imágenes que se correspondan con patrones de iluminación típicos para nuestro caso. Estas imágenes patrón se utilizarán como plantilla para fusionarla luego a cada una de nuestras imágenes de la base de datos del programa. Este método es el que más recursos consume al ejecutar el algoritmo, pues el tamaño de la base de datos crece proporcionalmente al número de plantillas utilizadas para la iluminación. Las máscaras utilizadas para nuestras pruebas y las imágenes posteriormente generadas son las siguientes:



Figura 6. Máscaras de iluminación utilizadas

Con este método se consigue un comportamiento parecido, pero algo mejorado comparándolo con el utilizado en la base sin aplicar las máscaras de iluminación. A continuación mostramos un ejemplo de la aplicación de las máscaras a una imagen de la base de datos.

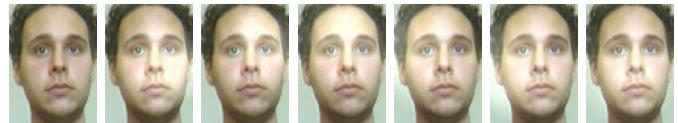


Figura 7. Imágenes combinadas con las máscaras de iluminación

IV-D. SSR (Single Scale Retinex)

Con este algoritmo se intenta convertir las imágenes que tenemos en la base de datos y las que nos entran en imágenes

tales que su variación respecto a cambios de luz sea mínima. El algoritmo Retinex [4] está basado en la respuesta logarítmica que posee el sistema visual humano. Con SSR (versión reducida de este método), se tratará de convertir una imagen en escala de grises que puede estar afectada de cualquier forma por la iluminación cambiante en otra imagen cuya variación con la iluminación sea mínima.

SSR se aplica a una imagen mediante la siguiente ecuación:

$$I_R = \log(A) - \log(A * F) \quad (7)$$

Siguiendo con la notación anterior, A es la imagen de entrada, * es el operador de convolución, y G es la gaussiana generada a partir de:

$$F(x, y) = ke^{-\frac{x^2+y^2}{c}} \quad (8)$$

Donde k es una constante que normaliza la gaussiana:

$$k = \frac{1}{\iint e^{-\frac{x^2+y^2}{c}} dx dy} \quad (9)$$

El resultado de aplicarlo a las imágenes de la base de datos con $c=4$, es el siguiente:



Figura 8. Ejemplos de imágenes tratadas con retinex

V. RESULTADOS

Para la realización de las distintas pruebas, se han utilizado 30 imágenes por cada usuario introducido en la base de datos, con 9 usuarios en total introducidos en la base como se ha podido comprobar anteriormente. Los umbrales utilizados en el decisor para los distintos métodos clasificación por distancia euclídea y distancia de Mahalanobis son los siguientes

Cuadro I

Umbrales utilizados en el decisor utilizando distancia euclídea

| Umbral | Original | Recortada | Ecualización | Máscaras | Retinex |
|--------|----------|-----------|--------------|----------|---------|
| u1 | 0 | 0 | 0 | 0 | 0 |
| u2 | 0.5 | 1 | 1 | 0.7 | 0.4 |
| u3 | 0.01 | 0.005 | 0.05 | 0.01 | 0.005 |

Pasamos a ilustrar el funcionamiento del decisor con las gráficas que se muestran en las siguientes imágenes

En las gráficas que se muestran en la figura 9 se observa como algunas de las distancias medidas con distancia euclídea

Cuadro II

Umbrales utilizados en el decisor utilizando distancia de Mahalanobis

| Umbral | Original | Recortada | Ecualización | Máscaras | Retinex |
|--------|----------|-----------|--------------|----------|---------|
| u1 | 1 | 1 | 0.5 | 1 | 0.2 |
| u2 | 1.8 | 2.1 | 1.2 | 1 | 0.4 |
| u3 | 0.03 | 0.01 | 0.05 | 0.01 | 0.005 |

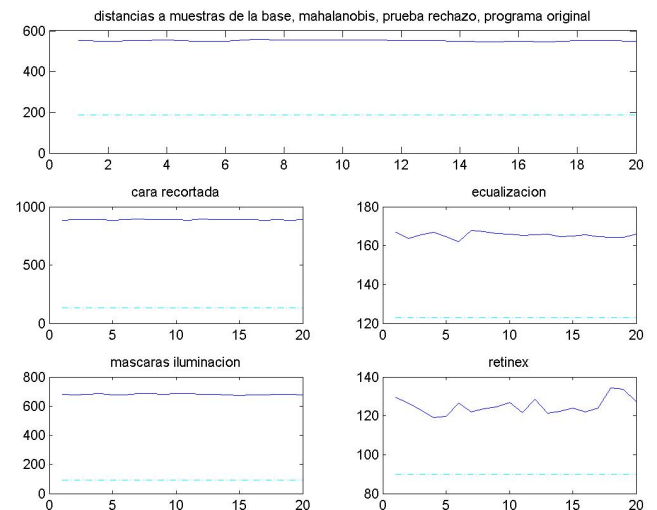
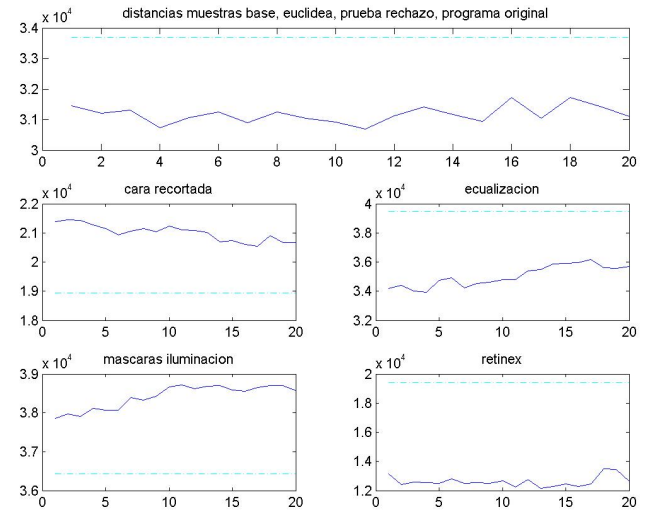


Figura 9. Ejemplo de rechazo con distancia euclídea y mahalanobis

y todas las distancias calculadas con Mahalanobis con los distintos métodos de tratamiento de imágenes empleados quedan por encima del umbral. Con esto se deduce que la distancia de Mahalanobis para este cometido es la mejor opción para implementar esta primera etapa de filtrado.

En las figuras 10, 11 y 12, mostramos el comportamiento del decisor en una prueba normal con imágenes capturadas distintas de las de la base de datos. Cada una de las gráficas de cada grupo se refiere a los distintos métodos de tratamiento de imágenes previo empleado. Empezamos mostrando cómo pasa la primera etapa de filtrado en el sistema.

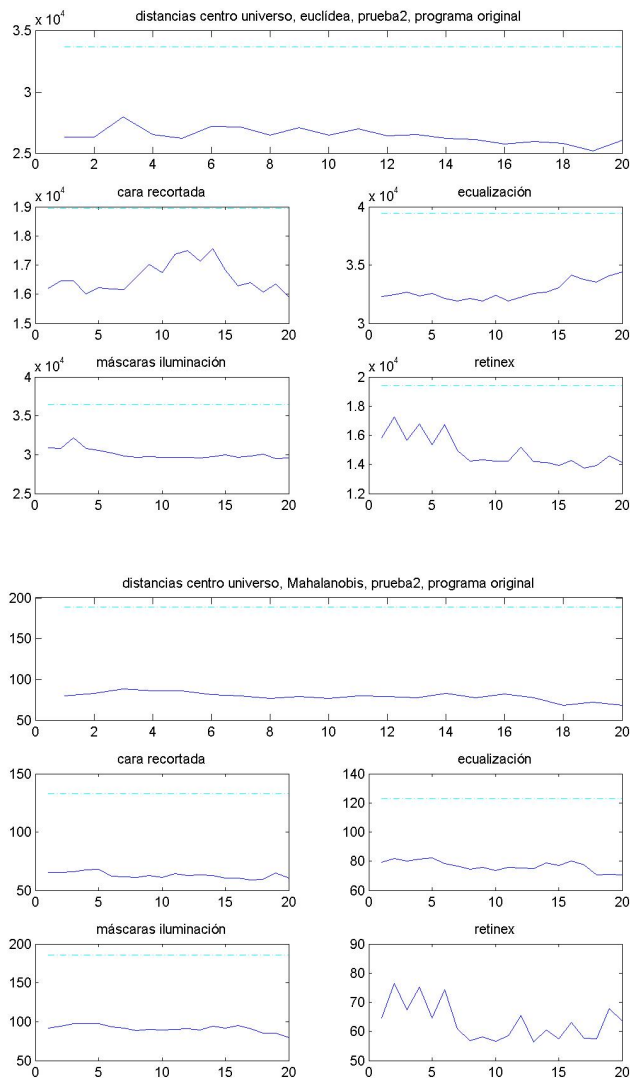


Figura 10. Distancias al centro del universo de una prueba realizada

A continuación, vemos cómo es capaz de salvar también la segunda etapa de filtrado con todos los métodos de tratamiento de imágenes empleados, con lo que tenemos la confirmación de que el conductor que tenemos en frente se encuentra en la base de datos.

Con las figuras 10 y 11 ha quedado claro que el usuario está en la base de datos. Ahora el problema está en identificar, dentro de la base de datos, cuál es el que más se parece. Para ello, se comprueba para cada usuario, cuál es la imagen dentro de su set que más se le parece.

A la vista de las figuras 10, 11 y 12 se puede ver que en algunas muestras tomadas el usuario quedaría rechazado o se confundiría de usuario. Sin embargo, haciendo una media de todas las distancias mínimas por cada usuario, se afina, eliminando los efectos de los reconocimientos fallidos si estos

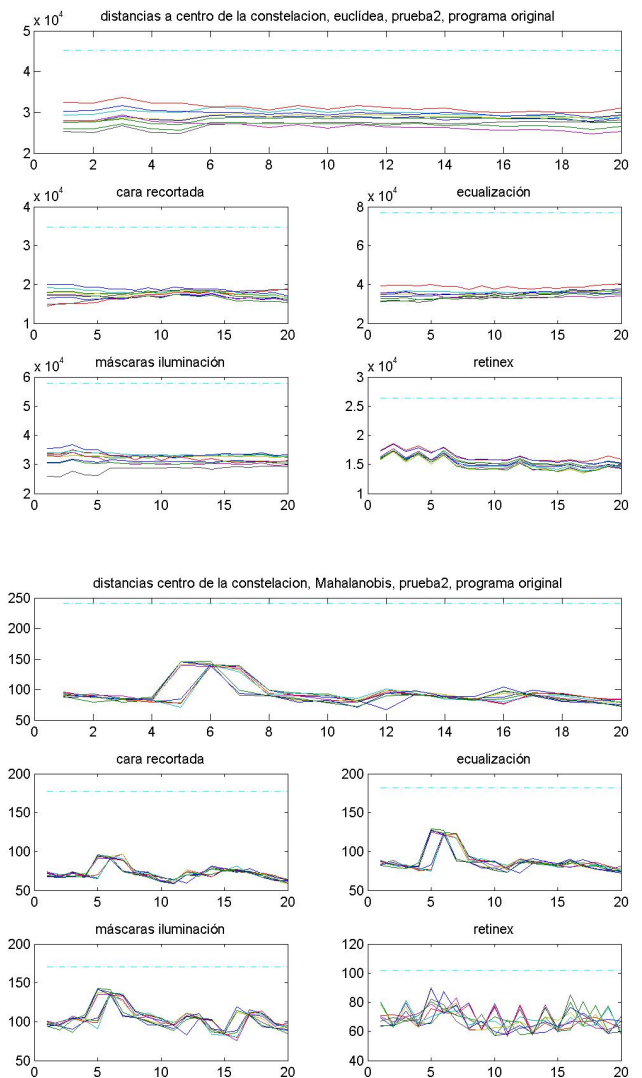


Figura 11. Distancias a los centros de las constelaciones de la prueba

no son muy frecuentes.

Por supuesto, en la figura 12, el usuario que está por debajo de todos los demás, es efectivamente el que estaba enfrente de la cámara.

Otras pruebas realizadas sobre otros vídeos, son mejores, en el caso de que la iluminación sea más o menos parecida a la que tengamos en la base y los gestos del usuario no se desvíen demasiado de los de la base de datos.

Además se han hecho otras pruebas con vídeos de peor calidad y con iluminación diferente en los que incluso se ha llegado a reconocer en varias ocasiones con éxito al usuario correspondiente y por último con variaciones de iluminación bruscas, en los que el sistema no era capaz de responder tan bien como en casos semi-controlados, como se ha expuesto en este trabajo.

VI. CONCLUSIONES

A la vista de los resultados, se deduce que la distancia de Mahalanobis es mejor para los propósitos finales que la

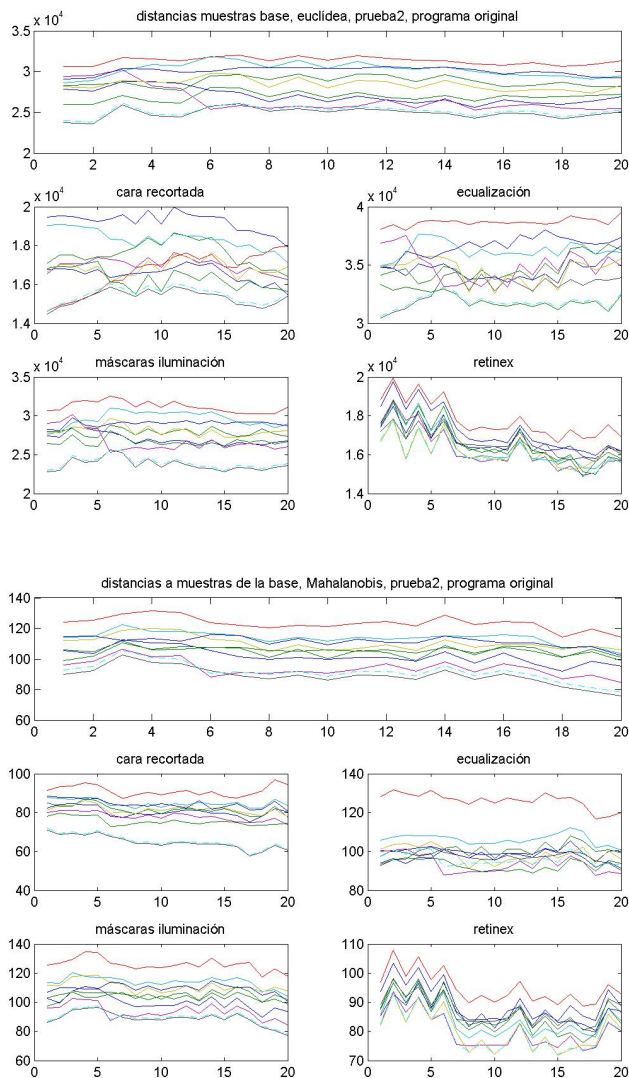


Figura 12. Distancias mínimas a usuarios de la BB.DD. en la prueba

euclídea, sobre todo para el rechazo de usuarios externos a la base de datos.

Además, se aprecia que la tendencia que siguen los resultados obtenidos por el programa original y por el de la base de datos incrementada con las máscaras de iluminación es muy parecida. Esto es debido a la naturaleza de funcionamiento de los mismos, pues parten de casi las mismas imágenes, con pequeñas variaciones en las zonas de la iluminación dadas por las máscaras aplicadas. Quizá la mejora obtenida con este método no compensa el tiempo de cómputo que se necesita, pues éste se dispara al tener que multiplicar el número de comparaciones a realizar por el número de máscaras aplicadas a cada imagen.

Respecto de la cara recortada servirá por ejemplo para la detección de gafas de sol, obligando al usuario del sistema a que se las quitase en el caso de ser detectadas. El sistema funcionando por sí solo tiene un porcentaje de éxito menor que con el programa original o el utilizado en las máscaras de iluminación. Hay que tener en cuenta que la cantidad de

información obtenida en este caso es inferior a la de la cara completa.

Con las pruebas de ecualización, se observa que no funciona como se esperaba, esto es, el índice de aciertos es muy bajo. Esto se debe a que este método distorsiona bastante la imagen, lo cual nos perjudica bastante. Por tanto, este método queda descartado.

Por último, aplicando retinex a las imágenes, el índice de aciertos es medio, ni tan bajo como lo obtenido en la ecualización, ni tan alto como los resultados del programa original o del de la base ampliada. Esto es debido a que la variación de niveles de gris en las imágenes tratadas no es tan rica como era anteriormente. Este aspecto se puede corregir ajustando el parámetro c , esto es, a medida que incrementamos éste, se introducen más sombras o luces en las imágenes, con lo que estamos en la misma situación que otros casos.

VII. TRABAJO FUTURO

Este sistema actual no es funcional en ausencia de iluminación, o sea, cuando tenemos que realizar un reconocimiento cuando se quiera arrancar el coche de noche.

Para solucionar este problema habrá que incluir en la base de datos imágenes de caras en estas condiciones reducidas de iluminación.

Además, el sistema aún no es capaz de reconocer a usuarios si las características de su cara cambian sustancialmente, esto es, el usuario se deja barba, perilla o bigote. Respecto al corte de pelo, aunque se intenta filtrar con en el recorte de la cara que excluye el pelo, si éste llega a tapar la cara, por ejemplo con un flequillo, el sistema no sería capaz de reconocer correctamente al usuario.

VIII. AGRADECIMIENTOS

Este trabajo ha sido financiado mediante fondos del Ministerio de Educación y Ciencia a través de los proyectos MOVICON (TRA2005 - 08529 - 02) y CABINTEC (PSE - 370100 - 2007 - 2), así como de la Comunidad de Madrid a través del proyecto Robocity2030 (CAM - S - 0505 / DPI / 000176).

REFERENCIAS

- [1] L. J.M.Buenaposada, E.Muñoz, "Efficiently estimating facial expression and illumination in appearance-based tracking," *Proceedings of BMVC, Edinburgh, UK*, September 2006.
- [2] M. R. G. Martin D. Levine and J. Bhattacharyya, "Image normalization for illumination compensation in facial images," *Department of Electrical & Computer Engineering & Center for Intelligent Machines. McGill University, Montreal, Canada*, August 2004.
- [3] E. L. M. S. E. J. A. H. L.M.Bergasa, R. Barea, "Sira, a robotic system for elderly assistance," *35th International Symposium on Robotics*, March 2004.
- [4] D. H. Brainard and B. A. Wandell, "Analysis of the retinex theory of color vision," *Journal Optical Society of America*, vol. 3, no. 10, p. 11, October 1986.
- [5] M. B. Ajmal S. Mian and R. Owens, "An efficient multimodal 2d-3d hybrid approach to automatic face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, November 2007.
- [6] L. Q. W. G. Bo Du, Shiguang Shan, "Empirical comparisons of several preprocessing methods for illumination insensitive face recognition," *Acoustics, Speech, and Signal Processing*, 2005.

- [7] W. G. Laiyun Qing, Shiguang Shan, "Eigen-harmonics faces: Face recognition under generic lighting," *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, 2004.
- [8] S. W. Lei Zhang, "Face synthesis and recognition from a single image under arbitrary unknown lighting using a spherical harmonic basis morphable mode," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 209–216, 2005.
- [9] O. Y.-P. Marius Herscovitz, Evgeny Artyomov, "Improving the global impression of brightness of the multi scale retinex algorithm for wide dynamic range pictures," *The VLSI Systems Center, Ben-Gurion University*.
- [10] T. O.-I. S. Y. S. Mihoko Shimano, Kenji Nagao, "Face recognition under varying illumination based on map estimation incorporating correlation between surface points," *Lecture notes in computer science*, 2006.
- [11] K. I.-M. S. Yasuyuki Matsushita, Ko Nishino, "Illumination normalization with time-dependent intrinsic images for video surveillance," *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2003.