

# DETECCIÓN DE CARAS EN IMÁGENES CAPTURADAS POR CÁMARAS WEB

Fernando Martín Rodríguez

Departamento de Teoría de la Señal y Comunicaciones

Universidad de Vigo

e-mail : [fmartin@tsc.uvigo.es](mailto:fmartin@tsc.uvigo.es)

Patricia Conde Pardo

e-mail : [patricia@tsc.uvigo.es](mailto:patricia@tsc.uvigo.es)

**Abstract-** This communications describes a face detector for WebCam images. The reason to implement this system is enhancing an e-learning platform with biometric security. The biometric features (face and voice) are used to verify each user login. Now, we want to monitor user activity at certain critical time intervals (for example, while answering an exam). Therefore we want to detect a computer user using “his” WebCam, knowing that perhaps user is not cooperating with the system. As we will see along the paper, we have implemented an skin detector followed by a simple “blob” processing stage.

## I. INTRODUCCIÓN

Como decíamos en el abstract, el objetivo es detectar (realmente, seguir) a un usuario de ordenador usando su propia cámara Web (el usuario está realizando un examen online y puede que no quiera colaborar con nuestro sistema).

Para detectar una cara podemos pasar primero por un detector de piel [1] (que seleccionará los puntos que tienen un color parecido al de la piel humana<sup>1</sup>) o bien se intentarán buscar patrones en imágenes en escala de grises [2].

De momento, sólo hemos realizado una primera aproximación. Veremos como las imágenes obtenidas difieren bastante en condiciones de iluminación (por la variabilidad de modelos y configuraciones de cámaras) y será necesaria una etapa de normalización. También veremos como las regiones detectadas (blobs) deben ser post-procesadas para obtener la posición de la cara.

## II. NORMALIZACIÓN DE COLORES

La normalización de colores es necesaria debido a las condiciones de iluminación impredecibles en cada imagen que, además, empeoran con la gran variabilidad de modelos y configuraciones de cámaras (la aplicación no tiene control sobre el tipo de cámara que se va a encontrar). En la figura 1, podemos ver dos ejemplos de imágenes capturadas con WebCams. La imagen de la izquierda tiene un contraste relativamente bueno mientras que la imagen de la derecha aparece con un brillo excesivo (gran parte de la imagen ha sido “quemada”, probablemente por un ajuste inadecuado del “blanco de referencia” por parte de la cámara).



Fig. 1. Imágenes obtenidas de cámaras Web.

Para realizar una normalización de los colores pensamos en usar el sistema coordinado HSV [3] ya que nos permite una mejor interpretación del significado de cada componente. Viendo que el problema principal era un rango inadecuado de la componente S en muchas imágenes decidimos ecualizar su histograma [4] (como si fuera una imagen en escala de grises). Con ello se logra una mayor variación de los valores de S, esto es: los colores “apagados” (cerca del blanco, S cercana a cero) van a ganar en viveza (la S crece). Como ejemplo, podemos ver las imágenes de la figura 1 después de haberles aplicado esta normalización (figura 2).



Fig. 2. Imágenes de la figura 1 con la saturación ecualizada.

Vemos que apenas hay efecto sobre la imagen bien contrastada (la piel pone de manifiesto su tinte naranja pero poco más) y los cambios son enormes sobre la imagen “quemada”. Véase que al crecer la saturación en la segunda imagen la piel también tiende al tono naranja pero aparecen colores azulados que no deberían estar ahí. Probablemente, la cámara captó un tinte incorrecto en estos puntos (haciendo dominante a la componente azul). En condiciones de baja saturación no se notaba pero la ecualización de S ha sacado este problema a la luz.

<sup>1</sup> Un buen detector intentará que las características que use sean independientes de la raza (por ejemplo, el tinte -H- la piel es naranja para todas las razas).

### III. EXTRACCIÓN DE CARACTERÍSTICAS

#### A. Características $[C_x \ C_y]$

Intuitivamente, debería ser el espacio de color HSV el mejor para reconocer la piel. Esto es así porque sabemos que el tinte (H) de la piel para todas las razas, se sitúa alrededor del color naranja. Por esto, H es el parámetro fundamental. La saturación (S) en el caso de la piel suele ser baja (la mayor sería la de las razas amarilla y amerindia). Por último, la V (brillo) parece “a priori” poco adecuada ya que difiere enormemente de la raza negra a la blanca y además es muy dependiente de la iluminación y de los ajustes de la cámara<sup>1</sup>.

Empezaremos utilizando, por tanto, las características H y S. Un posible problema viene dado porque el rango de H es circular (es un ángulo aunque se suele normalizar entre 0 y 1). Una forma de evitar ese problema es interpretar H y S como una representación en coordenadas polares donde S es el radio y H el ángulo. Pasando H y S a rectangulares tendremos un vector de características de tamaño 2 que no tendrá problemas con el rango circular de H. Esto es:

$$\begin{aligned} C_x &= S \cos(2\pi H) \\ C_y &= S \sin(2\pi H) \end{aligned} \quad (1)$$



Fig. 3. Círculo de color del sistema HSV (izquierda). Interpretación polar de H y S (derecha).

#### B. Características $[C_b \ C_r]$

Este es el vector más recomendado en la literatura [1,5] para reconocimiento de piel. Estas características proceden del espacio de color  $YC_bC_r$  que se diseñó para la televisión PAL [6]. Las fórmulas en este caso son:

$$\begin{aligned} Y &= 0.30R + 0.59B + 0.11G \\ C_b &= B - Y \\ C_r &= R - Y \end{aligned} \quad (2)$$

Estas fórmulas pretenden separar la parte en escala de grises (Y) de la parte de color ( $C_b$  y  $C_r$ ), igual que en la extracción anterior se desprecia la componente que representa el brillo (Y).

### IV. CLASIFICACIÓN

Sea cual sea el vector de características, hemos decidido usar un clasificador MAP (Máximo A Posteriori) de tipo

<sup>1</sup> Una vez implementado el sistema probamos a añadir V al vector de características y vimos que los resultados empeoraban incluso sólo con caras de individuos blancos europeos.

gaussiano. Este tipo de clasificador se basa en calcular varios representantes de cada clase (por ejemplo, con el algoritmo K-means) y centrar una gaussiana n-dimensional en cada representante (n es el tamaño del vector de características, en este caso n=2). La probabilidad a posteriori de la clase i-ésima sería entonces (K gaussianas):

$$P_i = \sum_{k=1}^K P(\mu_k^i) \frac{1}{(2\pi)^{n/2} \sqrt{\det(\Sigma_k^i)}} e^{-\frac{1}{2}(x-\mu_k^i)^T (\Sigma_k^i)^{-1} (x-\mu_k^i)} \quad (3)$$

Donde, en principio, se calcula una matriz de covarianzas independiente para cada gaussiana usando la conocida fórmula:

$$\Sigma_k = \frac{1}{N_k} \sum_{j=1}^{N_k} (x_j - \mu_k)(x_j - \mu_k)^T \quad (4)$$

Esta fórmula puede dar problemas si  $N_k$  (número de muestras por gaussiana) no es grande. En principio, basta que  $N_k$  sea igual a n para que la matriz sea no singular. Pero, en la práctica, si el valor de  $N_k$  no es mucho mayor que n; la matriz saldrá mal condicionada. Una solución, si no tenemos muchos puntos de muestra, es calcular una única matriz de covarianzas con todo el espacio muestral disponible.

En el caso de la piel sólo consideramos una clase. Esto es: sólo calculamos la probabilidad de que un punto sea “piel” y consideramos “no piel” para probabilidad baja. Entonces, para nosotros, desaparecen los índices “i” de la fórmula 3.

### V. PROCESADO POSTERIOR

El resultado de la etapa de clasificación será una matriz de valores de probabilidad (que, si queremos, podremos interpretar y representar como una imagen en escala de grises: figura 4).



Fig. 4. Mapas de probabilidad para las imágenes de la figura 1.

Para localizar la cara hemos decidido tratar los mapas de probabilidad como imágenes y aplicarles el algoritmo de Otsu [7]. El resultado es una imagen de blobs<sup>2</sup> binarios que definen posibles zonas de piel (figura 5). En esta imagen realizamos el siguiente procesado:

- Un cierre morfológico (o “closing” [8]) para unir blobs fragmentados.
- Una apertura morfológica (opening) para eliminar un poco el ruido.
- Elegimos el componente conectado mayor.

<sup>2</sup> Blob: componente conectado.



Fig. 5. Blobs de piel sin procesar.

Con el proceso anterior se consigue una máscara con la que ya podemos detectar la cara. En la figura 6 vemos las máscaras para las dos figuras de ejemplo. En la figura 7 hemos marcado en verde la envolvente convexa (“Convex Hull”) y en rojo la elipse de ajuste (“Best Fit Ellipse”) para las máscaras obtenidas.



Fig. 6. Blobs de piel ya procesados.

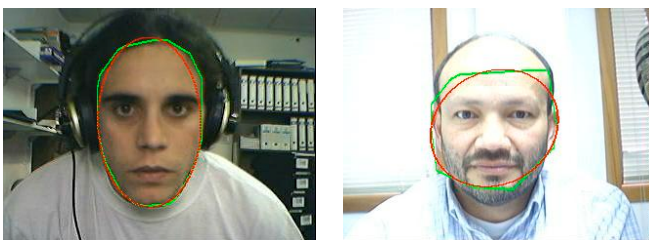


Fig. 7. Aproximaciones a la cara.

Podemos ver que en la cara mal iluminada hemos logrado una detección “regular”. Con esto queremos decir que la cara sería reconocible para un programa automático ya que hemos encontrado los puntos clave: los dos ojos, la nariz y la boca.

En este punto si estamos monitorizando a un alumno deberíamos comprobar si la zona obtenida es una cara o no (para saber si se va de la imagen, también nos interesaría saber si aparece una segunda persona, que podemos detectar con el componente conectado segundo en tamaño<sup>1</sup>). Para autentificar las caras podemos basarnos en el tamaño de la elipse de ajuste, su excentricidad y el porcentaje de la envolvente convexa ocupado por puntos que realmente son de la máscara (ese porcentaje será bajo si el sistema salta por ruido).

<sup>1</sup> Y realizando un seguimiento temporal.

## VI. PRUEBAS Y RESULTADOS

El sistema se ha probado con 124 imágenes de 12 individuos diferentes. Para estimar la calidad del modelo de probabilidades empleado en la clasificación hemos obtenido una estimación de las tasas de falso acierto y falso rechazo (TFA y TFR) sobre muestras de test para piel y fondo.

Además, se ha entrenado el clasificador MAP para distinto número de gaussianas. Hemos observado que aumentar el número de gaussianas no siempre incrementa la calidad.

Probando sin corrección del color, las características  $[C_x, C_y]$  (figura 8.a) aportan tasas elevadas de falso acierto y falso rechazo (en torno al 50%). Recuérdese que en nuestra base de pruebas las condiciones de iluminación no están controladas. Ese porcentaje descende a un 25% en el espacio  $[C_b, C_r]$  (figura 8.b).

Sin embargo, aplicando la corrección de color sobre el espacio  $[C_x, C_y]$  (figura 8.c) se obtiene una mejora de un 5% respecto al espacio  $[C_b, C_r]$ . De aquí deducimos que  $[C_x, C_y]$  resulta adecuado en el caso de imágenes de gran variabilidad, homogeneizando las muestras para la clasificación.

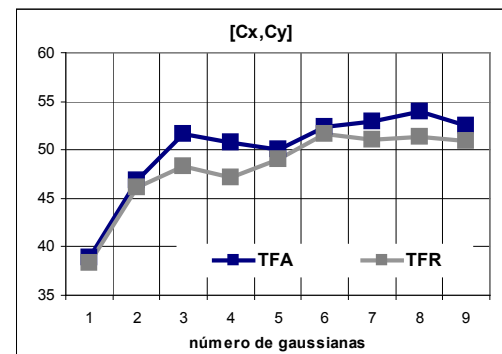


Fig. 8.a. Gráfica de TFA y TFR para  $[C_x, C_y]$  sin corrección.

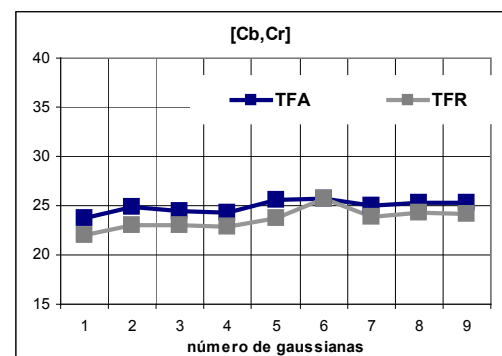


Fig. 8.b. Gráfica de TFA y TFR para  $[C_b, C_r]$  sin corrección.

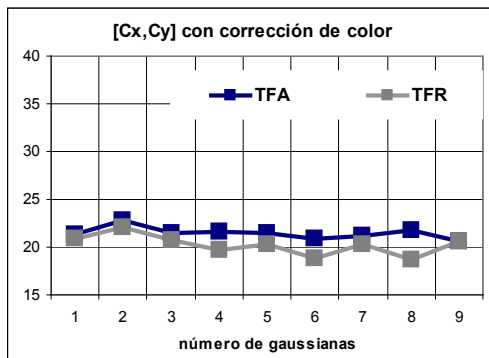


Fig. 8.c. Gráfica de TFA y TFR para  $[C_x, C_y]$  con corrección.

Finalmente, hicimos una prueba más global. Dejando funcionar al sistema y midiendo el porcentaje de solape entre la cara detectada y la cara real (obtenida manualmente). Nótese que las anteriores medidas (TFA y TFR) miden el acierto a nivel de pixel pero el sistema es capaz de superar los errores puntuales debido al post-procesado del mapa de probabilidad. Además añadimos una característica adicional: la corrección de color sólo se efectuó en aquellas imágenes donde la varianza de la saturación ecualizada era superior a 2.5 veces la obtenida antes de ecualizar. Esta estrategia permite entrenar dos reconocedores: uno para piel normal en condiciones buenas y otro para piel saturada (anaranjada) en peores condiciones. El porcentaje medio de solape para esta prueba fue del **80%** (figura 9). Además, se comprobó que en todas las imágenes los ojos, la nariz y la boca quedaron dentro de la envolvente convexa.

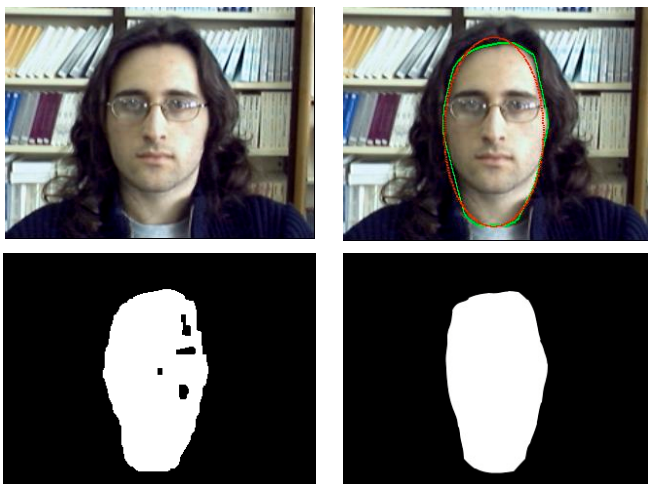


Fig. 9. Medición del porcentaje de solape. De izquierda a derecha y de arriba abajo: imagen original, cara localizada, máscara detectada automáticamente y máscara construida manualmente. Solape: 94%

(fórmula utilizada:  $s = \frac{\text{Área Común}}{\text{Área Máxima}} 100$ ).

## VII. CONCLUSIONES

Hemos visto como crear un sistema reconocedor de piel para imágenes de cámaras Web. Estas imágenes tienen unas condiciones que hacen muy difícil el trabajo (la iluminación

es incontrolable aunque suponemos ambiente de oficina, las características y configuración de la cámara también son incontrolables y con mucha variabilidad). A pesar de eso, hemos logrado resultados aceptables a los que ha ayudado una etapa de normalización basada en el concepto de saturación del color.

## VIII. LÍNEAS FUTURAS

En este sistema queda mucho por hacer, por ejemplo:

- Probar con nuevos espacios de color (y/o vectores de características mixtos como  $[C_b, C_r, S]^1$ ).
- Mejorar la etapa de procesamiento posterior.
- Combinar los métodos basados en color de la piel con otros basados en patrones de gris [2].
- Buscar una forma estándar para que la aplicación pueda "obligar" a las cámaras a adoptar una configuración parecida.
- Incluir individuos de raza no occidental en la base de pruebas.

## AGRADECIMIENTOS

Trabajo financiado por la Xunta de Galicia (código del proyecto: PGIDT02TIC32201PR).

## REFERENCIAS

- [1] R.-L. Hsu y otros, "Face Detection in Color Images", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.24, nº 5, Mayo 2002.
- [2] P. Viola, M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", Proceedings IEEE Conf. on Computer Vision and Pattern Recognition, 2001.
- [3] G. Wyszecki, W.S. Stiles, "Color Science", John Wiley & Sons, 2000.
- [4] A.K. Jain, "Fundamentals of Digital Image Processing", Prentice-Hall, 1989.
- [5] P. Santos, J.L. Alba, "Estudio de la detección de caras en imágenes en color con fondo complejo", Universidade de Vigo, 2003.
- [6] L. Torres y otros, "Sistemas Analógicos y Digitales de Televisión", Editorial UPC, 1993.
- [7] N. Otsu, "A Threshold Selection Method for Gray Level Histograms", IEEE Transactions on System, Man and Cybernetics, 1979.
- [8] R.C. González, R.E. Woods, "Digital Image Processing", Prentice-Hall, 2002.
- [9] L. Wiskott y otros, "Face Recognition by Elastic Bunch Graph Matching", Proceedings de CAPI'97 (Computer Analysis of Images and Patterns), 1997.

<sup>1</sup> Este vector ya se está probando.