

Comparación de métodos de detección de piel en modelos de color YCbCr y HSI para reconocimiento de caras

Ramello, Pedro Martín

Universidad Nacional del Litoral
Facultad de Ingeniería y Ciencias Hídricas
Captura y Procesamiento Digital de Señales e Imágenes

Resumen: En este trabajo se implementó una técnica basada en color para detectar caras humanas en posición frontal en imágenes utilizando dos modelos de color diferentes.

El método consiste en dos pasos de procesamiento de la imagen. Primero, se separan las regiones de piel, de aquellas regiones de no-piel. Luego, se detecta la cara humana en posición frontal en base a las regiones de piel.

A partir de la imagen en color original, se genera una imagen binaria para separar las regiones de piel de las regiones de no piel y se aplica correlación cruzada para determinar si una región de piel dada representa o no la cara humana frontal.

1. Introducción

El reconocimiento de caras, ha sido un tema de gran interés en los últimos años. Existen muchas aplicaciones en una variedad de campos tales como aplicación de la ley, identificación en bancos y sistemas de seguridad de acceso entre otros.



Figura 1: Etapas del proceso de reconocimiento de caras.

El propósito de este trabajo es tratar de realizar en una computadora lo que los seres humanos son capaces de hacer sin esfuerzo a cada momento de sus vidas, detectar la presencia o ausencia de caras en su campo de visión.

Mientras que este proceso es algo que a priori parece trivial, realizar los pasos necesarios para

una ejecución exitosa del mismo en un algoritmo es difícil y aún un problema no resuelto en visión computacional.

Segmentación basada en color

Asumiendo que la cara de una persona fotografiada, no puede tomar colores de una naturaleza tal como blanco, verde o rojo, y aunque los distintos grupos étnicos tienen diferentes niveles de melanina y pigmentación, el rango de colores que toman las caras humanas es claramente un subespacio del espacio de color total.

Asumiendo un escenario fotográfico típico, sería correcto tomar ventajas de la correlación de los colores de la piel para limitar la búsqueda de caras en áreas de la imagen de entrada que tengan, por lo menos, los componentes de color adecuados.

En búsqueda de este objetivo, se seleccionaron dos espacios de color, que han sido utilizados previamente en la bibliografía consultada, los espacios HSI e YCbCr.

Espacio de color HSI

Mientras que el espacio de color RGB, puede ser el más comúnmente usado para descripciones de color, tiene el aspecto negativo de que cada una de las coordenadas (rojo, verde y azul) están sujetos a los efectos de luminancia de la intensidad de luz del ambiente, un aspecto que no necesariamente provee información relevante acerca de cómo una región particular de la imagen representa una región de piel o de no-piel.

El modelo de color HSI, sin embargo, es mucho más intuitivo y provee la información del color en una manera más acorde a como los humanos piensan los colores y a como los artistas típicamente los mezclan.

El “tono” (hue) describe el color puro básico de la imagen, la “saturación” da la manera en que este color puro (tono) es diluido por la luz blanca y la “intensidad” provee una noción acromática de la intensidad del color. Son las dos primeras componentes (H y S) las que proveen información útil para la discriminación en la búsqueda de piel.

Las ecuaciones de conversión del espacio de color RGB a HSI son las siguientes:

$$\begin{cases} H &= \arctan\left(\frac{\beta}{\alpha}\right) \\ S &= \frac{\sqrt{\alpha^2 + \beta^2}}{(R + G + B)/3} \\ I &= (R + G + B)/3 \end{cases}$$

$$\begin{cases} \alpha = R - \frac{1}{2}(G + B) \\ \beta = \frac{\sqrt{3}}{2}(G - B) \end{cases}$$

Espacio de color YCbCr

Este modelo se basa en el modelo RGB, pero restringiendo y descomponiendo algunos valores del mismo. YCbCr es un espacio de color usado en los sistemas de video. Y es el componente de la luminancia y Cb y Cr los componentes de la crominancia.

Las ecuaciones de conversión del espacio de color RGB a YCbCr son las siguientes:

$$\begin{cases} Y &= 0.2989 \times R + 0.5866 \times G + 0.1145 \times B \\ Cb &= -0.1688 \times R - 0.3312 \times G + 0.5000 \times B \\ Cr &= 0.5000 \times R - 0.4184 \times G - 0.0816 \times B \end{cases}$$

2. Modelos de color de piel

Para segmentar regiones de piel humana basándose en color, es necesario un modelo de color conveniente, que sea adaptable a gente de diferente color de piel y a diferentes condiciones de iluminación.

Aunque los colores de piel de diferentes personas aparentan ser muy variados, estos difieren mucho menos en el color que en el brillo. El color de la piel de diferentes personas es muy parecido, diferenciándose en la intensidad.

De esta manera se procedió a construir un modelo de color de piel en el espacio de color cromático y en el espacio de color HSI.

Modelo de color de piel YCbCr

Un total de 30 imágenes color de caras de personas fueron utilizadas para determinar la distribución de color de la piel humana en el espacio de color cromático. Estas muestras fueron extraídas de imágenes color, por lo que fueron filtradas con un filtro pasa-bajos para reducir el efecto del ruido en las mismas.



Figura 2: Muestras de piel de caras, obtenidas de imágenes a color.

La distribución de color de la piel puede ser representada por un modelo Gaussiano, como en [2], donde la probabilidad de cualquier píxel de la imagen de representar piel humana puede ser calculada de la siguiente manera:

$$P(r, b) = \exp[-0.5(x - m)^T C^{-1}(x - m)]$$

where: $x = (r, b)^T$.

La **Figura 3** presenta el modelo Gaussiano que aproxima la distribución de colores en el modelo de color de piel YCbCr. Los ejes representan las componentes Cb y Cr, y la curva Gaussiana, los valores (probabilidad) que puede tomar un píxel de la imagen, de acuerdo al par de valores (Cb, Cr) de estas dos componentes.

Este modelo de color de piel puede transformar una imagen color en una imagen de grises, de manera que el valor de gris en cada píxel represente la probabilidad del mismo de pertenecer a una región de piel.

Con una apropiada umbralización, la imagen en tono de grises puede ser convertida a una imagen binaria mostrando las regiones de piel y de no-piel.

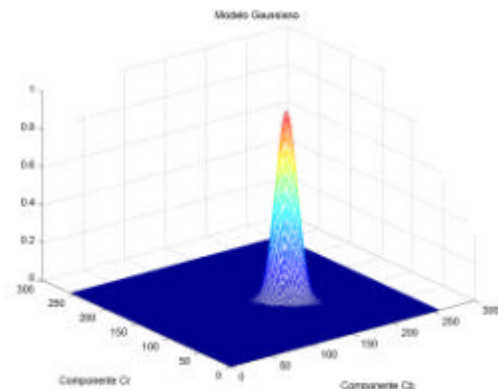


Figura 3: Vista perspectiva del modelo Gaussiano que aproxima la distribución de colores en el modelo de color de piel YCbCr.

Modelo de color de piel HSI

Las mismas imágenes color de caras de personas utilizadas en el espacio de color cromático (ver **Figura 2**), fueron utilizadas para tratar de determinar alguna tendencia útil en los valores observados en las componentes del modelo de piel HSI.

De estos resultados se observó que los valores de la componente H de las regiones de caras, tienden a ocupar un rango muy pequeño de valores. Esta tendencia fue usada para derivar la siguiente regla utilizada en el proceso de detección de piel:

Si ($H < 0.1$) Pixel de piel

En conjunto con esta regla se determinó que los valores de la componente S tienen una tendencia a tomar valores mayores a cierto límite y se creó una segunda regla:

Si ($S > 0.3$) Pixel de piel

Si el píxel analizado no cumple con las reglas este es descartado como perteneciente a una región de piel.

Este modelo de color de piel puede transformar una imagen color en una binaria directamente, mostrando las regiones de piel y de no-piel.

3. Segmentación de piel

A partir de una imagen color, la técnica basada en el modelo YCbCr la transforma en una imagen en escala de grises. Esto implica transformar cada píxel a la representación cromática y determinar la probabilidad basado en la ecuación dada anteriormente.

Dado que las regiones de piel son más brillosas que las otras partes de la imagen, son segmentadas del resto a través de un proceso de umbralización.

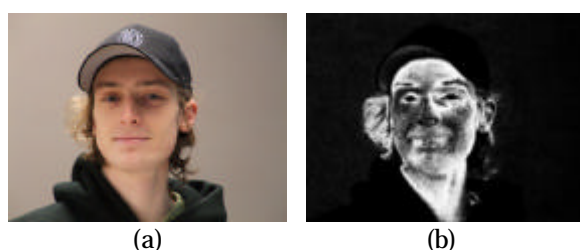


Figura 4: Imagen original RGB (a), imagen de grises (b), imagen segmentada por HSI (c) e imagen segmentada por YCbCr (d). En ambos modelos de segmentación, los resultados son convenientes.

Con la técnica basada en el modelo HSI, se transforma una imagen color en una imagen binaria sin el paso intermedio de la imagen de grises. Cada píxel se transforma a la representación tono-saturación-intensidad y se determinan las regiones de piel marcando los píxeles según cumplan o no con las reglas impuestas por la técnica.

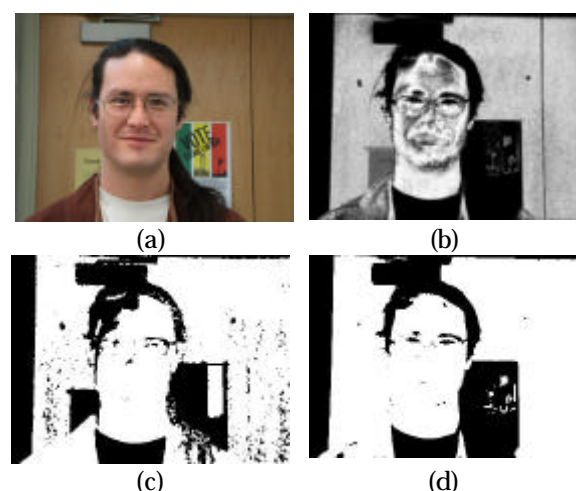


Figura 5: Imagen original RGB (a), imagen de grises (b), imagen segmentada por HSI (c) e imagen segmentada por YCbCr (d). En ambos modelos de segmentación, los resultados son insatisfactorios.

Esta claro, que no todas las regiones de piel detectadas contienen caras. Algunas corresponden a otras regiones de piel, como el cuello, brazos o manos y a otras partes expuestas del cuerpo, mientras que algunas corresponden a objetos con colores similares a la piel.

4. Búsqueda de caras en la escena.

Una aplicación muy útil y fácil de implementar es la búsqueda de un objeto particular (o plantilla) sobre la imagen (o escena), técnica conocida como "Template Matching".

La correlación permite desarrollar esta búsqueda, con la ventaja de la rapidez si se calcula en el dominio frecuencial.

Teorema de correlación

La correlación de dos funciones $f(x,y)$ y $h(x,y)$ esta definida como:

$$f(x, y) \circ h(x, y) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f^*(m, n) h(x + m, y + n)$$

Donde el símbolo $*$ representa el complejo conjugado de la función f . Normalmente al tratar con imágenes este valor es el mismo valor de f . La función de correlación tiene exactamente la misma forma que la función de convolución:

$$f(x, y) * h(x, y) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) h(x - m, y - n).$$

Con la excepción del valor complejo conjugado y el hecho que el segundo término de la sumatoria tiene signo positivo en vez de negativo. Esto significa que h no es espejada sobre su origen. Todo lo demás en la implementación de la correlación es igual a la convolución, incluyendo la necesidad de rellenado con ceros.

El teorema de correlación, análogo al teorema de convolución, determina que la correlación en el dominio espacial puede ser obtenida mediante la transformada inversa de Fourier de un producto:

$$f(x, y) \circ h(x, y) \Leftrightarrow F^*(u, v) H(u, v).$$

El principal uso de la correlación es en la “técnica de matching”. En esta técnica, $f(x,y)$ es una imagen que contiene objetos o regiones de interés. Si queremos determinar si f contiene un objeto o región particular, llamamos $h(x,y)$ a este objeto (normalmente esta imagen recibe el nombre de “template” o “plantilla”). Luego, si el objeto $h(x,y)$ se encuentra en $f(x,y)$ la correlación entre las dos funciones será máxima donde h encuentra una correspondencia en f . En la mayoría de las aplicaciones prácticas un pre-procesamiento como escalado y alineación es necesario, pero el punto clave es la correlación.

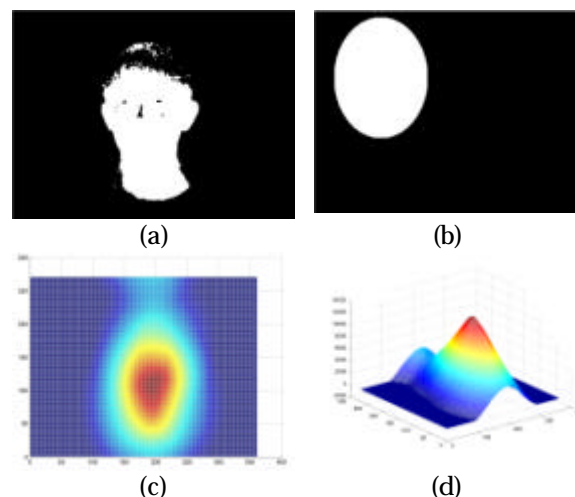


Figura 6: Imagen segmentada (a), elipse binaria (b), Vista superior (c) y vista perspectiva (d) de la matriz de correlación.

En este trabajo para la búsqueda o detección de caras se realiza el “Template Matching” con elipses binarias de distintos tamaños, asumiendo que las caras se encuentran en posición vertical, por lo que no es necesario realizar la rotación de la elipse en el proceso de correlación.

Se observa en la Figura 6 que el máximo valor de la función de correlación ocurre en el punto donde la plantilla se parece más a la forma elíptica de la cara segmentada. El valor de cada píxel es el valor de la correlación en cada punto (x,y) .

5. Resultados

Luego de la aplicación en serie de los dos pasos de procesamiento de la imagen (primero se separan las regiones de piel de aquellas regiones de no-piel, y luego, se obtiene el valor máximo del proceso de correlación) se genera como salida una imagen a color marcada con un cuadrado rojo en la posición donde se ha detectado la presencia de una cara humana frontal.

El proceso de detección por medio de “template matching” es dependiente de los resultados obtenidos en el primer paso del proceso. Una imagen mal binarizada puede producir (no necesariamente) la detección fallida de la cara humana en la imagen de entrada.

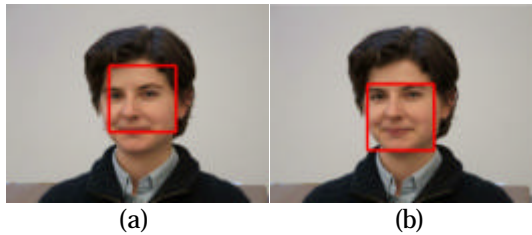


Figura 7: Resultados de detección de caras por el método HSI (a) y por YCbCr (b).



Figura 8: Imagen segmentada (a), resultado de detección erróneo por el método HSI debido a una mala segmentación (b).

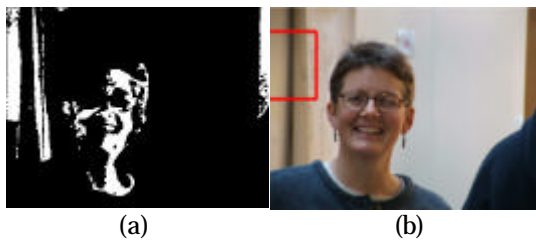


Figura 9: Imagen segmentada (a), resultado de detección erróneo por el método YCbCr debido a una mala segmentación (b).

La **Figura 7** muestra un ejemplo de detección correcta de ambos modelos de color. Es necesario destacar que ambos métodos no realizan la detección exactamente en la misma posición de la imagen original. Esto se debe a la sensibilidad que tiene el proceso a una segmentación adecuada.

Existen casos en que por uno de los métodos se detecta correctamente la cara y por el otro no. Son ejemplos de esto las **Figuras 8 y 9** donde se observa un caso de detección errónea para ambos modelos de color de piel. En la **Figura 8** se ha segmentado la pared junto con la cara, lo que produce la detección fallida, y en la **Figura 9** la segmentación de la cara ha sido muy débil.

Comparación de Resultados de ambos modelos

Las pruebas de ambos modelos se realizaron sobre un total de 22 imágenes entre las cuales se encuentran las mostradas previamente.

La **Tabla 1** muestra una comparación de los métodos de segmentación y de detección. Si la salida (imagen binaria) está correctamente segmentada, es decir que solo se ven en blanco las regiones pertenecientes a la cara o a la piel, el método obtuvo un resultado correcto (valor de 1), y si se han segmentado partes que no pertenecen a la piel el resultado es erróneo (valor de 0).

IMAGEN	Segmentación		Detección	
	HSI	YCbCr	HSI	YCbCr
foto_1	1	1	1	1
foto_2	0	0	1	0
foto_3	1	1	1	1
foto_4	0	0	0	0
foto_5	0	1	0	1
foto_6	0	1	1	1
foto_7	0	0	1	0
foto_8	0	0	1	1
foto_9	1	1	1	1
foto_10	1	1	1	1
foto_11	0	0	1	1
foto_12	1	1	1	1
foto_13	1	1	1	1
foto_14	0	0	0	0
foto_15	1	0	1	1
foto_16	0	0	1	0
foto_17	0	0	0	0
foto_18	0	0	1	1
foto_19	0	0	1	0
foto_20	0	0	1	1
foto_21	1	1	1	1
foto_22	1	0	1	1

Tabla 1: Si la segmentación y detección obtenida por cada modelo es correcta se asigna un valor de 1 y 0 si es errónea.

Se obtuvieron los siguientes porcentajes de segmentación correcta.

Segmentación HSI 40,9 %
Segmentación YCbCr 40,9 %

Y de detección correcta:

Detección HSI 81,8%
Detección YCbCr 68,18%

En la detección, los problemas encontrados en ambos modelos fueron debido a una mala segmentación. En algunos casos los colores de la ropa o incluso de la pared fueron segmentados

junto con la piel, lo que produjo la detección de caras en lugares incorrectos.

En el modelo YCbCr además de estos problemas se encontró que la segmentación de la piel de la cara se producía en pequeñas regiones en los contornos de la cara, lo que produjo una menor porcentaje de aciertos.

La **Figura 10** ilustra un problema solo detectado en la segmentación por YCbCr.

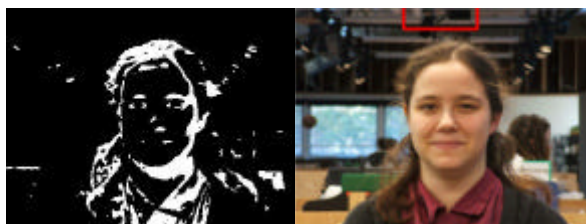


Figura 10: Mala detección por el método YCbCr.

Nótese que los porcentajes de detección son mayores a los de segmentación, esto se debe a que en la segmentación se evaluaron los casos en los que los resultados fueron de segmentación de solo piel. El método se correlaciona en algunos casos los problemas de la segmentación, ya que las formas de las regiones de no-piel que son segmentadas, obtienen un valor menor que las caras al realizar la comparación con las elipses blancas.

La **Tabla 2**, muestra una comparación de ambos métodos, para imágenes de diferentes tamaños en píxeles.

Tamaño (píxeles)	Tiempo HSI (segundos)	Tiempo YCbCr (segundos)
150 x 150	9.22	32.15
360 x 270	17.40	65.17
400 x 300	23.76	80.42
800 x 600	65.82	241.24

Tabla 2: Tiempos según tamaño de la imagen.

Se aprecia una notable diferencia en los tiempos de procesamiento de cada método. Es necesario destacar, que en el método YCbCr, son necesarios dos pasos en la obtención de la salida del proceso. Primero se obtiene la imagen de probabilidades y luego se realiza la segmentación, a diferencia del método HSI donde la segmentación es directa. En esto se encuentra la diferencia de tiempos, ya que el proceso de detección de caras es el mismo para ambos métodos.

A medida que se aumenta el tamaño de las imágenes a procesar el tiempo se dispara hacia

valores cada vez mayores. El procesamiento se realizó en una computadora personal con un microprocesador AMD K6-II de 533 MHz, con 128 MB de memoria RAM, bajo sistema operativo Windows XP; y para la implementación se utilizó MATLAB 6.

6. Conclusión

La comparación de ambos métodos sirve para poder analizar las ventajas y desventajas de cada uno.

Ambos son métodos sencillos, dado que se aplican a imágenes donde la persona se ubica a cierta distancia y se supone la posición vertical de las caras.

Los resultados obtenidos se pueden mejorar en un “ambiente controlado”, donde el fondo sea fijado de manera que no interfiera con la segmentación de la piel.

Puede ser interesante el desarrollo de un método que combine las características de ambos espacios de color de piel, para lograr una mejor segmentación de las regiones de piel. Una forma puede ser utilizando primero el método de segmentación YCbCr con un umbral más bajo en la binarización, y luego a estos resultados aplicarle el proceso de segmentación HSI.

Referencias

- [1] Rafael C. Gonzalez and Richard E. Woods, “Digital Image Processing”, 2nd. Edition, Prentice-Hall, 2002.
- [2] Henry Chang and Ulises Robles, “Face Detection, May 25, 2000. EE368 Final Project Report - Spring 2000”.
- [3] Michael Padilla and Zihong Fan. “Digital Image Processing Project –Automatic Face Detection Using Color Based Segmentation and Template/Energy Thresholding. EE368 - Dr. B. Girod, Spring 2002-2003”
- [4] Shinjiro Kawato and Jun Ohya. “Automatic Skin-color Distribution Extraction for Face Detection and Tracking”. ICSP2000: The 5th Int. Conf. on Signal Processing, vol.II.
- [5] Jianping Fan, David. K. Y. Yau, Ahmed. K. Elmagarmid and Walid G. Aref. “Automatic Image Segmentation by Integrating Color-Edge Extraction and Seeded Region Growing”. IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 10, NO. 10, OCTOBER 2001