

**UNIVERSIDAD AUTÓNOMA DE MADRID  
ESCUELA POLITÉCNICA SUPERIOR**



# **RECONOCIMIENTO DE CARAS FRONTALES MEDIANTE LA EXTRACCIÓN DE PUNTOS CARACTERÍSTICOS**

***-PROYECTO FIN DE CARRERA-***

**Miguel Cortés Marlia  
Marzo de 2009**



**RECONOCIMIENTO DE CARAS FRONTALES MEDIANTE LA EXTRACCIÓN  
DE PUNTOS CARACTERÍSTICOS**

**AUTOR: Miguel Cortés Marlia  
TUTOR: José María Martínez Sánchez**

**Grupo de Tratamiento e Interpretación de Vídeo (VPULab)  
Dpto. de Ingeniería Informática  
Escuela Politécnica Superior  
Universidad Autónoma de Madrid  
Marzo de 2009**

---

## **Palabras clave**

---

Reconocimiento de caras, Extracción de puntos característicos, Segmentación de rasgos faciales, Detección de ojos, Momentos de Hu, Modelo de Mezcla de Gaussianas,

## **Resumen**

---

El principal objetivo de este PFC es el diseño de un sistema para identificar personas (mediante el reconocimiento de caras frontales) y su aplicación a un sistema de videoseguridad orientado al reconocimiento de individuos en los accesos a un local.

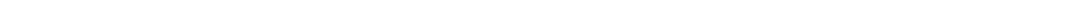
## **Abstract**

---

*The main objective of this PFC is the design of a system in order to identify peoples (using frontal faces recognition) and their use in videosecurity system orientated to the recognition of individuals in the accesses to a place.*

---

*Ese proyecto ha sido parcialmente financiado por la Cátedra Infoglobal-UAM para  
"Nuevas tecnologías de video aplicadas a sistemas de seguridad"*



## **Agradecimientos**

En primer lugar quería dar mi más sincero agradecimiento a Don José María Martínez Sánchez, por haberme dirigido este proyecto fin de carrera. Por su atención, su paciencia y por su apoyo a lo largo de este proyecto.

También quería agradecer a ATVS - *Biometric Recognition Group*, por la cesión de datos que ha permitido desarrollar el PFC.

A Juan Carlos, porque sin su incondicional apoyo y ayuda, la carrera no hubiera sido lo mismo.

Sólo me queda agradecer a toda mi familia, en especial a mis padres, por todo lo que han hecho por mí durante mi etapa de estudiante. Y como olvidarme de 'la pesada' de mi tía, que tenía más ganas de terminar que yo mismo.

*Miguel Cortés Marlia*  
*Marzo de 2009*

---



## INDICE DE CONTENIDOS

Palabras clave .....	4
Resumen .....	5
Abstract .....	6
1 Introducción .....	1
1.1 Motivación .....	1
1.2 Objetivos .....	2
1.3 Organización de la memoria .....	2
2 Estado del arte .....	5
2.1 Reconocimiento Biométrico .....	5
2.1.1 Introducción .....	5
2.2 Técnicas de Reconocimiento Facial en 2D .....	6
2.2.1 <i>Principal Components Analysis (PCA) vs Linear Discriminant Analysis (LDA)</i> .....	6
2.2.2 <i>Elastic Bunch Graph Matching (EBGP)</i> .....	8
2.2.3 <i>Active Appearance Model (AAM)</i> .....	10
2.3 Técnicas de Reconocimiento en 3D .....	12
2.3.1 <i>3D Eigenfaces</i> .....	13
2.3.2 <i>Modelos Deformables: 3D Morphable Models</i> .....	14
2.4 Sistemas Multimodales .....	15
3 Diseño .....	17
3.1 Introducción .....	17
3.2 Diseño Técnico .....	18
3.2.1 Introducción .....	18
3.2.2 BBDD de caras .....	19
3.2.2.1 BBDD de Test .....	19
3.2.2.2 BBDD Biosec .....	20
3.2.3 Aplicación de Identificación y Modelado .....	22
3.2.3.1 Localización de los ojos .....	23
3.2.3.2 Identificación de puntos característicos .....	28
3.2.3.3 Ratios – Distancias relativas entre puntos .....	30
3.2.4 Aplicación de Reconocimiento .....	30
3.2.4.1 Reconocimiento .....	31
4 Desarrollo .....	33
4.1 Introducción .....	33
4.2 Desarrollo de la aplicación de Identificación y Modelado .....	33
4.2.1 Verificación de los límites de las imágenes .....	34
4.2.2 Binarización ( <i>Thresholding</i> ) .....	35
4.2.2.1 Método 1 .....	35
4.2.2.2 Método 2 .....	37
4.2.3 Identificación de Objetos .....	42
4.2.3.1 Algoritmo de 8-conectividad .....	43
4.2.3.2 Localización de los objetos .....	44

4.2.4	Rotación de la imagen.....	54
4.2.5	Identificación de Puntos Característicos.....	55
4.2.5.1	Segmentación de zonas características de imágenes de caras frontales e identificación de puntos .....	56
4.2.6	Ratios – Distancias Relativas entre Puntos .....	75
4.2.7	Modelado.....	77
4.3	Desarrollo de la aplicación de reconocimiento.....	77
4.3.1	El módulo de reconocimiento emplea la técnica de reconocimiento de mezcla de gaussianas. Módulo de Reconocimiento.....	78
5	Integración, pruebas y resultados.....	79
5.1	Introducción.....	79
5.2	Sistema completo .....	79
5.2.1	Prefiltrado de imágenes.....	79
5.2.2	Sistema Localización de Ojos .....	81
5.2.3	Pruebas de reconocimiento y resultados .....	82
6	Conclusiones y trabajo futuro.....	87
6.1	Conclusiones .....	87
6.2	Trabajo futuro.....	87
	Referencias .....	I
	Glosario.....	VII
	Apéndices .....	IX
A	Conceptos Básicos.....	IX
B	Modelo de mezcla de Gaussianas – GMM (Gaussian Mixture Models) .....	X
	Anexos .....	XI
A	Manual de instalación y usuario.....	XI
B	Manual del programador .....	XVII
B.1	Aplicación de Identificación y Modelado.....	XVII
B.2	Aplicación de Reconocimiento .....	XXI
C	Generación de humanoides mediante descripciones de alto nivel en el marco de estándares MPEG.....	XXIV
C.1	Descripción de las Relaciones Corporales.....	XXIV
	Presupuesto.....	XXVII
	Pliego de Condiciones.....	XXIX
	Condiciones generales.....	XXIX
	Condiciones particulares .....	XXXII

## INDICE DE FIGURAS

FIGURA 1: <i>EIGENFACES</i> DE UNA SERIE DE IMÁGENES (IMAGEN OBTENIDA DE [6]).....	7
FIGURA 2: CARA MEDIA DE UNA BBDD (IMAGEN OBTENIDA DE [6]).....	7
FIGURA 3: REPRESENTACIÓN INTERNA DE LA APROXIMACIÓN EBGM. (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 <i>JET</i> QUE MEJOR SE ADAPTA. (IMAGEN OBTENIDA DE [13]).....	8
FIGURA 4: PARA ACLARAR LA DEFINICIÓN DE <i>JET</i> , SÓLO SE MUESTRAN EN LA FIGURA, 3 FRECUENCIAS Y 4 ORIENTACIONES. UNA COLECCIÓN DE ESTOS <i>JETS</i> , JUNTO CON INFORMACIÓN SOBRE SU LOCALIZACIÓN CONSTITUYE UN <i>IMAGE GRAPH</i> , USADO PARA REPRESENTAR UN COMO OBJETO, COMO POR EJEMPLO, UNA CARA (IMAGEN OBTENIDA DE [14]).....	9
FIGURA 5: REPRESENTACIÓN DE UN <i>FACE BUNCH GRAPH</i> (FBG) (IMAGEN OBTENIDA DE [14]).....	10
FIGURA 6: CARA CON <i>LANDMARKS</i> (IMAGEN OBTENIDA DE [16]).....	11
FIGURA 7: PARTICIÓN DE LA TEXTURA EN ZONAS (IMAGEN OBTENIDA DE [18]).....	12
FIGURA 8: MODELO DE CARA REPRESENTADO POR UNA NUBE DE PUNTOS (IMAGEN OBTENIDA DE [22]).....	13
FIGURA 9: LA RED MODELADA EN DIFERENTES NIVELES. (A) RED BÁSICA. (B) DE PRIMER NIVEL (C) DE SEGUNDO NIVEL. (D) DE TERCER NIVEL. (E) DE CUARTO NIVEL (IMAGEN OBTENIDA DE [22]).....	13
FIGURA 10: PARA EL RECONOCIMIENTO, LOS COEFICIENTES A Y B DE LA IMAGEN DE PRUEBA, SE COMPARAN CON LOS COEFICIENTES DE LA COLECCIÓN DE IMÁGENES DE ENTRENAMIENTO (IMAGEN OBTENIDA DE [17]).....	15
FIGURA 11: COMPARATIVA ENTRE EL RECONOCIMIENTO 2D, 3D Y EL HÍBRIDO (IMAGEN OBTENIDA DE [32]).....	16
FIGURA 12: PUNTOS CARACTERÍSTICOS DE UNA CARA NEUTRAL (IMAGEN OBTENIDA DE [27]).....	17
FIGURA 13: ARQUITECTURA GENERAL DEL SISTEMA.....	19
FIGURA 14: IMÁGENES DE LA BBDD DE TEST.....	20
FIGURA 15: A LA IZQUIERDA LA IMAGEN ORIGINAL. A LA DERECHA SE MUESTRA LA ZONA DE LA IMAGEN QUE CONTIENE COLOR CARNE.....	21
FIGURA 16: EXTRACCIÓN DE LA CARA USANDO LA DETECCIÓN DE COLOR CARNE.....	21
FIGURA 17: EXTRACCIÓN DE LA CARA EN UNA IMAGEN DE LA BBDD.....	22
FIGURA 18: APLICACION DE IDENTIFICACIÓN.....	22

FIGURA 19: LOCALIZACIÓN DE OJOS .....	24
FIGURA 20: A LA IZQUIERDA LAS IMÁGENES DE LA BBDD. EN EL CENTRO EL RESULTADO DE LA UMBRALIZACIÓN POR EL PRIMÉR MÉTODO. A LA DERECHA, LAS MISMAS IMÁGENES PERO ESTA VEZ UMBRALIZÁNDOLAS POR EL SEGUNDO MÉTODO. ....	25
FIGURA 21: A LA IZQUIERDA LA IMAGEN RESULTANTE DE LA ETAPA DE UMBRALIZACIÓN. A LA DERECHA SE MUESTRA LOS DISTINTOS OBJETOS IDENTIFICADOS COMO CANDIDATOS A SER LOS OJOS. CADA COLOR REPRESENTA UN OBJETO DISTINTO. ....	26
FIGURA 22 : A LA IZQUIERDA LA IMAGEN ORIGINAL, A LA DERECHA LA MISMA IMAGEN DESPUÉS DE SER ROTADA.....	28
FIGURA 23: REGIONES DE INTERÉS DE UNA CARA.....	29
FIGURA 24: A LA IZQUIERDA LA IMAGEN ORIGINAL, A LA DERECHA LA NUBE DE PUNTOS EXTRAIDA DE LA MISMA IMAGEN.....	29
FIGURA 25 : A) LA IMAGEN ORIGEN. B) EL RECTANGULO ROJO REPRESENTA LA IMAGEN DESTINO QUE SE OBTENDRÁ AL EXTRAER DICHA PORCIÓN DE LA IMAGEN ORIGEN. C) LA ZONA SOMBREADA REFLEJA QUE LA IMAGEN DESTINO SE ENCUENTRA FUERA DE LOS LÍMITES ESTABLECIDOS DE LA IMAGEN ORIGEN.....	34
FIGURA 26: ORIGEN DE COORDENADAS DE LAS IMÁGENES.....	35
FIGURA 27: HISTOGRAMA DE LA IMAGEN. EN VERDE, EL VALOR DEL PÍXEL QUE MÁS SE REPITE; EN ROJO, EL VALOR UMBRAL .....	36
FIGURA 28: A LA DERECHA SE MUESTRA LA IMAGEN BINARIA OBTENIDA A PARTIR DEL PRIMER MÉTODO DE UMBRALIZACIÓN SOBRE LA IMAGEN DE LA IZQUIERDA. ....	37
FIGURA 29 : A LA DERECHA UNA IMAGEN DE LA BBDD. EN EL CENTRO, LA IMAGEN DE LUMINANCIA. A LA DERECHA, EL HISTOGRAMA DE LA IMAGEN DE LUMINANCIA .....	38
FIGURA 30: A LA DERECHA, LA IMAGEN UMBRALIZADA POR EL PRIMER MÉTODO.....	39
FIGURA 31: HISTOGRAMA DE UNA IMAGEN DE LA CARA. M1 Y M2 SON LOS DOS MÁXIMOS DEL HISTOGRAMA; Y TH ES EL VALOR DE LA INTENSIDAD ESCOGIDO PARA LA UMBRALIZACIÓN O <i>THRESHOLDING</i> (IMAGEN OBTENIDA DE [28]). ....	40
FIGURA 32: ARRIBA A LA IZQUIERDA, LA IMAGEN DE LUMINANCIA DE IMAGEN ORIGINAL. ARRIBA A LA DERECHA, EL HISTOGRAMA DE ESTA IMAGEN. ABAJO A LA IZQUIERDA LA IMAGEN DE LUMINANCIA SUAVIZADA. ABAJO A LA DERECHA, EL HISTOGRAMA SUAVIZADO.....	41
FIGURA 33: DE IZQUIERDA A DERECHA, LA IMAGEN DE LUMINANCIA ORIGINAL (A), LA ORIGINAL SUAVIZADA (B), LA IMAGEN UMBRALIZADA SIN SUAVIZAR (C) Y LA UMBRALIZADA SUAVIZADA (D).....	42
FIGURA 34: RESULTADO DEL <i>THRESHOLDING</i> DE 16 IMÁGENES DE LA BBDD DE TEST. ....	42
FIGURA 35: TRES TIPO DE CONECTIVIDAD: A) 4-CONECTIVIDAD, B) 8-CONECTIVIDAD Y C) 6-CONECTIVIDAD .....	43

FIGURA 36: RESULTADO DEL ALGORITMO DE CONECTIVIDAD. A LA IZQUIERDA 4-CONECTIVIDAD. CADA COLOR REPRESENTA UN OBJETO DISTINTO. A LA DERECHA 8-CONECTIVIDAD.....	44
FIGURA 37: EJEMPLO DE <i>Skew</i> EN UNA IMAGEN.....	46
FIGURA 38: EL CONJUNTO DE LOS SIETE MOMENTOS INVARIANTE DE <i>Hu</i> APLICADOS SOBRE LA IMAGEN DE LA FIGURA 37. CADA GRÁFICA CONTIENE 3 ETIQUETAS QUE REPRESENTAN CADA PALO DE LA BARAJA: EL TRÉBOL (C), EL CORAZÓN (H) Y LOS DIAMANTES (D) (IMAGEN OBTENIDA DE [31]). .....	48
FIGURA 39: A LA IZQUIERDA LA IMAGEN ORIGINAL. EN EL CENTRO, LA IMAGEN UMBRELIZADA. A LA DERECHA, EN VERDE, LOS OBJETOS QUE SUPERAN EL LÍMITE MÁXIMO A SER LOS OJOS; EN AZUL, AQUELLOS QUE NO SUPERAN LA TALLA MÍNIMA; EN ROJO, LOS CANDIDATOS A SER LOS OJOS. ....	49
FIGURA 40: DETALLE DE UN OBJETO. X REPRESENTA LAS FILAS EN LAS QUE ESTÁ EL OBJETO EN LA IMAGEN ORIGINAL Y LAS Y, LAS COLUMNAS. ....	50
FIGURA 41 : MUESTRA DE 16 IMÁGENES DE LA BBDD DE TEST .....	52
FIGURA 42: RESULTADOS DE LA DETECCIÓN DE OJOS, SIN LA ETAPA DE VERIFICACIÓN.....	53
FIGURA 43: RESULTADOS DE LA DETECCIÓN DE OJOS, CON LA ETAPA DE VERIFICACIÓN .....	54
FIGURA 44: CÁLCULO DEL ÁNGULO DE ROTACIÓN DE UNA IMAGEN .....	54
FIGURA 45 : A) IMAGEN SIN ROTAR. B) IMAGEN ROTADA .....	55
FIGURA 46: PUNTOS CARACTERÍSTICOS DE LA CARA.....	55
FIGURA 47: DIAGRAMA DE BLOQUES DE LA EXTRACCIÓN DE IMÁGENES .....	57
FIGURA 48: DEFINICIÓN DE LA MÁSCARA INICIAL CORRESPONDIENTES A LOS OJOS.....	58
FIGURA 49: DETALLE DEL PUNTO DE REFERENCIA DE LAS IMÁGENES EXTRAÍDAS.....	59
FIGURA 50: LAS IMÁGENES DE LA DERECHA SON LAS IMÁGENES BINARIAS RESULTANTES DEL PROCESO DE BINARIZACIÓN. LAS IMÁGENES DE LA DERECHA SON EL RESULTADO DE LA ETAPA ANTERIOR. ....	59
FIGURA 51: BARRIDO PROGRESIVO.....	60
FIGURA 52: VENTANA DE ANÁLISIS PARA ESTA IMAGEN.....	60
FIGURA 53: VENTANA DE ANÁLISIS AJUSTADA .....	60
FIGURA 54: EL FIN DE LAS PESTAÑAS COINCIDE CON EL EXTREMO BUSCADO.....	61
FIGURA 55: ARRIBA, LA IMAGEN HISTOGRAMA DE CADA OJO; ABAJO, LA ROI A ESTUDIAR .....	62
FIGURA 56: PUNTOS CARACTERÍSTICOS DE LOS OJOS .....	62

FIGURA 57: A LA IZQUIERDA LA IMAGEN DE LA LUMINANCIA DE LA NARIZ. A LA DERECHA, EL HISTOGRAMA CORRESPONDIENTE .....	64
FIGURA 58: EXTRACCIÓN DE LOS PUNTOS DE LA NARIZ PARA UNA IMAGEN DE LA BBDD. A LA IZQUIERDA SE MUESTRAN LA IMAGEN ORIGINAL YA ROTADA Y A LA DERECHA UN PRIMER PLANO DE LA NARIZ CON LOS PUNTOS EXTRAÍDOS.....	66
FIGURA 59: IMAGEN DE LA BOCA Y SU TRANSFORMACIÓN CROMÁTICA.....	68
FIGURA 60: TRANSFORMACIÓN CROMÁTICA Y SU PERÍMETRO.....	68
FIGURA 61: RESULTADO DE LA EXTRACCIÓN DE PUNTOS DE LA BOCA PARA UNA IMAGEN DE LA BBDD BIOSEC. A LA IZQUIERDA, LA IMAGEN DE LA BBDD; EN EL CENTRO, EL PERÍMETRO DE LA TRANSFORMACIÓN CROMÁTICA DE LA BOCA; A LA DERECHA, LOS PUNTOS CARACTERÍSTICOS EXTRAÍDOS DE LA BOCA.....	69
FIGURA 62: DE IZQUIERDA A DERECHA, LA IMAGEN DE BORDES DE LA COMPONENTE VERDE, LA IMAGEN DE BORDES DE LA COMPONENTE AZUL Y LA IMAGEN HISTOGRAMA H. ....	70
FIGURA 63 : EN A) LA IMAGEN DE BORDES CON EL VECTOR A ANALIZAR EN ROJO. EN B) DETALLE DEL VECTOR A ANALIZAR. ....	71
FIGURA 64: A LA IZQUIERDA, LA IMAGEN DE LA BBDD CON LOS PUNTOS DETECTADOS. A LA DERECHA, EN ROJO Y AZUL, LAS RECTAS QUE DELIMITAN EL VECTOR A ANALIZAR; EN AMARILLO, EL VECTOR A ANALIZAR. ....	75
FIGURA 65 : PROPORCIONES CARACTERÍSTICAS DE UNA CARA. ....	76
FIGURA 66: EJEMPLO DE IMAGEN DONDE LOS PUNTOS ESTÁN MAL UBICADOS .....	80
FIGURA 67: RESULTADO DE LA LOCALIZACIÓN DE LOS OJOS EN UN CONJUNTO DE 12 IMÁGENES DE LA BBDD DE <i>BIOSEC</i> .....	81
FIGURA 68: POSIBLES CASOS EN LA LOCALIZACIÓN: FIGURA A) DETECCIÓN ERRÓNEA, NO SE HA DETECTADO BIEN LOS OJOS; FIGURA B) AUNQUE SE DETECTA UNA PARTE DEL OJO, LA LOCALIZACIÓN NO SE CONSIDERA CORRECTA; FIGURA C) DETECCIÓN CORRECTA, SE DETECTA LA TOTALIDAD DE LOS OJOS .....	82
FIGURA 69: RESULTADO DE IDENTIFICACIÓN TOMANDO 4 MUESTRAS POR MODELO. EL RESULTADO SE MUESTRA TANTO SI SE APLICA EL SEXTO FILTRO COMO SI NO.....	84
FIGURA 70: RESULTADO DE LA IDENTIFICACIÓN DE 13 INDIVIDUOS DE LA BBDD DE <i>BIOSEC</i> MODELÁNDOLOS CON DISTINTO NÚMERO DE MUESTRAS.....	85

## INDICE DE TABLAS

TABLA 1: DEFINICIÓN DE LOS MOMENTOS DE HU .....	47
TABLA 2: DEFINICIÓN DE LOS PUNTOS CARACTERÍSTICOS ( <i>KEYPOINTS</i> ) DE LA CARA.....	56
TABLA 3 : RATIOS.....	76
TABLA 4 : DESCRIPCIÓN DE LOS RATIOS .....	83
TABLA 5.- DIMENSIONES DE LOS PRISMAS CONTENEDORES DE LOS ELEMENTOS PRINCIPALES DE LA CARA .....	XXVI





# 1 Introducción

---

## 1.1 Motivación

El reconocimiento de caras es una tarea que los seres humanos realizamos de forma rutinaria y sin esfuerzo. Tal es así que la ejecutamos casi todos los días sin darnos cuenta.

El bajo coste de los ordenadores y el avance tecnológico ha permitido que haya un enorme interés en automatizar procesos de tratamiento de imágenes y video. La investigación en el mundo del reconocimiento ha sido motivada por el número de aplicaciones que demandan la identificación del individuo.

El interés sobre el reconocimiento ha ido creciendo a medida que han ido surgiendo nuevos avances tecnológico como las cámaras digitales, dispositivos móviles, Internet y aumento de la demanda en temas de seguridad. A día de hoy podemos encontrar dispositivos digitales que detectan el número de caras de una escena para establecer el mejor enfoque de la misma[1]; buscadores de imágenes en Internet [2] que permiten identificar las distintas webs que hacen uso de la imagen. Cabe destacar que el mayor avance en estos últimos años se ha producido sobre todo en el ámbito de la seguridad: acceso personal autorizado, reconocimiento de matrículas, reconocimiento de somnolencia al volante [3], etc. Una de las principales características del reconocimiento facial es que se trata de un método no intrusivo; es decir, los datos pueden ser adquiridos si que el sujeto se percate de ello. Además, el aspecto facial es el método más utilizado de manera natural por el ser humano para reconocerse unos a otros.

Los desarrollos sobre los sistemas de reconocimiento de caras han ido mejorándose significativamente desde el primer sistema de reconocimiento automático de caras implementado por Kanade [4].

Aunque los progresos en el reconocimiento de caras han sido alentadores, el reconocimiento ha resultado ser una tarea difícil, debido a que las imágenes varían considerablemente según el punto de vista, iluminación, expresión, la pose, los accesorios, etc.

Uno de los casos de éxito más reciente en el reconocimiento de caras, es el sistema de autenticación implantado en el mayor casino del mundo, "Venetian" con el fin de autenticar la entrada de forma segura y eficiente a los 12000 empleados del *resort* [5].

Otro ejemplo a destacar es la compra de una compañía centrada en programas de reconocimiento de fotografías, Neven Vision, por parte de Google en Agosto 2006. La intención de Google es integrar la tecnología de reconocimiento de caras a sus álbumes de fotos de "Picasa" con el fin de catalogar las fotos automáticamente.

La motivación de este PFC viene alentada por el hecho de estudiar y analizar una de las técnicas de reconocimiento más complejas (por las distintas variables que alteran una imagen) y con mayor futuro en el ámbito de la seguridad como la identificación de potenciales sospechosos (fichados) en un recinto público.

## 1.2 Objetivos

El principal objetivo de este PFC es desarrollar un sistema de reconocimiento facial en posición frontal (o cuasi-frontal) basado en las aproximaciones más "prometedoras" de las encontradas durante la fase de estudio del arte.

La solución seleccionada se basa en la detección de un subconjunto de los puntos característicos de la cara (*keypoints*, KP) y definir una serie de características relativas a los mismos que posteriormente permitan evaluar las capacidades discriminatorias entre las diversas personas a reconocer. Para ello se segmentará el rostro a reconocer en las distintas partes que lo componen para posteriormente identificar estos puntos característicos. Una vez obtenidos estos puntos, se calcularán relaciones entre medidas (*ratios*), como por ejemplo, la distancia de ojos entre la altura de la nariz; con el fin de recharacterizar a las personas por varias medidas denominadas ratios.

## 1.3 Organización de la memoria

La memoria consta de los siguientes capítulos:

- **Capítulo 1.** Introducción, objetivos y motivación del proyecto.
- **Capítulo 2.** Estado del arte. Análisis de las principales técnicas de reconocimiento de caras.
- **Capítulo 3.** Descripción del diseño y arquitectura del sistema. Además también se analiza el formato de los datos a usar por dicho sistema.
- **Capítulo 4.** Descripción del desarrollo e implementación del diseño del sistema. Problemas en el desarrollo del sistema.
- **Capítulo 5.** Resultados obtenidos con las pruebas realizadas al sistema completo que se ha implementado en el apartado anterior.
- **Capítulo 6.** Conclusiones obtenidas tras el desarrollo del sistema. Relación de posibles líneas futuras de desarrollo y mejoras del sistema.
- **Glosario.** Catálogo de términos empleados a lo largo de esta memoria.
- **Apéndice A.** Conceptos Básicos sobre el tratamiento digital de una imagen.
- **Apéndice B.** Breve descripción del modelo de mezclas de gaussianas (GMM – Gaussian Mixture Model)

- **Anexo A.** Breve manual con instrucciones de instalación del software utilizado por el sistema.
- **Anexo B.** Breve guía de las funciones que conforman el sistema.
- **Anexo C.** Descripción del estudio sobre relaciones de la cara realizado en el trabajo de iniciación a la investigación de Víctor Fernández-Carbajales Cañete.



## **2 Estado del arte**

---

### **2.1 Reconocimiento Biométrico**

#### **2.1.1 Introducción**

El reconocimiento biométrico se define como el reconocimiento automático de individuos basados en sus características fisiológicas o de comportamiento.

Las huellas dactilares, las retinas, el iris, los patrones faciales de venas de la mano o la geometría de la palma de la mano, representan ejemplos de características físicas (estáticas), mientras que entre los ejemplos de características del comportamiento se incluye la firma, el paso y el tecleo (dinámicas). La voz se considera una mezcla de características físicas y del comportamiento, pero todos los rasgos biométricos comparten aspectos físicos y del comportamiento.

Las características deben ser invariantes en el tiempo, representadas a través de una medición numérica, distinguibles entre una persona y otra, y universales, es decir que cada persona la tenga. Adicionalmente, los sistemas biométricos deben contar con una serie de requisitos: deben ser seguros de tal forma que no puedan ser burlados usando métodos fraudulentos; deben tener un alto rendimiento trabajando a velocidades razonables, con bajas tasas de error; deben tener una alta aceptabilidad por los usuarios ya que voluntariamente ellos deben aceptar someterse a ser medidos.

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 tratamiento 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 que utilizan imágenes que luego pueden ser mostradas para su cotejo por parte de dicho operador.

De todos los rasgos anatómicos, el rostro es el elemento que con más frecuencia utilizamos los seres humanos para identificar a otro individuo. Para ello, el cerebro comienza por establecer los aspectos físicos de una cara, a continuación determina si estas facciones son conocidas o no y, por último, procede a otorgar un nombre a lo que ve.

Este proceso tan aparentemente sencillo para nosotros puede resultar muy difícil para una máquina. Además, nuestra habilidad es tal que somos capaces de reconocer la fisonomía de cualquier persona obviando detalles como la longitud del pelo, ciertos complementos (gafas, sombreros,...) o incluso los cambios que experimenta el rostro por el paso del tiempo.

Desde 1960 se lleva investigando en el reconocimiento de caras, y hasta el momento no existe un sistema lo suficientemente fiable como para sustituir a las actuales tarjetas magnéticas o claves (métodos no biométricos). Los sistemas biométricos de reconocimiento facial, al cometer errores considerables, no han podido reemplazar a los no biométricos basados en tarjetas magnéticas o claves [5]. Sin embargo, cada día está siendo más evidente, incluir información biométrica en sistemas de reconocimiento de personas.

## 2.2 Técnicas de Reconocimiento Facial en 2D

Varios métodos de reconocimiento facial han sido desarrollados hasta el momento, la mayoría de ellos basados en imágenes 2D: *eigenfaces*, *fisherfaces*, *Active Appearance Models*...

### 2.2.1 *Principal Components Analysis (PCA) vs Linear Discriminant Analysis (LDA)*

El método más utilizado, tanto por su simplicidad como por los buenos resultados que proporciona, consiste en proyectar las imágenes en un espacio unidimensional.

Se transforma la imagen [ $n \times m$  píxeles] en un vector de dimensiones [ $n \cdot m$ ]. Este espacio  $n \cdot m$  es demasiado grande como para permitir un reconocimiento de objetos rápido y robusto. Una manera rápida de resolver este problema consiste en usar técnicas que permitan reducir la dimensión del problema. Las dos técnicas más conocidas son: PCA[6] (*Principal Components Analysis*) y LDA[7] (*Linear Discriminant Analysis*, también conocida como *Fisher Discriminant Analysis*, FDA).

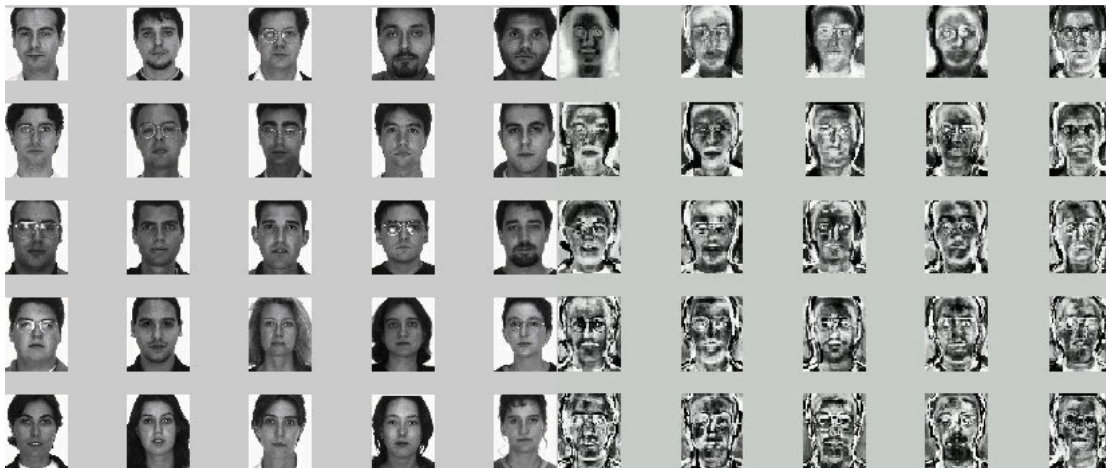
PCA es el más utilizado en la técnica de reconocimiento conocida como autocaras (*Eigenfaces*) [8] [9] mientras que FDA o LDA es utilizado en *Fisherfaces*.

El análisis PCA deriva de la Transformada de Karhunen-Loewe (KL). Dada una representación vectorial  $s$ -dimensional de cada cara, en un conjunto de imágenes de entrenamiento, el análisis PCA tiende a encontrar un subespacio  $t$ -dimensional ( $t < s$ ) cuya base de vectores corresponde a las direcciones de máxima varianza del espacio de la imagen original. Este nuevo subespacio es normalmente de dimensión inferior ( $t < s$ ). La base de vectores del PCA son definidas como autovectores de la matriz de covarianza (matriz de dispersión)[8].

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 1 muestra los autovectores de un conjunto de entrenamiento

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.



**Figura 1: Eigenfaces de una serie de imágenes (Imagen obtenida de [6])**

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.



**Figura 2: Cara media de una BBDD (Imagen obtenida de [6])**

Una variación del Análisis de Componentes Principales denominada “*Fisherfaces*” [11] fue propuesto como una mejora del PCA. 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.

El objetivo de LDA es encontrar los vectores en el espacio subyacente que mejor discriminen entre clases. Para todas las imágenes de todas las clases, se definen dos tipos de matrices de dispersión, una matriz de dispersión entre clases (*inter-class*),  $S_B$ , y otra dentro de las clases (*intra-class*),  $S_W$ . El objetivo es maximizar

$S_B$  reduciendo al mínimo  $S_W$ , en otras palabras, maximizar la relación

$\frac{\det|S_B|}{\det|S_W|}$ . Se consigue maximizar esta relación haciendo que los vectores

columna de la matriz de proyección sean los autovectores de  $S_W^{-1}xS_B$ .

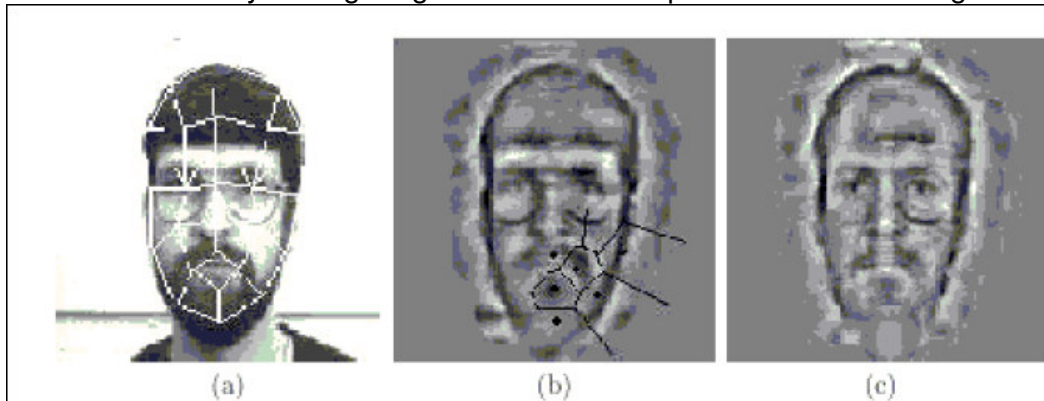
El usar uno u otro método dependerá de lo grande que sea la Base de Datos (BBDD) de entrenamiento. En una BBDD no muy grande es más robusto el análisis PCA que el LDA [10][11]

La principal ventaja de estos dos análisis es que no es necesario crear modelos de los objetos a reconocer, mientras que la desventaja radica en que al añadir/borrar una imagen de la BBDD de entrenamiento es necesario entrenar de nuevo toda la BBDD, con el correspondiente coste computacional que ello conlleva.

### 2.2.2 Elastic Bunch Graph Matching (EBGM)

Los métodos basados en ondículas (*wavelets*), utilizan esta transformación para identificar los bordes de las características faciales[13]. El análisis EBGM (*Elastic Bunch Graph Matching*) parte de la idea de que todas las caras comparten una estructura topológica similar.

El algoritmo EBGM (ver Figura 3) 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.

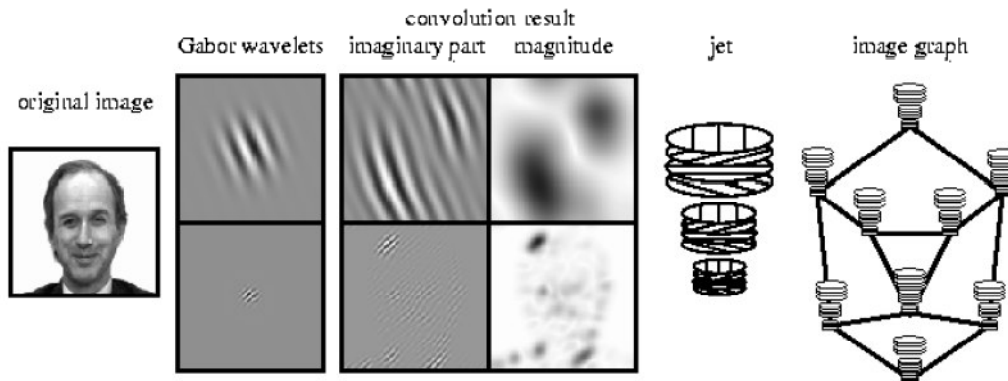


**Figura 3: Representación interna de la aproximación EBGM. (a) representa un grafo sobre una cara. (b) es la reconstrucción de la cara a partir del grafo. (c) representa la reconstrucción de la cara a partir de un grafo conjunto en el que se ha considerado en cada nodo el *jet* que mejor se adapta. (Imagen obtenida de [13])**



Las caras son representadas como grafos, con nodos situados en puntos de la cara, llamados “*fiducial points*” (ojos, nariz...); y por bordes, etiquetados por vectores de distancia. Cada nodo contiene un conjunto de 40 coeficientes complejos de la transformada *wavelet* de Gabor (5 frecuencias x 8 orientaciones), a diferentes escalas y orientaciones (fase, amplitud). A estos coeficientes les denominan “*jets*” [14] [15](ver Figura 4).

El reconocimiento está basado en gráficos etiquetados (*face graph*). Un gráfico etiquetado es un conjunto de nodos conectados por bordes, los nodos están etiquetados por jets y los bordes etiquetados por distancias.



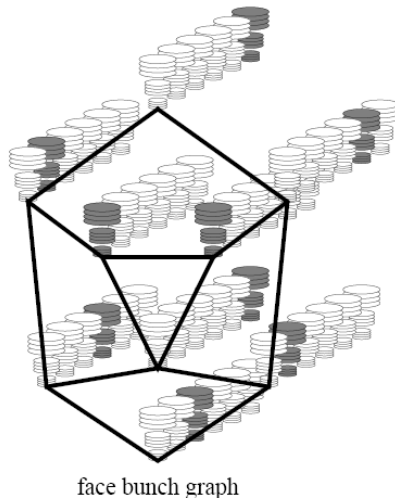
**Figura 4:** Para aclarar la definición de *Jet*, sólo se muestran en la figura, 3 frecuencias y 4 orientaciones. Una colección de estos *jets*, junto con información sobre su localización constituye un *image graph*, usado para representar un como objeto, como por ejemplo, una cara (Imagen obtenida de [14])

*Face Bunch Graph* (FBG) (ver Figura 5) es una representación general de las caras a reconocer. Se diseña para cubrir todas las posibles variaciones de la cara. FBG combina la información de un gran número de *face graphs*. En la comparación de una imagen con el modelo, el mejor ajuste de cada *jet*, se realiza por separado, y se indica en gris

Una pregunta que cabe hacerse es ¿Cuánto de grande tiene que ser FBG?  
 ¿Cuántos modelos de caras tienen que ser utilizados? La respuesta a ambas preguntas, depende en un primer momento de las variaciones que se quieran reconocer en una cara.

Si se quiere cubrir, caras de diferente raza, expresión, edad, etc., FBG deberá contener diferentes modelos para cubrir este abanico de posibilidades.

El tamaño de FBG crece a medida que se quiere aumentar la exactitud para encontrar los puntos en una nueva cara. Esta exactitud se puede estimar aplicando FBG a una cara en la que los *fiducial points* han sido extraídos de forma manual.



**Figura 5: Representación de un *Face Bunch Graph* (FBG) (Imagen obtenida de [14]).**

El tamaño de FBG no depende de número de imágenes de entrenamiento, en general, los modelos FBG pueden ser tan diferentes como se quiera, con el fin de reducir la redundancia y maximizar la variabilidad entre las imágenes.

La construcción del FBG, se realiza de forma manual. Inicialmente, cuando FBG contiene sólo unas pocas caras, es necesario revisar y por lo tanto corregir el resultado del *matching*; pero cuando FBG es suficientemente rico (aproximadamente 70 *graphs*), uno puede confiar en la correspondencia (*matching*) y generar un gran número modelos de forma automática.

La definición de un *face graph*, se realiza en tres etapas: primero se marcan en la imagen los *fiducial points* de una imagen dada; a continuación, se dibujan las líneas que unen los *fiducial points* y se generan etiqueta automáticamente con la diferencia entre la posición de los nodos. Finalmente, se aplica la transformada *wavelet de Gabor* para obtener los jets de esos nodos.

En general, el conjunto de los *fiducial points* debería cubrir uniformemente la cara, pero dependiendo la tarea, puede ser apropiado acentuar ciertas regiones por nodos adicionales. Por ejemplo, en el reconocimiento de caras colocaríamos más puntos dentro de la cara, debido a su importancia para el reconocimiento.

El objetivo del EGBP, es encontrar los *fiducial points* y así, extraer de la imagen, un grafo que maximice la semejanza con el FBG, para lograr encontrar esta semejanza, se emplea una función de similitud.

### **2.2.3 Active Appearance Model (AAM)**

El problema clave de un modelo de cara es la variabilidad y obviamente, para que un modelo sea útil y válido necesita ser específico, es decir, que sea capaz de representar sólo ejemplos válidos del tipo de objeto que hemos modelado. Estos modelos deben ser capaces de representar la variabilidad, así que tienen que ser modelos deformables.

Un modelo de este tipo es el AAM (Modelo Activo de Apariencia o *Active Appearance Models*). AAM propuesto por Cootes et al. [16][17].

Estos modelos se construyen a partir del análisis estadístico de la estructura interna de un objeto cuyo contorno está marcado por puntos, estos puntos reciben el nombre de *landmarks*.

*Active Appearance Model*, es un modelo estadístico integrado que combina un modelo de variaciones de formas con un modelo de variaciones de apariencia. El modelo estadístico, en el que las formas y variaciones del nivel de gris del objeto de interés, pueden generalizarse para la mayoría de los objetos.

Aplicado a una imagen, significa encontrar los parámetros del modelo que generan una imagen sintética lo más cerca posible a la imagen objetivo

Los tres pasos para generar el modelo en el sistema AAM son:

CAPTURA → NORMALIZACIÓN → ANÁLISIS ESTADÍSTICO

Estos tres pasos se aplican a las distintas imágenes que tengamos en nuestro conjunto de entrenamiento.

Captura:

- Forma: se define un número determinado de puntos sobre el contorno del objeto que los interesa para determinar la forma (ver Figura 6).



Figura 6: Cara con *Landmarks* (Imagen Obtenida de [16])

- Textura: se muestrea una imagen usando antes una función de ajuste (*warping function*) como *piece-wise affine*[19], *thin-plate*[20]... y separando, en zonas a comparar, la textura mediante un esquema de triangulación.

Normalización:

- Forma: Se busca un espacio (escenario) normalizado, para ello se alinean la formas eliminando la posición, escalado y orientación.

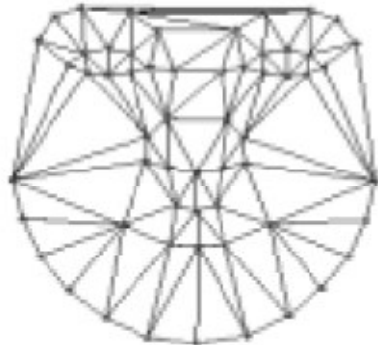
- Textura: Se quita de la imagen los efectos de la iluminación, etc, normalizándola, buscando un contraste y brillo estándar para todas las imágenes.

Análisis estadístico:

- Forma y textura: en ambos casos se utiliza PCA para conseguir una descripción compacta y restrictiva, obteniéndose  $b_s$ , vector de dimensión  $2n$ , diferente para cada una de las caras de nuestro conjunto de entrenamiento

Para generar un modelo completo AAM debe considerarse también la textura. Con la forma, la adquisición de los datos era directa porque los *landmarks* de la forma eran los propios datos. En el caso de la textura se necesita un método consistente para coger la información de textura que hay entre los *landmarks*, es decir, se necesita dividir las imágenes en zonas iguales que nos permitan comparar los valores de textura de distintas caras de forma ordenada en todo el conjunto de entrenamiento.

Cada una de las formas de cada imagen es ajustada a una forma de referencia (ver Figura 7) que, generalmente, es la forma media y luego es muestreada, con ello se consigue, para cada imagen, un trozo de textura con el mismo número de puntos, y ya es posible procesarla.



**Figura 7: Partición de la textura en zonas (Imagen obtenida de [18])**

Llegado a este punto, se realiza una normalización de valores y se aplica sobre la textura PCA, obteniendo un parámetro  $b_G$ .

El sistema decisor compara los parámetros  $b_s$  y  $b_G$  de la imagen capturada con la foto con los parámetros almacenados.

### 2.3 Técnicas de Reconocimiento en 3D

Los métodos basados en análisis 2D están sujetos a limitaciones, como las condiciones ambientales en el momento de capturar la imagen

Algunos intentos para salvaguardar estos inconvenientes vienen de la mano de las redes neuronales, que aplican la modelización de cerebros biológicos para crear sistemas más inteligentes.

### 2.3.1 3D Eigenfaces

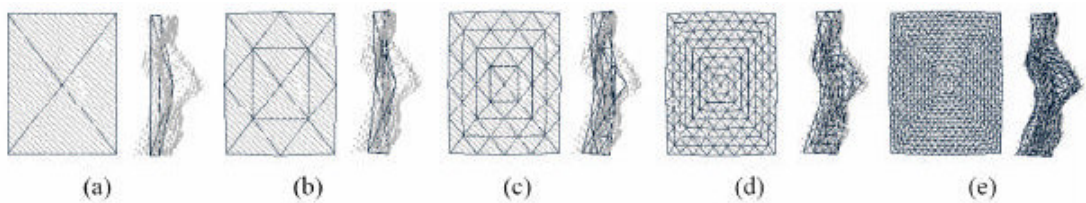
Este método[22], se aplica sobre los modelos de cara de la base de datos 3D\_RMA (BBDD de caras de la *Royal Military Academy*) [38], en la cual los modelos son representados por una nube de puntos (ver Figura 8).



**Figura 8: Modelo de Cara representado por una nube de puntos (Imagen obtenida de [22]).**

El primer problema que hay que mencionar consiste en construir la nube de puntos. Esto se hace mediante un algoritmo iterativo donde lo primero, es localizar la punta de la nariz, que corresponde con el punto más prominente de la nube de puntos, entonces, se alinea el punto encontrado con una red básica. Esta red, se subdivide en cada iteración, como se muestra en la Figura 9.

Se considera el proceso suficientemente preciso con cuatro niveles.



**Figura 9: La red modelada en diferentes niveles. (a) red Básica. (b) de Primer Nivel (c) de Segundo Nivel. (d) de Tercer Nivel. (e) de Cuarto Nivel (Imagen obtenida de [22]).**

Las nubes de puntos tienen diferentes orientaciones, conservando las redes resultantes esta orientación. Se calcula un modelo medio y se alinean a este modelo, todas las redes, existiendo para ello, seis parámetros para las rotaciones y seis para las traslaciones.

Debido al ruido, a veces, no se pueden construir algunos modelos de red que permitan describir la forma geométrica del individuo. Estos modelos se denominan modelos de no caras.

Cada red contiene 545 nodos (puntos), donde los puntos son usados como una imagen de intensidad bidimensional, en donde la intensidad del píxel es la tercera coordenada (coordenada z) de cada nodo. La técnica *eigenfaces* se aplica a estas

imágenes de intensidad. Para el entrenamiento se utiliza un subconjunto de las imágenes calculadas.

Sean  $M_1, M_2, \dots, M_n$  las imágenes de red de entrenamiento y sea  $M_{aver}$  la imagen media de la red y definiendo  $\Phi_i = M_i - M_{aver}$ , entonces la matriz de covariancia se calcula según la expresión:

$$C = \frac{1}{n} \sum_{i=1}^n \Phi_i \Phi_i^T = AA^T$$

Se calculan los autovalores y autovectores de la matriz C, guardando sólo los autovalores más grandes, es decir; aquellos autovectores ortogonales,  $u_1, u_2, \dots, u_e$ , tales que  $e < n$ .

### 2.3.2 Modelos Deformables: 3D Morphable Models

Este sistema de reconocimiento de caras [23] combina modelos deformables en 3D con una simulación gráfica de la proyección y la iluminación realizada por ordenador.

Dada una imagen del usuario, el algoritmo automáticamente estima la forma, la textura y el resto de parámetros relevantes de la representación 3D.

El modelo deformable de la cara, se basa en una representación vectorial de las caras. Este espacio se construye tal que, cualquier combinación convexa que pertenezca al espacio de  $S_i$  y  $T_i$  describe una cara:

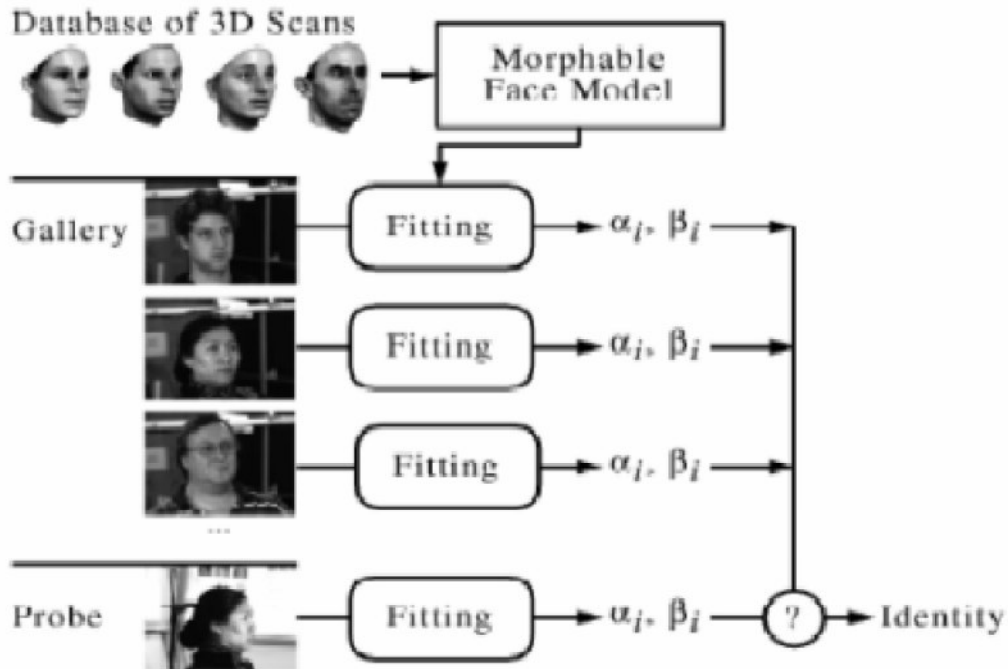
$$S = \sum_{i=1}^m a_i S_i \quad y \quad T = \sum_{i=1}^m a_i T_i$$

Para asegurar que los continuos cambios en  $a_i$  y  $b_i$  representen una transición de una cara a otra, se debe garantizar que no existe nada que coaccione la correspondencia punto a punto. Esto se hace mediante la generalización de la técnica de flujo óptico [39] (al ser un modelo deformable, se usa para ajustar la máscara a la cara, en una secuencia de video), aplicada sobre imágenes tridimensionales en escala de grises.

Los vectores S y T (\*\*) se extraen directamente del modelo 3D, donde el vector S representa la concatenación de los ejes cartesianos (x, y, z) de los puntos; y el vector T es la concatenación de la textura (R, G, B). Se aplica el análisis PCA a los vectores  $S_i$  y  $T_i$ , usando el modelo de Phong [36] para describir la reflexión difusa y especular de la superficie; con el fin de obtener un modelo deformable promedio.

Mediante una función de coste, el algoritmo de ajuste (*fitting*) optimiza una serie de coeficientes de forma y coeficientes de textura, usando para ello 22 parámetros de *rendering*. Estos parámetros se recogen en un vector de características, vector r, agrupando en él, datos relativos al ángulo de la pose, la

traslación en 3D, la distancia focal, ... Se emplean dos paradigmas para probar este método. En el primero de todos, se analizan toda la colección de imágenes con el algoritmo de ajuste (*fitting*), almacenando la forma y la textura. De la misma manera, para una imagen de prueba, se calculan todos los coeficientes y se compara con la colección de datos, obtenida anteriormente, para encontrar la mejor representación gráfica, como se muestra en Figura 10 [17].



**Figura 10:** Para el reconocimiento, los coeficientes  $\alpha$  y  $\beta$  de la imagen de prueba, se comparan con los coeficientes de la colección de imágenes de entrenamiento (Imagen obtenida de [17]).

El segundo de los paradigmas, usa la reconstrucción 3D de la cara, para generar, sintéticamente, una serie de imágenes 2D del sujeto, con el fin de transferirlas a un segundo sistema de reconocimiento dependiente del punto de vista.

## 2.4 Sistemas Multimodales

Los sistemas multimodales tienen como objetivo mejorar la tasa de reconocimiento de los sistemas 2D existentes. Para ello, combinan los resultados de los sistemas de reconocimiento 3D y 2D.

Existen varios estudios, que demuestran que la tasa de reconocimiento obtenida usando un modelo multimodal es mejor que la que se puede obtener usando un único sistema unimodal (ya sea 2D o 3D):

- El método descrito en [22] emplea tanto el rango de datos como la textura de la imagen.

En el dominio tridimensional, los *Point Signature*, se usan para describir los contornos de la cara (se define la forma), en cambio,

para la textura, se aplican los filtros de Gabor para localizar y caracterizar diez puntos de control (los extremos de los ojos, la punta de la nariz, los extremos de la boca,...). Obtenidos los vectores 2D y 3D, se aplica el análisis PCA, para el posterior reconocimiento.

- En [32], el trabajo que ahí se presenta, es un informe sobre la técnica de reconocimiento PCA.

El estudio se hizo sobre una BBDD de 200 personas, usando tanto la forma de la cara como la textura. Las imágenes 2D y 3D obtenidas en el proceso, se analizan por separado y se fusionan usando la distancia Euclídea o la distancia de Mahalanobis, según convenga (se analizan las dos y se escoge la mejor).

Los resultados de este experimento, se muestran en la Figura 11

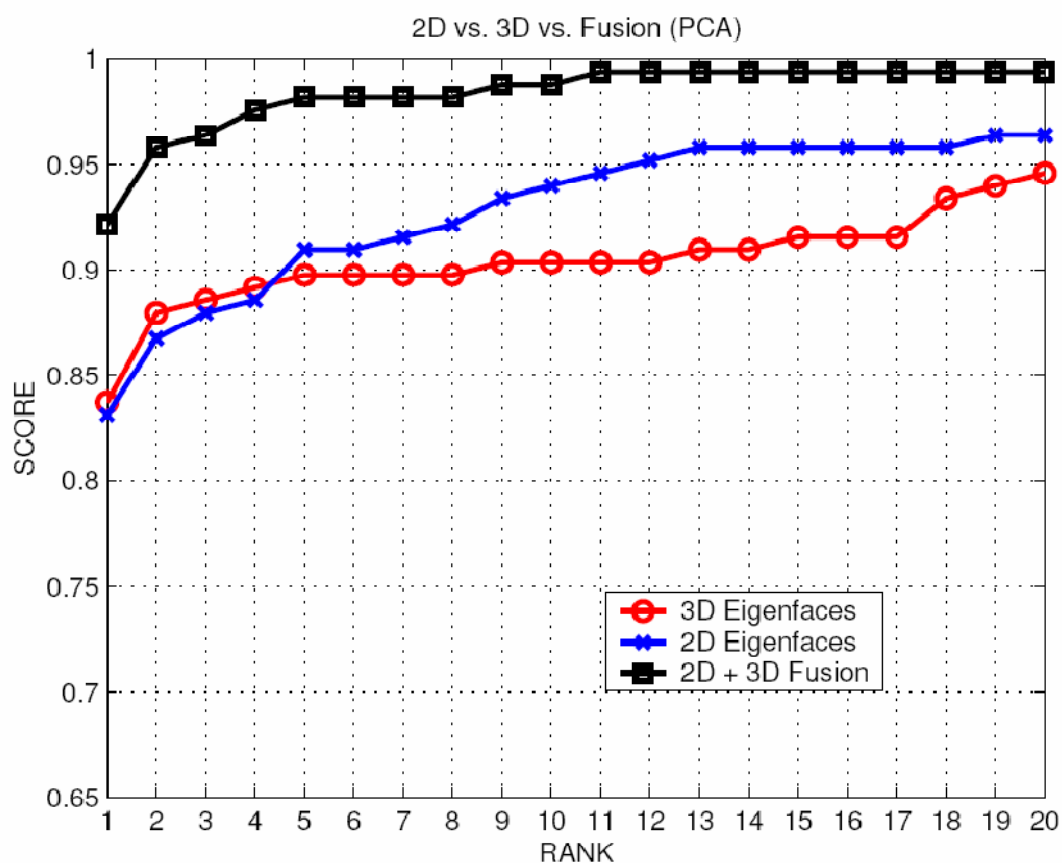


Figura 11: Comparativa entre el reconocimiento 2D, 3D y el Híbrido (Imagen obtenida de [32]).



## 3 Diseño

### 3.1 Introducción

A lo largo de este capítulo vamos a abordar diferentes aspectos involucrados en el diseño del sistema y las razones por las cuales se ha adoptado una u otra tecnología.

Para ello iremos analizando la arquitectura del sistema completo y de sus diferentes bloques funcionales y lo relacionaremos con las tecnologías usadas en cada módulo.

El principal objetivo de este PFC es desarrollar un sistema que sea capaz de reconocer a las personas por las características que presenta su cara con la restricción, entre otras, del tamaño, la luminosidad y la posición frontal o cuasi-frontal de la cara.

El sistema se fundamenta en una técnica basada en la obtención de puntos característicos de la cara para reconocer al individuo. Esta técnica, distinta a cualquier otra comentada en el estado del arte, parte de la definición de puntos característicos de la cara del estándar MPEG-4[27] (ver Figura 12). En este PFC se han diseñado e implementado métodos para encontrar de forma automática estos puntos.

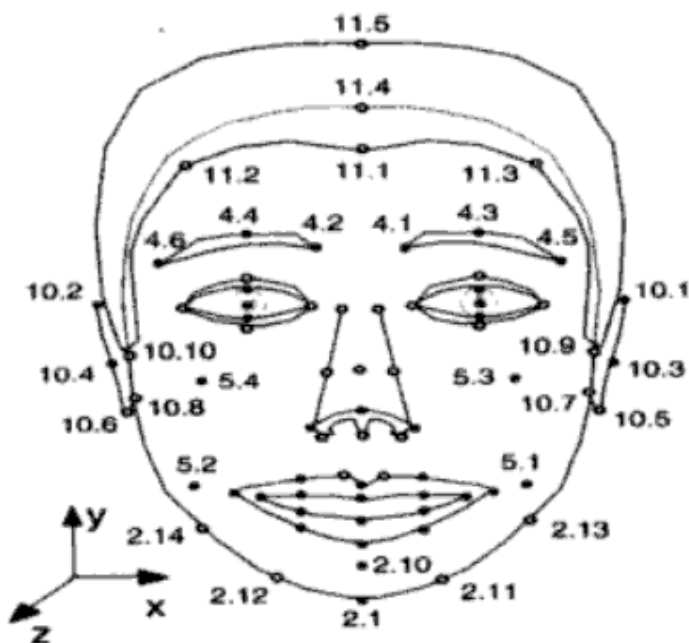


Figura 12: Puntos característicos de una cara neutral (imagen obtenida de [27])

El sistema se compone de dos aplicaciones, una que modela la imagen de la cara para extraer los puntos característicos y otra que a partir de los puntos de la

nueva imagen y de los almacenados en la BBDD, intentará reconocer/identificar al individuo.

La máscara de puntos que se obtiene del primer sistema, así como las relaciones entre estos puntos y las imágenes del individuo, son almacenadas para su posterior utilización en el sistema reconocedor.

Se ha utilizado Matlab como lenguaje para programar las aplicaciones, debido a la facilidad que presenta para trabajar con matrices de datos y a su adecuación para desarrollo de algoritmos de tratamiento de señal de forma rápida y flexible.

Los módulos de dicha aplicación serán descritos con detalle en las posteriores secciones.

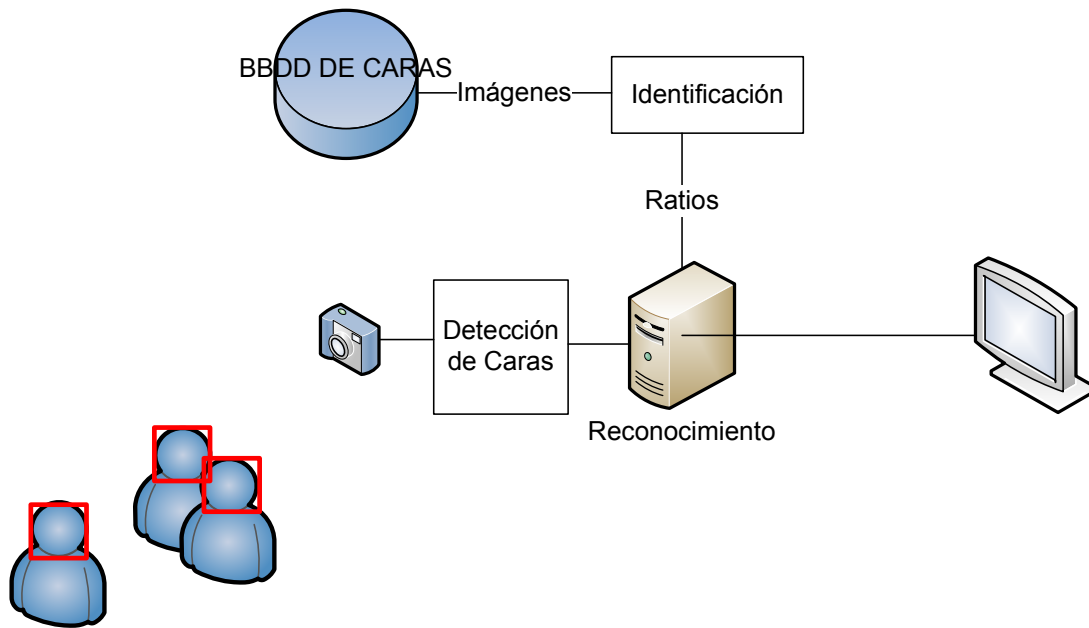
Las imágenes empleadas para modelar al individuo y las empleadas en el reconocimiento deben de seguir uno de los siguientes formatos de codificación: JPEG, BMP o PNG. En el caso de que las imágenes a utilizar estén en otro formato, se deberán de procesar para convertirlas a uno de los formatos descritos anteriormente. Para ello se deberá añadir un módulo, entre la cámara que capture la imagen y el sistema de reconocimiento, que hiciese la transcodificación de un formato a otro.

## 3.2 Diseño Técnico

### 3.2.1 Introducción

El sistema constará de los siguientes elementos:

- BBDD de caras: contendrá fotografías de caras frontales de las personas que se deseen reconocer. Las imágenes de la BBDD son imágenes preprocesadas. El preprocesamiento se basa en aplicar un detector de caras a cada una de las imágenes con el fin de concentrarnos con lo que realmente nos interesa de la imagen, la cara.
- Aplicación de Identificación y Modelado: a partir de las imágenes de la BBDD se extraerá los ratios correspondientes de los 'keypoints' (puntos característicos) seleccionados y se modelará empleando la técnica GMM (*Gaussian Mixture Model*- Modelo de Mezcla de Gaussianas).
- Aplicación de Reconocimiento: se alimenta de las imágenes de caras extraídas del detector. Calcula los ratios de cada imagen y realiza un reconocimiento de la imagen de entrada. Para el reconocimiento se han empleado 10 ratios. El detector está formado por
  - Cámara de vídeo: que capta una secuencia de imágenes.
  - Detector de caras: a partir de las secuencias de imágenes, recogidas de la cámara, extraerá las caras de las distintas personas que aparezcan en la escena. Los tamaños de las imágenes, ya recortadas, con las que se ha trabajado son aproximadamente 200x300 píxeles.



**Figura 13: Arquitectura general del sistema**

### **3.2.2 BBDD de caras**

#### **3.2.2.1 BBDD de Test**

Para el desarrollo de los distintos módulos se ha empleado una BBDD de Test. Esta BBDD consta de 16 imágenes, una por cada individuo, tomadas aleatoriamente con cámaras digitales de calidad megapixel (ver Figura 14).

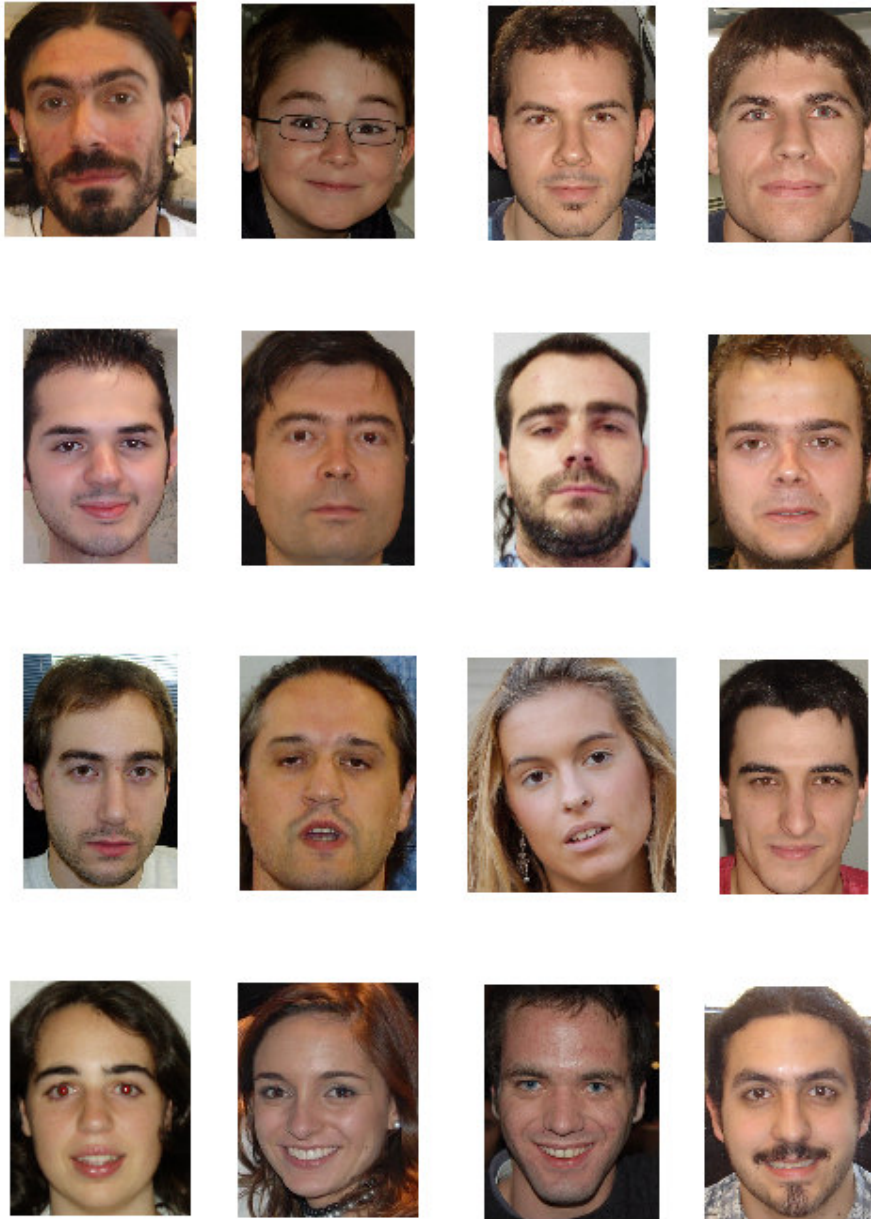


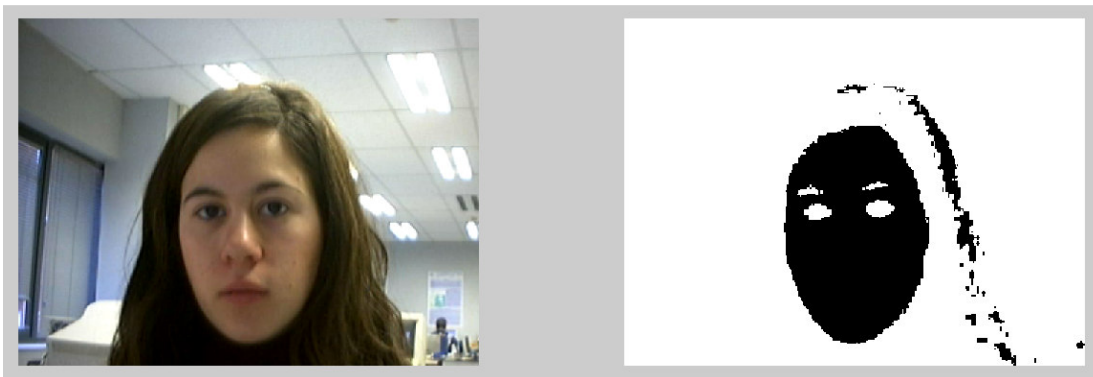
Figura 14: Imágenes de la BBDD de Test.

### 3.2.2.2 BBDD *Biosec*

Los experimentos se han realizado sobre la base de datos BIOSEC[30] [Se trata de una base de datos biométrica compuesta por datos de 200 personas. De entre los rasgos biométricos que componen la base de datos (huellas dactilares, imágenes del Iris, imágenes de caras y grabaciones de voz), los únicos utilizados en este proyecto son los correspondientes a las imágenes frontales de la cara en posición neutral que fueron tomadas a una distancia de unos 30cm de la cámara. Las imágenes de cada persona fueron adquiridas en 2 sesiones diferentes, separadas en el tiempo.

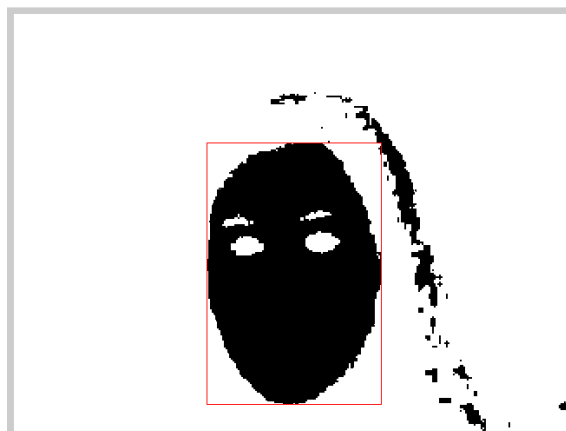
Todas las imágenes de la BBDD han sido preprocesadas para identificar en ellas las caras que aparecen en la escena. La técnica de detección de caras que se ha empleado se ha basado en la detección del color carne.

En la siguiente figura se muestra las zonas de la imagen que el algoritmo detecta como caras.



**Figura 15:** A la izquierda la imagen original. A la derecha se muestra la zona de la imagen que contiene color carne.

La cara se extrae partiendo de la suposición de que sólo hay una cara por imagen. Con esta premisa la cara corresponde con la zona más grande de la imagen blanco y negro resultante del detector de color carne (ver Figura 16).



**Figura 16:** Extracción de la cara usando la detección de color carne.

El resultado final de este preprocesamiento de la BBDD se muestra a continuación:

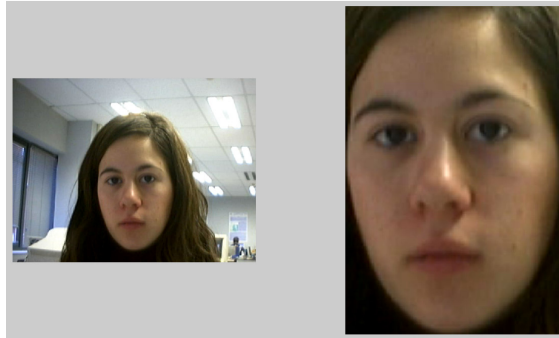


Figura 17: Extracción de la cara en una imagen de la BBDD

### 3.2.3 Aplicación de Identificación y Modelado

Esta aplicación es la encargada de generar toda la información necesaria para el reconocimiento. Analiza cada imagen de forma separada para generar una máscara de puntos característicos que defina la cara.

Dichas imágenes de tamaño variable y en color deben de estar en formato JPG, BMP o PNG (si la imagen tiene cualquier otro formato, existirá un módulo de transcodificación para cambiar el formato de la misma). El esquema de esta aplicación se describe a continuación (Figura 18):

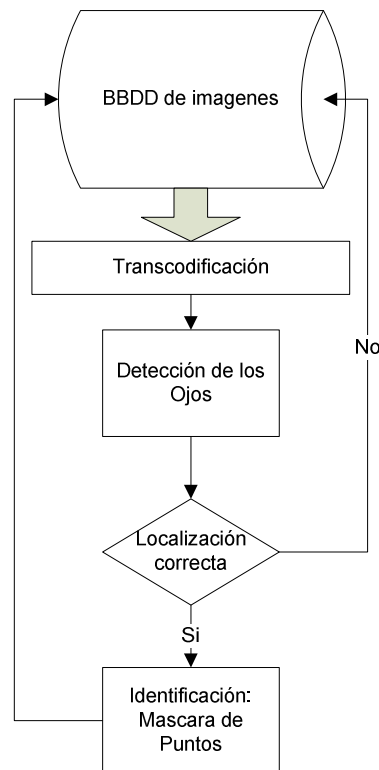


Figura 18: Aplicación de Identificación

Los dos módulos principales son la detección de los ojos y la identificación de la máscara de puntos.

La localización de los ojos se compone de los siguientes bloques funcionales:

- Binarización (*Thresholding*).
- Identificación de objetos.
- Localización.
  - Momentos de Hu.
  - Reglas de Localización.

Los bloques funcionales que definen la identificación de la máscara de puntos son los siguientes:

- Segmentación de rasgos: ojos, nariz, boca.
- Análisis de imágenes y extracción de sus puntos.
- Análisis del resto de Puntos: contorno de la cara.

La aplicación genera los siguientes datos:

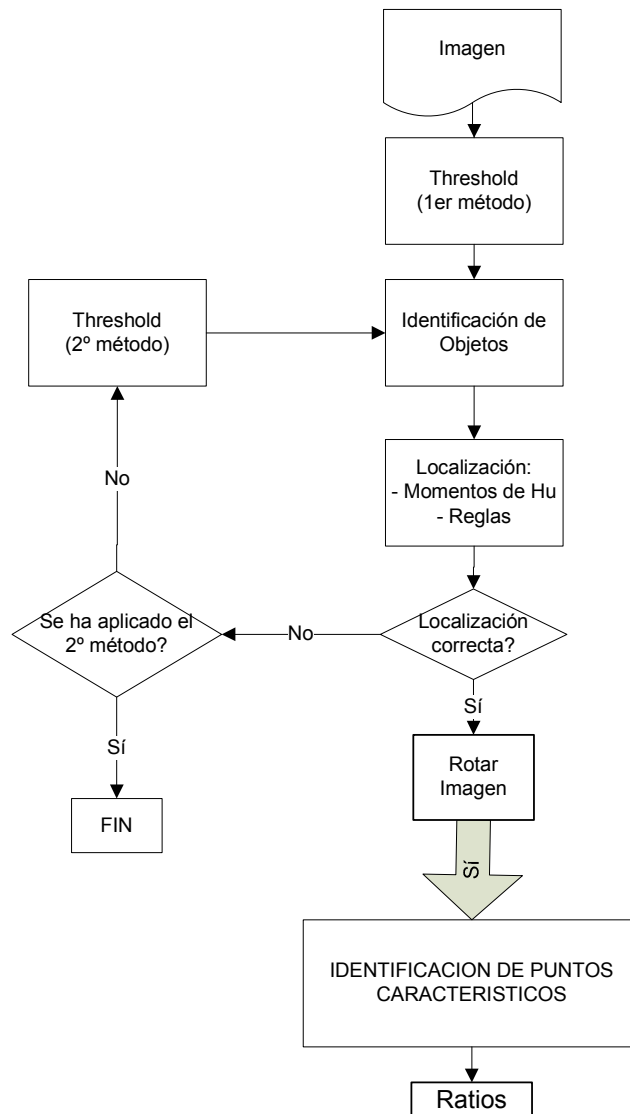
- Imagen rotada en formato JPEG.
- Fichero con la máscara de puntos de cada imagen (en formato .mat)
- Fichero con los ratios calculados (en formato .mat).

### **3.2.3.1 Localización de los ojos**

Para resolver la problemática de encontrar los ojos en una imagen, se divide en tres etapas[28]:

- 1) Primero se binariza la imagen de la cara de tal forma que queden plasmados en la nueva imagen los rasgos más característicos del rostro (boca, nariz, ojos, pelo)
- 2) Se identifican los objetos (ojos, nariz y boca) de la etapa anterior
- 3) Identificados los objetos, se calculan los momentos de Hu y se aplican ciertas reglas.

La arquitectura de este módulo es la siguiente (Figura 19):



**Figura 19: Localización de Ojos**

### 3.2.3.1.1 Binarización (*Thresholding*)

El objetivo de binarizar la imagen es encontrar los rasgos característicos que identifican una cara, como son: los ojos, la nariz, la boca,... El resultado es una imagen en blanco y negro que deja al descubierto algunos rasgos de la cara.

Se proponen dos métodos: El método 1 es algo más rápido ya que no requiere de excesivos cálculos; pero en algunas ocasiones no nos sirve para llegar a nuestro objetivo, localizar los ojos.

Debido a que el método 1 no cubre el 100% de nuestras necesidades, se implementó el segundo método con el fin de subsanar las deficiencias del primero. El paso de un método a otro se refleja en la Figura 19. Si con el primer



método de binalización no es posible encontrar los ojos, se binariza la imagen con el segundo método.

Tanto el primer método como el segundo, se basan en umbralizar la imagen a partir de un cierto valor umbral. El primer método toma como valor umbral la mitad del valor que más se repite en el histograma. El segundo método toma como valor umbral un valor entre dos máximos del histograma. En relación a la rapidez, el primer método encuentra el valor umbral algo más rápido que el segundo. Los resultados de ejecución son:

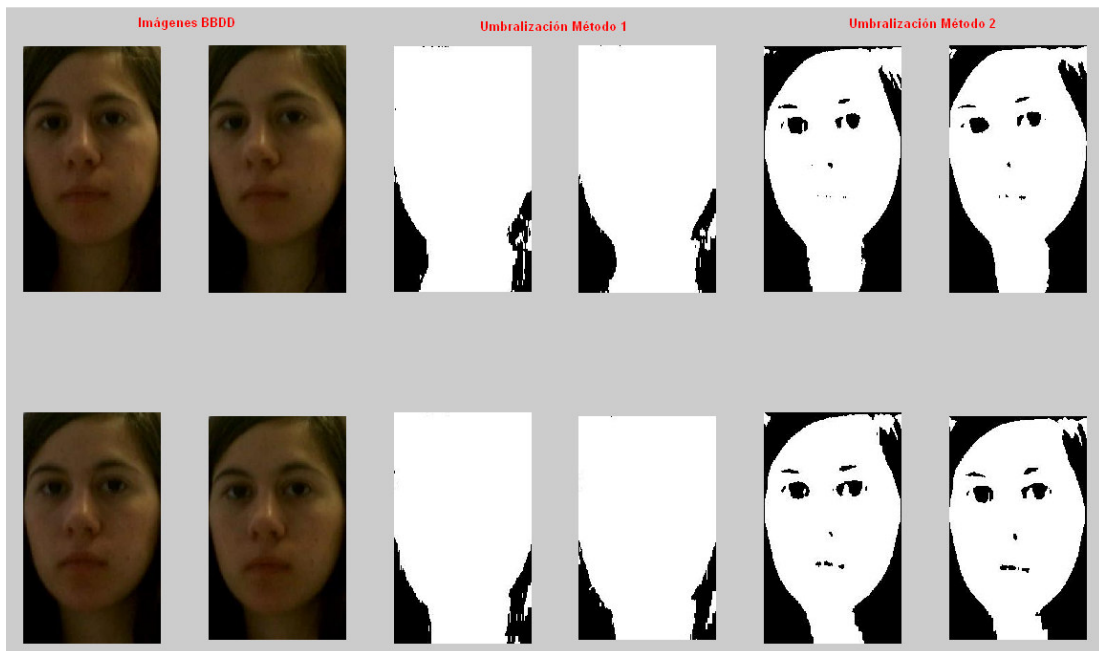
El primer método tarda una media de 0.7530s en encontrar el valor umbral

El segundo método tarda una media de 0.8770s en encontrar el valor umbral.

Los resultados se han realizado sobre 933 imágenes que corresponden con el total de imágenes de la BBDD eliminando aquellas imágenes que no superan el primer filtrado.

De las 933 imágenes si únicamente usamos el primer método se encuentran los ojos en 454. Si empleamos a continuación el segundo método aumentamos el número de imágenes dónde encontramos los ojos en 137.

A continuación se muestra la necesidad de emplear el 2º Método en caso de que el primero no sea capaz de encontrar los ojos.



**Figura 20:** A la izquierda las imágenes de la BBDD. En el centro el resultado de la umbralización por el primér método. A la derecha, las mismas imágenes pero esta vez umbralizándolas por el segundo método.

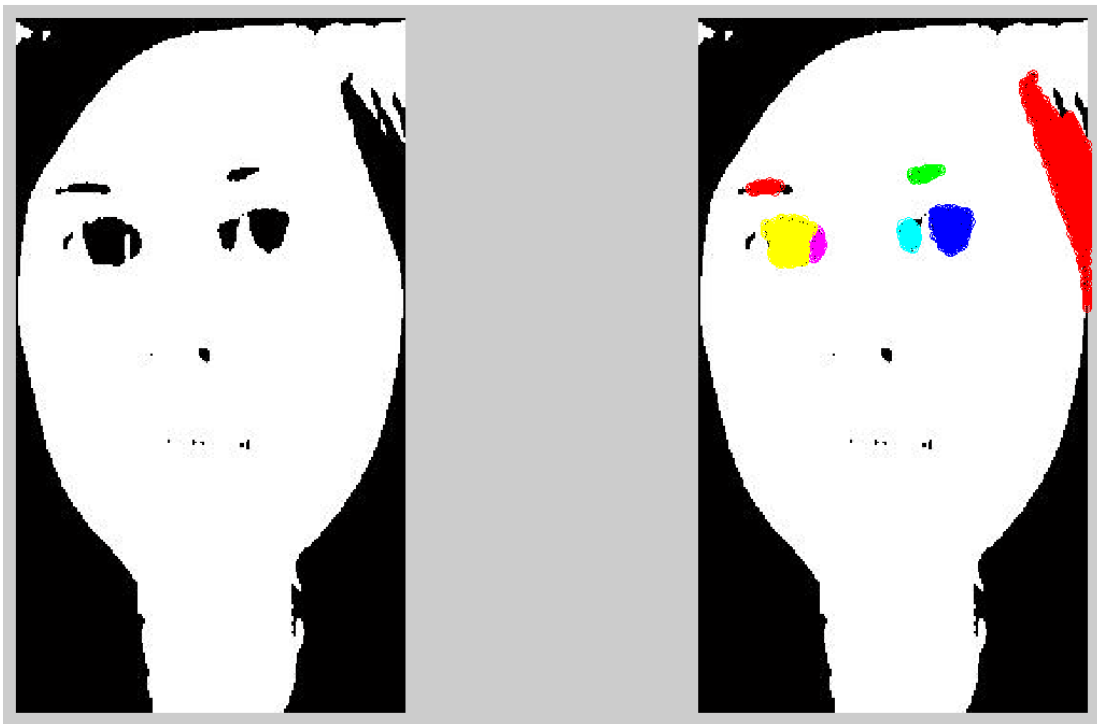
### 3.2.3.1.2 Identificación de los objetos

Esta etapa trata de etiquetar cada uno de los distintos objetos (manchas negras) que aparecen tras obtener la imagen binaria de la etapa anterior.

Para etiquetar los distintos objetos, se ha empleado el algoritmo 8-conectividad[47] cuyo propósito es separar los objetos de la escena.

Este algoritmo permite segmentar la imagen binaria en los distintos objetos que aparecen en ella. La segmentación aplicada elimina de la imagen aquellas zonas negras lo suficientemente pequeñas o lo suficientemente grandes como para no ser los ojos.

Cada uno de los objetos encontrados se identifica mediante un identificador único. En la siguiente imagen se muestra el resultado de la identificación de los objetos para una imagen binaria procedente de la etapa de segmentación.



**Figura 21:** A la izquierda la imagen resultante de la etapa de umbralización. A la derecha se muestra los distintos objetos identificados como candidatos a ser los ojos. Cada color representa un objeto distinto.

### 3.2.3.1.3 Localización de los objetos

Para poder decir dónde se encuentran localizados los ojos es necesario caracterizar cada objeto encontrado en la etapa anterior. Para ello se emplean los momentos de Hu[29][30][31].

Gracias a la información que aporta estos momentos y a la aplicación de ciertas reglas se puede detectar los ojos en una imagen.

### **3.2.3.1.3.1 Momentos de Hu**

¿Qué se entiende por Momento?

El concepto de momento en matemáticas evoluciona del concepto de momento en física. En física, el momento de la fuerza, es una cantidad que representa la magnitud resultante del producto del valor de una fuerza por su distancia a un punto de referencia. Así, en matemáticas se podría definir un momento como la magnitud que representa una característica del objeto de estudio.

Aplicado a una imagen, describen el contenido de una imagen o su distribución en lo que concierne a sus ejes.

Hu describió[40] un conjunto de 6 momentos invariantes a la rotación, la escala y la traslación. Estos momentos son aplicados a imágenes binarias para caracterizar sus formas. Al ser invariantes a la rotación y la escala, se ha aplicado para encontrar los ojos en la imagen binaria obtenida según se indica en el (apartado 4.2.2).

En [29], los momentos de Hu se aplican para identificar la palma de la mano.

### **3.2.3.1.3.2 Reglas de Localización**

Una vez identificados los objetos y calculados los momentos de Hu, se comparan los objetos por los momentos que lo definen. Los objetos que definen a los ojos suelen ser objetos simétricos entre sí con lo que al compararlos entre sí estos suelen presentar una diferencia de momentos muy pequeña.

Los objetos se empiezan a comparar por orden de aparición en la imagen. El orden de aparición de los objetos es idéntico al orden de las palabras en un texto, es decir el primer objeto será el que este situado lo más arriba y a la izquierda posible y el último el que esté situado lo más a la derecha y abajo posible.

Además de que los ojos son simétricos, otro objeto que suele encontrarse antes y que también suele ser simétrico son las cejas.

Para poder distinguir en si los objetos identificados son los ojos o las cejas es necesarios aplicar ciertas reglas que nos permita asegurar que realmente son los ojos.

Las reglas que se han aplicado son de dos tipos:

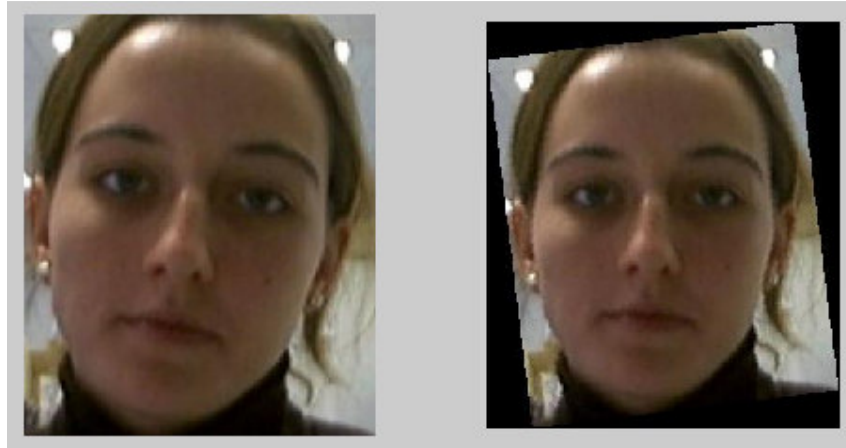
- reglas de distribución espacial definen como de separados tienen que estar estos los objetos para ser considerados ojos.
- Las reglas de localización definen la zona más probable donde encontrar los ojos.

Estas reglas vienen definidas con más detalle en el siguiente capítulo (ver sección 4.2.3.2.2).

#### 3.2.3.1.4 Rotación de la imagen

Una vez localizados los ojos se procede a rotar la imagen de tal manera que los ojos queden alineados en una misma horizontal.

En la siguiente figura se muestra el proceso de rotación aplicado a una imagen de la BBDD.



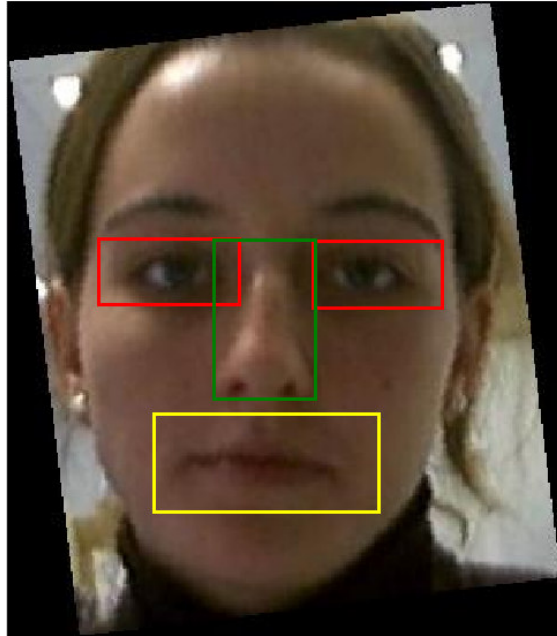
**Figura 22 : A la izquierda la imagen original, a la derecha la misma imagen después de ser rotada.**

La etapa de rotación es básica para tener normalizadas las imágenes a un mismo patrón. Con esto se consigue olvidarse del ángulo de inclinación de la cara que lo único que nos aporta la complejidad en los cálculos a la hora de seguir estudiando la imagen.

#### 3.2.3.2 Identificación de puntos característicos

Con la imagen rotada y tomando como referencia los objetos considerados como los ojos, se procede a analizar la cara con el fin de extraer los puntos que la caracterizan. El proceso de identificación de los puntos, parte siempre de un punto de referencia para estudiar una cierta región de interés. El resultado de este estudio permite extraer esos puntos de referencia para cada región de interés:

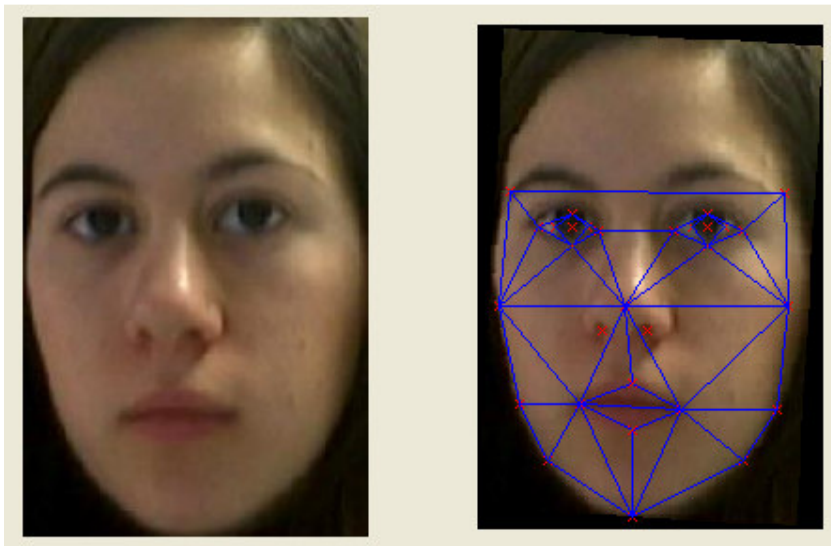
- Ojos.
- Nariz.
- Boca.



**Figura 23: Regiones de interés de una cara**

Los algoritmos desarrollados permiten extraer un total de 30 puntos por cara pero no todos los puntos encontrados se emplean para el modelado y posterior reconocimiento del individuo.

En la siguiente figura se puede ver la nube de puntos extraída para una determinada cada de la BBDD.



**Figura 24: A la izquierda la imagen original, a la derecha la nube de puntos extraída de la misma imagen.**

### 3.2.3.3 Ratios – Distancias relativas entre puntos

La definición de los ratios se ha basado en el estudio [13] que emplea un total de 14 ratios para reconocer a una persona. Para este PFC se han definido un total de 10 ratios por imagen. Estos diez ratios corresponden con las relaciones entre las características faciales (ojos, nariz y boca). No se han empleado los otros cuatro ratios debido a que la probabilidad de que otros puntos estén correctamente ubicados con los algoritmos desarrollados, es baja.

Los ratios seleccionados para modelar cada individuo son:

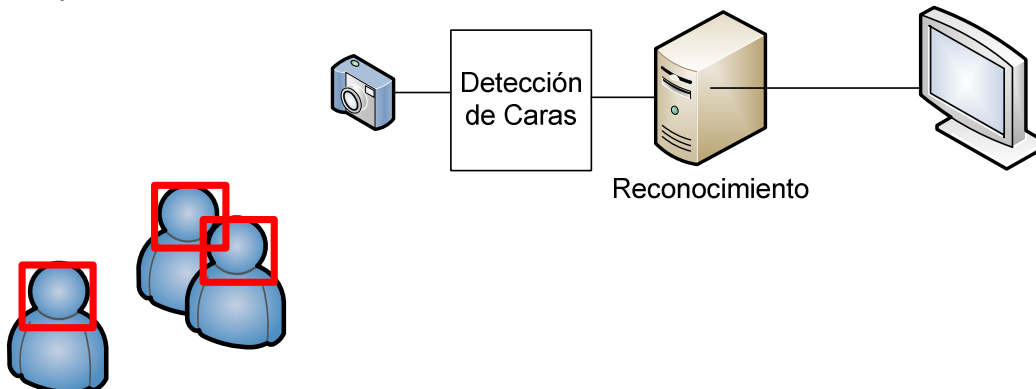
- ‘Separación entre pupilas’ entre ‘Ancho de la boca’.
- ‘Separación entre pupilas’ entre ‘Distancia del ojo izquierdo a la boca’.
- ‘Separación entre pupilas’ entre ‘Distancia del ojo derecho a la boca’.
- ‘Separación entre pupilas’ entre ‘Altura de la nariz’.
- ‘Ancho de la boca’ entre ‘Distancia del ojo izquierdo a la boca’.
- ‘Ancho de la boca’ entre ‘Distancia del ojo derecho a la boca’.
- ‘Ancho de la boca’ entre ‘Altura de la nariz’.
- ‘Distancia del ojo izquierdo a la boca’ entre ‘Distancia del ojo derecho a la boca’.
- ‘Distancia del ojo izquierdo a la boca’ entre ‘Altura de la nariz’.
- ‘Distancia del ojo derecho a la boca’ entre ‘Altura de la nariz’.

### 3.2.4 Aplicación de Reconocimiento

La aplicación de reconocimiento, recibe imágenes preprocesadas capturadas directamente de una fuente de video. El preprocesado consiste en encontrar todas las posibles caras en el fotograma.

Para éste PFC, al contar con imágenes con una sola cara, el preprocesado consiste en eliminar de la imagen todo lo que no sea la cara. Se ha empleado un detector de color carne para localizar la cara dentro del fotograma.

La arquitectura de esta aplicación es igual a la aplicación de modelado salvo que incorpora el módulo de reconocimiento.



### **3.2.4.1 Reconocimiento**

Este módulo emplea la técnica de mezcla de Gaussianas (GMM) para identificar a las personas localizadas dentro de la escena.

Se ha empleado esta técnica por predominar en la mayoría de los sistemas de reconocimiento.

Se describió por primera vez en una serie de artículos publicados por Reynolds y Rose a partir de 1990[41][42]. Desde entonces, han aparecido en múltiples publicaciones de congresos[43][44][45][46] y numerosos sistemas basados en GMMs han participado en las competiciones anuales NIST (*National Institute of Standards and Technology*)





## 4 Desarrollo

---

### 4.1 Introducción

Este capítulo se centra en cómo se ha llevado a cabo la implementación de los diversos subsistemas que se han detallado en el capítulo anterior.

### 4.2 Desarrollo de la aplicación de Identificación y Modelado

La aplicación de modelado trata de extraer de la imagen de la cara una serie de puntos que la caracterizan. Para poder llevar a cabo esta extracción la aplicación de ha dividido la aplicación en una serie de módulos:

- Binarización (*Thresholding*)
- Identificación de Objetos
- Rotación de la imagen
- Identificación de Puntos Característicos
- Cálculo de Ratios
- Modelado

La etapa “Binarización” convierte la imagen en color en una imagen en blanco y negro.

La etapa “Identificación de Objetos” analiza e identifica en cada uno de los objetos resultantes de la imagen en blanco y negro, aquellos que corresponden con los ojos, la nariz y la boca.

La etapa de “Rotación” permite disponer de las imágenes simétricas respecto al eje vertical. El objetivo es conseguir que las imágenes sigan un mismo patrón con el fin de que la rotación de la cara afecte lo menos posible al modelado y reconocimiento del individuo.

La etapa “Identificación de Puntos Característicos” extrae aquellos puntos que identifican de forma unívoca al usuario.

La etapa “Ratios” calcula las relaciones entre los distintos puntos característicos encontrados.

Por último, la etapa de modelado, permite estimar, a partir de los ratios calculados, el modelo que más se parezca al individuo.

La localización de los ojos es la etapa principal de la aplicación. Si esta localización falla, la imagen se desecha por no tener información suficiente para continuar con el procesado de la cara.

Para localizar los ojos, primero se convierte la imagen origen en una imagen en blanco y negro. Con la imagen en blanco y negro, se analizan las manchas que aparecen en esta nueva imagen (objetos de color negro) y aplicando los momentos de Hu y unas ciertas reglas, es posible localizar los ojos dentro de la imagen.

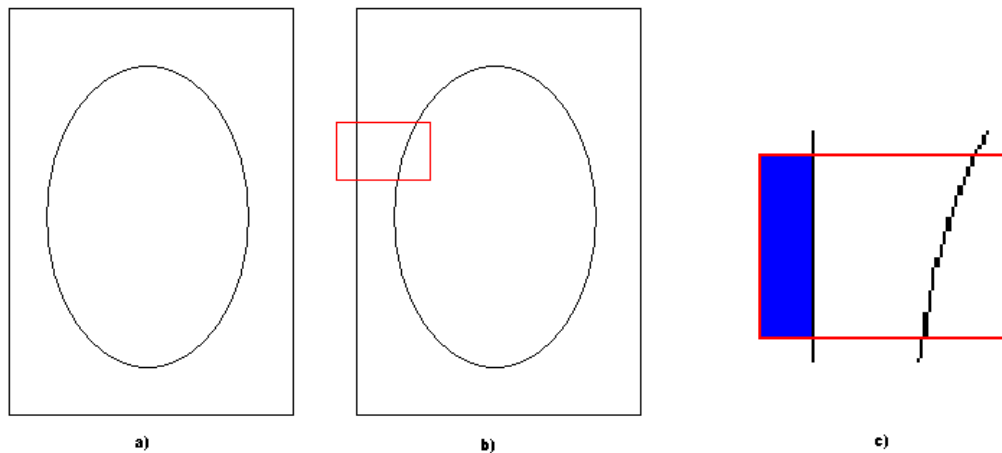
En la Figura 19 se puede ver el esquema seguido para desarrollar este módulo.

Cada uno de estos módulos se trata con más detalle a continuación.

#### 4.2.1 Verificación de los límites de las imágenes

Hay que distinguir entre dos tipos de imágenes: la imagen origen (original) será aquella de la que extraigamos la información; y la imagen destino, que se corresponde con los datos extraídos de la imagen origen.

Debido a que el sistema desarrollado no es perfecto, y a que en algunas ocasiones, intentamos acceder a zonas no definidas dentro de la imagen, es necesaria la incorporación de un módulo que permita verificar en todo momento si los índices, tanto en filas como en columnas, de las imágenes que se van a analizar, realmente existe. Este chequeo siempre se realiza cuando a partir de la imagen origen extraemos una porción de ésta. En la mayoría de los casos la porción que se extrae es de unas dimensiones fijas lo que provoca que a veces estemos fuera de los límites de la imagen origen (ver Figura 25).



**Figura 25 :** a) La imagen origen. b) El rectángulo rojo representa la imagen destino que se obtendrá al extraer dicha porción de la imagen origen. c) La zona sombreada refleja que la imagen destino se encuentra fuera de los límites establecidos de la imagen origen.

Este módulo, debe conocer de antemano las dimensiones de la imagen origen, para comprobar que los índices a los que se intentan acceder existen; es decir, no se quiere extraer nada fuera de la imagen.

Siempre que se intenta acceder a una posición de la imagen origen para obtener una nueva imagen, se deberá llamar a este método y se deberá comprobar si se recibe algún error.

Entendiendo la imagen como una matriz de datos bidimensional, los datos de las filas y las columnas, empiezan a contar desde la esquina superior izquierda, según se ilustra en la Figura 26. Los valores de las filas corresponden con la coordenada Y, y los valores de las columnas con la X. Para chequear que estamos dentro de los límites definidos en la imagen origen es necesario comprobar lo siguiente:

- La esquina superior izquierda no tenga ningún valor en la de coordenada,  $x$  o  $y$  negativo, en cuyo caso, la coordenada o coordenadas, que tengan el valor negativo su valor será 1.
- La coordenada  $X$  de la esquina superior derecha no sea inferior que su homólogo (esquina superior izquierda). En este caso se devuelve un error.
- La coordenada  $X$  de la esquina superior derecha no sea superior al ancho de la imagen origen, en cuyo caso, el valor de esta coordenada será el ancho de la imagen origen.
- La coordenada  $Y$  de la esquina inferior derecha no sea inferior que su homólogo (esquina inferior izquierda). En este caso se devuelve un error.
- La coordenada  $Y$  de la esquina inferior derecha no sea superior al alto de la imagen origen, en cuyo caso, el valor de esta coordenada será el alto de la imagen origen.



**Figura 26: Origen de Coordenadas de las imágenes**

## **4.2.2 Binarización (*Thresholding*)**

El objetivo de transformar la imagen en otra en blanco y negro es encontrar los rasgos que caracterizan una cara, como son: los ojos, la nariz y la boca.

Se han implementado dos métodos, el primer método umbraliza la imagen a un valor umbral que corresponde con la mitad del valor máximo del histograma y el segundo método calcula el valor umbral a partir de los dos primeros máximos del histograma.

El motivo de implementar dos algoritmos se debe a que el primer método no es capaz de realizar una binarización perfecta de todas las imágenes de la BBDD porque la iluminación de cada cara varía de una imagen a otra.

### **4.2.2.1 Método 1**

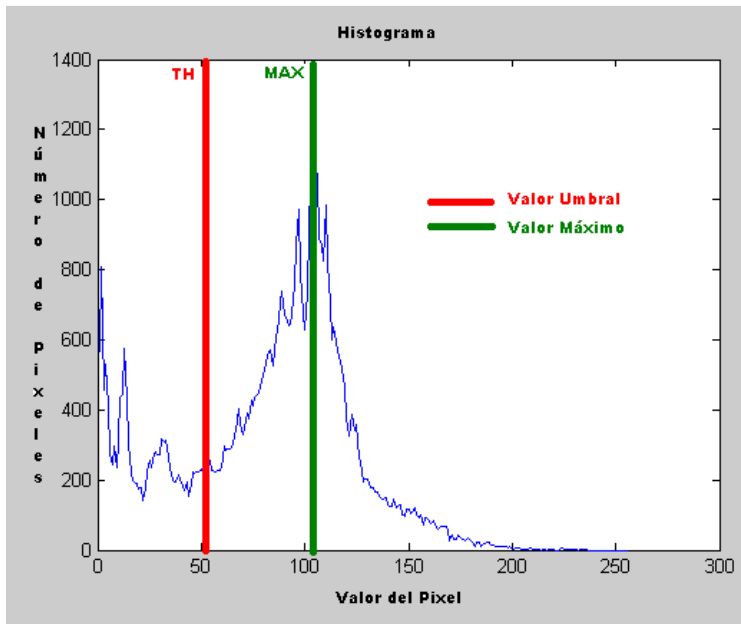
Este método se aplica para conseguir una imagen binaria (blanco y negro) a partir de un cierto valor umbral.

El primer paso, consiste en calcular la imagen de luminancia a partir de la imagen a color como sigue:

$$Y = 0,299R + 0,587G + 0,114B$$

siendo R, G y B las componentes de color de la imagen

Con la imagen de luminancia, Y, se calcula su histograma correspondiente y se toma como umbral, para binarizar, la mitad del valor del píxel que más se repite en la imagen; es decir, se calcula el valor máximo del histograma (corresponde con el valor del píxel que más se repite), y el umbral es la mitad del valor del píxel. En la Figura 27 se muestra tanto el histograma de una imagen, como el valor máximo el valor umbral utilizado para binarizar.



**Figura 27: Histograma de la imagen. En verde, el valor del píxel que más se repite; en rojo, el valor umbral**

La imagen así obtenida, es una imagen monocromática, binaria (por contener sólo valores 0 y 1), como la mostrada en la Figura 28

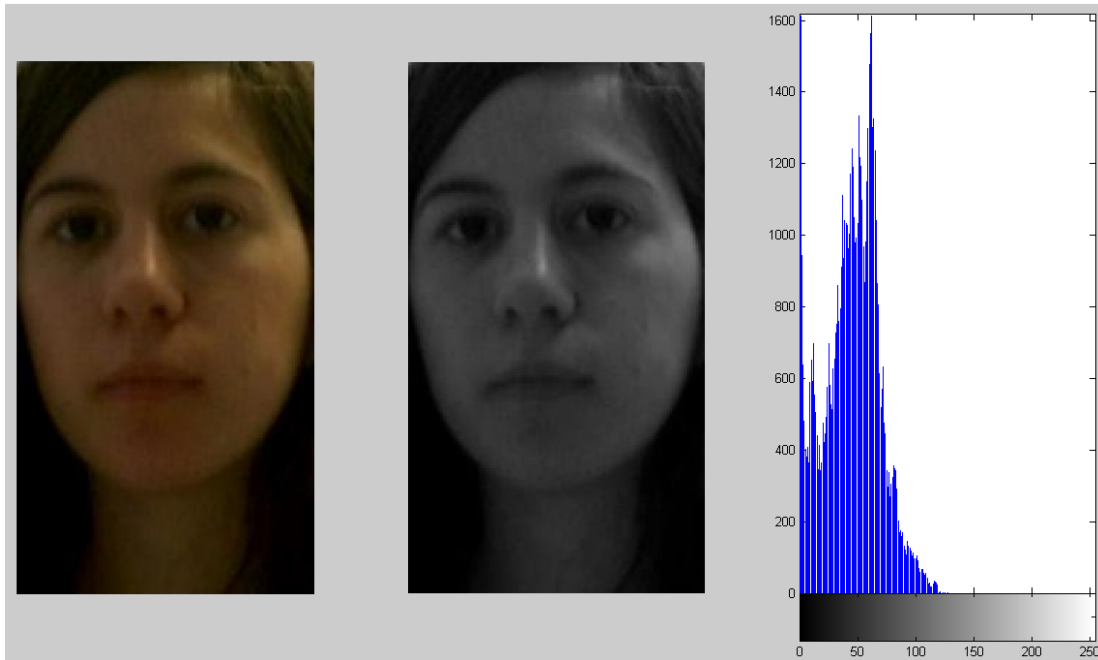


**Figura 28:** A la derecha se muestra la imagen binaria obtenida a partir del primer método de umbralización sobre la imagen de la izquierda.

Con la imagen binaria, se procede a localizar los ojos (ver sección 4.2.3.2.2), si no se encontrasen, se aplicaría el segundo método para binarizar la imagen.

#### **4.2.2.2 Método 2**

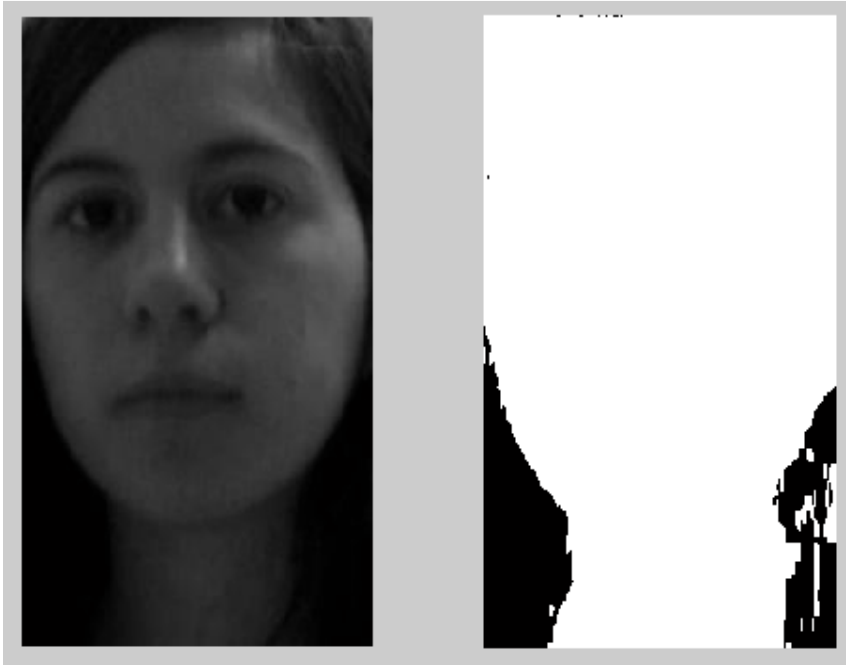
El primer método es un método rápido para calcular el valor umbral de una imagen (La mitad del valor máximo del histograma). En muchas ocasiones, el valor umbral así calculado no permite identificar los ojos debido a que parte del histograma de la imagen de luminancia sin aplicar un prefiltrado para eliminar el ruido existente en la imagen. En la Figura 29 se muestra el histograma de una imagen de la BBDD donde el primer método no permite la localización de los ojos.



**Figura 29 : A la derecha una imagen de la BBDD. En el centro, la imagen de luminancia. A la derecha, el histograma de la imagen de luminancia**

Se puede observar en la Figura 29 que el valor umbral calculado con el primer método es muy bajo, en concreto el valor máximo del histograma está en el índice 1 y por lo tanto el valor umbral es el entero más próximo de la mitad de este índice; es decir, 1.

Si se umbraliza la imagen con este valor umbral, el resultado es tal que no permite identificar los ojos en la imagen y por lo tanto la imagen sería descartada (ver Figura 30).

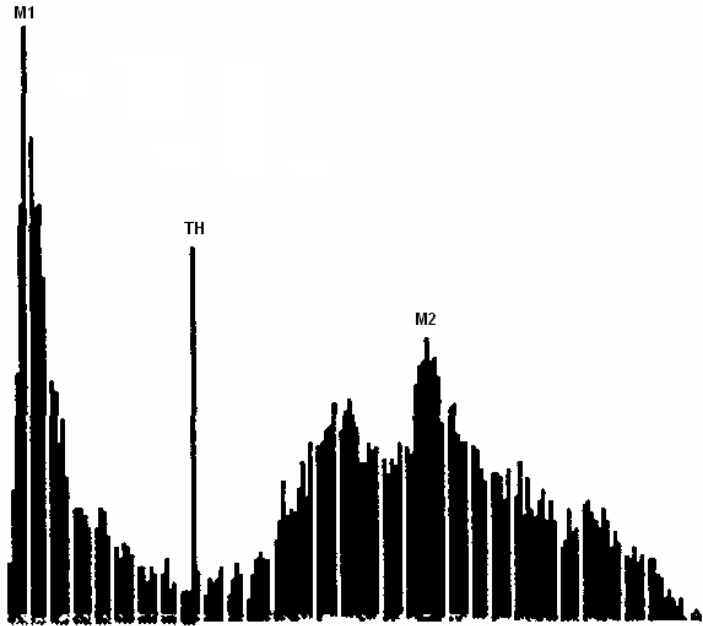


**Figura 30:** A la derecha, la imagen umbralizada por el primer método.

Por este motivo se implementó un segundo método que realizara un cálculo más exhaustivo del valor umbral.

Este segundo método se emplea únicamente si con el primero no fuese posible localizar los ojos.

En la Figura 31 se muestra el histograma de una imagen (más o menos, el los histogramas de una cara siguen la misma envolvente) correspondiente a una cara, donde el área del primer máximo, representa las regiones oscuras de la cara como el pelo, ojos, cejas y boca. La técnica de thresholding propuesta en [21], propone escoger un valor de intensidad entre las dos primeras áreas máximas.

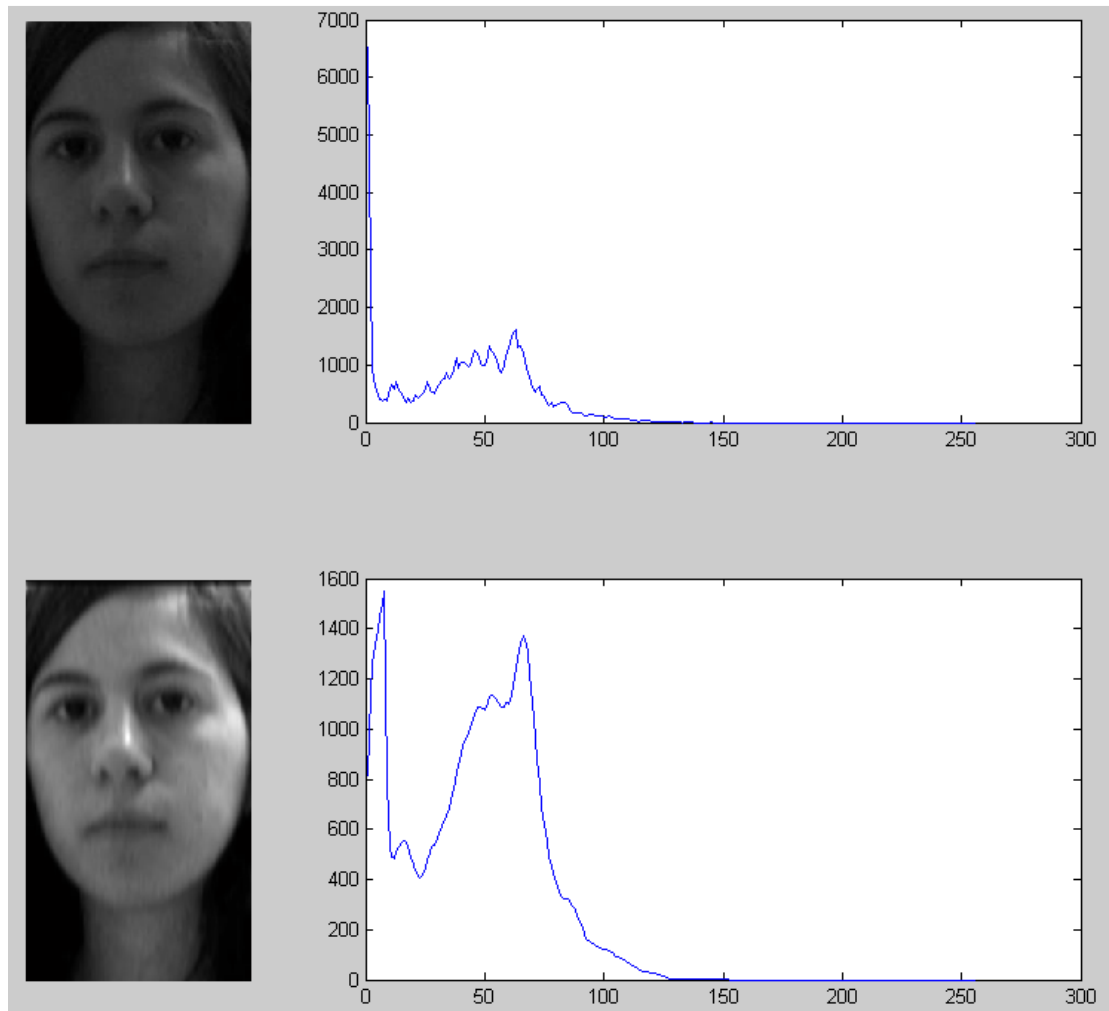


**Figura 31: Histograma de una imagen de la cara. M1 y M2 son los dos máximos del histograma; y TH es el valor de la intensidad escogido para la umbralización o *Thresholding* (Imagen obtenida de [28]).**

Es necesario eliminar el ruido de la imagen antes de seleccionar el umbral (TH) del histograma. Para eliminar el ruido suavizamos el histograma, consiguiendo además, que resulte más fácil encontrar los dos máximos del histograma, tanto visualmente como computacionalmente. Un ejemplo de suavizado del histograma se muestra en la Figura 32, donde se ha aplicado un filtro de según la siguiente fórmula:

$$y(n) = \frac{1}{8}x(n) + \frac{1}{8}x(n-1) + \frac{1}{8}x(n-2) + \frac{1}{8}x(n-3) + \frac{1}{8}x(n-4) + \frac{1}{8}x(n-7)$$





**Figura 32:** Arriba a la izquierda, la imagen de luminancia de imagen original. Arriba a la derecha, el histograma de esta imagen. Abajo a la izquierda la imagen de luminancia suavizada. Abajo a la derecha, el histograma suavizado.

El promediado del histograma da lugar a que la imagen resultante del proceso de umbralización salga menos nítida y por lo tanto con menos ruido, con lo que cumplimos el objetivo de eliminar el ruido de la imagen. En la Figura 33 (se muestra el resultado de la umbralización de la imagen).

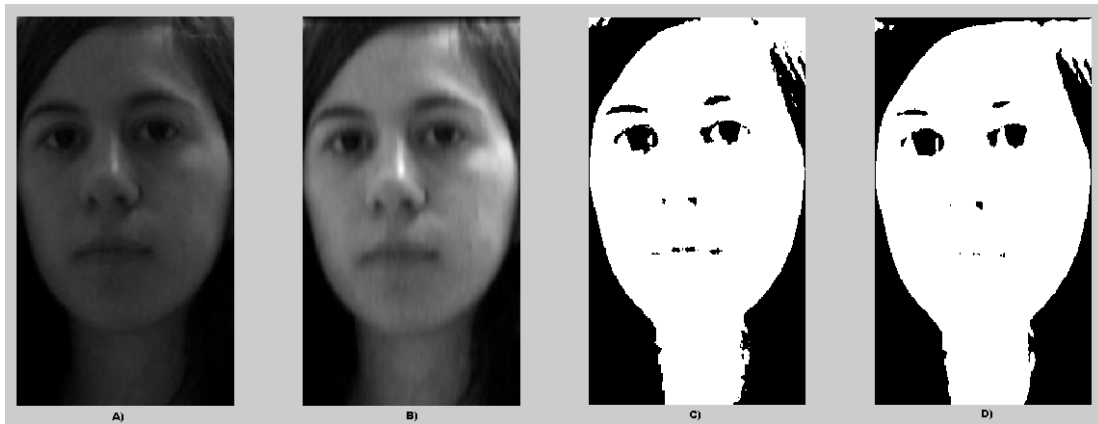


Figura 33: De izquierda a derecha, la imagen de luminancia original (A), la original suavizada (B), la imagen umbralizada sin suavizar (C) y la umbralizada suavizada (D).

Este método, se ha probado con 16 caras de la BBDD de Test, obteniéndose el siguiente resultado:

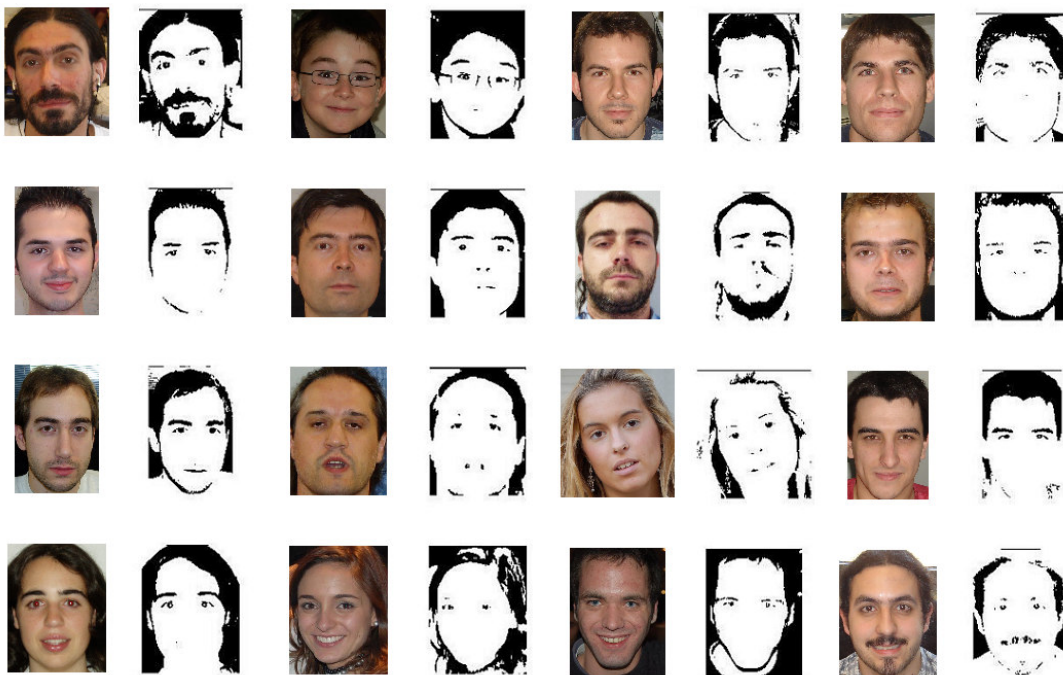


Figura 34: Resultado del *thresholding* de 16 imágenes de la BBDD de Test.

#### 4.2.3 Identificación de Objetos

Obtenida la imagen binaria, localizar los ojos no es una tarea fácil. A simple vista a un ser humano le puede parece sencillo localizar los ojos, pero la dificultad radica en cómo transferir ese conocimiento a una máquina.

Para ayudarnos a resolver el problema, el primer paso consiste en etiquetar cada una de las manchas de color negro (objetos) que aparecen en la imagen binaria, Una vez terminado este paso, se calculan los momentos de Hu[40], y comparando el resultado de los siete momentos, invariantes entre sí, y aplicando ciertas reglas, es posible localizar los ojos, con una baja probabilidad de error.

Para etiquetar cada uno de los rasgos que aparecen como consecuencia del proceso de umbralización o binarización, se ha empleado el algoritmo 8-conectividad[47] cuyo propósito es separar los objetos:

- del fondo
- unos de otros (e.g. separar un ojo de una ceja, ...)

Dentro de los distintos objetos que aparecen en la imagen binaria, hay muchos que directamente son rechazados de ser analizados, por su tamaño. En concreto, se realiza un filtrado previo de aquellos objetos que son o lo suficientemente pequeños (la mayoría ruido producido por una iluminación irregular de la cara) o lo suficientemente grandes (como puede ser el pelo) como para ser posibles candidatos a ser los ojos.

El concepto “pequeño” o “grande” viene determinado por el área del objeto en cuestión por lo tanto, todos los objetos que aparecen en la imagen binaria son identificados para después tomar la decisión de si continuar con ellos para el análisis o no.

Un objeto se considera pequeño si el área no supera el 0.05% del área total de la imagen. Por el contrario, se considera grande si el área supera el 2,5% del área total de la imagen.

#### 4.2.3.1 Algoritmo de 8-conectividad

El algoritmo 8-conectividad trata de etiquetar cada región de píxeles contiguos con un valor diferente. El problema viene, en saber cuántos objetos hay en una imagen binaria. La respuesta, depende de lo que se considere adyacente a un píxel (conectividad); y de lo que se considere interior o exterior; entre otros aspectos.

El concepto de conectividad aplicado en este contexto, es la proximidad espacial entre píxeles de la imagen binaria. Existen, al menos, tres posibles formas de considerar regiones conexas:

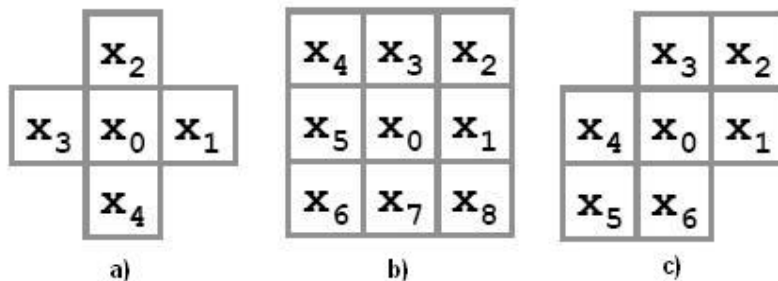
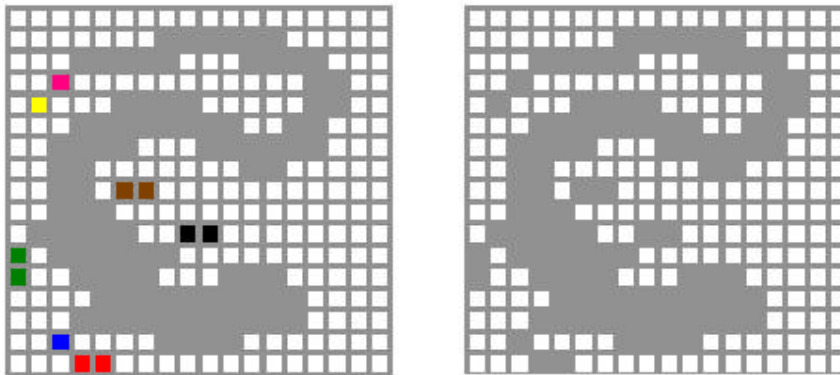




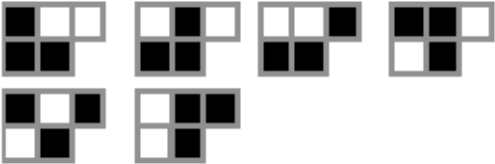
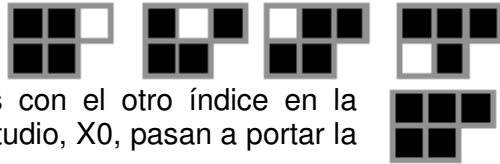
Figura 35: Tres tipo de conectividad: a) 4-conectividad, b) 8-conectividad y c) 6-conectividad

Según lo que se considere por región conexas, se suele usar uno u otro algoritmo. En la Figura 36 se muestra el resultado de la conectividad aplicando 4-conectividad y 8-conectividad. .



**Figura 36: Resultado del algoritmo de conectividad. A la izquierda 4-conectividad. Cada color representa un objeto distinto. A la derecha 8-conectividad.**

Para etiquetar los objetos de la imagen binaria, se ha implementado un algoritmo secuencial basado en la idea de 8-conectividad, definiéndose para ello las siguientes reglas para etiquetar:

- Si los vecinos X2, X3, X4, X5 (ver la **Figura 35b**) no tienen etiqueta, al píxel X0 se le asigna una nueva etiqueta. 
- Si los vecinos X2, X3, X4 o X5, tienen una etiqueta, al píxel X0 se le asigna esa etiqueta. 
- Si hay dos o más vecinos etiquetados, existen dos subcasos:
  - Tengan la misma etiqueta, y por lo tanto al píxel X0 se le asigna esa etiqueta 
  - Tengan distintas etiquetas, entonces se busca la etiqueta con índice menor y todos los píxeles etiquetados con el otro índice en la imagen y el píxel objeto de estudio, X0, pasan a portar la etiqueta seleccionada. 

#### 4.2.3.2 Localización de los objetos

Una vez separados los objetos de la escena, la siguiente etapa trata de identificar de todos los objetos encontrados, cuales corresponden con los ojos. Para llegar a tal fin se ha empleado los momentos de Hu por ser invariantes ante la rotación, la escala y la traslación.

Como ya se comentó en la sección 3.2.3.1.3.1, estos momentos se aplican a imágenes binarias para caracterizar sus formas.

Se han definido un total de 7 momentos para caracterizar lo mejor posible a los objetos encontrados dentro de la imagen.

Dada la simetría casi perfecta de una cara frontal, los objetos (de la imagen binaria) correspondientes a los ojos, son objetos simétricos el uno del otro; de ahí que se haya empleado esta técnica y no otra.

El ruido de la imagen, la iluminación irregular, la no perfecta simetría de la cara y otros factores hacen que además de los momentos de Hu, haya que emplear unas ciertas reglas para localizar los ojos dentro de la imagen.

#### 4.2.3.2.1 Momentos de Hu

##### 4.2.3.2.1.1 Introducción

En términos de las integrales de Riemann [48], un momento 2-dimensional de orden  $(p+q)$ -ésimo de la función de distribución  $\rho(x, y)$ , se define como:

$$M_{p,q} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q \rho(x, y) dx dy, \quad p, q = 0, 1, 2, 3, \dots$$

Algunos momentos característicos son:

- $M_{0,0}$  es el área del objeto
- $(M_{0,1}, M_{1,0})$  es el centro de masas del objeto

Los momentos centrales son momentos definidos por:

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q \rho(x, y) d(x - \bar{x}) d(y - \bar{y})$$

Donde

$$\bar{x} = M_{1,0} / M_{0,0}, \quad \bar{y} = M_{0,1} / M_{0,0}$$

Estos momentos presentan la característica de ser invariantes frente a una traslación de las coordenadas del objeto.

Como estos momentos son invariantes a la traslación, se puede expresar fácilmente en términos de los momentos ordinarios  $(M_{p,q})$ . Así, los cuatro primeros momentos quedarían:

$$\mu_{00} = M_{00} \equiv \mu$$

$$\mu_{10} = \mu_{01} = 0$$

$$\mu_{20} = M_{20} - \mu \bar{x}^2 \quad \mu_{11} = M_{11} - \mu \bar{x} \bar{y} \quad \mu_{02} = M_{02} - \mu \bar{y}^2$$

$$\mu_{30} = M_{30} - 3M_{20} \bar{x} + 2\mu \bar{x}^3 \quad \mu_{21} = M_{21} - M_{20} \bar{y} - 2M_{11} \bar{x} + 2\mu \bar{x}^2 \bar{y}$$

$$\mu_{12} = M_{12} - M_{20} \bar{x} - 2M_{11} \bar{y} + 2\mu \bar{x} \bar{y}^2 \quad \mu_{03} = M_{03} - 3M_{02} \bar{y} + 2\mu \bar{y}^3$$

Las definiciones físicas del momento central se organizan según la siguiente tabla:

$\mu_{20}$  : Varianza de la Abscisa

$\mu_{02}$  : Varianza de la Ordenada

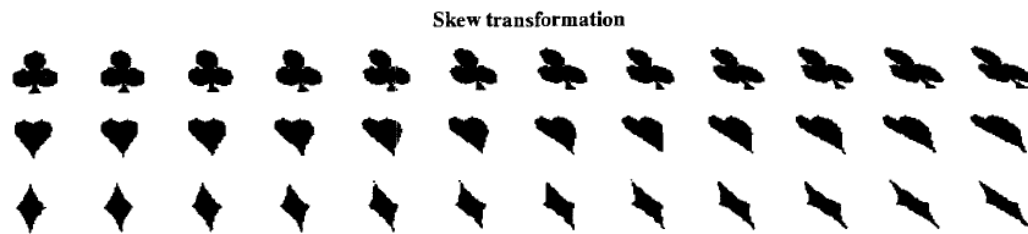
$\mu_{11}$  : Covarianza de la Abscisa y de la Ordenada

$\mu_{12}$  : Comparación en la Abscisa de la intensidad de distribución del lado izquierdo con la del lado derecho

$\mu_{21}$  : Comparación en la Ordenada de la intensidad de distribución del lado inferior con la del lado superior

$\mu_{30}$  : Intensidad de la deformación (*skew*) en la Abscisa (ver Figura 37)

$\mu_{03}$  : Intensidad de la deformación (*skew*) en la Ordenada



**Figura 37: Ejemplo de *Skew* en una imagen**

#### 4.2.3.2.1.2 Momentos de Hu

$$\begin{aligned}\phi_1 &= \mu_{20} + \mu_{02} \\ \phi_2 &= (\mu_{20} + \mu_{02})^2 + 4\mu_{11}^2 \\ \phi_3 &= (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2 \\ \phi_4 &= (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2 \\ \phi_5 &= (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] \\ &+ (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \\ \phi_6 &= (\mu_{20} - \mu_{02})[(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \\ &+ 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03})^2 \\ \phi_7 &= (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] \\ &+ (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2]\end{aligned}$$

El significado de cada momento de Hu, definidos según la ecuación anterior, se muestra en la siguiente tabla:

$\phi_1$	La suma de varianza horizontal y vertical, indica cuánto más distribuido esta el objeto, hacia los ejes horizontales y verticales.
$\phi_2$	El valor de covarianza de los ejes vertical y horizontal, cuando el valor de la varianza de los ejes vertical y horizontal es similar.
$\phi_3$	Valor que acentúa la inclinación según esté más inclinado a la derecha o a la izquierda
$\phi_4$	Valor que acentúa la inclinación según esté más inclinado arriba o abajo
$\phi_5, \phi_6$	Otros valores invariantes al tamaño, la rotación y la traslación
$\phi_7$	Momento que es invariante al <i>skew</i>

**Tabla 1: Definición de los Momentos de Hu**

Por ejemplo, en [31] se muestra un estudio sobre como de invariantes son estos momentos si se aplica a una imagen una transformación de tipo *skew* o una perspectiva. Los resultados obtenidos para la transformación tipo *skew*, se muestran en la Figura 38.

## Skew transformation

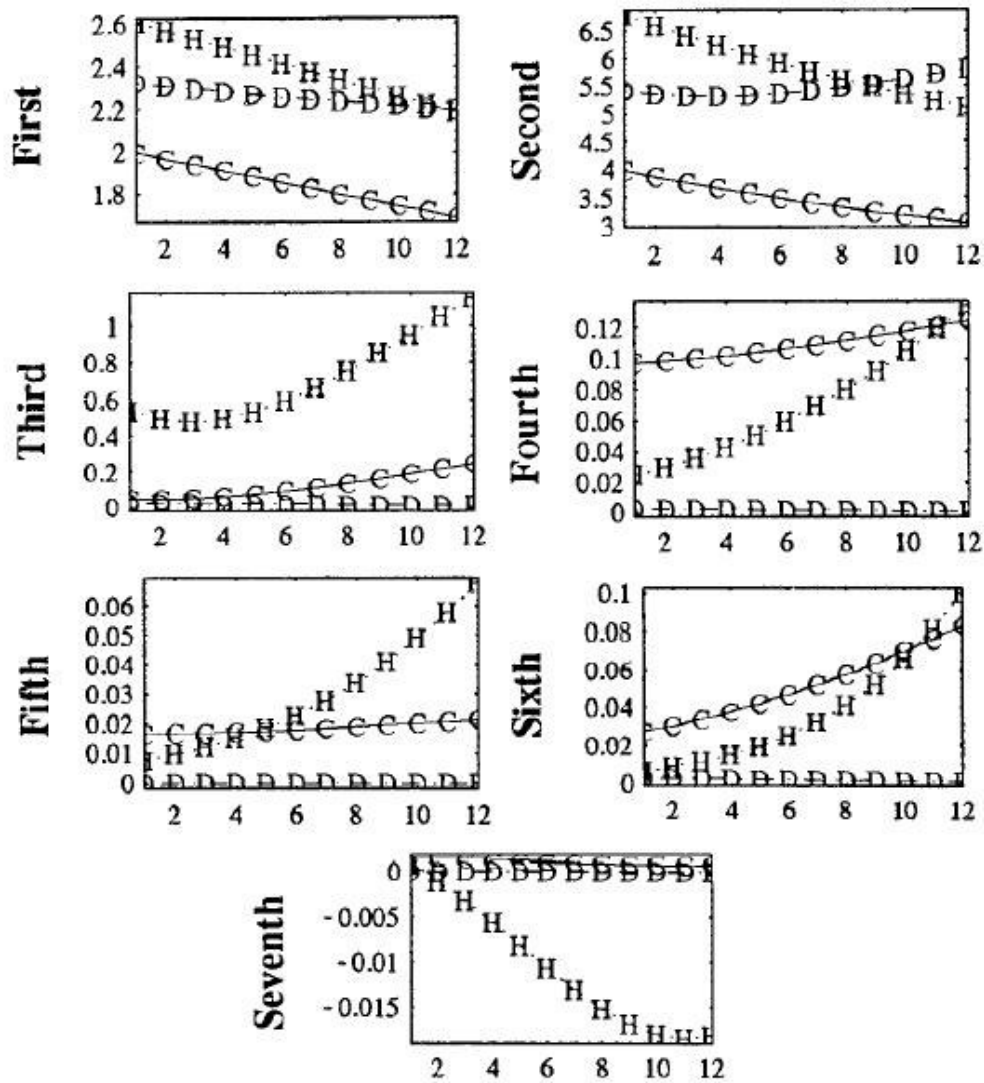


Figura 38: El conjunto de los siete momentos invariante de Hu aplicados sobre la imagen de la Figura 37. Cada gráfica contiene 3 etiquetas que representan cada palo de la baraja: el trébol (C), el corazón (H) y los diamantes (D) (Imagen obtenida de [31]).

### 4.2.3.2 Reglas de Localización de los ojos

Caracterizados los objetos con los siete momentos de Hu, hay que aplicar ciertas reglas para localizar los ojos. El motivo es que la imagen sigue conteniendo un ruido "intrínseco" al fotograma que no se puede eliminar. Este ruido lo compone:

- Iluminación irregular de la imagen → da lugar a que la imagen binaria no sea simétrica.
- No frontalidad de la cara → Cuando la cara está ligeramente inclinada o girada (imagen no simétrica) los "objetos" tienden a deformarse no siendo perfectamente simétricos unos de otros.

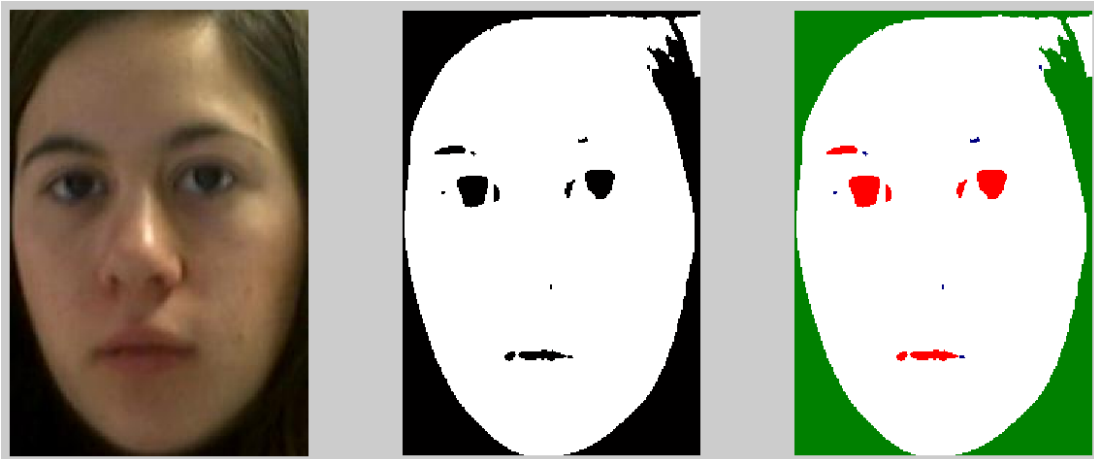


- Ruido de la imagen → Aún aplicando el suavizado del histograma (ver Figura 33) el ruido de la imagen no se consigue eliminar del todo. Esto da lugar a distintos objetos que se parecen mucho entre ellos.

La localización se divide en dos etapas: la primera etapa consiste en calcular la distancia euclídea entre los siete momentos de Hu; y la segunda trata de analizar estos datos para encontrar los ojos en la imagen.

Aplicando el algoritmo 8-conectividad descrito en el apartado 4.2.3, se extraen los objetos que aparecen en la imagen binaria, excluyendo aquellos lo suficientemente pequeños por considerarse ruido (se considera suficientemente pequeños aquellos cuya área no supera el 0.05% del área de la imagen original). También son descartados aquellos objetos cuyo tamaño es muy grande (aquellos cuya área supera el 2.5% del área de la imagen original); con esto, se consigue eliminar entre otras cosas, el pelo de la imagen que se está tratando.

En la siguiente imagen se muestra los objetos que han sido excluidos por su tamaño. En color verde se eliminan aquellos cuyo tamaño supera el permitido y en azul, los objetos que no superan la talla mínima exigida. El resto de objetos, color rojo, representan el conjunto de candidatos a ser los ojos.



**Figura 39:** A la izquierda la imagen original. En el centro, la imagen umbrelizada. A la derecha, en verde, los objetos que superan el límite máximo a ser los ojos; en azul, aquellos que no superan la talla mínima; en rojo, los candidatos a ser los ojos.

De los objetos que pasan estos filtrados se calculan sus momentos de Hu.

Siendo N el número de objetos de la imagen y  $\phi_{x,y}$  el y-momento de Hu del objeto; la distancia euclídea entre el objeto i y el objeto j,  $d_{i,j}$ , se expresa de la siguiente manera:

$$d_{i,j} = \sqrt{\sum_{n=1}^7 (\phi_{j,n} - \phi_{i,n})^2} \quad i = 1 \dots N - 1$$

Si la cara estuviese iluminada uniformemente y la simetría de la cara fuese perfecta, se podría afirmar que los momentos de Hu correspondientes a los objetos que representan los ojos tendría el mismo valor; y su diferencia sería cero. Dado que las condiciones anteriormente citadas, no suelen reproducirse en la realidad, no basta con buscar aquella diferencia que cuyo valor sea los más próximo a cero; sino que es necesario aplicar ciertas reglas que nos permitan, con un elevado porcentaje acierto, afirmar que los objetos seleccionados corresponden a los ojos de la imagen.

Estas reglas de localización se dividen en reglas de distribución espacial y reglas de localización:

- Las reglas de distribución espacial definen como de separados tienen que estar estos los objetos para ser considerados ojos. Estas reglas son las siguientes:



**Figura 40: Detalle de un objeto. X representa las filas en las que está el objeto en la imagen original y las Y, las columnas.**

- La diferencia entre las coordenadas Xmin (ver Figura 40) de los objetos que se comparan, sea menor o igual que la décima parte<sup>1</sup> del número de filas de la imagen original. Dicho de otra manera, que la diferencia de altura entre los dos objetos no sea mayor que la décima parte del ancho de la imagen. Según esto, una cara que este lo suficientemente inclinada, el sistema sería incapaz de encontrar los ojos, por no cumplir con las condiciones iniciales.
- La diferencia entre las coordenadas Ymin (ver Figura 40) de los objetos que se comparan, sea mayor o igual que la veintiocho centésimas partes<sup>2</sup> del número de columnas de la imagen original; o lo que es lo mismo, que los objetos a estudio, deben de estar separados como mínimo una cierta distancia, en concreto, el 28% de ancho de la imagen facial.

---

<sup>1</sup> El valor de “la décima parte del número de columnas de la imagen”, se ha obtenido como consecuencia de un estudio empírico, que se ha realizado para que el porcentaje de aciertos, a la hora de localizar los ojos, sea suficientemente alto.

<sup>2</sup> El valor de “la veintiocho centésimas partes del número de columnas de la imagen”, se ha obtenido como consecuencia de un estudio empírico, que se ha realizado para que el porcentaje de aciertos, a la hora de localizar los ojos, sea suficientemente alto.

- Las reglas de localización definen la zona más probable donde se encuentran los ojos. Un objeto que se encuentra fuera o entre medias de esta zona, no es considerado para la búsqueda. Las reglas son las siguientes:
  - El valor de yMin del objeto debe de ser mayor que la décima parte del ancho de la imagen.
  - El valor de yMax del objeto debe de ser menor que las nueve décimas partes del ancho de la imagen.
  - El valor de xMin del objeto debe de ser menor que las quintas octavas partes del alto de la imagen.
  - El valor de xMax del objeto debe de ser menor que las sextas octavas partes del alto de la imagen.

Todas estas reglas se aplican a cada pareja de objetos.

Basta con que cualquiera de los dos objetos, no cumpla alguna de estas reglas, para que la pareja de objetos sea descartada como los posibles candidatos a ser los ojos; si ocurre esto, se realiza el mismo procedimiento para la siguiente pareja (aquella que tiene la segunda menor distancia euclídea).

Se procede de esta manera, hasta encontrar una pareja que cumpla con todas las reglas.

A la primera pareja de objetos que cumplan las reglas, se la somete a una etapa de verificación.

La imagen binaria es analizada de arriba hacia abajo. Como los primeros objetos corresponden con los objetos que aparecen en la zona más elevada de la cara, se puede dar el caso de que la primera pareja de objetos en cumplir las reglas no sean en realidad los ojos sino las cejas; de ahí que, pasen por una etapa de verificación.

La etapa de verificación consiste en comprobar si los objetos encontrados son los ojos o las cejas: para ello, para cada objeto de esta dupla, se mira si hay algún objeto que este colocado por encima o por debajo de éste. Si hay algún objeto por encima, el objeto encontrado encima puede ser la ceja y por lo tanto el objeto de la dupla es un ojo. Si hay algún objeto por debajo, este nuevo objeto puede ser el ojo; entonces el objeto de la dupla era la ceja. En ambas situaciones se procede a intercambiar el objeto erróneo por el correcto

Se ha tomado una muestra de 16 imágenes para exponer de forma gráfica el motivo de incluir esta etapa de verificación.



**Figura 41 : Muestra de 16 imágenes de la BBDD de Test**

La siguiente figura, muestra los resultados si se obviase la etapa de verificación.



**Figura 42: Resultados de la detección de ojos, sin la etapa de verificación**

Sin esta etapa de verificación, de la colección de 16 imágenes, un cuarto de las imágenes tratadas, fallan en la localización.

Con la etapa de verificación, el porcentaje de acierto es mayor. La Figura 43, muestra el resultado con la etapa de verificación sobre la misma BBDD de 16 caras.



Figura 43: Resultados de la detección de ojos, con la etapa de verificación

#### 4.2.4 Rotación de la imagen

Localizados los ojos, se procede a rotar la imagen de tal manera que la máscara de puntos que se está buscando sea prácticamente simétrica respecto al eje vertical.

El ángulo de rotación se calcula a partir del centro de masas de los objetos ojos. Se forma un triángulo rectángulo con los centros de masa calculados (como se muestra en la Figura 44) y se aplica el teorema de Pitágoras para obtener el valor del ángulo de rotación.

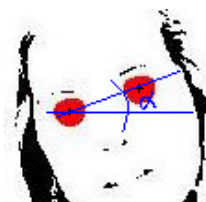


Figura 44: Cálculo del ángulo de rotación de una imagen

Se ha empleado la función de Matlab "*imrotate*" para rotar la imagen. En la siguiente figura se muestra el resultado de la rotación:



Figura 45 : a) Imagen sin rotar. b) Imagen rotada

#### 4.2.5 Identificación de Puntos Característicos

En esta sección se va a explicar en detalle los pasos que se han seguido para la extracción de los puntos característicos. En la Figura 46 se muestran los puntos característicos seleccionados para ser extraídos de cada imagen a analizar.

Tomando como referencia el estudio [37] en el cual reconocían al individuo a través de sus rasgos empleando descripciones MPEG7 (color, textura,...), los puntos aquí seleccionados tratan de caracterizar al individuo a través de las relaciones entre los puntos que definen cada rasgo.

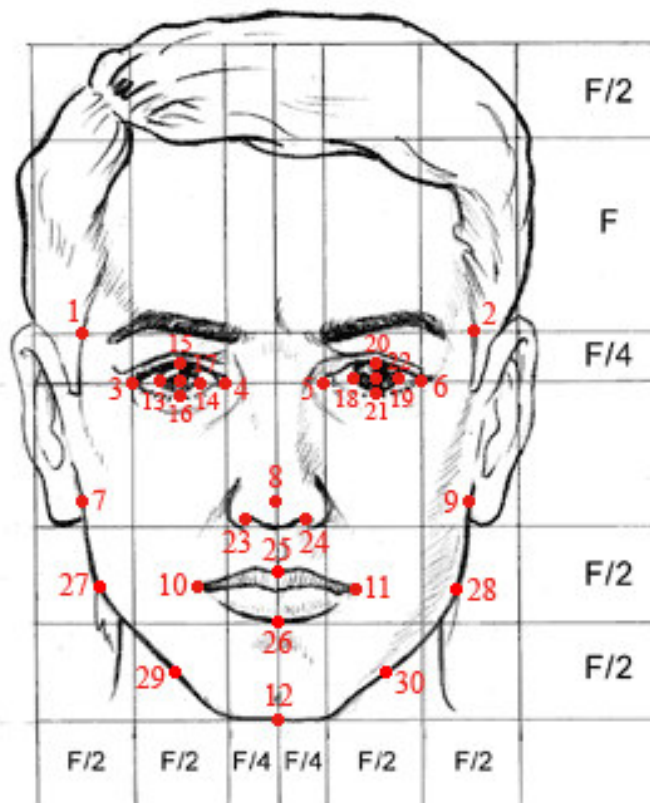


Figura 46: Puntos Característicos de la Cara

<b>Numero de KeyPoint</b>	<b>Definición</b>
Punto 1	Sien derecha
Punto 2	Sien Izquierda
Punto 3	Extremo derecho del ojo derecho
Punto 4	Extremo izquierdo del ojo derecho
Punto 5	Extremo derecho del ojo izquierdo
Punto 6	Extremo izquierdo del ojo izquierdo
Punto 7	Extremo derecho de la mandíbula
Punto 8	Punta de la Nariz
Punto 9	Extremo izquierdo de la mandíbula
Punto 10	Extremo derecho de la comisura de los labios
Punto 11	Extremo izquierdo de la comisura de los labios
Punto 12	Barbilla
Punto 13	Extremo derecho de la Pupila Derecha
Punto 14	Extremo izquierdo de la Pupila Derecha
Punto 15	Extremo superior de la Pupila Derecha
Punto 16	Extremo inferior de la Pupila Derecha
Punto 17	Centro Pupila Ojo Derecho
Punto 18	Extremo derecho de la Pupila Izquierda
Punto 19	Extremo izquierdo de la Pupila Izquierda
Punto 20	Extremo superior de la Pupila Izquierda
Punto 21	Extremo inferior de la Pupila Izquierda
Punto 22	Centro Pupila Ojo Izquierdo
Punto 23	Orificio Nasal Derecho
Punto 24	Orificio Nasal Izquierdo
Punto 25	Labio Superior
Punto 26	Labio Inferior
Punto 27	Borde Derecho de la cara a la altura de la Boca
Punto 28	Borde Izquierdo de la cara a la altura de la Boca
Punto 29	Borde Derecho de la cara entre la Boca y la Barbilla
Punto 30	Borde Izquierdo de la cara entre la Boca y la Barbilla

Tabla 2: Definición de los Puntos Característicos (*KeyPoints*) de la cara

La técnica de extracción de puntos pasa por segmentar cada imagen para extraer nuevas imágenes con los rasgos faciales. Esta técnica trata de forma distinta un rasgo de otro. Con este se consigue reducir, entre otras cosas, el coste computación debido a que en lugar de tratar con la imagen completa, se trabaja con una zona reducida de esta.

#### 4.2.5.1 Segmentación de zonas características de imágenes de caras frontales e identificación de puntos

Una vez localizados los ojos, se procede a obtener, de cada una de las regiones de interés de la cara (ojos, nariz, boca), los puntos que caracterizan a ese individuo, conocidos como *KeyPoints* (o Puntos Característicos).

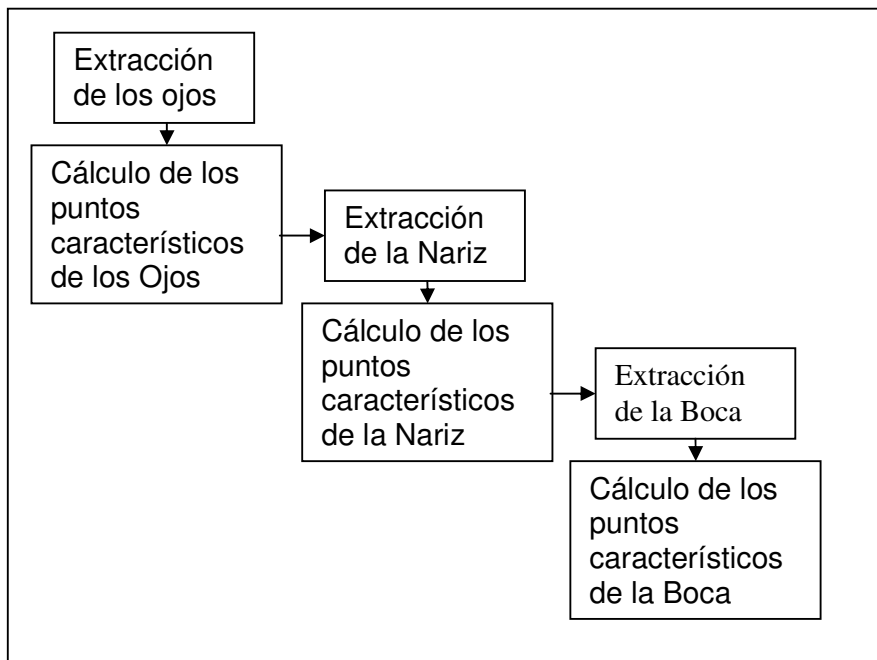
Previamente a la identificación de estos puntos, se extraen las imágenes, correspondientes a cada uno de rasgos que vamos a estudiar. La extracción de imágenes, nos permite poder aplicar un análisis localizado y personalizado para



cada una de las zonas a tratar. Estas zonas, como se ha mencionado anteriormente, son: los ojos, la nariz y la boca.

La extracción de los rasgos faciales es secuencial, es decir, exige que terminen la etapa anterior para extraer la siguiente región de interés consiguiendo así más precisión a la hora de analizar la nueva imagen. Viéndolo desde otro punto de vista, el resultado de la localización de puntos de la etapa anterior es el que se emplea para extraer y analizar la siguiente región.

El orden a seguir se refleja en el esquema de la Figura 47.



**Figura 47: Diagrama de bloques de la extracción de imágenes**

Para la extracción de los rasgos, se toma una primera máscara de puntos de referencia tomando el modelo ideal en el que la imagen es perfectamente simétrica respecto a los ejes horizontal y vertical (ver Figura 46).

A continuación se detallan cada una de estas seis etapas expuestas en la Figura 47

#### **4.2.5.1.1 Extracción de las imágenes de los Ojos**

Localizados los ojos, se extraen las imágenes correspondientes a éstos, basándonos en el centro de masas de los objetos etiquetados como tales.

Teniendo esta estimación, y la distancia que separa los centros de los ojos,  $F$  (relación de proporción, ver Figura 46) podemos extraer las imágenes correspondientes de ambos ojos.

Como se comentó anteriormente, se define la máscara de puntos inicial donde deberían de estar los ojos (resultado de la etapa de localización de los ojos). La máscara inicial para los ojos, se define en la Figura 48.

Se define la máscara y se extraen de la imagen los ojos que corresponden con el trozo de la cara que se ve englobado por esta máscara.

Como la localización de los ojos no es perfecta, hay que ser flexibles a la hora de definir la máscara. En la Figura 49 se muestra como la máscara definida engloba una superficie mayor que la del ojo. El error cometido se tiene en cuenta para definir los límites de lo que será las imágenes correspondientes a los ojos.

```

% -----
%                                     MASCARA INICIAL OJOS
% -----

mascaraCara(3,:)=[alturaOjos(1) centroOjos(1)-round(F/4)]; % OJO IZQUIERDO EXT
IZQ
mascaraCara(4,:)=[alturaOjos(1) centroOjos(1)+round(F/4)]; % OJO IZQUIERDO EXT
DCH
mascaraCara(5,:)=[alturaOjos(2) centroOjos(2)-round(F/4)]; % OJO DERECHO EXT IZQ
mascaraCara(6,:)=[alturaOjos(2) centroOjos(2)+round(F/4)]; % OJO DERECHO EXT DCH

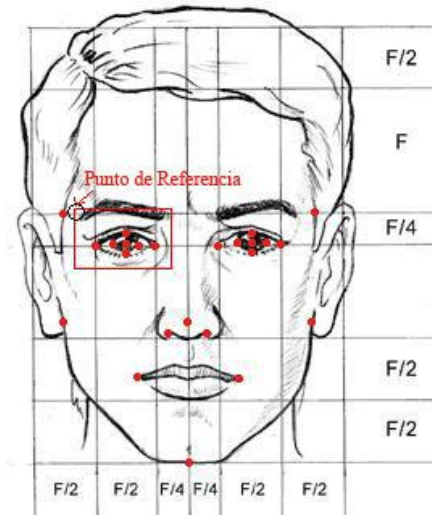
mascaraCara(17,:)=[alturaOjos(1) centroOjos(1)]; % CENTRO OJO IZQUIERDO
mascaraCara(22,:)=[alturaOjos(2) centroOjos(2)]; % CENTRO OJO DERECHO

mascaraCara(13,:)=mascaraCara(17,:); % OJO IZQUIERDO SUP
mascaraCara(14,:)=mascaraCara(17,:); % OJO IZQUIERDO INF
mascaraCara(15,:)=mascaraCara(17,:); % PUPILA EXTREMO IZQ
mascaraCara(16,:)=mascaraCara(17,:); % PUPILA EXTREMO DCH

mascaraCara(18,:)=mascaraCara(22,:); % OJO DERECHO SUP
mascaraCara(19,:)=mascaraCara(22,:); % OJO DERECHO INF
mascaraCara(20,:)=mascaraCara(22,:); % PUPILA EXTREMO IZQ
mascaraCara(21,:)=mascaraCara(22,:); % PUPILA EXTREMO DCH

```

**Figura 48: Definición de la máscara inicial correspondientes a los ojos**



**Figura 49: Detalle del punto de Referencia de las imágenes extraídas**

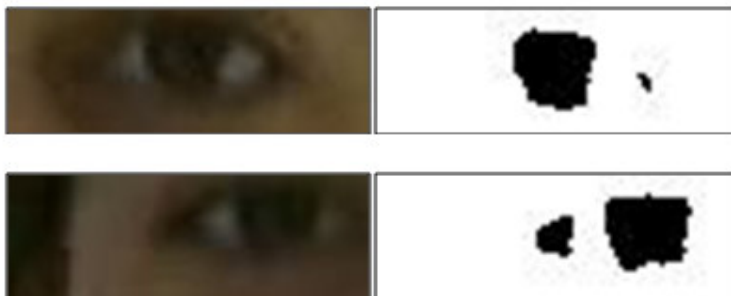
De esta etapa se extraen dos imágenes por ojo; una extraída de la imagen en color, y otra, extraída de la imagen binaria de la etapa de umbralización (Ver Figura 50).

La imagen binaria del ojo, se emplea para localizar la pupila del ojo. Detectada la pupila, se extraen los cinco puntos que la definen (los puntos 13, 14, 15, 16 y 17, del ojo izquierdo; y 18, 19, 20, 21 y 22, del ojo derecho).

La imagen en color, nos permite detectar los otros dos puntos (los puntos 3 y 4, del ojo izquierdo; y 5 y 6, del ojo derecho). Con estos 7 puntos, se termina por definir la máscara de puntos del ojo.

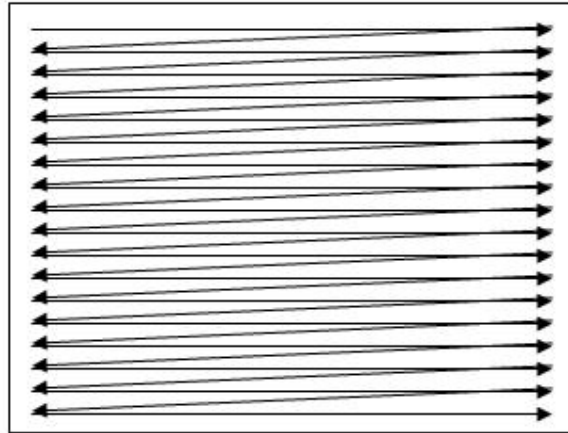
#### 4.2.5.1.1 Extracción de puntos de los Ojos

Para obtener los puntos correspondientes a la pupila, 13, 14, 15, 16, 17, 18, 19, 20, 21 y 22 (definidos en la Tabla 2), se hace uso de la imagen binaria (imágenes de la derecha de Figura 50)



**Figura 50: Las imágenes de la derecha son las imágenes binarias resultantes del proceso de binarización. Las imágenes de la izquierda son el resultado de la etapa anterior.**

La pupila corresponde con la zona más densa o “más negra” de la imagen binaria. Para hallar esta zona, se analiza toda la imagen con una ventana de análisis de la que se saca información sobre el área que esta ventana abarca; es decir, del número de píxeles de color negro de la zona de la imagen englobada por la ventana. Para analizar la imagen, se realiza un barrido progresivo de la imagen (ver Figura 51). El desplazamiento de la ventana de análisis es píxel a píxel con lo que se obtiene una mayor precisión en la búsqueda.



**Figura 51: Barrido Progresivo**

El barrido píxel a píxel se puede realizar debido a que al tratarse de una imagen pequeña, el coste computacional no es excesivamente elevado. En la Figura 52 se puede ver la ventana de análisis aplicada sobre la imagen binaria.



**Figura 52: Ventana de análisis para esta imagen**

Analizada toda la imagen, se calcula la posición en la que la ventana engloba el área máxima. Obtenida la posición de la fila y la columna en la que se aplica la ventana, se procede a ajustar su tamaño hasta eliminar posibles filas y columnas que no tengan información; es decir, que no tengan ningún píxel negro. El resultado de este proceso se muestra en la Figura 53.



**Figura 53: Ventana de Análisis ajustada**

Encontrada la región de interés (ROI), se determinan los cinco primeros puntos de la siguiente forma:

- El centro de la pupila, punto 17 ó 22, corresponde con el centro de masas de la ROI.

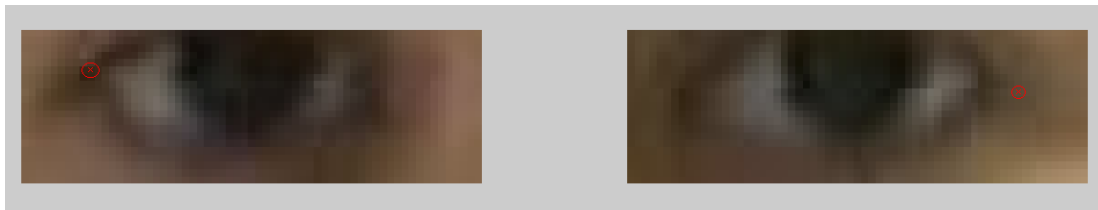
- Los puntos, 13, 14, 15 y 16 ó 18, 19, 20, y 21, se obtienen, prolongando el punto del centro de la pupila en la direcciones de los ejes coordenados, hasta encontrar el borde de la pupila
- Los otros dos puntos que faltan, para terminar de definir el ojo, son los correspondientes a los extremos: los puntos 3 y, 4 ó, 5 y 6 de la Figura 46

Como se mencionó anteriormente, para la obtención de estos puntos se utiliza la imagen color (imágenes de la izquierda de la Figura 50).

Independientemente del ojo (izquierdo o derecho), primero se calcula el extremo que está más alejado de la nariz y a continuación el más cercano. El punto correspondiente al más alejado, coincide con el fin de la línea de las pestañas y ésta, al ser una zona oscura del ojo, se puede extraer umbralizando la imagen a color. Cabría pensar que la imagen binaria con la que se han extraído los anteriores puntos, vale para extraer esta información, pero en algunos casos, la imagen anterior únicamente contiene la información de la pupila.

La técnica usada para umbralización de esta imagen a color es la misma que la usada para extraer los puntos de la boca (ver apartado 4.2.5.1.3.1 ó [33]). Dado que esta técnica se suele emplear para definir los puntos que caracterizan la boca, será en ese apartado cuando se defina con más detalle.

Obtenida la imagen binaria, el extremo más alejado del ojo, coincide con la zona más alejada de esta imagen binaria. El resultado de este proceso se muestra en la Figura 54.



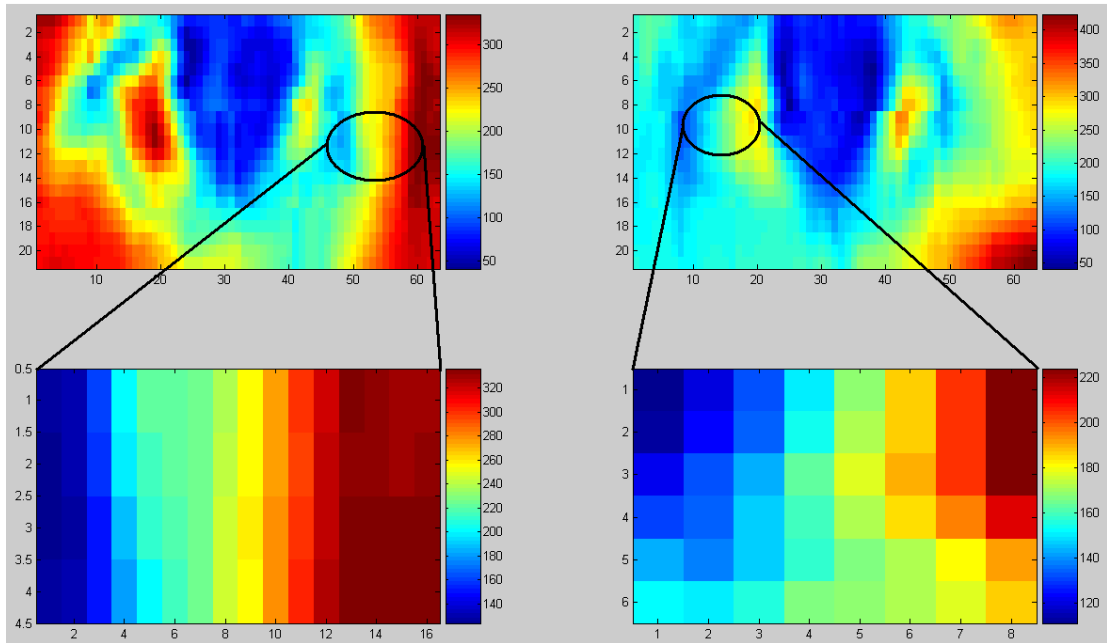
**Figura 54: El fin de las pestañas coincide con el extremo buscado**

El otro extremo es más complejo de detectar, ya que a simple vista no existe ninguna información que sirva de referencia.

Aquí, es cuando realmente se necesita la imagen en color. A partir de un estudio empírico se ha comprobado que si se suma las tres componentes de color de la imagen

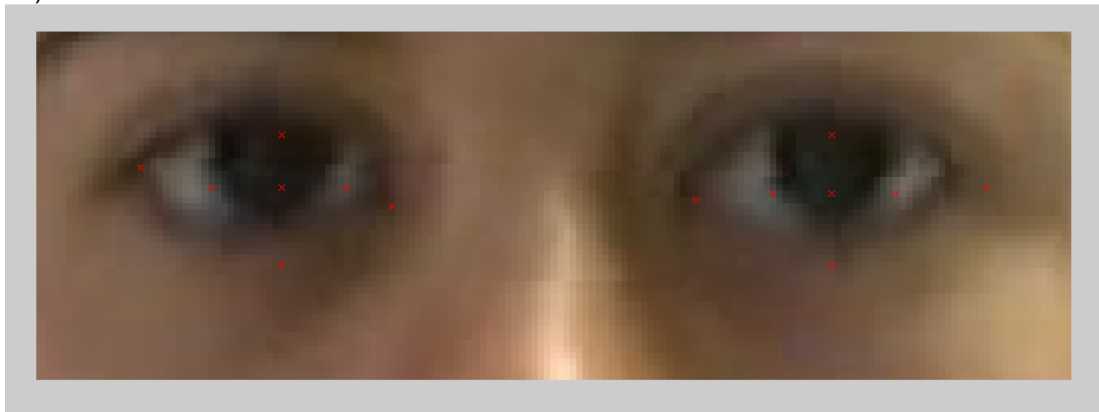
$$H=R+G+B$$

La imagen resultante, H contiene la información que se necesita. En la Figura 55 se muestra la imagen Histograma (H), en la parte superior; y la ROI (Región de Interés) a estudiar, en la inferior. Como se aprecia, el punto buscado coincide con el mínimo valor de la ROI.



**Figura 55:** Arriba, la imagen Histograma de cada ojo; abajo, la ROI a estudiar

A continuación, se muestra el resultado de este proceso en una imagen (Figura 56).



**Figura 56:** Puntos Característicos de los Ojos

#### 4.2.5.1.2 Extracción de la imagen de la Nariz

La imagen que se obtiene de la nariz, depende de los resultados de la etapa anterior.

Como en los ojos, se define una máscara inicial, en concreto, se establece (usando como referencia la imagen de la Figura 46) que la punta de la nariz tiene que estar a  $F/2$ , en el eje horizontal; y a  $3F/4$  en dirección vertical, del centro de la pupila del ojo izquierdo.

Con estas suposiciones, se extrae la imagen de la nariz de la siguiente manera:

- Coordenadas del punto de referencia (PR) (Figura 49) son:
  - Eje Horizontal (X): coordenada X del Punto 17
  - Eje Vertical (Y): coordenada Y del Punto 4 menos F/6.
  - El ancho va desde la coordenada X del PR anterior hasta la coordenada X del Punto 8 más F/8.
  - El alto de la imagen va desde la coordenada Y del PR hasta la coordenada Y del Punto 5 más F/6.

Al igual que en los ojos, es necesario comprobar que la imagen esta está dentro de los límites permitidos

#### **4.2.5.1.2.1 Extracción de puntos de la nariz**

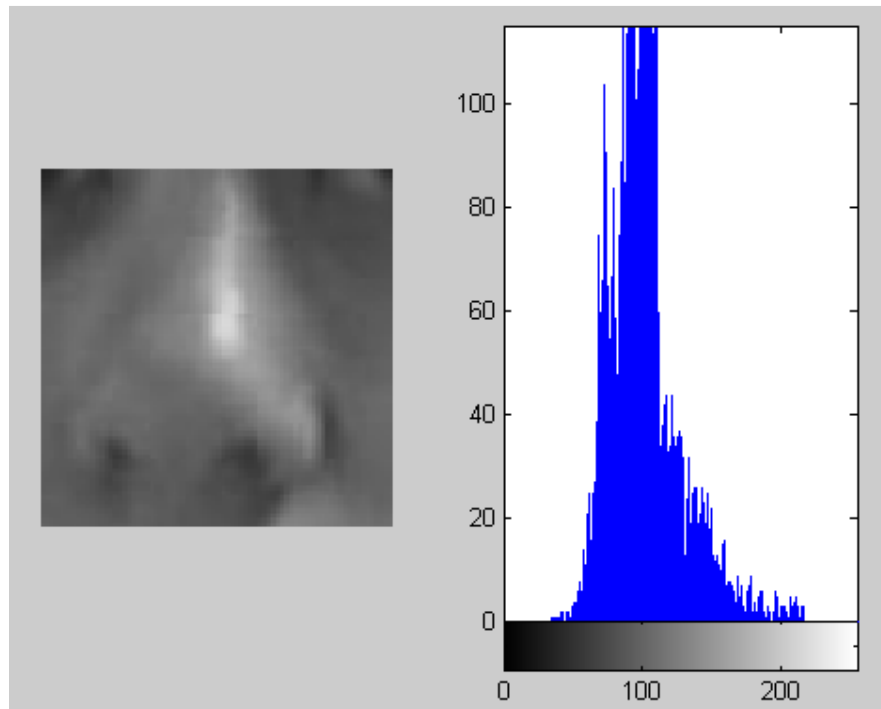
El punto correspondiente a la punta de la nariz, Punto 8 de la Tabla 2; es un punto difícil de hallar. En algunas ocasiones, este punto se corresponde con la zona de más brillo de la nariz; pero por desgracia no es siempre generalizable.

Se ha probado a binarizar la imagen y a analizar la imagen histograma, H; pero de ninguna de las formas, se puede extraer este punto de forma directa.

La solución que se ha adoptado, ha sido, hallar los orificios nasales; y a partir de esta información, extraer el Punto 8.

#### **4.2.5.1.2.2 Orificios Nasales**

Antes de hallar los orificios, conviene detenerse un poco en el histograma de la imagen. La Figura 57 muestra el histograma de la luminancia de la Nariz. A primera vista, se observa que el histograma tiende a concentrar toda la información a la mitad del histograma. Esto se debe a que la imagen de una nariz, es plana en cuanto al color; es decir, no presenta unos cambios bruscos de color en toda la imagen.



**Figura 57: A la izquierda la imagen de la luminancia de la nariz. A la derecha, el histograma correspondiente**

La escala de grises debajo del histograma, indica, aproximadamente, el color correspondiente a cada una de las líneas verticales de la imagen histograma. Se observa, que las zonas oscuras están al principio del mismo. Este hecho, es el que nos permitirá encontrar los orificios nasales.

El primer paso consiste en calcular la imagen de luminancia.

$$Y=0,3R + 0,59G + 0,11B$$

Obtenida la luminancia, umbralizamos esta imagen con un valor suficientemente bajo para que únicamente se obtengan las zonas oscuras de la imagen (suelen coincidir con los orificios). El valor umbral elegido para convertir en blanco y negro la imagen, es la mitad de índice (eje x del histograma de la Figura 57 ) que tiene mayor concentración; es decir, del histograma, calculamos el valor máximo, vemos que índice tiene y ponemos el umbral como la mitad de ese índice. En concreto, para la Figura 57 ; el índice del valor máximo es 105, por lo tanto, el umbral es 53.

Con la imagen binaria, clasificamos los objetos resultantes como se explicó en el apartado 4.2.3. Con los objetos etiquetados, calculamos los momentos de Hu de estos objetos y se escogen la pareja de objetos con momentos parecidos. Si esta pareja cumple con ciertas reglas de verificación serán los orificios; sino, se escoge la siguiente pareja. Así sucesivamente, hasta acabar con todas las posibles combinaciones de objetos. La primera pareja que cumpla con las reglas de verificación, serán los escogidos como los orificios nasales.



Las reglas de verificación que se aplican son las siguientes:

- No son “nostriles”, aquellos objetos que se encuentren localizados en el primer cuarto de la imagen (cuarto superior).
- No son “nostriles”, aquellos objetos cuyo ancho y/o alto sea igual a un píxel.
- Tampoco, aquella pareja de objetos que se encuentren en el mismo lado de la imagen (lado izquierdo o lado derecho). El lado izquierdo corresponde con la mitad izquierda de la imagen si dividiésemos ésta, en dos trozos iguales, trazando una línea vertical; y el derecho, con la mitad derecha.
- Tampoco lo son, los objetos que cumpliendo las anteriores premisas, tienen una diferencia en altura, superior a la separación, en el eje horizontal.

Si no se encontrasen los orificios, se aumenta el umbral para obtener la imagen binaria, en 10 unidades y se procede como anteriormente. Si aún así se sigue sin encontrar los nostriles, aumentamos de nuevo el umbral en 7 unidades más.

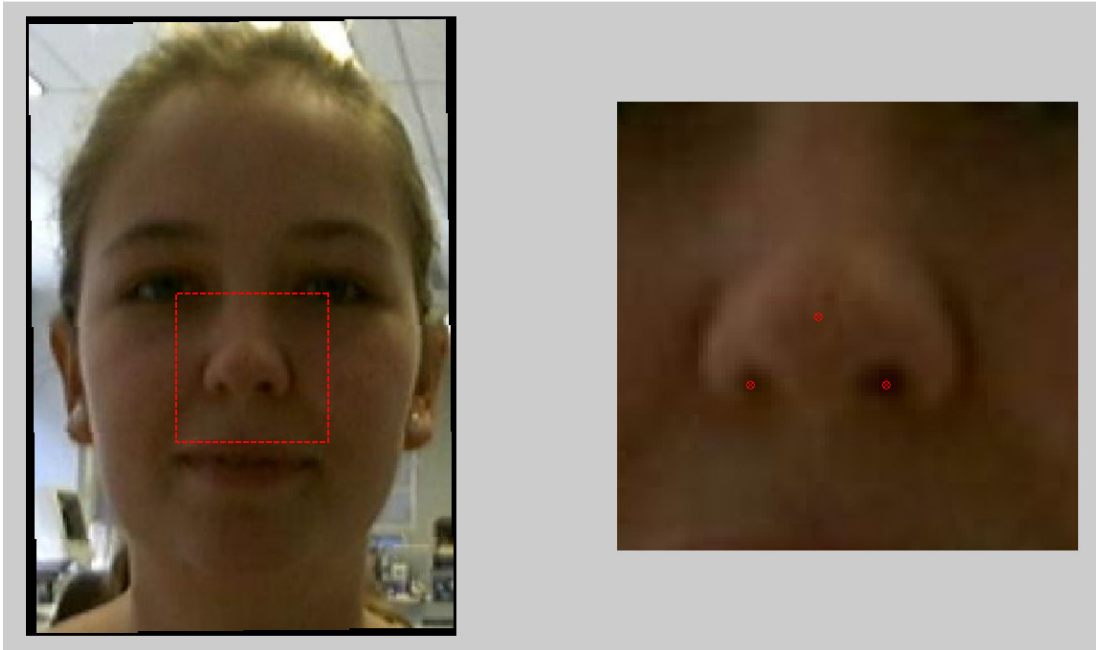
Si después de este tercer intento no se encuentran, la imagen es eliminada de la BBDD.

No se sigue aumentando más el umbral porque según las distintas pruebas realizadas se comprueba que si se amplía demasiado, la imagen se vuelve completamente negra.

Al encontrarse los orificios nasales, los Puntos correspondiente (Puntos 23 y 24 de la Tabla 2), se sitúan en el centro de masas de cada objeto.

El Punto 8, punta de la nariz, se sitúa entre medias de los orificios y elevado de estos puntos, una distancia igual a la mitad de la distancia de separación de los nostriles.

En la siguiente figura se puede ver el resultado de la extracción de puntos de una imagen de la BBDD de *Biosec*.



**Figura 58:** Extracción de los puntos de la nariz para una imagen de la BBDD. A la izquierda se muestra la imagen original ya rotada y a la derecha un primer plano de la nariz con los puntos extraídos.

#### **4.2.5.1.3 Extracción de las imágenes de la Boca**

A partir de los nostriles, se define un punto de referencia para la imagen de la boca a analizar. A partir de este punto y tomando un área lo suficientemente grande como para contener la boca se extrae la imagen de la boca.

El área se define tomando como base Figura 46.

El punto tomado como referencia es el orificio izquierdo de la nariz. Cabe pensar que el punto óptimo sea la punta de nariz, pero como se ha visto en el apartado anterior, este punto se calcula estadísticamente por lo que resulta poco fiable.

Si en la etapa anterior, no se encontrasen los orificios nasales, la imagen es eliminada de la BBDD y se interrumpe el proceso de detección de puntos.

##### **4.2.5.1.3.1 Extracción de puntos de la Boca**

Para encontrar los Puntos 10 y 11 de la Tabla 2, se utiliza la técnica de umbralización expuesta en [6] y [33].

Este mismo método se ha empleado en el apartado 4.2.5.1.1 para la obtención de los puntos 3 y 6.

El método consiste en hacer una transformación del color, de tal forma que se queden bien definidas las zonas que presentan cambios bruscos de color.

En el espacio de color RGB, los píxeles de la piel y los de los labios tienen componentes de color bastante diferentes (ver Figura 59)

Tanto en el espacio de color que define la piel como el que define los labios, el color rojo es el más frecuente. Además, la piel, presenta una mezcla de azul y verde, teniendo más componente de verde que de azul; y en los labios, estas dos componentes son prácticamente iguales.

La diferencia entre el rojo y el verde es mayor en los labios que en la piel, por eso Hulbert y Poggio [35] propuso una definición de una matriz, pseudo hue, que expone esta diferencia y que se calcula como sigue:

$$h(x, y) = \frac{R(x, y)}{G(x, y) + R(x, y)}$$

Esta matriz, cuantifica la diferencia entre la componente Roja y la Verde.

Si normalizamos a 1 la matriz de pseudo-hue ( $hN$ ) y la luminancia de la imagen ( $YN$ ); la transformación cromática se define como la diferencia de las dos matrices:

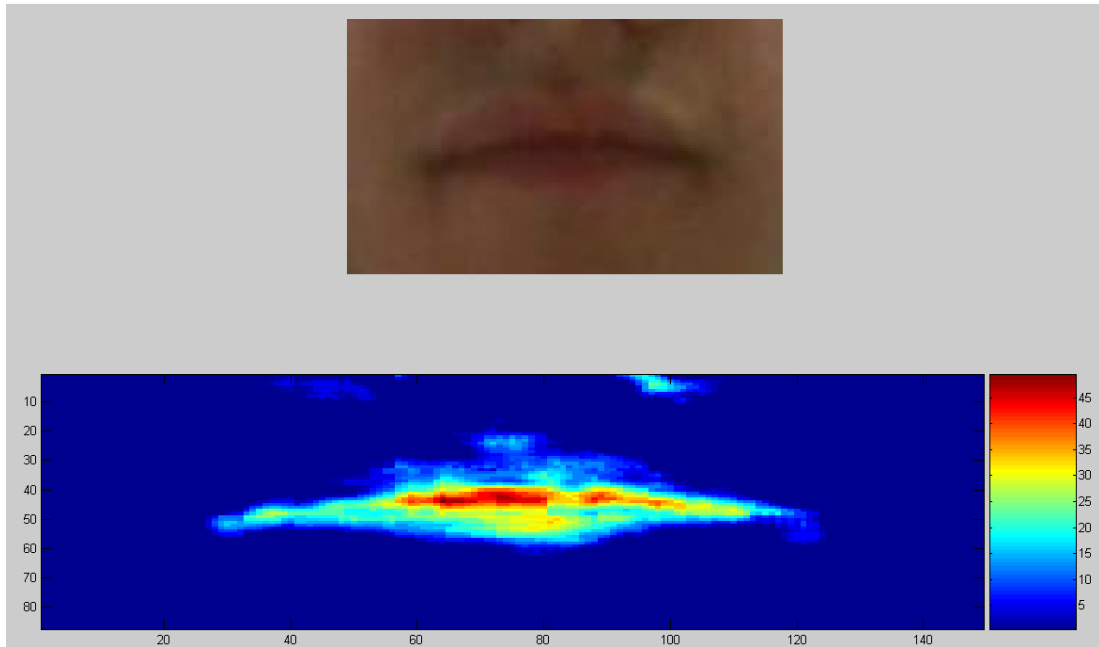
$$R_{cor} = hN - YN$$

En la Figura 59 se muestra la transformación cromática,  $R_{cor}$  aplicada a una imagen.

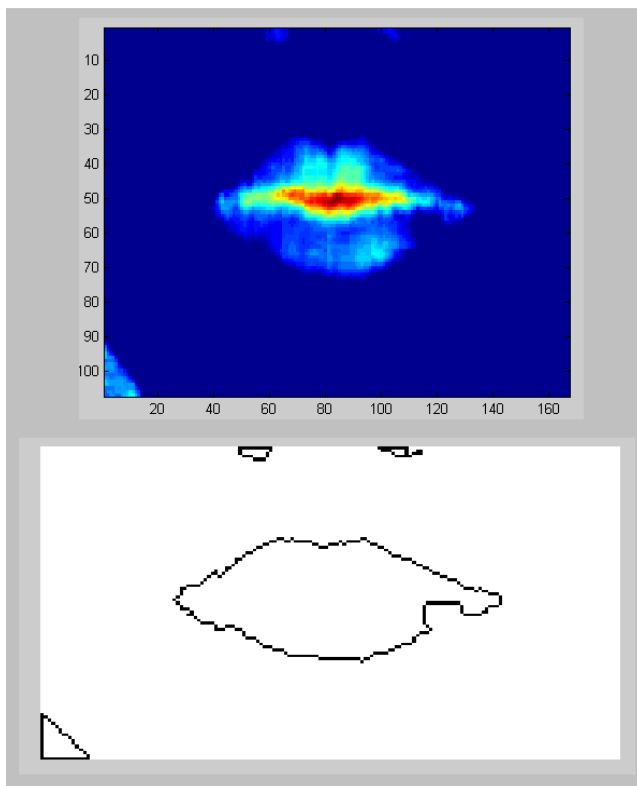
Obtenida la transformación cromática, se calcula el perímetro de la imagen (ver Figura 60). Para la obtención del perímetro se ha empleado la función "*bwperim*" del *Image Processing Toolbox* de Matlab.

Con la imagen binaria resultante (perímetro de la imagen), se estudian los objetos, usando la clasificación de objetos del apartado 4.2.3. De los objetos resultantes se escoge aquel que ocupe la mayor parte de la imagen.

Los puntos de la boca corresponden con el valor máximo y mínimo de la coordenada Y del objeto, y sus correspondientes valores en X.

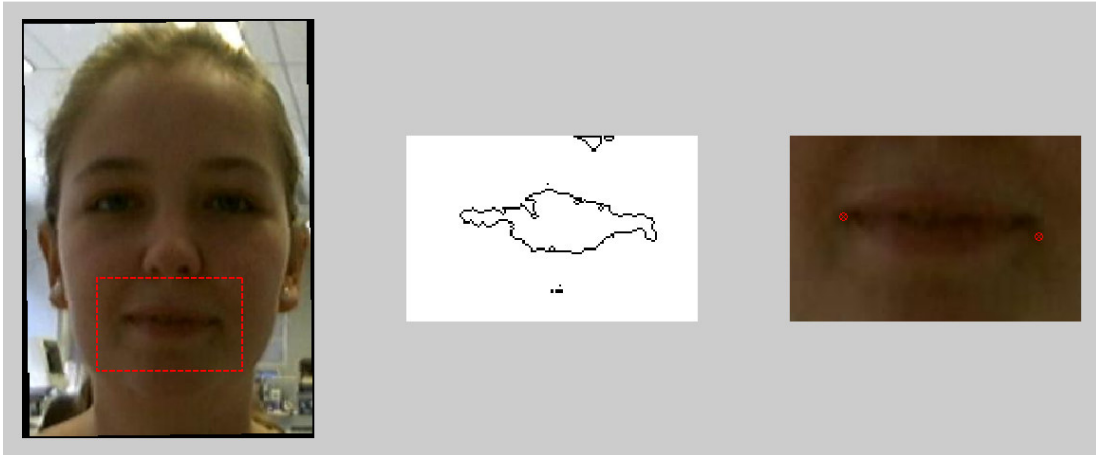


**Figura 59: Imagen de la Boca y su transformación cromática**



**Figura 60: Transformación cromática y su perímetro**

En la siguiente figura, se puede observar el proceso de extracción de los puntos característicos de la boca para una imagen de la BBDD de *Biosec*.



**Figura 61: Resultado de la extracción de puntos de la Boca para una imagen de la BBDD Biosec. A la Izquierda, la imagen de la BBDD; en el centro, el perímetro de la transformación cromatica de la boca; a la derecha, los puntos característicos extraídos de la boca.**

#### **4.2.5.1.4 Otros puntos: Barbilla, Extremos y Labios**

En este apartado se va a definir la obtención de los Puntos 1, 2, 7, 9, 12, 25, 26, 27, 28, 29 y 30 de la Tabla 2. Previa localización de estos 11 Puntos, se debe haber obtenido los puntos que definen los rasgos faciales de los ojos, nariz y la boca.

Para hallar cada punto se emplean tres “transformaciones” de la imagen:

- Una imagen de bordes, obtenida a partir de la componente Verde (G) usando la detección de bordes de Canny [49].
- Otra imagen de bordes, usando Canny pero en la componente Azul (B).
- Una imagen histograma, H, obtenida a partir de la Imagen en Color, aplicando (YYYY).

Según se expuso en 4.2.5.1.3.1, la piel presenta una mezcla de azul y verde, teniendo más componente de verde que de azul; por este motivo se utilizan estas dos componentes para calcular la imagen de bordes. Para calcular esta imagen de bordes, se ha empleado la función “*edge*” de Matlab.

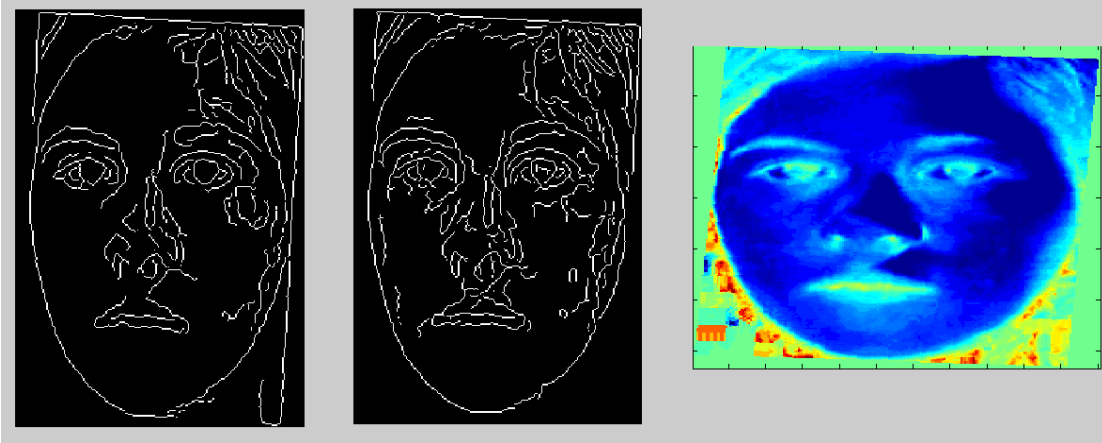
La primera imagen (imagen de bordes de la componente Verde) se emplea para la localización de los puntos que definen el contorno de la cara, exceptuando la barbilla (punto 12).

La tercera imagen sólo se emplea para la localización del punto correspondiente a los labios (puntos 25 y 46) y a la barbilla (punto 12).

Para la barbilla, se emplea la imagen H para obtener una primera estimación del punto, a continuación se emplea la imagen de bordes, para un ajuste más fino. La imagen de bordes, de la componente Azul, suele mostrar con más precisión, el contorno de la barbilla, por eso, se emplea esta imagen y no la imagen de bordes

de la componente verde para detectar el punto que define la barbilla (ver Figura 62).

En la Figura 62 se muestran las tres imágenes que se van a emplear.



**Figura 62: De izquierda a derecha, la imagen de bordes de la componente Verde, la imagen de bordes de las de la componente Azul y la imagen histograma H.**

Las dos imágenes de bordes, son imágenes binarias, con valor 0 ó 1. La línea que define los bordes tiene valor 1.

Ninguno de los puntos hallados a continuación se emplea ni para el modelado ni para el reconocimiento del individuo. El motivo de no usar estos puntos, se debe principalmente a que el “ruido” que presenta la imagen hace que la localización del punto no sea 100% efectiva. Si se emplease este punto a la hora de modelar y reconocer al individuo, se estaría introduciendo una incertidumbre en el sistema que provocaría que la tasa de reconocimiento baje.

#### **4.2.5.1.4.1 Punto 1 (Sien Derecha)**

Para obtener el Punto 1 correspondiente, se parte de la suposición de que el punto se encuentra a  $F/5$  a la izquierda y a  $F/4$  hacia arriba del Punto 3 (Extremo Derecho del Ojo Derecho).

Con la primera imagen de bordes y a partir del punto de referencia se van cogiendo valores a la izquierda y a la derecha de ese Punto; es decir, se hace un barrido horizontal que va desde el principio de la imagen, hasta el Punto 1 más F/32 a la derecha (ver Figura 63).



**Figura 63 :** En a) la imagen de bordes con el vector a analizar en rojo. En b) detalle del vector a analizar.

Este punto, Punto 1, al igual que el resto de puntos que forman parte del borde de la cara (puntos 1, 2, 7, 9, 27, 28, 29 y 30), tienen la peculiaridad de que definen una zona de la imagen que contienen información de alta frecuencia, es decir, entre estos puntos y un punto anterior o posterior, según corresponda, existe un cambio, cuanto menos notable, de luminancia y crominancia. Este cambio notable, es el que se muestra en la imagen de bordes empleada.

Partiendo de esta hipótesis, se analiza el vector, de derecha a izquierda. El Punto 1, corresponde con el primer píxel cuyo valor sea igual a 1 (color blanco de la Figura 63).

#### **4.2.5.1.4.2 Punto 2 (Sien Izquierda)**

Este punto, parte del mismo principio que el Punto 1. Se define un punto de referencia, PR, situado F/5 a la derecha y F/4 arriba del Punto 6; y se hace un

barrido horizontal como antes; pero esta vez, desde el final de la imagen hasta el PR menos  $F/32$ .

Con el vector formado, se analiza de izquierda a derecha y se busca el primer valor que sea igual a 1, es decir, la primera línea que defina un borde.

#### **4.2.5.1.4.3 Punto 7 (Extremo derecho de la mandíbula)**

Este punto también parte de un punto de referencia, PR, que se define a partir del Punto 1 y del Punto 8. Este punto, se sitúa, en el eje horizontal, a la altura de la nariz, punto 8; y en el eje vertical, diez píxeles a la izquierda del Punto 1.

Partiendo el PR definido, se realiza un barrido horizontal de la imagen, con el fin de recoger la información de los píxeles que se encuentren definidos en el intervalo que va desde inicio de la imagen hasta el Punto 1 más  $F/8$ .

El vector formado, se analiza de derecha a izquierda al fin de encontrar el primer píxel que defina un borde, cuyo valor es 1. Ese píxel, será el punto buscado.

#### **4.2.5.1.4.4 Punto 9 (Extremo izquierdo de la mandíbula)**

Este punto también parte de un punto de referencia, PR, que se define a partir del Punto 2 y del Punto 8. Este punto, se sitúa, en el eje horizontal, a la altura de la nariz, punto 8; y en el eje vertical, diez píxeles a la derecha del Punto 2.

Partiendo el PR definido, se realiza un barrido horizontal de la imagen, con el fin de recoger la información de los píxeles que se encuentren definidos en el intervalo que va desde el punto 2 menos  $F/8$  hasta final de la imagen.

El vector formado, se analiza de izquierda a derecha, al fin de encontrar el primer píxel que defina un borde, cuyo valor es 1. Ese píxel, será el punto buscado.

#### **4.2.5.1.4.5 Punto 25 (Labio Superior)**

Para este Punto, se emplea la imagen H. El PR se sitúa entre medias de la boca; es decir, entre medias de los puntos 10 y 11. Y a una altura  $F/4$  superior al punto 10.

A continuación se realiza un barrido vertical desde el PR hasta el punto medio entre los puntos 10 y 11; y se busca, en el vector formado, el primer punto que superando un determinado umbral, el siguiente valor del píxel que no lo supere.

El vector se analiza de abajo hacia arriba. El umbral definido corresponde con la cuarta parte de la suma del triple del valor mínimo más el valor máximo del vector.



#### **4.2.5.1.4.6 Punto 26 (Labio Inferior)**

Al igual que el anterior punto, se emplea la imagen H y se realiza también un barrido vertical desde el PR definido a partir del Punto 10 (Extremo derecho de la comisura de la boca) y del Punto 11 (Extremo izquierdo de la comisura de la boca). El PR se sitúa entre medias de la boca; es decir, entre medias de los puntos 10 y 11. Y a una altura  $F/4$  inferior al punto 10.

A continuación se realiza un barrido vertical desde el punto medio entre los puntos 10 y 11 hasta el PR; y se busca, en el vector formado, el primer punto que superando un determinado umbral, el siguiente valor del píxel no lo supere.

El vector se analiza de arriba hacia abajo. El umbral definido corresponde con la cuarta parte de la suma del triple del valor mínimo más el valor máximo del vector.

#### **4.2.5.1.4.7 Punto 12 (Barbilla)**

El punto de la barbilla es un poco más complicado debido a que no se aprecia tanto el cambio brusco de luminancia y crominancia (debajo de la barbilla está el cuello, que suele ser del mismo color que la barbilla); por lo tanto, el procedimiento a seguir difiere de los demás puntos.

El Punto de la barbilla se puede hallar gracias a que debajo de la barbilla, suele existir una zona más oscura que el resto de la barbilla, que aparece como consecuencia de que la propia cara hace sombra.

El PR, que se utiliza en este caso, se sitúa en la vertical del Punto 25 a una altura  $3F/4$  inferior a la altura del Punto 10.

Debido a que la imagen de bordes de la componente azul, no siempre contiene la información de contorno de la barbilla, se hace una primera aproximación a partir de la imagen H, en la que se realiza un barrido vertical desde el PR hasta 10 píxeles más abajo del PR. Con el vector formado, se analiza de arriba hacia abajo y se busca el valor máximo del vector (zona oscura de debajo de la barbilla). Este punto, se toma como primera aproximación del punto.

A continuación, se realiza el mismo barrido vertical pero con la imagen de bordes de la componente azul, si no se encontrase ninguna línea de borde en el rango definido en este barrido vertical, el punto de la barbilla será el definido antes. En cambio, si se encontrase algún píxel cuyo valor es 1, el punto correspondiente a la barbilla corresponde con este nuevo punto.

#### **4.2.5.1.4.8 Punto 27 (Borde Derecho de la cara a la altura de la Boca)**

Se parte de la imagen de bordes de la componente verde y de un punto de referencia, PR, que se define a partir del Punto 10 y del Punto 7. El PR se

sitúa, en el eje horizontal, a la altura de la boca, punto 10; y en el eje vertical, trece píxeles a la derecha del Punto 7.

Partiendo el PR definido, se realiza un barrido horizontal de la imagen, con el fin de recoger la información de los píxeles que se encuentren definidos en el intervalo que va desde el inicio de la imagen hasta el punto medio entre el Punto 7 y el 10.

El vector formado, se analiza de derecha a izquierda:

- Si se encuentra un píxel con valor 1 que está situado a la derecha del punto 7, se almacena la posición de este punto y se sigue analizando el vector en busca de otro posible candidato.
- Si se encuentra un píxel con valor 1 que este situado a la izquierda del punto 7:
  - Se detiene la búsqueda.

#### **4.2.5.1.4.9 Punto 28 (Borde Izquierdo de la cara a la altura de la Boca)**

Al igual que el anterior punto, se define un punto de referencia, PR, a partir del Punto 11 y del Punto 9. Este punto, se sitúa, en el eje horizontal, a la altura de la boca, punto 11; y en el eje vertical, trece píxeles a la izquierda del Punto 9.

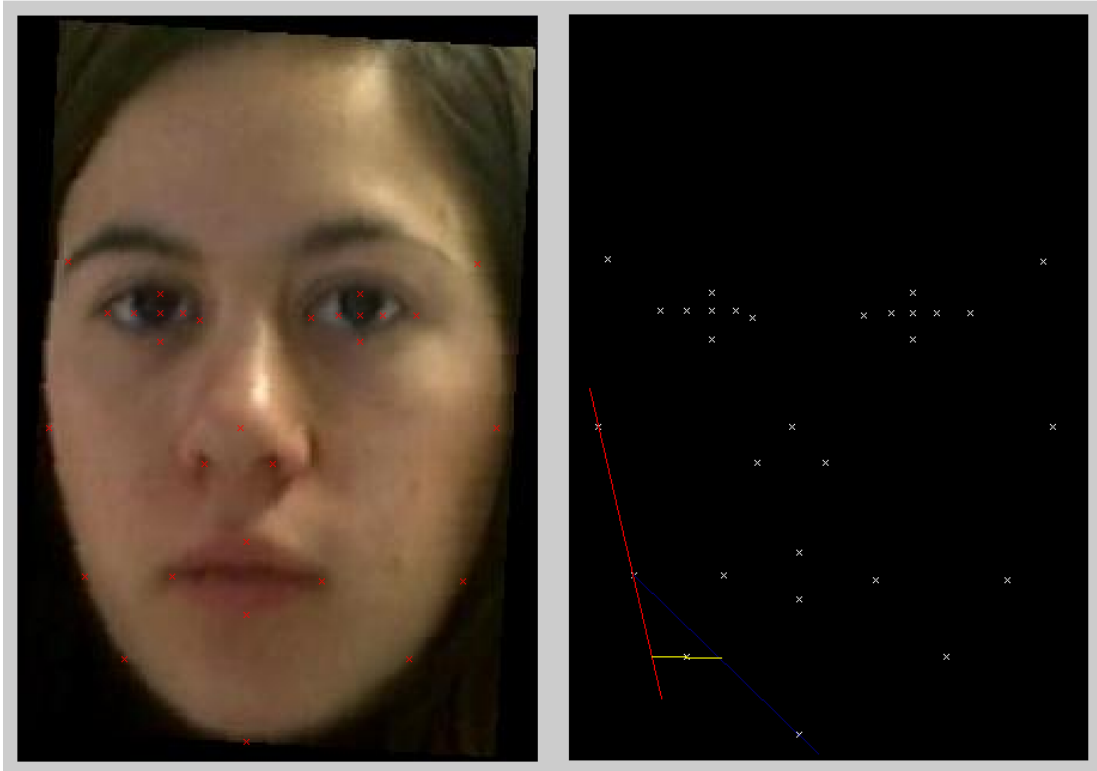
Partiendo el PR definido, se realiza un barrido horizontal de la imagen, con el fin de recoger la información de los píxeles que se encuentren definidos en el intervalo que va desde el punto medio entre el Punto 9 y el 11 hasta el final de la imagen.

Siguiendo un proceso similar al anterior se analiza de izquierda a derecha el vector formado como sigue:

- Si se encuentra un píxel con valor 1 que está situado a la izquierda del punto 9, se almacena la posición de este punto y se sigue analizando el vector en busca de otro posible candidato.
- Si se encuentra un píxel con valor 1 que este situado a la derecha del punto 9:
  - Se detiene la búsqueda

#### **4.2.5.1.4.10 Punto 29 (Borde Derecho de la cara entre la Boca y la Barbilla)**

Este punto se sitúa en el eje vertical, a una altura entre la boca y la barbilla. El vector formado (recta de color amarillo de la Figura 64), recoge los puntos definidos entre dos rectas a la altura mencionada. La primera recta es la recta que pasa por los puntos 7 y 27 (recta de color rojo de la Figura 64); y la segunda la que pasa por los puntos 12 y 27 (recta de color azul de la Figura 64).



**Figura 64:** A la izquierda, la imagen de la BBDD con los puntos detectados. A la derecha, en rojo y azul, las rectas que delimitan el vector a analizar; en amarillo, el vector a analizar.

El vector se analiza de derecha a izquierda, hasta encontrar el primer píxel del vector, cuyo valor sea 1. Ese será el punto buscado

#### **4.2.5.1.4.11 Punto 30 (Borde izquierdo de la cara entre la Boca y la Barbilla)**

Este punto se sitúa en el eje vertical, a una altura entre la boca y la barbilla. El vector que formamos, recoge los puntos definidos entre dos rectas a la altura mencionada. La primera recta es la recta que pasa por los puntos 9 y 28; y la segunda la que pasa por los puntos 12 y 28.

Se analiza el vector de izquierda a derecha, y se busca el primer píxel cuyo valor sea 1. Ese será el punto buscado

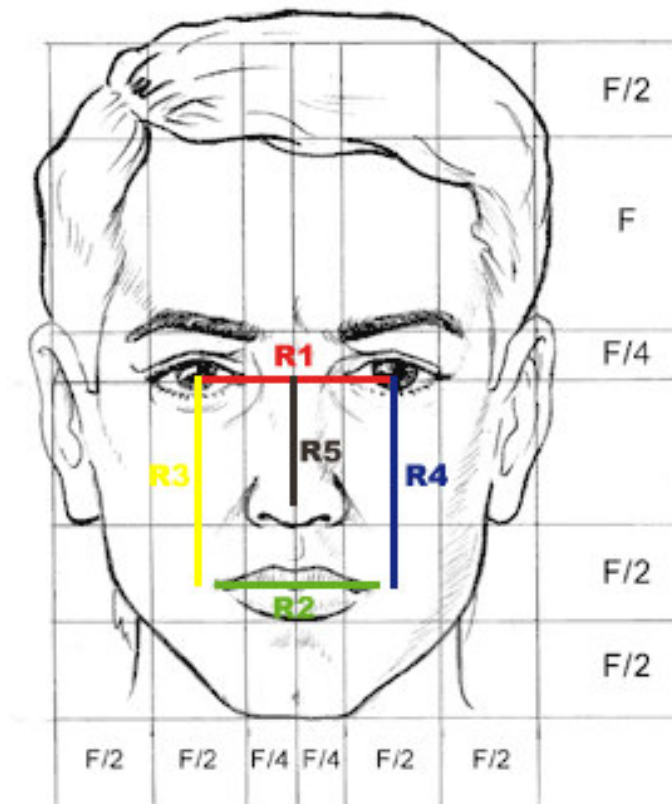
### **4.2.6 Ratios – Distancias Relativas entre Puntos**

La definición de los ratios se ha basado en el estudio [13] que emplea un total de 14 ratios para reconocer a una persona. Para este PFC se han definido un total de 10 ratios por imagen. No se han definido 14 ratios como en [13], debido a que en este estudio emplean ratios que vienen definidos por puntos que se han desechado para el modelado y reconocimiento. La información de los ratios que se emplea para reconocer, se muestra tabulada a continuación:

Ratios	Relación
Ratio 1	$r1 / r2$
Ratio 2	$r1 / r3$
Ratio 3	$r1 / r4$
Ratio 4	$r1 / r5$
Ratio 5	$r2 / r3$
Ratio 6	$r2 / r4$
Ratio 7	$r2 / r5$
Ratio 8	$r3 / r4$
Ratio 9	$r3 / r5$
Ratio 10	$r4 / r5$

$r1$  = separación entre pupilas  
 $r2$  = ancho boca  
 $r3$  = distancia ojo izquierdo boca  
 $r4$  = distancia ojo derecho boca  
 $r5$  = altura nariz

**Tabla 3 : Ratios**



**Figura 65 : Proporciones características de una cara.**

Calculada las 5 proporciones características de una cara ( $r1$ ,  $r2$ ,  $r3$ ,  $r4$  y  $r5$  de la Figura 65), se calculan los ratios dividiendo entre sí cada una de estas proporciones. El resultado de esta división, son los llamados ratios que se muestran en la Tabla 3.

Los resultados del reconocimiento, se recogen en el apartado 5.2.3.

## 4.2.7 Modelado

Una vez analizada la BBDD, con los puntos extraídos se procede a modelar a los diferentes individuos. Se ha empleado la técnica de modelado Mezcla de Gaussianas (GMM) (ver apéndice B) por ser una de las técnicas más utilizadas los actuales sistemas de reconocimientos; además, presenta una serie de ventajas como son la eficiencia computacional, la facilidad de implementación y una estructura probabilística robusta, entre otras.

Para el modelado se ha empleado estimaciones de máxima verosimilitud (ML) [51][50]. El método ML es un procedimiento que permite estimar los parámetros de un modelo probabilístico, o los coeficientes de un modelo matemático, de tal manera que sean los más probables a partir de los datos obtenidos.

Se parte de un conjunto de  $T$  vectores  $X = \{x_1, \dots, x_t\}$  siendo  $X_i$  el vector formado con los 10 ratios calculados. Considerando que el peso o contribución de cada  $X_i$  es el mismo, se obtienen los valores iniciales del vector de medias y la matriz de covarianzas y se calcula  $p(X|\lambda)$  (ver apéndice B).

Con este primer modelo,  $\lambda$ , se emplea el algoritmo Estimación-Maximización (EM) para se estima un nuevo modelo  $\lambda'$  que se parezca más al buscado.

El algoritmo EM [51] proporciona una técnica iterativa (se repiten iterativamente hasta alcanzar el umbral de convergencia definido) para realizar una estimación de máxima verosimilitud de un conjunto de parámetros en problemas en los que existen *datos ocultos* (valores de ratios no contemplados en el modelo por tener este un número limitado de muestras) que dependen estadísticamente de los datos a estimar y de los datos observados.

La iteración del algoritmo EM consiste en un paso de estimación (paso E), en el que se obtienen los valores esperados de las posiciones del entorno, seguido de un paso de maximización (paso M). En cada iteración la verosimilitud aumenta de forma monótona.

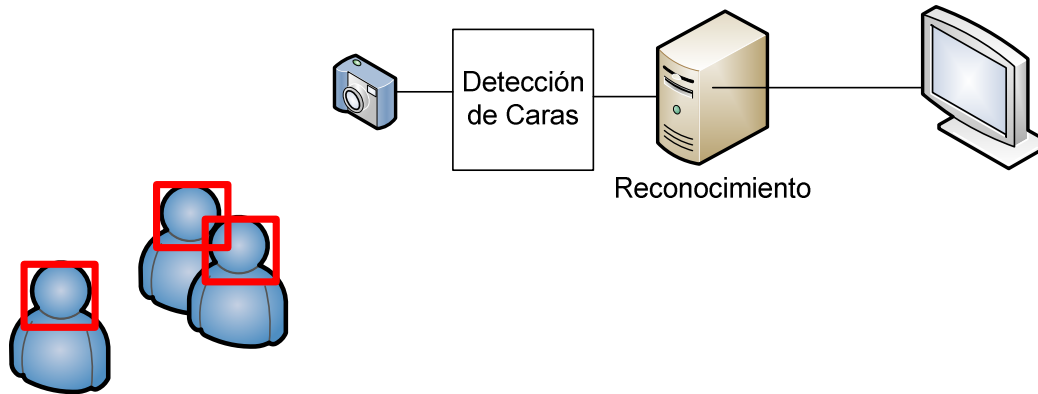
Para realizar el modelado se ha hecho uso de las funciones implementada por la Universidad Politécnica de Virginia [52].

## 4.3 Desarrollo de la aplicación de reconocimiento

La aplicación de reconocimiento, recibe imágenes preprocesadas capturadas directamente de una fuente de video. El preprocesado consiste en encontrar todas las posibles caras del fotograma o imagen.

Para éste PFC, al contar con imágenes con una sola cara, el preprocesado consiste en eliminar de la imagen todo lo que no sea la cara. Se ha empleado un detector de color carne para localizar la cara dentro del fotograma.

La arquitectura de esta aplicación es similar a la aplicación de modelado; la diferencia radica en que incorpora el módulo de reconocimiento en lugar del módulo para modelar al individuo.



#### 4.3.1 El módulo de reconocimiento emplea la técnica de reconocimiento de mezcla de gaussianas. Módulo de Reconocimiento

El proceso de reconocimiento es similar al proceso de modelado expuesto en el capítulo anterior. Toda imagen que entra en el sistema, pasa por el módulo de extracción de puntos característicos para a continuación calcular los ratios que definen la cara.

Con los ratios calculados, el módulo de reconocimiento trata de buscar el modelo que más se ajuste al vector de ratios calculado.

Para calcular qué modelo es el que mejor se ajusta al vector de ratios, se calcula la función de densidad de probabilidad de la mezcla de gaussianas. Siendo  $X$  el vector de ratios de la imagen a identificar, y  $\lambda_k$  los modelos de cada sujeto de la BBDD, se busca el modelo que tiene la máxima probabilidad como sigue:

$$\hat{S} = \arg \max_{1 \leq k \leq S} \{Pr(\lambda_k|X)\}$$

Suponiendo que los modelos de la BBDD son equiprobables y aplicando el Teorema de Bayes en la expresión anterior, la ecuación queda como sigue:

$$\hat{S} = \arg \max_{1 \leq k \leq S} \{p(X|\lambda_k)\}$$

Se ha empleado la función `gmm1pdf` del *Toolbox VoiceBox* de Matlab para realizar este cálculo.

## **5 Integración, pruebas y resultados**

---

### **5.1 Introducción**

En este capítulo se presentará la integración de las dos aplicaciones desarrolladas (detección de puntos/ratios y reconocimiento/identificación). Posteriormente se irán explicando los distintos problemas que han ido surgiendo mientras se integraban los distintos módulos (si aplica).

En el desarrollo de los distintos módulos se ha empleado la BBDD de Test, de forma que así los test son más reales, al ser el conjunto de desarrollo y prueba distintos. Los resultados que se presentan a continuación corresponden únicamente con las imágenes contenidas dentro de la BBDD de *Biosec*.

### **5.2 Sistema completo**

#### **5.2.1 Prefiltrado de imágenes**

Como en cualquier sistema real que se precie, se ha implementado un módulo para filtrar del sistema aquellas imágenes que no son procesables por el mismo.

En un sistema de identificación de acceso, este prefiltrado podría realizarse interactivamente indicando al usuario que volviera a posar delante de la cámara, en el caso de que la muestra no fuese válida.

Para este sistema, se han definido 6 filtros para eliminar de la BBDD, aquellas imágenes que no se deben tener en cuenta en el sistema.

El primer filtro, filtro de dimensiones, elimina aquellas imágenes que devuelve el detector de caras, cuya anchura es mayor o igual a la altura. Con esas dimensiones, se considera que la imagen que ha recibido el sistema, no es una cara, y si lo es, o no contiene toda la información de la cara, o la información que aporta, recoge la cara y otras cosas.

El segundo Filtro, filtro detección de ojos, elimina aquellas imágenes de las que no ha sido posible la detección de ojos.

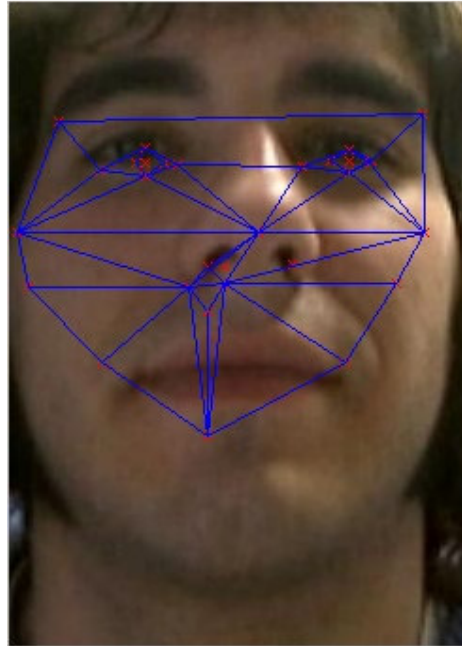
El tercer filtro, filtro de orificios nasales, elimina las imágenes en las que la detección del punto correspondiente a la nariz, no es muy fiable; es decir, elimina aquellas imágenes en las que no se detectan los orificios nasales.

El cuarto filtro, elimina aquellas imágenes cuya detección de puntos, ha dado como resultado algún tipo de error, del tipo, no se encuentran los puntos.

El quinto filtro elimina aquellas imágenes donde la máscara de puntos detectada, tiene algún punto fuera de los límites definidos de la imagen; en concreto, aquellas imágenes que tiene algún punto donde alguna de sus coordenadas o tiene valor negativo o tiene una valor superior a alguna de las dimensiones de la imagen (la componente x del punto no puede ser superior a

ancho de la imagen; y la componente y, no puede ser superior al alto de la imagen).

El sexto filtro elimina aquellas imágenes que habiendo detectado satisfactoriamente los puntos característicos, éstos no estén correctamente definidos; e.g. la boca situada en el eje izquierdo/derecho de la cara.



**Figura 66: Ejemplo de imagen dónde los puntos están mal ubicados**

Este Sexto filtrado ha sido incluido con posterioridad a raíz de los resultados obtenidos. Se ha observado que incluyendo este filtro se aumentaba considerablemente las probabilidades de reconocimiento.

De las 1144 imágenes de la BBDD, el resultado de los filtros aplicado ha sido el siguiente:

- 211 imágenes han sido eliminada por el primer filtro. En concreto un 18,44% del total de la BBDD
- 342 imágenes han sido eliminada por el segundo filtro. En concreto un 30,07% del total de la BBDD
- 95 imágenes han sido eliminada por el tercer filtro. En concreto un 8,13% del total de la BBDD
- 16 imágenes han sido eliminada por el cuarto filtro. En concreto un 1,14% del total de la BBDD
- 116 imágenes han sido eliminada por el quinto filtro. En concreto un 10,05% del total de la BBDD
- 32 imágenes han sido eliminada por el sexto filtro. En concreto un 0,61% del total de la BBDD



El tiempo de procesamiento para el prefiltrado de las 1144 imágenes ha sido de 45m 33,672s. Se ha empleado para ello una maquina con 2Gb de Ram y un procesador Intel Corel Duo 2 7200.

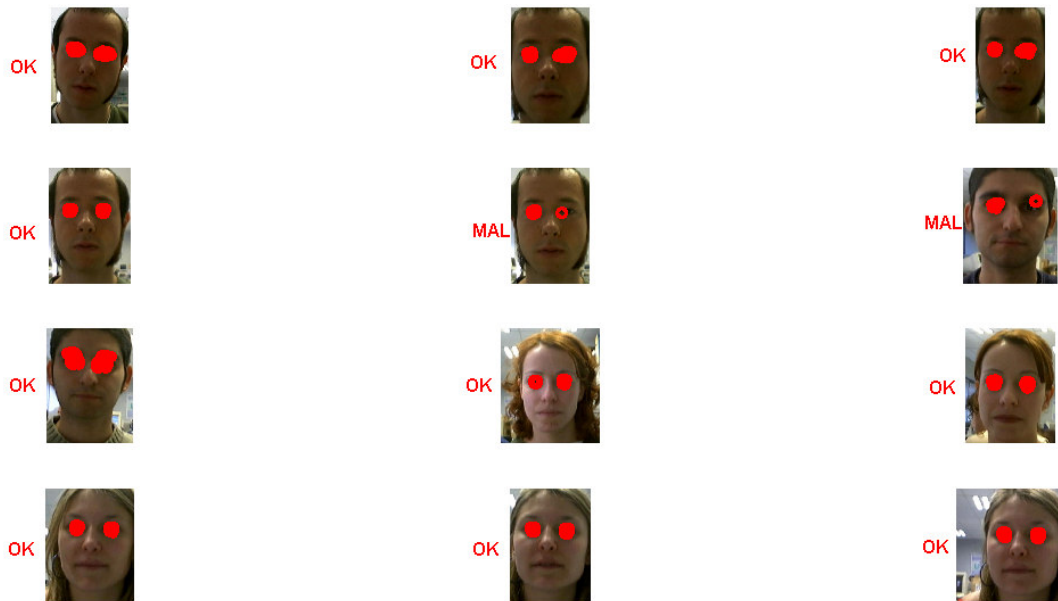
Después de estos filtrados, de las 1144 imágenes queda un total 332 imágenes correspondientes a 118 individuos distintos frente a los 200 originales.

A la hora de modelar a los 118 individuos restantes, sólo se procesan aquellos que presentan al menos 4 muestras. Con esta última premisa únicamente se consiguen modelar 34 de los 118 individuos

## 5.2.2 Sistema Localización de Ojos

Del total de imágenes restantes contando con las imágenes eliminadas por el tercer, cuarto y quinto filtro (se cuentan, porque su eliminación es posterior a la localización de los ojos), un total de 589, el porcentaje de acierto en la localización es de un 83,65%. El análisis para determinar cuando la localización es correcta, se ha realizado manualmente como se describe a continuación.

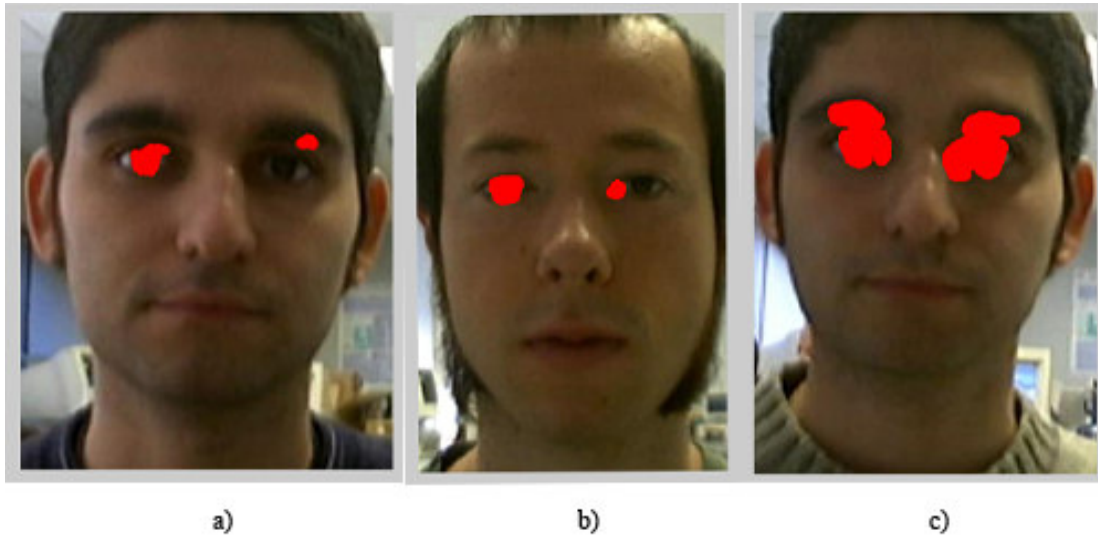
En la Figura 67, se muestra el resultado de la localización sobre una muestra reducida de la BBDD de *Biosec*.



**Figura 67: Resultado de la localización de los ojos en un conjunto 12 imágenes de la BBDD de *Biosec***

Se considera que la detección es errónea, no sólo cuando detecta algo que no son los ojos, caso de la Figura 68 a), sino también, cuando lo que detecta,

aunque sea una parte del ojo, no sea ni la pupila (ver Figura 68 b)), ni la totalidad del ojo (ver Figura 68 c)).



**Figura 68: Posibles casos en la localización: Figura a) Detección errónea, no se ha detectado bien los ojos; Figura b) Aunque se detecta una parte del ojo, la localización no se considera correcta; Figura c) Detección correcta, se detecta la totalidad de los ojos**

El tiempo de ejecución para la identificación de los ojos en las 1144 imágenes de la BBDD (sin filtrar), ha sido de: 26m, 22,688s. Se ha empleado para ello una maquina con 2Gb de Ram y un procesador Intel Corel Duo 2 7200.

### 5.2.3 Pruebas de reconocimiento y resultados

Se han realizado pruebas tomando distintos número de gaussianas por modelo.

Las imágenes usadas para modelar cada usuario se han escogido de manera secuencial; es decir, si consideramos 1 gaussiana por modelo la primera imagen se usa para el modelado y el resto para el reconocimiento, si tomamos 2 gaussianas por modelo, las dos primeras imágenes se utilizan para el modelado y el resto para el reconocimiento; así sucesivamente.

Como se ha comentado en el apartado 5.2 se ha incluido un nuevo filtro que permite mejorar considerablemente la tasa de acierto. Este filtro, lo único que hace es eliminar aquellas imágenes que aún teniendo detectados todos los puntos, estos no estén correctamente ubicados.

Los resultados que a continuación se muestran corresponden con pruebas realizadas únicamente con 10 ratios. Los ratios empleados son los siguientes:

Ratios	Descripción
<b>Ratio 1</b>	Separación de ojos/Longitud de la Boca
<b>Ratio 2</b>	Separación de ojos/Altura de la Nariz
<b>Ratio 3</b>	Separación de ojos/Distancia del ojo izquierdo a la comisura izquierda de la boca
<b>Ratio 4</b>	Separación de ojos/Distancia del ojo derecho a la comisura derecha de la boca
<b>Ratio 5</b>	Longitud de la Boca/Altura de la Nariz
<b>Ratio 6</b>	Longitud de la Boca/Distancia del ojo izquierdo a la comisura izquierda de la boca
<b>Ratio 7</b>	Longitud de la Boca/Distancia del ojo derecho a la comisura derecha de la boca
<b>Ratio 8</b>	Altura de la Nariz/Distancia del ojo izquierdo a la comisura izquierda de la boca
<b>Ratio 9</b>	Altura de la Nariz/Distancia del ojo derecho a la comisura derecha de la boca
<b>Ratio 10</b>	Distancia del ojo izquierdo a la comisura izquierda de la boca/Distancia del ojo derecho a la comisura derecha de la boca

**Tabla 4 : Descripción de los Ratios**

Siendo  $KP(X)$  (*KeyPoint*) el punto característico  $X$ , cada numerador y denominador se calcula de la siguiente manera:

*Separación ojos*

$$= \frac{1}{3} \{ [KP(5) - KP(3)] + [KP(6) - KP(4)] + [KP(22) - KP(17)] \}$$

$$\text{Longitud de la boca} = KP(11) - KP(10)$$

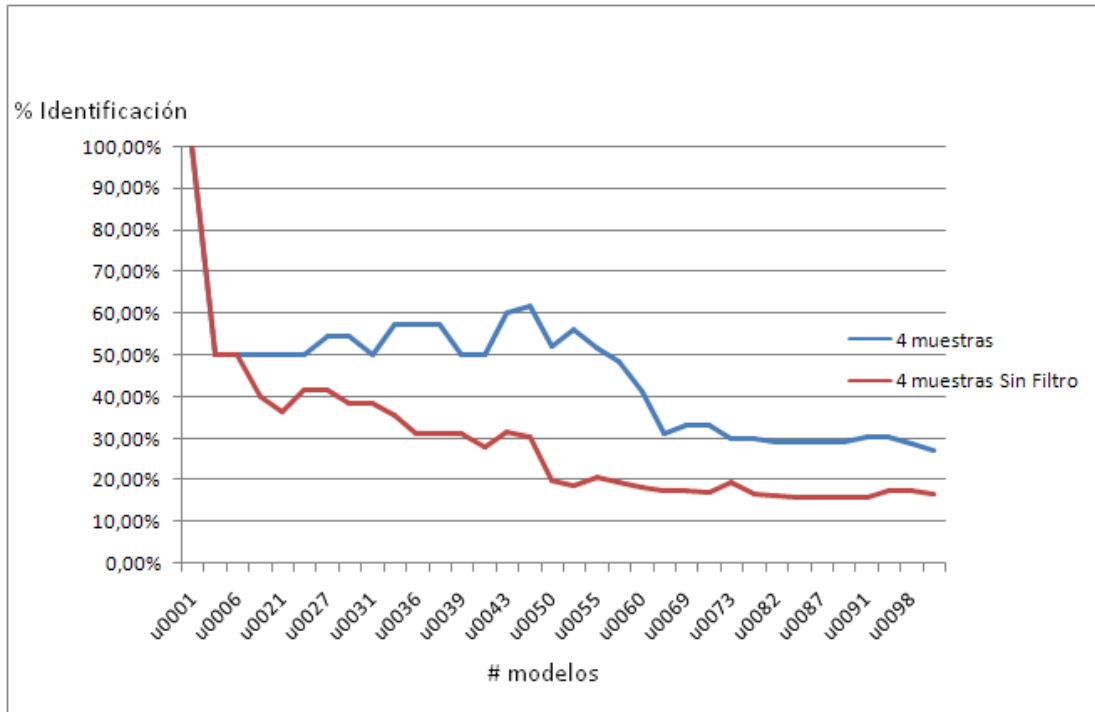
$$\text{Altura de la Nariz} = KP(8) - KP(17)$$

$$\text{Distancia del ojo izquierdo a la comisura izquierda de la boca} = KP(10) - KP(17)$$

$$\text{Distancia del ojo derecho a la comisura derecha de la boca} = KP(11) - KP(22)$$

En la Figura 69 se muestra el resultado de la identificación aplicado sobre la BBDD de *Biosec*, en el que se han tomado 4 muestras para modelar cada sujeto de la BBDD. Se puede observar que a medida que aumenta el número de modelos que se tienen para identificar a una persona (un total de 34), la probabilidad de acierto va disminuyendo. En la misma figura, se muestra la

diferencia en el resultado de identificación si se obviase el sexto filtrado (línea roja de la Figura 69).



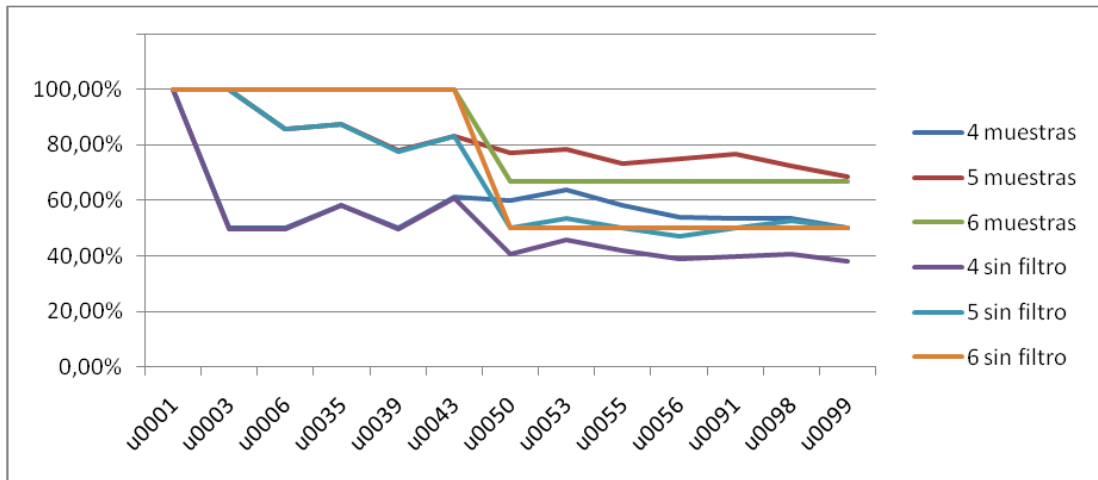
**Figura 69: Resultado de identificación tomando 4 muestras por modelo. El resultado se muestra tanto si se aplica el sexto filtro como si no.**

Los resultados que se exponen a continuación se han realizado sobre un total de 13 individuos frente a los 34 que quedan en la BBDD después de pasar por el prefiltrado.

Se han elegido estos 13 individuos debido a que de ellos se tiene al menos 6 muestras válidas con lo que permite comparar la tasa de acierto a medida que se aumenta el número de muestras que se emplean para modelar al sujeto.

En el gráfico de la Figura 70 se muestra el resultado de la identificación cogiendo para estos sujetos 4, 5 y 6 muestras para modelarlos y el resto de imágenes de la BBDD para el reconocimiento.

En el mismo gráfico se ha muestra el resultado tanto si se aplica el sexto filtro como si no.



**Figura 70: Resultado de la identificación de 13 individuos de la BBDD de *Biosec* modelándolos con distinto número de muestras.**

De la Figura 70 se puede observar que la probabilidad de acierto aumenta cuantas más muestras se empleen para modelar al individuo.



## **6 Conclusiones y trabajo futuro**

---

### **6.1 Conclusiones**

En el presente proyecto se han estudiado los sistemas de reconocimiento de caras basados en técnicas de reconocimiento 2D, 3D y 2D+3D; dado que de los sistemas estudiados, ninguno se adaptaba al objetivo del presente proyecto, extracción de puntos característicos de forma automática, se ha optado por desarrollar una serie de algoritmos de tratamiento digital de imagen que, partiendo de los puntos característicos definidos en el estándar MPEG-4 [27], extraiga un conjunto de estos puntos de forma automática.

Esta técnica, a diferencia de las otras estudiadas en el proyecto, permite una extracción automática de los puntos que definen una cara. Esta extracción automática de puntos se consigue, a base de dividir el problema, en otros más sencillos; para ello, cada imagen, es segmentada en los rasgos que la definen: los ojos, la nariz y la boca.

Otra de las ventajas que presenta, es que cada individuo se puede modelar con independencia del resto, lo que permite, actualizar la BBDD de modelos con una cierta flexibilidad y rapidez al no haber dependencias entre ellos.

El pilar sobre el que se sustenta la aplicación, es la detección de los ojos en la imagen. El sistema de detección desarrollado, basado en la utilización de los momentos de Hu, ha permitido obtener una tasa de acierto en la localización de los ojos bastante elevada, en torno al 83%. Se puede afirmar, de acorde a los resultados objetivos, que aunque la aplicación se sostenga en la detección de los ojos, éstos no representan un hándicap en la misma.

A partir de los puntos característicos extraídos en la imagen, se ha empleado un sistema de reconocimiento de caras basado en GMM por ser ésta técnica la más extendida en los actuales sistemas de reconocimiento.

Pese a que los resultados obtenidos, no sean equiparables a otras técnicas de reconocimiento; en general, se podría decir, que el sistema desarrollado cumple con los objetivos propuestos.

### **6.2 Trabajo futuro**

Gracias al desarrollo modular del sistema, cualquier parte de ambas aplicaciones (identificación/modelado y reconocimiento) podría ser susceptible de ser mejorada de forma independiente.

Como toda nueva técnica que se desarrolla, presenta una serie de debilidades frente a las ya instauradas y estudiadas en el pasado. Uno de los puntos débiles más importantes de esta técnica, es la dependencia que tiene con las condiciones del entorno en el que se toma la fotografía. Una iluminación irregular de la cara, la expresión, la pose, los accesorios, el punto de vista, etc; hacen que la extracción de los puntos sea una tarea ardua y costosa que en la mayoría de los casos, impiden una correcta localización de éstos.

Aunque los resultados no sean equiparables a los actuales sistemas de reconocimiento de caras, se podría continuar trabajando en el mismo, perfeccionando el sistema para validar la correcta colocación de la máscara de puntos, así como, el sistema para detectar los ojos en la imagen; siempre y

cuando se orientase el PFC hacia otras áreas que no sean el reconocimiento o la identificación de determinados individuos en una escena, e.g., detección de estados de ánimo, detector de sonrisas, etc.



## Referencias

---

- [1] [http://www.europe-nikon.com/downloads/Q==/DaJquIALgFokITW7BKBjltroeYJ91rnB/beryRVQ8FFk6s5QNjZJHyhwPphxVyLHNO/KZ5oxYHH~sMbtWvm0~293/L2L3L4\\_8pg\\_FINAL\\_ES.pdf](http://www.europe-nikon.com/downloads/Q==/DaJquIALgFokITW7BKBjltroeYJ91rnB/beryRVQ8FFk6s5QNjZJHyhwPphxVyLHNO/KZ5oxYHH~sMbtWvm0~293/L2L3L4_8pg_FINAL_ES.pdf)
- [2] [www.Tineye.com](http://www.Tineye.com)
- [3] <http://www.motorpasion.com/2007/08/04-un-prototipo-de-nissan-detecta-si-estas-borracho-y-no-te-deja-conducir>
- [4] T. Kanade. *Picture Processing by Computer Complex and Recognition of Human Faces- Ph. D. thesis, Kyoto University, 1973*
- [5] <http://casino-group.blogspot.com/2007/12/el-venetian-macao-resort-hotel-utiliza.html>
- [6] M. Turk, A. Pentland, *Eigenfaces for Recognition, Journal of Cognitive Neuroscience*, Vol. 3, No. 1, 1991, pp. 71-86
- [7] J. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, *Face Recognition Using LDA-Based Algorithms, IEEE Trans. on Neural Networks*, Vol. 14, No. 1, January 2003, pp. 195-200
- [8] <http://www.pages.drexel.edu/~sis26/Eigenface%20Tutorial.htm>
- [9] Zhujie, Y. L. Yu, "Face Recognition with Eigenfaces", *Industrial Technology*, 1994. *Proceedings of the IEEE International Conference on 5-9 Dec. 1994 pp. 434 – 438*
- [10] A.M. Martinez, A.C. Kak, *PCA versus LDA, IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 2, 2001, pp. 228-233
- [11] Peter N. Belhumeur, João P. Hespanha, and David J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, July 1997
- [12] Lizama, E., Waldoestl, D., Nickolay, B, "An eigenfaces-based automatic face recognition system", *Systems, Man, and Cybernetics*, 1997. 'Computational Cybernetics and Simulation'., 1997 *IEEE International Conference on Volume 1, 12-15 Oct. 1997 pp. 174 - 177 vol.1*
- [13] A. Amira, P. Farrell, "An automatic Face Recognition System Based on Wavelet Transforms", *School of Computer Science, Institute of Electronics, Communications and Information Technology (ECIT), Queen Univerity of Belfast, 2005*

- [14] L. Wiskott, J.-M. Fellous, N. Krueger, C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching", Chapter 11 in *Intelligent Biometric Techniques in Fingerprint and Face Recognition*, eds. L.C. Jain et al., CRC Press, 1999, pp. 355-396
- [15] L. Wiskott, J.-M. Fellous, N. Krueger, C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, 1997, pp. 776-779
- [16] T.F. Cootes, C.J. Taylor, "Statistical Models of Appearance for Computer Vision", *Technical Report, University of Manchester*, 125 pages
- [17] T.F. Cootes, K. Walker, C.J. Taylor, "View-Based Active Appearance Models", *Proc. of the IEEE International Conference on Automatic Face and Gesture Recognition*, 26-30 March K, Grenoble, France, pp. 227-232
- [18] M. Íñiguez, E. Bernués, I. Gil Lorente, "Aplicación para control de acceso mediante reconocimiento facial", Centro Politécnico Superior, Departamento de Electrónica y Comunicaciones, Universidad de Zaragoza.
- [19] S.V.Rakovic, P. Grieder, M. Kvasnica, D. Q. Mayne, M. Morari, "Computation of Invariant Sets for Piecewise Affine Discrete Time Systems subject to Bounded Disturbances", *43rd IEEE Conference on Decision and Control December 14-17, 2004 Atlantis, Paradise Island, Bahamas*
- [20] F.L. Bookstein, "Principal Warps: Thin-Plate Splines and the Decomposition of Deformations", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 11, Issue 6, June 1989 pp. 567 - 585
- [21] M. B. Stegmann, "Active Appearance Models: Theory, Extensions and Cases", *Informatics and Mathematical Modelling, Technical University of Denmark, DTU*, 2000.
- [22] Xu C., Wang Y., Tan T., Long Q., "A new Attempt to Face Recognition Using 3D Eigenfaces", *the 6th Asian Conference on Computer Vision (ACCV)*, Vol. 2, pp.884-889, 2004.
- [23] Blanz V. and Vetter T., "Face Recognition Based on Fitting a 3D Morphable Model", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063-1074, September 2003.
- [24] Ansari A., Abdel-Mottaleb M., "3D Face Modeling Using Two Orthogonal Views and a Generic Face Model", *Proc. Of the International Conference on Multimedia and Expo (ICME '03)*, Vol. 3, pp.: 289-292, July 2003.

- [25] Wang Y, Chua C., Ho Y., "Facial Feature Detection and Facial Recognition from 2D and 3D images", *Pattern Recognition Letters*, 23:1191-1202, 2002
- [26] Chang K. I., Bowyer K. W. and Flynn P. J., "Multi-Modal 2D and 3D Biometrics for Face recognition", *Proc. of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG'03)*, pp. 187– 194, October 2003.
- [27] F. Pereira, T. Ebrahimi, "The MPEG-4 Book", Prentice Hall, Upper Saddle River, 2002.
- [28] L. Zhang and P. Lenders, "Knowledge-Based Eye Detection for Human Face Recognition", *Fourth International conference on Knowledge-Based Intelligent Engineering Systems 6 AIM Tecluwbgws, 3@ Aug-1 Sept 2000, Bngkrm.UK, School of Mathematical and Computer Sciences, University of New England, Armidale, NSW 2351, Australia, (liming, pat)@mcs.une.edu.au*
- [29] J. S. Noh and K. H. Rhee, "Palmpoint Identification Algorithm using Hu Invariant Moments", *Dept. of Electronic Eng., College of Elec-Info Eng., Chosun University njinsoo@hanmail.net and [khrhee@chosun.ac.kr](mailto:khrhee@chosun.ac.kr)*
- [30] J. Fierrez, J. Ortega-Garcia, D. Torre-Toledano and J. Gonzalez-Rodriguez, "BioSec baseline corpus: A multimodal biometric database", *Pattern Recognition*, Vol. 40, n. 4, pp. 1389-1392, April 2007.
- [31] R. Sivaramakrishna, "Hu's moment invariants: How invariant are they under skew and perspective transformations?", *Department of Electrical & Computer Engineering and Biosystems Engineering, University of Manitoba, Winnipeg, Manitoba, Canada*
- [32] Chang K. I., Bowyer K. W. and Flynn P. J., "Multi-Modal 2D and 3D Biometrics for Face recognition", *Proc. of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG'03)*, pp. 187– 194, October 2003.
- [33] N. Eveno, A. Caplier, P.-Y. Coulon, "A new color transformation for lips segmentation", *Laboratoire des Images et Signaux, Institut National Polytechnique de Grenoble LIS, INPG, 46 av. FClx Viallet, 38031 Grenoble Cedex, France.*
- [34] N. Eveno, A. Caplier, P.Y.Coulon, "Accurate and Quasi-Automatic Lip Tracking", *ieee transactions on circuits and systems for video technology*, vol. 14, no. 5, May 2004

- [35] A. Hulbert and T. Poggio, "Synthesizing a Color Algorithm From Examples", *Science*, Vol. 239, pp 482-485, 1998
- [36] Phong, Bui-Tuong, "Illumination for Computer-Generated Pictures", *Communications of the ACM*, vol. 18, no. 3, pp. 311-317, 1975
- [37] Jens-Rainer Ohm and Bela Makai, "Feature-Similarity Retrieval of Face Images Based on 2D Model Description", *Heinrich-Hertz-Institut, Image Processing Department, Einsteinufer 37, D-10587 Berlin*
- [38] [http://www.sic.rma.ac.be/~beumier/DB/3d\\_rma.html](http://www.sic.rma.ac.be/~beumier/DB/3d_rma.html)
- [39] B. Horn y B. Schunck. "Determining Optical Flow." *Artificial Intelligence*, No. 17, pp. 185-203, 1981.
- [40] M. K. Hu, "Pattern recognition by moment invariants," *Proc. IEEE*, Vol. 49, No. 9, pp. 1428, Sept. 1961.
- [41] D. A. Reynolds, R. C. Rose, and J. T. Smith. "PC based TMS320c30 implementation of the Gaussian mixture model text-independent speaker recognition system". *Int. Conf. Sig. Proc. Applications & Technology*, November 1992.
- [42] D. A. Reynolds and R. C. Rose. "Robust text-independent speaker identification using gaussian mixture speaker models". *IEEE Trans. on Speech, Audio Processing*, 3(1):72-83, January 1995.
- [43] Yang Cao, Yupin Luo, Shiyuan Yang, "Image Denoising With Gaussian Mixture Model," *cisp*, pp. 339-343, 2008 *Congress on Image and Signal Processing*, Vol. 3, 2008.
- [44] Jing Wu, William A. P. Smith, Edwin R. Hancock, "Gender Classification Using Principal Geodesic Analysis and Gaussian Mixture Models", *11th Iberoamerican Congress in Pattern Recognition, CIARP 2006, Cancun, Mexico, November 14-17, 2006*, pp. 58-67.
- [45] Jiri Grim, Petr Somol, Michal Haindl, Pavel Pudil, "Color Texture Segmentation by Decomposition of Gaussian Mixture Model", *11th Iberoamerican Congress in Pattern Recognition, CIARP 2006, Cancun, Mexico, November 14-17, 2006*, pp. 287-296
- [46] Xue Jianru, Geng Xinli, Zheng Nanning, "TWS tracking techniques based on adaptive Gaussian mixture model in phased array radar" *Intelligent Control and Automation, 2002. Proceedings of the 4th World Congress on Vol. 4, Issue, 2002 pp. 3166 - 3170.*
- [47] <http://webdiis.unizar.es/~neira/12082/conectividad.pdf>
- [48] [http://es.wikipedia.org/wiki/Integraci3n\\_de\\_Riemann](http://es.wikipedia.org/wiki/Integraci3n_de_Riemann)

- [49] [http://www.pages.drexel.edu/~weg22/can\\_tut.html](http://www.pages.drexel.edu/~weg22/can_tut.html)
- [50] Fisher, R. A., "*On the Mathematical Foundations of Theoretical Statistics.*" *Philosophical Transactions of the Royal Society of London, A* 222:309-368, 1922.
- [51] Dempster, A. P., Laird, N. M. Rubin, D. B., "*Maximum likelihood from incomplete data via the EM algorithm*", *Journal of the Royal Statistical Society, B*, 39, 1-38, 1977.
- [52] <http://mcliff.cob.vt.edu/>



## Glosario

---

PFC	Proyecto Fin de Carrera
BBDD	Base de Datos
JPG	<i>Joint Photographic Experts Group</i>
BMP	<i>Bits Maps Protocole</i>
PNG	<i>Portable Network Graphics</i>
GMM	<i>Gaussian Mixture Model</i>
KP	<i>KeyPoint</i>
PCA	<i>Principal Component Analysis</i>
LDA	<i>Linear Discriminant Analysis</i>
FDA	<i>Fisher Discriminant Analysis</i>
KL	<i>Karhunen-Loewe</i>
EBGP	<i>Elastic Bunch Graph Matching</i>
FBG	<i>Face Bunch Graph</i>
AAM	<i>Active Appearance Model</i>
RMA	<i>Royal Militar Academy</i>
NIST	<i>National Institute of Standards and Technology</i>
MPEG	Grupo de Expertos de Imágenes en Movimiento
ROI	Región de Interés
PR	Punto de Referencia
F	Constante de Proporcionalidad
ML	Máxima Verosimilitud
EM	Estimación-Maximización





## Apéndices

---

### A Conceptos Básicos

En el campo del tratamiento de imágenes, se han ido acuñando una serie muy amplia de términos, principalmente en inglés, y que han sido trasladados al español con menor o mayor acierto. A continuación se definen algunos de esos términos y algunos conceptos relacionados con reconocimiento de caras, mostrando los términos tanto en español como en inglés:

- **Imagen (*Image*):** Es la proyección en perspectiva en el plano bidimensional de una escena tridimensional en un determinado instante de tiempo  $t_0$ .
- **Pixel (*Pixel: Picture Element*):** Es cada una de las posiciones en que es discretizada una imagen, o lo que es lo mismo, cada una de las posiciones de la imagen, cada uno con un cierto valor lumínico.
- **Imagen binaria (*Binary image*):** Son aquellas imágenes cuyos *pixels* sólo tienen dos valores: cero y uno.
- **Histograma (*Histogram*):** Es la representación gráfica de las frecuencias relativas con las que aparecen los distintos colores en una determinada imagen. Proporciona información sobre el brillo y el contraste de la imagen, y puede ser utilizado para ajustar estos parámetros, eliminar ciertas tonalidades molestas, etc...
- **Umbral (*Threshold*):** La forma más sencilla de obtener una imagen monocromática es fijar un nivel de gris que constituye el umbral que nos indica si el punto se pinta. Por ejemplo, si pintamos puntos blancos sobre fondo negro, una vez fijado el umbral, los puntos cuya intensidad lo sobrepase se pintarán de blanco, los demás no aparecerán, por lo tanto un umbral bajo producirá una imagen muy clara.
- **Flujo óptico (*Optical flow*):** Es la velocidad aparente de las estructuras de niveles de gris. Es un *array* bidimensional de vectores. El flujo óptico puede considerarse como una aproximación del campo de movimiento, pero a diferencia de aquel puede ser obtenido de una secuencia de imágenes. Las variaciones en los niveles de gris pueden ser debidas al movimiento de los objetos en la secuencia de imágenes, por ello a partir del flujo óptico es posible determinar el movimiento de los objetos en secuencias de imágenes

## B Modelo de mezcla de Gaussianas – GMM (*Gaussian Mixture Models*)

Los sistemas de reconocimiento de locutor basados en GMMs asumen que la Probabilidad viene dada por una mezcla de distribuciones Gaussianas.

Dado un sujeto  $X$  y su correspondiente vector de características  $X = \{X_t, t = 1 \dots N\}$  de dimensión  $D$ , la densidad mezcla de Gaussianas utilizada como función de verosimilitud se define como:

$$\rho(x|\lambda) = \sum_{i=1}^M w_i p_i(x)$$

Siendo  $w_i$  el peso de cada una de las mezcla de gaussianas y  $p_i(x)$  las  $M$  densidades componentes, la función de verosimilitud se trata como de una suma ponderada de las  $M$  densidades componentes.

Los pesos cumplen que  $\sum_{i=1}^M w_i = 1$

Cada gaussiana viene definida por:

$$p_i(x) = \frac{1}{(2\pi)^{D/2} \cdot |\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(x - \mu_i)' \Sigma_i^{-1} (x - \mu_i)\right\}$$

donde  $\lambda_i$  es el vector de medias es el vector de medias de dimensión  $D \times 1$  y  $\Sigma_i$  es la matriz  $D \times D$  de covarianzas.

Definida cada gaussiana, el entrenamiento se ha realizado mediante estimaciones de máxima verosimilitud (*Maximum Likelihood*, ML) a través del algoritmo EM (Estimación-maximización). Con este método de manera iterativa se refinan los parámetros del GMM para que aumente la probabilidad de generar el vector de características  $X$  dado el modelo:

$$\rho(x|\lambda^{(k+1)}) \geq \rho(x|\lambda^k)$$

## Anexos

---

### A Manual de instalación y usuario

Para la instalación del sistema completo, se necesitara el sistema operativo Windows XP con Matlab instalado en el PC.

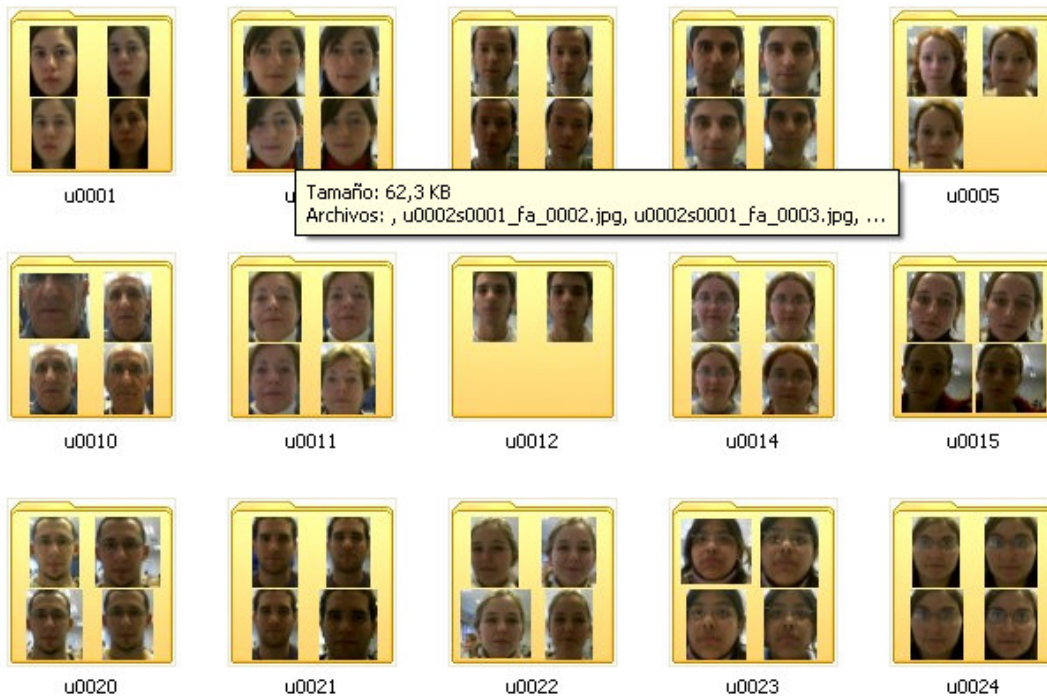
La aplicación de modelado solamente necesitara el fichero ejecutable ("interfazModelado".exe) que contendrá todas las rutinas para realizar el modelado del individuo/s indicado/s.



**Ilustración 1: Vista de la Aplicación de Modelado**

Para realizar el modelado de la BBDD, es necesario indicar en el cuadro de texto el *path* completo de la BBDD.

Los datos de entrada de la aplicación, imágenes de la BBDD, deberán estar separadas en distintas carpetas, siendo cada carpeta un individuo distinto a modelar.



**Ilustración 2: Datos de entrada**

Los únicos formatos de codificación de imagen permitidos son BMP, JPG y PNG.

La BBDD podrá estar colgada en el mismo directorio que la aplicación, en este caso con escribir el nombre de la carpeta que contiene la BBDD es suficiente para arrancar el programa; o bien puede estar ubicada en cualquier directorio del PC, en este caso es necesario escribir la ruta completa de la BBDD.

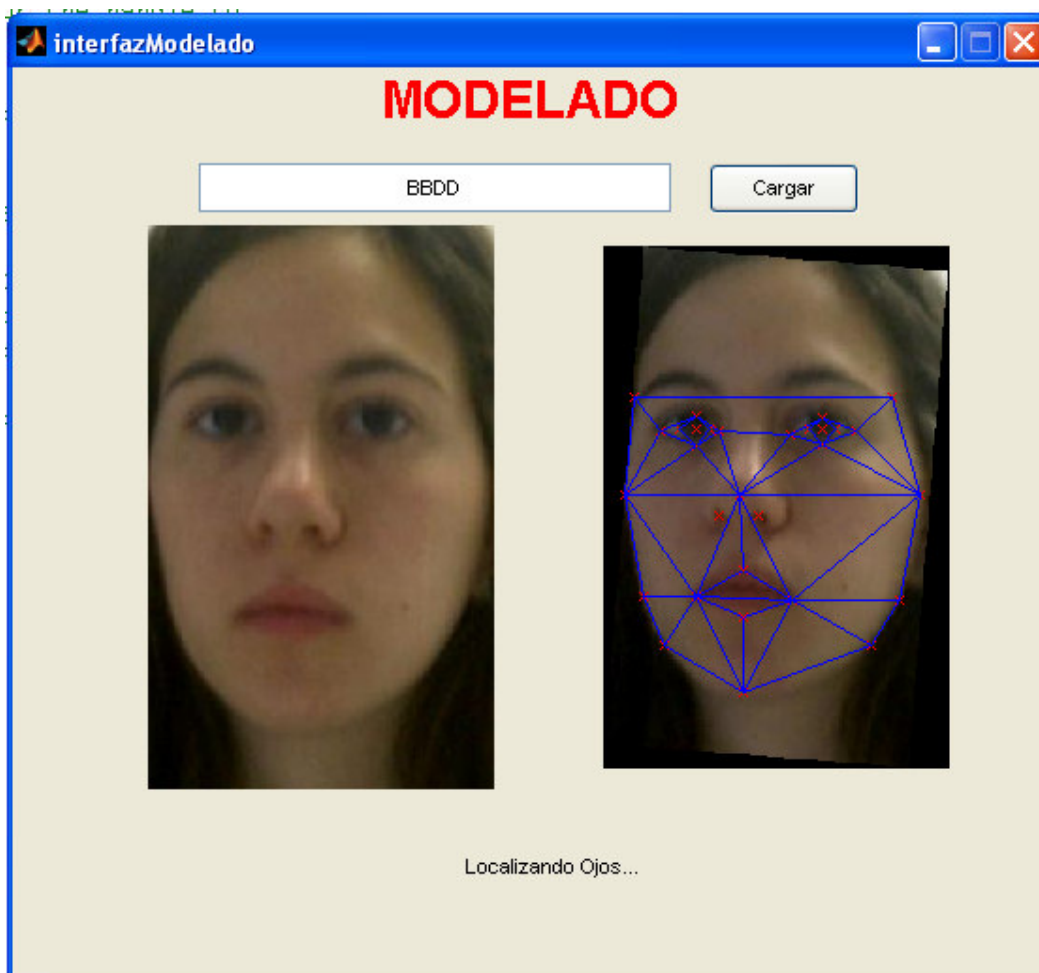
Los datos de salida que genera la aplicación se localizaran en distintas carpetas. En una carpeta temporal "tmp/" se van guardando las imágenes de la boca, los ojos, y la nariz extraídas de la imagen que se está analizando. En la carpeta "RotatesFaces/" se guardan las imágenes rotadas de la BBDD (si se han encontrado los ojos), clasificadas por cada individuo. En la carpeta "Keypoints/" se guardan la matriz de puntos característicos de cada imagen de la BBDD clasificados también por cada individuo. En la carpeta "ratios/" se almacenan los distintos ratios calculados, clasificados también por cada individuo. Por último, en la carpeta "models/" se almacenan el modelo de mezcla de gaussianas de cada individuo de la BBDD.

A lo largo de la ejecución del programa, se van mostrando tanto la imagen que se está analizando como el resultado de la extracción de puntos.

Debajo de las imágenes, se va mostrando las distintas etapas por las que va pasando la imagen que se está analizando. Los mensajes que aparecen son:

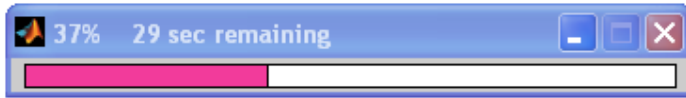
- Leyendo Imagen... → Inicio del proceso de Lectura de la imagen
- Leyendo Imagen... OK → Indicar que el proceso de Lectura ha sido satisfactorio
- Umbralizando Imagen... → Inicio del proceso de Umbralización.

- Umbralizando Imagen...OK → Fin del proceso de Umbralización.
- Localizando Ojos... → Inicio del proceso de Localización de los Ojos.
- Localizando Ojos... FAIL → Fallo en el proceso de localización de los Ojos.
- Localizando Ojos... OK → Fin del proceso de Localización de los Ojos.
- Cara Inclínada. Rectificando... → Inicio del proceso para rotar la imagen.
- Cara Inclínada. Rectificando... OK → Fin del proceso para rotar la imagen.
- Error en la extracción de puntos de la Boca → Error a la hora de localizar los puntos de la boca.
- Error en la extracción de puntos del Labio → Error a la hora de localizar los puntos de los labios.
- Error en la extracción de puntos de los ojos → Error a la hora de localizar los puntos de los ojos.
- Error en la extracción de puntos de la Nariz → Error a la hora de localizar los puntos de la nariz.
- Error en el proceso → Otro tipo de errores en la localización de los puntos.



**Ilustración 3: Captura del proceso de modelado de un individuo de la BBDD**

Existe además, una barra de progreso que permite ver, el tiempo estimado para terminar de modelar las distintas imágenes del mismo individuo.



**Ilustración 4: Barra de progreso de la aplicación**

La aplicación de identificación se lanza a través de otro ejecutable, "Identificación.exe".

Esta aplicación es necesaria configurarla con una serie de parámetros para indicar dónde tiene que ir a buscar los datos.

Para realizar la identificación del individuo, es necesario indicar en el cuadro de texto el *path* dónde se ubica las imágenes de las distintas personas que se desean identificar.

Los datos de entrada de la aplicación, imágenes a identificar, deberán estar ubicadas todas en la misma carpeta.



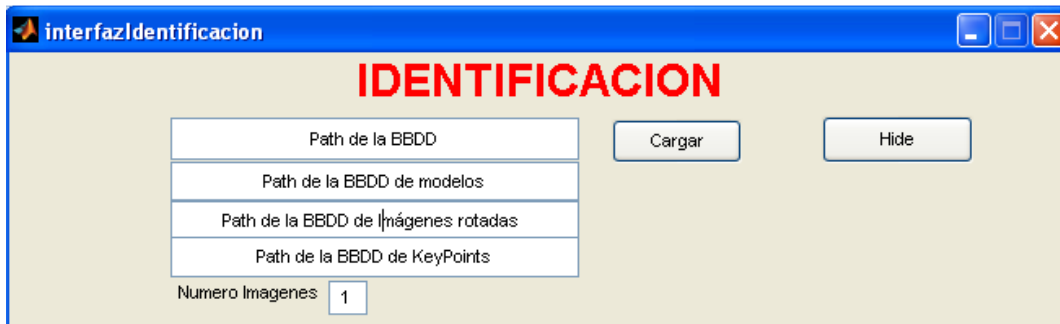
**Ilustración 5: Datos de entrada**

La vista se sencilla se presenta como sigue.



**Ilustración 6: Vista sencilla de la aplicación de Identificación**

Si es necesario modificar los parámetros de las rutas dónde la aplicación tiene que ir a buscar la información, es necesario pulsar el botón “Advance” y modificar las rutas necesarias. Además de los parámetros de las rutas, se puede parametrizar los ‘n’ individuos que más se parecen.



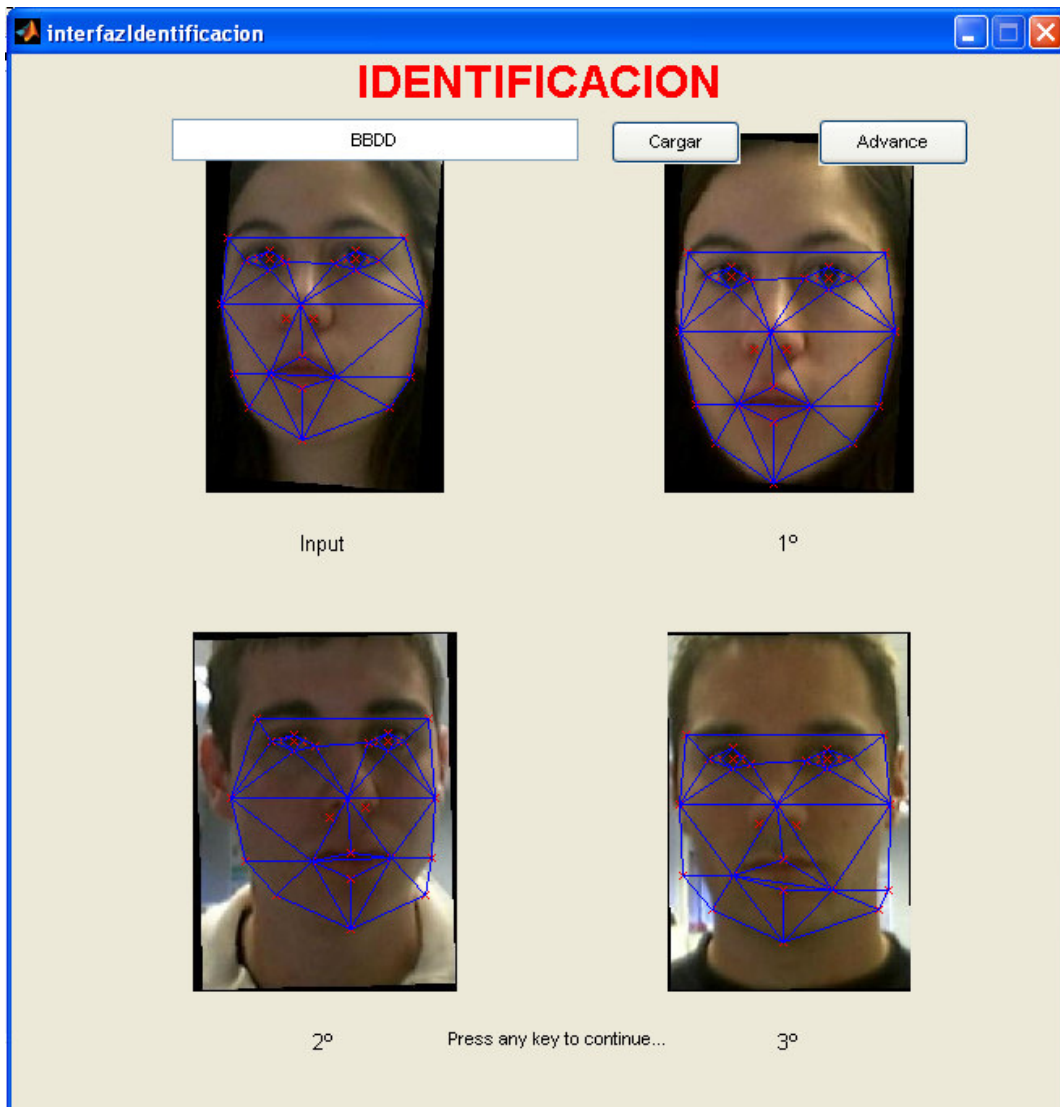
**Ilustración 7: Vista avanzada de la aplicación de Identificación.**

Los parámetros avanzados son:

- *Path* de la BBDD de Modelos: Ruta dónde se encuentra los modelos generados por la aplicación de Modelado.
- *Path* de la BBDD de imágenes Rotadas: Ruta dónde se encuentran las imágenes rotadas generadas por la aplicación de Modelado.
- *Path* de la BBDD de *KeyPoints*: Ruta dónde se encuentran los *keypoints* generados por la aplicación de Modelado.
- Número de imágenes: permite mostrar los ‘n’ primeros individuos que más se parecen.

Al igual que en la aplicación de modelado, a lo largo de la ejecución del programa, se van mostrando las distintas etapas por las que pasa la imagen que se está analizando: Leyendo Imagen..., Leyendo Imagen... OK, Umbralizando Imagen... Umbralizando Imagen...OK, Localizando Ojos..., Localizando Ojos... FAIL, Localizando Ojos... OK, Cara Inclinada. Rectificando..., Cara Inclinada. Rectificando... OK, Error en la extracción de puntos de la Boca, Error en la extracción de puntos del Labio, Error en la extracción de puntos de los ojos, Error en la extracción de puntos de la Nariz, Error en el proceso.

Una vez analizada la imagen, se procede a identificar ésta dentro de la BBDD. En función del número de imágenes que hemos querido mostrar, aparecerán en pantalla las 'n' primeros individuos que más se parecen a la imagen dada.



**Ilustración 8: Resultado de la Identificación. Se han mostrado los 3 primeros individuos de la BBDD que más se parecen. Arriba a la izquierda, la imagen a analizar.**



## B Manual del programador

El propósito de este manual del programador es dar a conocer al lector una pequeña guía de las principales funciones que conforman el sistema.

La estructura del manual se ha dividido en dos, para separar claramente las dos aplicaciones generadas: Aplicación de Identificación y Modelado; y la Aplicación de Reconocimiento.

### B.1 Aplicación de Identificación y Modelado.

```
function identificar(path,handles)
```

IDENTIFICAR

IDENTIFICAR (PATH,HANLDES) lee cada una de las imágenes contenida en el directorio indicado, las analiza y cuando termina con todas, las modela empleando la técnica de mezcla de Gaussianas, GMM.

Parámetros de entrada:

PATH --> Ruta o directorio que contiene la BBDD de imágenes. Los únicos formatos soportados son: JPG, BMP y PNG.  
HANDLES --> Puntero de control del GUI.

El diagrama de flujo que sigue esta función es el siguiente:

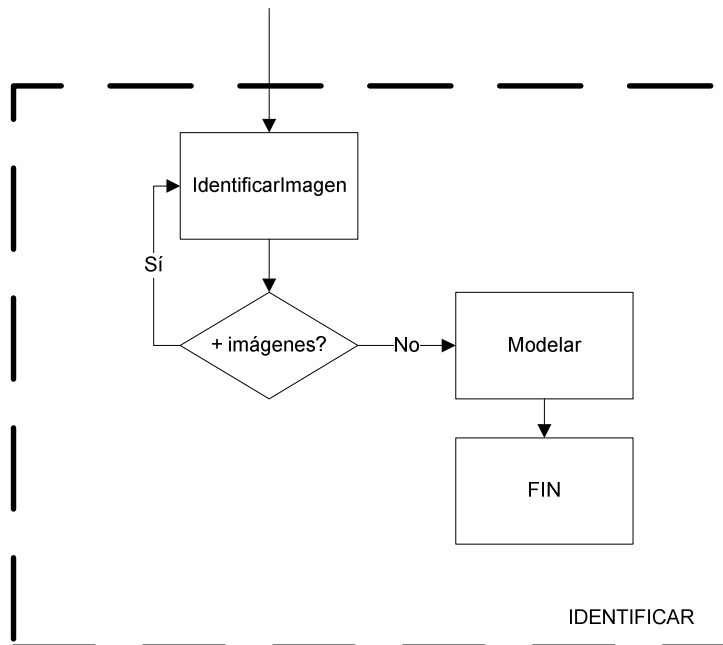


Ilustración 9: Diagrama de flujo de la función IDENTIFICAR

```
function identificarImagen(path,FILENAME,EXTENSION,handles)
```

#### IDENTIFICARIMAGEN

IDENTIFICARIMAGEN (PATH,FILENAME,EXTENSION,HANLDES) Analiza la imagen indicada para extraer de ella la información relativa sus ratios. Analizada la imagen se almacena la información acerca de sus Puntos Característicos, la Imagen Rotada y sus Ratios.

Parámetros de entrada:

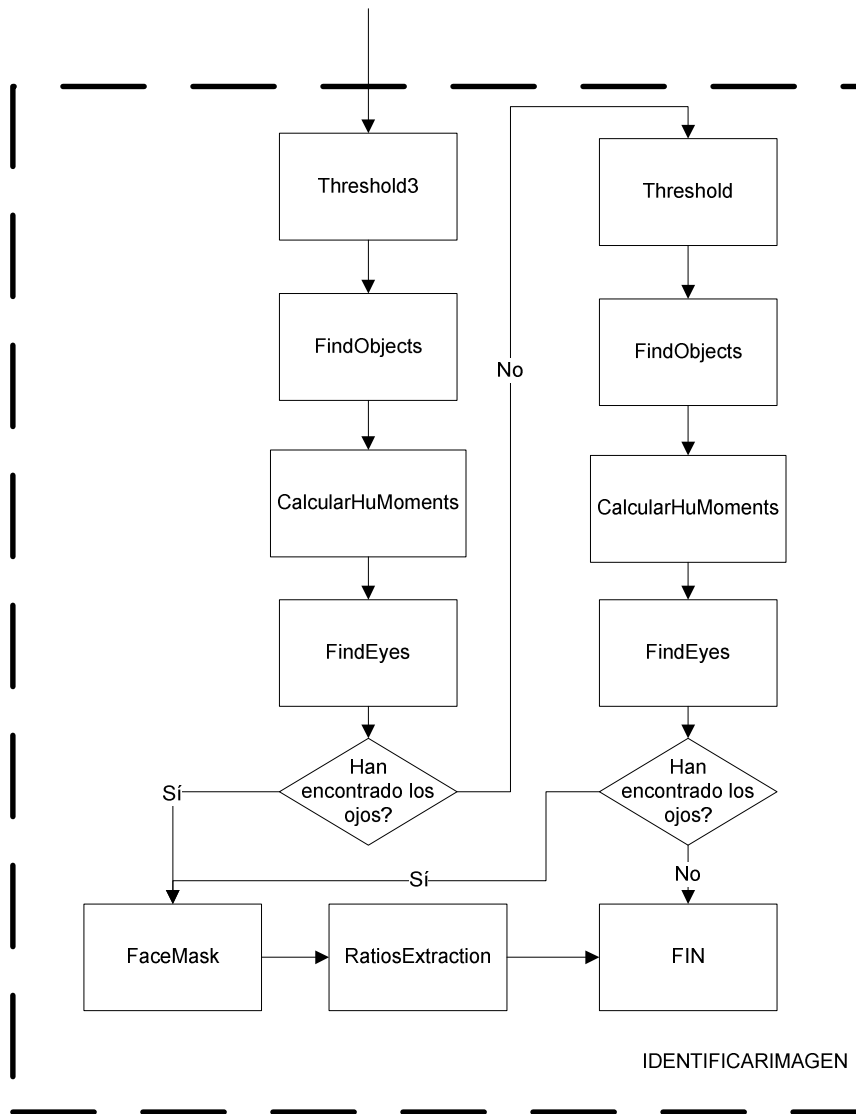
PATH --> Ruta donde está ubicada la imagen

FILENAME --> Nombre de la imagen.

EXTENSION --> Extensión (JPG, BMP o PNG).

HANDLES --> Puntero de control del GUI.

El diagrama de flujo que sigue esta función es el siguiente:



**Ilustración 10: Diagrama de Flujo de la función IDENTIFICARIMAGEN**

```
function modelar()
```

```
MODELAR
```

```
MODELAR () Lee los ratios almacenados en el directorio correspondiente (carpeta 'Ratios') y modela cada uno de los individuos empleando la técnica de Mezcla de Gaussianas. El modelo de cada individuo se almacena en la carpeta 'Models'.
```

```
function [th_value] = threshol3(Y)
```

```
THRESHOLD3
```

```
THRESHOLD3 (Y) Método rápido de umbralización. El valor umbral corresponde con la mitad del valor máximo del histograma.
```

```
Parámetros de entrada:
```

```
Y --> Luminancia de la imagen.
```

```
Parámetros de salida:
```

```
TH_VALUE --> Valor umbral calculado.
```

```
function[objetos, matrixNumeros]=findObjects(ima, Ojos);
```

```
FINDOBJECTS
```

```
FINDOBJECTS (IMA,SELECTLIM,OJOS) Analiza la imagen en BW para identificar en ella los distintos objetos que tiene("manchas negras"). Emplea la técnica de 8-conectividad para analizar e identificar los objetos de la imagen.
```

```
Parámetros de entrada:
```

```
IMA --> Imagen en BW a analizar
```

```
OJOS --> OJOS = 1 indica que se emplea para buscar ls ojos y que por lo tanto es necesario aplicar reglas para discriminar objetos por su tamaño. OJOS = 0 indica que no se emplea para buscar los ojos y por lo tanto no se excluyen objetos por su tamaño.
```

```
Parámetros de salida:
```

```
OBJETOS --> Estructura de datos que contiene la siguiente información: índice que identifica el objeto, Area del objeto, dimensiones del objeto y las coordenadas del centro de masas del objeto.
```

```
MATRIXNUMEROS --> Imagen con los objetos identificados. Cada objeto lleva un número que lo identifica.
```

```
function [MomentsHu]=calcularHuMoments(ima, objetos)
```

```
CALCULARHUMOMENTS
```

```
CALCULARHUMOMENTS (IMA,OBJETOS) Extrae de la imagen los momentos de Hu de cada uno de los objetos identificados.
```

```
Parámetros de entrada:
```

IMA --> Imagen en BW a analizar  
OBJETOS --> Estructura de datos que contiene la siguiente información: índice que identifica el objeto, Área del objeto, dimensiones del objeto y las coordenadas del centro de masas del objeto.

Parámetros de salida:

MOMENTSHU --> Matriz con los 7 momentos de Hu de cada uno de los objetos identificados en la imagen.

```
function[indCejas, indOjos]=findEyes(objetos,MomentsHu,dimensiones)
    [indCejas(i) indOjos(i)]=OjosCejas(indOjos(i),F);
```

FINDEYES

FINDEYES (OBJETOS,MOMENTOSHU,DIMENSIONES) Identifica los ojos en la estructura de objetos indicada.

Parámetros de entrada:

OBJETOS --> Estructura de datos que contiene la siguiente información: índice que identifica el objeto, Área del objeto, dimensiones del objeto y las coordenadas del centro de masas del objeto.  
MOMENTSHU --> Matriz con los 7 momentos de Hu de cada uno de los objetos identificados en la imagen.  
DIMENSIONES --> Dimensiones de la imagen en la cual se quiere encontrar los ojos.

Parámetros de salida:

INDCEJAS --> Índices de los objetos considerados como las cejas  
INDOJOS --> Índices de los objetos considerados como los ojos

```
function [th_value] = threshold(Y)
```

THRESHOLD

THRESHOLD (Y) Método lento de umbralización. El valor umbral corresponde con un valor entre los dos máximos del histograma.

Parámetros de entrada:

Y --> Luminancia de la imagen.

Parámetros de salida:

TH\_VALUE --> Valor umbral calculado.

```
function[estado,mascaraCara]=faceMask(Y,imaColor,centroOjos,alturaOjos)
)
```

FACEMASK

FACEMASK (Y, IMACOLOR, CENTROOJOS, ALTURAOJOS) Analiza la imagen y devuelve la máscara de puntos que la define

Parámetros de entrada:

Y --> Luminancia de la imagen.  
IMACOLOR --> Imagen a color a analizar.  
CENTROOJOS --> Coordenada X del centro de masas de cada Ojo.  
ALTURAOJOS --> Coordenada Y del centro de masas de cada Ojo.

Parámetros de salida:

ESTADO --> Indica si se ha habido errores en el proceso o si se han encontrado los puntos correctamente.  
MASCARACARA --> Máscara con los puntos característicos de la imagen.

```
function[ratios]=ratiosExtraction(mascaraCara)
```

RATIOSEXTRACTION

RATIOSEXTRACTION (MASCARACARA) Calcula los ratios de la imagen a partir de la máscara de puntos que la define.

Parámetros de entrada:

MASCARACARA --> Máscara con los puntos característicos de la imagen.

Parámetros de salida:

RATIOS --> Ratios de la imagen.

## B.2 Aplicación de Reconocimiento

```
function analisisIden  
(path,pathModels,handles,pathImaRot,pathKeyPointsRot,nResults)
```

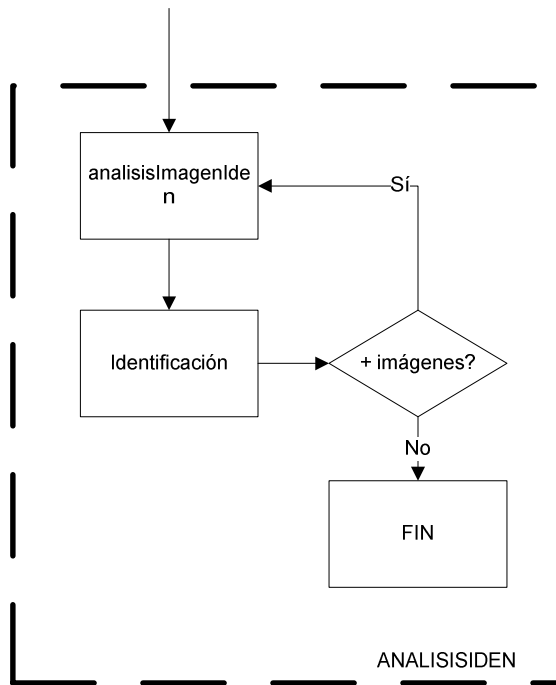
ANALISISIDEN

ANALISISIDEN (PATH,PATHMODELS,HANLDES, PATHIMAROT, PATHKEYPOINTSROT,NRESULTS) Lee cada una de las imágenes contenida en el directorio indicado (PATH), las analiza e identifica el modelo/s que más se parezca.

Parámetros de entrada:

PATH --> Ruta donde está ubicada las imágenes a reconocer  
PATHMODELS --> Nombre de ruta dónde se encuentran los modelos de cada uno de los usuarios registrados en la BBDD.  
HANDLES --> Puntero de control del GUI.  
PATHIMAROT --> Nombre de ruta dónde se encuentran las imágenes rotadas de cada uno de los usuarios registrados en la BBDD.  
PATHKEYPOINTSROT --> Nombre de ruta dónde se encuentran los puntos característicos de cada uno de los usuarios registrados en la BBDD.  
NRESULTS --> Número de resultados a mostrar en la ventana

El diagrama de flujo que sigue esta función es el siguiente:



**Ilustración 11: Diagrama de flujo de la función ANALISISIDEN**

```
function analisisImagenIde n(path,FILENAME,EXTENSION,handles)
```

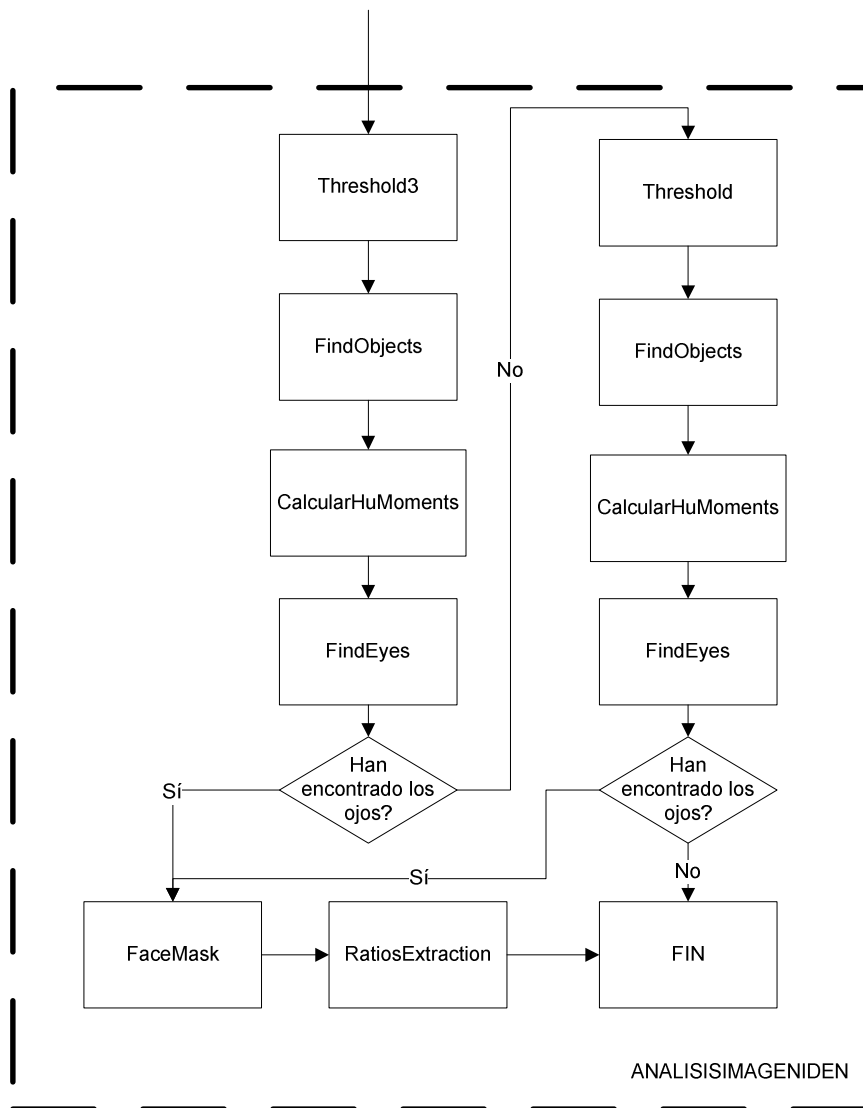
```
ANALISISIMAGENIDEN
```

```
ANALISISIMAGENIDEN(PATH,FILENAME,EXTENSION,HANLDES) Analiza la imagen indicada para extraer de ella la información relativa sus ratios. Analizada la imagen se almacena la información acerca de sus Puntos Característicos (carpeta 'KeyPoints'), la Imagen Rotada (carpeta 'RotatesFaces') y sus Ratios (carpeta 'Ratios').
```

```
Parámetros de entrada:
```

```
PATH --> Ruta donde está ubicada la imagen
```

FILENAME --> Nombre de la imagen.  
 EXTENSION --> Extensión (JPG, BMP o PNG).  
 HANDLES --> Puntero de control del GUI.



**Ilustración 12: Diagrama de flujo de la función ANALISISIMAGENIDEN**

```
function final_res = identificacion(ratios,models,nResults)
```

IDENTIFICACION

IDENTIFICACION(RATIOS,MODELS,NRESULTS) Identifica el usuario/s que más se parezca.

Parámetros de entrada:

RATIOS --> Ratios de la imagen a identificar.  
 MODELS --> Modelos de cada uno de los usuarios registrados en la BBDD.  
 NRESULTS --> Número de resultados a mostrar en la ventana

Parámetros de salida:

FINAL\_RES --> Nombre del usuario/s que más se parece.

## **C Generación de humanoides mediante descripciones de alto nivel en el marco de estándares MPEG**

Este anexo hace referencia al trabajo de iniciación a la investigación de Víctor Fernández-Carbajal Cañete:

- *Generación de humanoides mediante descripciones de alto nivel en el marco de estándares MPEG*, Víctor Fernández-Carbajales Cañete, Escuela Politécnica Superior, Universidad Autónoma de Madrid, curso 2005-2006, Junio 2006

. De este trabajo se ha hecho uso del estudio realizado sobre las relaciones de la cara.

### **C.1 Descripción de las Relaciones Corporales**

Tras el estudio de estos temas se han podido clasificar las relaciones corporales en tres grandes grupos: longitudinales, de anchura y de grosor.

Las relaciones longitudinales son aquellas que interconectan la longitud de un parte del cuerpo con otras (como puede ser la longitud del brazo con la longitud de la cabeza), la clave para estas relaciones es la altura del ser humano.

Las relaciones de anchura son aquellas que permiten obtener la anchura de una parte corporal mediante la anchura de otras (como poder obtener la anchura del torso a través de la anchura de otra/s), la clave para estas relaciones es la anchura de la cabeza, así como la altura del humanoide.

Las relaciones del grosor son aquellas que permiten obtener el grosor de una parte corporal mediante el estudio de cómo afecta el sistema muscular y la acumulación de grasa en esta, la clave de estas relaciones es la propia subdivisión corporal y como afecta los músculos y la acumulación de grasa corporal en la misma.

También tenemos que indicar otra división en lo que respecta a las relaciones corporales y es la división entre el tratamiento que tiene la cabeza con el resto del cuerpo. La cabeza posee sus propias relaciones corporales que son bastante independientes del resto del cuerpo, por la razón que la mayoría de la expresividad de un dibujo se encuentra en la cara del ser humano, de aquí esta haya recibido un tratamiento especial, por esta razón empezaremos indicando las relaciones corporales que tenemos en la cara del ser humano y posteriormente trataremos el cuerpo del mismo.

Todas las relaciones que tenemos en la cabeza del humanoide se obtienen a través de un valor F o proporción de frente, siendo este la distancia en centímetros entre el nacimiento y la línea imaginaria que cortaría el nacimiento de las cuencas oculares.

Pero antes de indicar las proporciones tenemos que definir una serie de términos para permitir una comprensión de las mismas:



- El origen de coordenadas se encontrara en la intersección de tres líneas imaginarias una que cortaría el final de la mandíbula del humanoide, otra que pasa por la parte más posterior del cráneo y la última que toca el final de la oreja derecha del humanoide.
- Los ejes quedan definidos de la siguiente forma:
  - El eje X estaría situada en la base de la cara frontal de la cabeza.
  - El eje Z estaría situada en la base de la cara de perfil de la cabeza.
  - El eje Y estaría situada como la altura de la cabeza (tanto de frente como de perfil).
- Definimos la anchura como cualquier medición en el eje X.
- Definimos la altura como cualquier medición en el eje Y.
- Definimos el fondo como cualquier medición en el eje Z.

Con este valor podemos indicar que la cabeza tiene las siguientes relaciones corporales:

- La cabeza de frente queda inscrita en un rectángulo cuya base (anchura de la cara) es de  $(2 + \frac{1}{2}) * F$  y de una altura (longitud de la cara) de  $(3 + \frac{1}{2}) * F$ .
- La cabeza de perfil queda inscrita en un cuadrado de  $(3 + \frac{1}{2}) * F$  unidades de lado.
- La oreja se encuadraría en un prisma de anchura  $\frac{1}{4}$  de F, un fondo de  $\frac{1}{2}$  de F y una altura de F. La posición inicial de la oreja izquierda se encontrara a una anchura de  $\frac{1}{4}$  de F, a una altura de F y a una fondo de  $(1 + \frac{1}{2})$  de F. En el caso de la oreja derecha solo cambia que se inicia a una anchura de  $(2 + \frac{1}{4})$  de F.
- La boca queda englobada en un prisma de una anchura aproximada de  $\frac{1}{2}$  de F, una altura de  $\frac{1}{4}$  de F y una fondo de  $\frac{1}{3}$  de F. La posición del extremo derecho de la boca es de una anchura de F, una altura de  $\frac{1}{2}$  de F y un fondo de  $(2 + \frac{5}{6})$  de F, mientras que el extremo izquierda queda a igual fondo y altura, pero con un ancho de  $(1 + \frac{1}{2})$  de F.
- La nariz queda encuadrada en un prisma de una anchura de  $\frac{1}{2}$  de F, una altura de F y un fondo de  $\frac{1}{3}$  de F. El punto medio de la nariz se encuentra a una anchura de  $(1 + \frac{1}{4})$  de F, una altura de F y un fondo de  $(2 + \frac{1}{2})$  de F.
- Cada uno de los ojos se puede introducir en un primas de anchura  $\frac{1}{2}$  de F, una altura de  $\frac{1}{4}$  de F y un fondo de  $\frac{2}{3}$  de F. Si los ojos están mirando al frente, la pupila del ojo derecho cae a una altura de  $(1 + \frac{7}{8})$  de F, un fondo de  $(2 + \frac{5}{6})$  de F y a un anchura de  $\frac{3}{4}$  de F, en el caso del ojo izquierdo solo varia su anchura, en este caso la anchura es  $(1 + \frac{3}{4})$  de F.
- Los pómulos del humanoide quedarían sitiados a una anchura de  $\frac{3}{4}$  de F el derecho y  $(1 + \frac{3}{4})$  de F el izquierdo, a una altura de  $(1 + \frac{1}{2})$  de F y a una fondo de  $(2 + \frac{5}{6})$  de F.
- La frente puede encuadrarse en un prisma cuadrado de una altura F, una anchura  $2 F$  y un fondo de  $1 + \frac{1}{2}$  de F.
- La barbilla del humanoide tiene una anchura de  $\frac{1}{2}$  de F en su parte más fina, hasta de  $2 F$  al final de la mandíbula, la altura de la misma se encuadra en un F, mientras que el fondo es de  $(1 + \frac{2}{3})$  de F.

Estas relaciones corporales quedan indicadas en la Ilustración 13 y en la Tabla 5.

Parte de la Cabeza	Anchura	Altura	Fondo
Cabeza	$(2 + \frac{1}{2}) F$	$(3 + \frac{1}{2}) F$	$(3 + \frac{1}{2}) F$
Oreja	$\frac{1}{4} F$	F	$\frac{1}{2} F$
Boca	$\frac{1}{2} F$	$\frac{1}{4} F$	$\frac{1}{3} F$
Nariz	$\frac{1}{2} F$	F	$\frac{1}{3} F$
Ojo	$\frac{1}{2} F$	$\frac{1}{4} F$	$\frac{2}{3} F$
Frente	2 F	F	$(1 + \frac{1}{2}) F$
Mandíbula	$\frac{1}{2} F$ a 2 F	F	$(1 + \frac{2}{3}) F$

Tabla 5.- Dimensiones de los prismas contenedores de los elementos principales de la cara

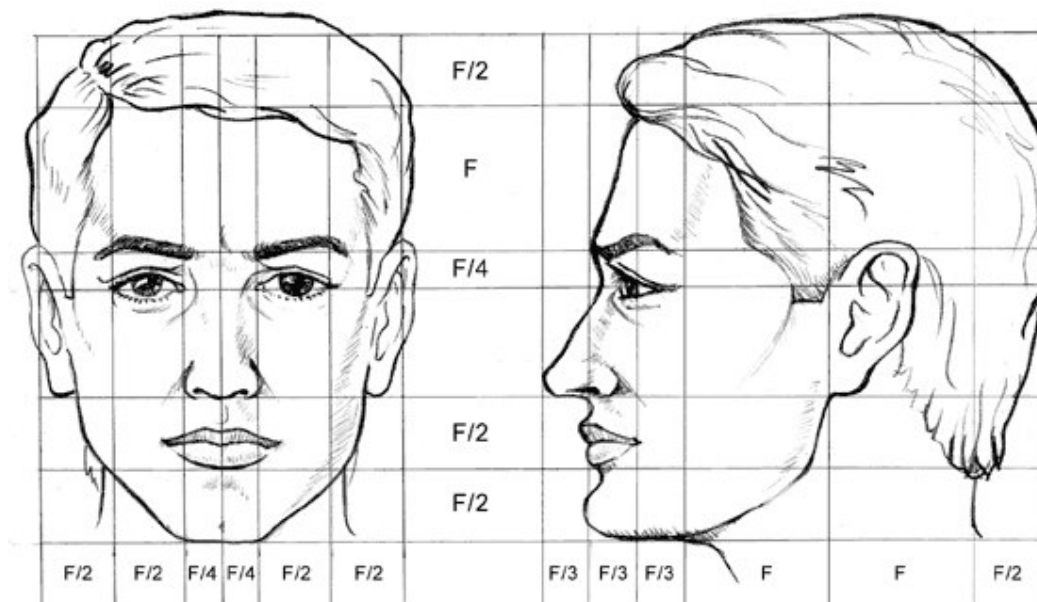


Ilustración 13.- Relaciones corporales de la cabeza. La proporción F.

Como podemos ver de esta forma podemos obtener la localización y tamaño de los elementos principales de la cara, así como de las medidas de la cabeza. Sin embargo el valor de F, todavía no se a relacionado con otras medidas corporales, por lo que esta indefinida. El valor de F se obtiene a través de las relaciones corporales de la cabeza tiene con el resto del cuerpo y por lo tanto se explicará en el apartado de relaciones longitudinales del ser humano.

## Presupuesto

---

### 1) Ejecución Material

- Compra de ordenador personal (Software incluido)..... 1.800 €
- Material de oficina ..... 150 €
- Total de ejecución material ..... 1.950 €

### 2) Gastos generales

- 16 % sobre Ejecución Material ..... 312 €

### 3) Beneficio Industrial

- 6 % sobre Ejecución Material ..... 117 €

### 4) Honorarios Proyecto

- 840 horas a 15 € / hora ..... 12600 €

### 5) Material fungible

- Gastos de impresión..... 90 €
- Encuadernación ..... 220 €

### 6) Subtotal del presupuesto

- Subtotal Presupuesto ..... 15289 €

### 7) I.V.A. aplicable

- 16% Subtotal Presupuesto ..... 2446,24 €

### 8) Total presupuesto

- Total Presupuesto ..... 17735,24 €

Madrid, Marzo de 2009  
El Ingeniero Jefe de Proyecto

Fdo.: Miguel Cortés Marlia  
Ingeniero Superior de Telecomunicación



## **Pliego de Condiciones**

---

Este documento contiene las condiciones legales que guiarán la realización, en este proyecto, de un sistema de reconocimiento facial “Reconocimiento de caras frontales mediante la extracción de puntos característicos”. En lo que sigue, se supondrá que el proyecto ha sido encargado por una empresa cliente a una empresa consultora con la finalidad de realizar dicho sistema. Dicha empresa ha debido desarrollar una línea de investigación con objeto de elaborar el proyecto. Esta línea de investigación, junto con el posterior desarrollo de los programas está amparada por las condiciones particulares del siguiente pliego.

Supuesto que la utilización industrial de los métodos recogidos en el presente proyecto ha sido decidida por parte de la empresa cliente o de otras, la obra a realizar se regulará por las siguientes:

### **Condiciones generales**

1. La modalidad de contratación será el concurso. La adjudicación se hará, por tanto, a la proposición más favorable sin atender exclusivamente al valor económico, dependiendo de las mayores garantías ofrecidas. La empresa que somete el proyecto a concurso se reserva el derecho a declararlo desierto.

2. El montaje y mecanización completa de los equipos que intervengan será realizado totalmente por la empresa licitadora.

3. En la oferta, se hará constar el precio total por el que se compromete a realizar la obra y el tanto por ciento de baja que supone este precio en relación con un importe límite si este se hubiera fijado.

4. La obra se realizará bajo la dirección técnica de un Ingeniero Superior de Telecomunicación, auxiliado por el número de Ingenieros Técnicos y Programadores que se estime preciso para el desarrollo de la misma.

5. Aparte del Ingeniero Director, el contratista tendrá derecho a contratar al resto del personal, pudiendo ceder esta prerrogativa a favor del Ingeniero Director, quien no estará obligado a aceptarla.

6. El contratista tiene derecho a sacar copias a su costa de los planos, pliego de condiciones y presupuestos. El Ingeniero autor del proyecto autorizará con su firma las copias solicitadas por el contratista después de confrontarlas.

7. Se abonará al contratista la obra que realmente ejecute con sujeción al proyecto que sirvió de base para la contratación, a las modificaciones

autorizadas por la superioridad o a las órdenes que con arreglo a sus facultades le hayan comunicado por escrito al Ingeniero Director de obras siempre que dicha obra se haya ajustado a los preceptos de los pliegos de condiciones, con arreglo a los cuales, se harán las modificaciones y la valoración de las diversas unidades sin que el importe total pueda exceder de los presupuestos aprobados. Por consiguiente, el número de unidades que se consignan en el proyecto o en el presupuesto, no podrá servirle de fundamento para entablar reclamaciones de ninguna clase, salvo en los casos de rescisión.

8. Tanto en las certificaciones de obras como en la liquidación final, se abonarán los trabajos realizados por el contratista a los precios de ejecución material que figuran en el presupuesto para cada unidad de la obra.

9. Si excepcionalmente se hubiera ejecutado algún trabajo que no se ajustase a las condiciones de la contrata pero que sin embargo es admisible a juicio del Ingeniero Director de obras, se dará conocimiento a la Dirección, proponiendo a la vez la rebaja de precios que el Ingeniero estime justa y si la Dirección resolviera aceptar la obra, quedará el contratista obligado a conformarse con la rebaja acordada.

10. Cuando se juzgue necesario emplear materiales o ejecutar obras que no figuren en el presupuesto de la contrata, se evaluará su importe a los precios asignados a otras obras o materiales análogos si los hubiere y cuando no, se discutirán entre el Ingeniero Director y el contratista, sometiéndolos a la aprobación de la Dirección. Los nuevos precios convenidos por uno u otro procedimiento, se sujetarán siempre al establecido en el punto anterior.

11. Cuando el contratista, con autorización del Ingeniero Director de obras, emplee materiales de calidad más elevada o de mayores dimensiones de lo estipulado en el proyecto, o sustituya una clase de fabricación por otra que tenga asignado mayor precio o ejecute con mayores dimensiones cualquier otra parte de las obras, o en general, introduzca en ellas cualquier modificación que sea beneficiosa a juicio del Ingeniero Director de obras, no tendrá derecho sin embargo, sino a lo que le correspondería si hubiera realizado la obra con estricta sujeción a lo proyectado y contratado.

12. Las cantidades calculadas para obras accesorias, aunque figuren por partidaalzada en el presupuesto final (general), no serán abonadas sino a los precios de la contrata, según las condiciones de la misma y los proyectos particulares que para ellas se formen, o en su defecto, por lo que resulte de su medición final.

13. El contratista queda obligado a abonar al Ingeniero autor del proyecto y director de obras así como a los Ingenieros Técnicos, el importe de sus respectivos honorarios facultativos por formación del proyecto, dirección técnica y administración en su caso, con arreglo a las tarifas y honorarios vigentes.

14. Concluida la ejecución de la obra, será reconocida por el Ingeniero Director que a tal efecto designe la empresa.

15. La garantía definitiva será del 4% del presupuesto y la provisional del 2%.

16. La forma de pago será por certificaciones mensuales de la obra ejecutada, de acuerdo con los precios del presupuesto, deducida la baja si la hubiera.

17. La fecha de comienzo de las obras será a partir de los 15 días naturales del replanteo oficial de las mismas y la definitiva, al año de haber ejecutado la provisional, procediéndose si no existe reclamación alguna, a la reclamación de la fianza.

18. Si el contratista al efectuar el replanteo, observase algún error en el proyecto, deberá comunicarlo en el plazo de quince días al Ingeniero Director de obras, pues transcurrido ese plazo será responsable de la exactitud del proyecto.

19. El contratista está obligado a designar una persona responsable que se entenderá con el Ingeniero Director de obras, o con el delegado que éste designe, para todo relacionado con ella. Al ser el Ingeniero Director de obras el que interpreta el proyecto, el contratista deberá consultarle cualquier duda que surja en su realización.

20. Durante la realización de la obra, se girarán visitas de inspección por personal facultativo de la empresa cliente, para hacer las comprobaciones que se crean oportunas. Es obligación del contratista, la conservación de la obra ya ejecutada hasta la recepción de la misma, por lo que el deterioro parcial o total de ella, aunque sea por agentes atmosféricos u otras causas, deberá ser reparado o reconstruido por su cuenta.

21. El contratista, deberá realizar la obra en el plazo mencionado a partir de la fecha del contrato, incurriendo en multa, por retraso de la ejecución siempre que éste no sea debido a causas de fuerza mayor. A la terminación de la obra, se hará una recepción provisional previo reconocimiento y examen por la dirección técnica, el depositario de efectos, el interventor y el jefe de servicio o un representante, estampando su conformidad el contratista.

22. Hecha la recepción provisional, se certificará al contratista el resto de la obra, reservándose la administración el importe de los gastos de conservación de la misma hasta su recepción definitiva y la fianza durante el tiempo señalado como plazo de garantía. La recepción definitiva se hará en las mismas condiciones que la provisional, extendiéndose el acta correspondiente. El Director Técnico propondrá a la Junta Económica la devolución de la fianza al contratista de acuerdo con las condiciones económicas legales establecidas.

23. Las tarifas para la determinación de honorarios, reguladas por orden de la Presidencia del Gobierno el 19 de Octubre de 1961, se aplicarán sobre el denominado en la actualidad "Presupuesto de Ejecución de Contrata" y

anteriormente llamado "Presupuesto de Ejecución Material" que hoy designa otro concepto.

## **Condiciones particulares**

La empresa consultora, que ha desarrollado el presente proyecto, lo entregará a la empresa cliente bajo las condiciones generales ya formuladas, debiendo añadirse las siguientes condiciones particulares:

1. La propiedad intelectual de los procesos descritos y analizados en el presente trabajo, pertenece por entero a la empresa consultora representada por el Ingeniero Director del Proyecto.

2. La empresa consultora se reserva el derecho a la utilización total o parcial de los resultados de la investigación realizada para desarrollar el siguiente proyecto, bien para su publicación o bien para su uso en trabajos o proyectos posteriores, para la misma empresa cliente o para otra.

3. Cualquier tipo de reproducción aparte de las reseñadas en las condiciones generales, bien sea para uso particular de la empresa cliente, o para cualquier otra aplicación, contará con autorización expresa y por escrito del Ingeniero Director del Proyecto, que actuará en representación de la empresa consultora.

4. En la autorización se ha de hacer constar la aplicación a que se destinan sus reproducciones así como su cantidad.

5. En todas las reproducciones se indicará su procedencia, explicitando el nombre del proyecto, nombre del Ingeniero Director y de la empresa consultora.

6. Si el proyecto pasa la etapa de desarrollo, cualquier modificación que se realice sobre él, deberá ser notificada al Ingeniero Director del Proyecto y a criterio de éste, la empresa consultora decidirá aceptar o no la modificación propuesta.

7. Si la modificación se acepta, la empresa consultora se hará responsable al mismo nivel que el proyecto inicial del que resulta el añadirla.

8. Si la modificación no es aceptada, por el contrario, la empresa consultora declinará toda responsabilidad que se derive de la aplicación o influencia de la misma.

9. Si la empresa cliente decide desarrollar industrialmente uno o varios productos en los que resulte parcial o totalmente aplicable el estudio de este proyecto, deberá comunicarlo a la empresa consultora.



10. La empresa consultora no se responsabiliza de los efectos laterales que se puedan producir en el momento en que se utilice la herramienta objeto del presente proyecto para la realización de otras aplicaciones.

11. La empresa consultora tendrá prioridad respecto a otras en la elaboración de los proyectos auxiliares que fuese necesario desarrollar para dicha aplicación industrial, siempre que no haga explícita renuncia a este hecho. En este caso, deberá autorizar expresamente los proyectos presentados por otros.

12. El Ingeniero Director del presente proyecto, será el responsable de la dirección de la aplicación industrial siempre que la empresa consultora lo estime oportuno. En caso contrario, la persona designada deberá contar con la autorización del mismo, quien delegará en él las responsabilidades que ostente.