



# UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

DETECCIÓN RÁPIDA DE ROSTROS EN TIEMPO REAL, BASADO EN EL COLOR DE LA PIEL

# T E S I S

QUE PARA OBTENER EL TÍTULO DE LICENCIADO EN CIENCIAS DE LA COMPUTACIÓN

P R E S E N T A:

HUMBERTO FERRER CAMACHO

DIRECTORA DE TESIS:

MAT. ANA LUISA SOLÍS GONZÁLEZ COSÍO



2005



m342265



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



LIBERTAD NACIONAL  
AYUNTAMIENTO  
MEXICO

**ACT. MAURICIO AGUILAR GONZÁLEZ**  
**Jefe de la División de Estudios Profesionales de la**  
**Facultad de Ciencias**  
**Presente**

Comunicamos a usted que hemos revisado el trabajo escrito:

"DETECCIÓN RÁPIDA DE ROSTROS EN TIEMPO REAL, BASADO EN EL COLOR DE LA PIEL"

realizado por FERRER CAMACHO HUMBERTO

con número de cuenta 093219002 , quien cubrió los créditos de la carrera de:

LICENCIATURA EN CIENCIAS DE LA COMPUTACIÓN

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis  
Propietario

MAT. ANA LUISA SOLÍS GONZÁLEZ COSÍO

Propietario

M. EN C. LUIS MIGUEL DE LA CRUZ SALAS

Propietario

M. EN C. MIGUEL MIRANDA MIRANDA

Suplente

DRA. MA. ELENA MARTÍNEZ PÉREZ

Suplente

M. EN C. ALEJANDRO AGUILAR SIERRA

Consejo Departamental de MATEMÁTICAS

DR. FRANCISCO HERNÁNDEZ QUITROZ

FACULTAD DE CIENCIAS  
CONSEJO DEPARTAMENTAL DE  
MATEMÁTICAS

Dedicado con todo cariño aquellas personas que  
siempre me apoyaron y creyeron en mi.  
Mis papás.

Gracias a todas las personas que hicieron esto posible especialmente  
a quienes siempre estuvieron allí y que dejaron  
algo que hacer por ayudarme.

# Índice

## Capítulo 1

1.	INTRODUCCIÓN	
1.1	Antecedentes -----	2
1.2	Objetivo de la tesis -----	3
1.3	Organización de esta tesis -----	3

## Capítulo 2

2.	ANTECEDENTES	
2.1	Introducción -----	5
2.2	Segmentación de cara -----	5
2.2.1	Segmentación de caras desde imágenes fijas -----	6
2.2.2	Segmentación de caras desde video -----	7
2.3	Utilidad de los detectores de piel -----	7

## Capítulo 3

3.	TEORIA DEL COLOR Y ESPACIOS DEL COLOR	
3.1	Introducción -----	9
3.2	Teoría del Color -----	9
3.3	Bases de la colorimetría tricromática -----	10
3.4	Leyes de Grassman -----	10
3.5	Síntesis tricromática o modelo RGB -----	11
3.5.1	Representación cúbica - cubo de Maxwell -----	11
3.6	Sistema Munsell -----	13
3.7	Modelo HSL -----	15
3.8	Medida normalizada del color -----	17
3.9	RGB Normalizado -----	18
3.9.1	Representación triangular de los colores -----	18
3.10	Espacio YCbCr -----	19

## Capítulo 4

4.	FUNDAMENTOS DE PROBABILIDAD PARA EL ANÁLISIS DE IMÁGENES	
4.1	Función de densidad de probabilidad normal unidimensional -----	21
4.2	Función densidad de probabilidad normal multidimensional -----	23
4.3	Distancia de Mahalanobis -----	24
4.4	Correlación de variables -----	26
4.5	Algoritmo de Camshift -----	29

## Capítulo 5

5.	SEGMENTACIÓN DEL COLOR DE LA PIEL	
5.1	Introducción -----	31
5.2	Compensación de luz -----	31
5.3	Representación del color de la piel -----	34
5.4	Optimización del detector del color de la piel -----	39

## Capítulo 6

### 6. IMPLEMENTACIÓN DE LA DETECCIÓN DE PIEL APLICANDO A ROSTROS

6.1	Introducción -----	41
6.2	Descripción del problema -----	41
6.3	Etapas del desarrollo -----	41
6.4	Descripción de la interfaz -----	42
6.5	Comparativa de modelos de color -----	44
6.6	Obtención de ventana de búsqueda -----	45
6.7	Resultados -----	46
	CONCLUSIONES -----	52
A.	ANEXO	
A.1	Filtro Gussiano -----	54
	BIBLIOGRAFÍA -----	56

## Lista de Figuras

1.1	Imágenes requeridas para el análisis de expresiones... -----	1
2.1	Diagrama general sobre el procesamiento de FACs -----	5
3.1	Longitud de onda (nm) -----	9
3.2	Cubo de Maxwell -----	12
3.3	Cubo de color RGB. La diagonal principal indica los niveles de gris... -----	12
3.4	A) Imagen original en color. B) Plano de rojo.... -----	13
3.5	Representación del espacio de Musnsell -----	14
3.6	Distribución del los colores mas usado en el sistema de Munsell.-----	15
3.7	En la parte superior un ejemplo de claridad y abajo un ejemplo de cromat- 15	
3.8	Modelo HSL -----	16
3.9	De lado izq. una paleta de colores basada en la Saturación ... -----	16
3.10	El área hexagonal corresponde a un valor de $V=1$ ... -----	17
3.11	El tono se mide como el ángulo alrededor del eje S.... -----	17
3.12	Triangulo equilátero de Maxwell -----	18
3.13	Obtención de la mezcla de colores en el triangulo de Maxwell -----	19
3.14	Ejemplo del plano u-v del color, valor de $Y = 0,5$ -----	20
4.1	Ejemplo de distribución normal -----	22
4.2	Áreas bajo la curva de la función de probabilidad gaussiana... -----	23
4.3	Representación de una función de probabilidad normal bidimensional -----	24
4.4	A) Densidad de probabilidad. B) Diagrama de dispersión -----	25
4.5	Dos distribuciones normales con igual media... -----	25
4.6	A) Alta covarianza. B) Baja covarianza -----	26
4.7	En la imagen se muestran unos ejemplos de correlaciones nulas -----	28
4.8	Ejemplos de correlación positiva -----	28
4.9	Ejemplos de correlación lineal -----	29
4.10	Ejemplos de correlación negativa -----	29
4.11	De lado derecho se muestra el diagrama de flujo. -----	30
5.1	Los puntos representan los distintos tonos color de la piel.-----	31
5.2	Resultado después de la compensación de luz -----	33
5.3	Representación del modelo cromático de los colores puros. -----	35
5.4	Distribución del color de la piel en el espacio normalizado -----	35
5.5	Muestran de diferentes tipos de iluminación -----	36
5.6	Resultados al seleccionar el color de la piel con esta técnica -----	38
5.7	Función para calcular la probabilidad de un píxel -----	39
5.8	Resultado del proceso de detección de piel -----	40
6.1	Diagrama de etapas del desarrollo -----	42
6.2	Aplicación desarrollada en esta Tesis -----	43
6.3	Muestra de piel tomada para el proceso de muestreo -----	44
6.4	Comparación de resultado entre RGB Normalizado y YCrCb -----	45
6.5	Obtención de la ventana de búsqueda utilizando el algoritmo Camshift ----	46
6.6	Gráfica comparativa de tiempos entre las dos versiones -----	46
6.7	Resultados de la aplicación -----	47
6.8	Resultados de la aplicación -----	48
6.9	Resultados de la aplicación -----	48
6.10	Resultados de la aplicación -----	49
6.11	Resultados de la aplicación -----	49
6.12	Resultados de la aplicación -----	50
6.13	Resultados de la aplicación -----	50



6.14 Resultados de la aplicación -----	51
6.15 Resultados de la aplicación -----	51

## Lista de Tablas

1.1	Comparación de métodos biométricos -----	1
2.1	Investigaciones relacionadas con los detectores del piel -----	8
3.1	Modelo HSL -----	16
6.1	Comparativa de tiempos de respuesta en cuadros por segundo -----	46

# Capítulo 1

## INTRODUCCIÓN

La habilidad de la visión humana para reconocer rostros y expresiones (alegría, tristeza, miedo), parece iniciarse desde que las personas nacen. Esta habilidad es robusta, ya que la gente puede reconocer cientos de caras, identificar caras familiares aún a pesar de haber pasado mucho años. Además los cambios de tono de piel no parecen afectar, el estilo de cabello, la expresión ni las condiciones de luz, entre otros. El percibir una cara humana es actualmente un área activa de investigación en el campo de visión por computadora.

El lograr dotar a las máquinas de una forma de visión es deseable en tareas que se puedan automatizar. Por ejemplo:

- La vigilancia: Donde se puede colocar una cámara la cual pueda verificar si las personas no son criminales.
- Control de acceso: Aunque actualmente hay diversas técnicas (tarjetas de acceso, claves de acceso, NIP, etc.), de identificación de usuarios cada vez se tiende a usar sistemas que sean más amigables con el usuario. Estos sistemas se basan en los rasgos de las personas (ojos, huellas, etc.), y están basados en características físicas del usuario a identificar. El reconocimiento de formas, la inteligencia artificial y el aprendizaje son las ramas de la computación que desempeñan el papel más importante en los sistemas de identificación biométricos; la criptología se limita aquí a un uso secundario, como el cifrado de una base de datos de patrones retinales, o la transmisión de una huella dactilar entre un dispositivo analizador y una base de datos. [26]

	Ojo - Iris	Ojo - Retina	Huellas dactilares	Geometría de la mano	Escritura - Firma	Voz	Reconocimiento facial
Fiabilidad	Muy alta	Muy alta	Alta	Alta	Alta	Alta	Alta
Facilidad de uso	Media	Baja	Alta	Alta	Alta	Alta	Alta
Prevención de ataques	Muy alta	Muy alta	Alta	Alta	Media	Media	Baja
Aceptación	Media	Media	Media	Alta	Muy alta	Alta	Media
Estabilidad	Alta	Alta	Alta	Media	Media	Media	Media
Identificación y autenticación	Ambas	Ambas	Ambas	Autenticación	Ambas	Autenticación	Ambas

*Tabla 1.1: Comparación de métodos biométricos.*

- Interacción con ambientes virtuales en los cuales se supervise las acciones que realiza el usuario mediante una cámara, reconociendo algunos patrones o movimientos de usuario.
- Laboratorios virtuales: En estos se impartirían tutoriales. Se utilizaría una cámara situada de frente al usuario mediante la información capturada se reconocen las expresiones faciales y apoyados con esta información puede cambiar el ritmo del curso de forma automática.
- Interfaces hombre máquina: Haciendo que la computadora pueda responder incluso a las emociones del usuario.
- Vídeo conferencias: En lugar de transmitir todos los cuadros que salen de la cámara de video como actualmente se hace se pueden tomar únicamente los rasgos de los rostros y mandar estos en un pequeño ancho de banda, haciendo que la máquina que recibe esta información haga el modelado, aplique texturas y despliegue la imagen.

## ***1.1 Antecedentes***

La idea de localizar los rostros de una forma sencilla y rápida surge de la necesidad de que la computadora pueda interactuar de una forma automática con el usuario, mediante una cámara.

Si una persona está estudiando una lección en la computadora y la computadora puede supervisar al usuario, puede detectar si éste está poniendo atención, si se tiene que modificar el ritmo del curso ya sea que se acelere o que disminuya la velocidad del curso. [20] [21] y [23].

Para esto se requiere un sistema que supervise al usuario en tiempo real<sup>1</sup>. Los sistemas utilizados en visión, requieren de condiciones muy especiales para el reconocimiento y el entendimiento de los rostros y de sus expresiones faciales, estas condiciones son:

- tener las caras centradas en la imagen,
- que esta este ajustada al tamaño del rostro y
- en algunos casos que la imagen se encuentre en escala de grises (figura 1.1),

Aunque esta última condición se omitirá para poder regresar imágenes a color. Si no se tienen estas condiciones los algoritmos se tardan demasiado tiempo localizando y analizando la imagen. Esto cuando fuese capaz de llegar a un resultado exitoso. Lo que convierte los procesos de localización y seguimiento un prerrequisito de los programas de visión dedicados al análisis de expresiones faciales y reconocimiento que requieren capturar las imágenes de una cámara o secuencia de video.

---

<sup>1</sup> Tiempo real: Que el procesamiento sea tan rápido como suceden los eventos en la realidad.



*Figura 1.1 Imágenes requeridas para el análisis de expresiones faciales o reconocimiento de rostros*

## ***1.2 Objetivo de la tesis***

La primera etapa para analizar las expresiones faciales de una persona en tiempo real, es localizar la cara en la escena (imagen captada por la cámara). Este trabajo realiza ésta tarea a través del tono de piel.

## ***1.3 Organización de la tesis***

La organización de la tesis esta dividida en seis capítulos. En el primer capítulo se presenta la introducción al presente trabajo. En el segundo capítulo se muestran varios de los trabajos realizados en el área y se presentan un recopilado de otros trabajos próximos al área como son la identificación de personas, detección y análisis de expresiones faciales, etc. A los que este sistema puede apoyar como paso inicial y reducir el tiempo de procesamiento o dar las condiciones iniciales para su desempeño. En el capítulo tres se considera a el color como un rasgo importante en el proceso de detección de rostros en especial en el color de la piel. Se inicia esta parte con la teoría del color donde se dan conceptos básico sobre el color, enseguida se ven bases de la colorimetría tricromática y leyes de Grassman que es el antecedente del modelo de RGB y las reglas con las que se

rige. Enseguida se muestra el sistema de Munsell que es donde surge el modelo HSI el cual comparamos los resultado en el capítulo tres, con el modelo RGB normalizado. El RGB normalizado es uno de los modelos a color que corresponden a la medida normalizada del color en cual también se muestra el espacio  $YCbCr$  que en el otro modelo de color utilizado en el ultimo acercamiento. De aquí seleccionamos un modelo para obtener las ventanas de las imágenes que contienen el rostro.

En el cuarto capítulo se muestra la teoría probabilística necesaria para comprender e implementar los algoritmos detectores de piel; se inicia explicando la función de densidad normal unidimensional que es utilizada en la obtención de las ventana de búsqueda a partir de las imágenes obtenidas del proceso de filtrado del color de la piel. Después se explica la función de densidad normal multivariada (FDNM) que es una generalización de la unidimensional pero para  $n$  variables. Esta FDNM es utilizada para calcular la probabilidad de que los pixeles sean del conjunto color de la piel. Posteriormente se explica la distancia de Mahalanobis que se utiliza como método de clasificación de color de la piel en lugar de la FDNM, esto debido a que es menor el tiempo para calcularla y que muestra resultados equivalentes. Al usar la FDNM obtenemos una imagen con la probabilidad de que los pixeles pertenezcan al grupo color de la piel y posteriormente se tiene que utilizar una función umbral para obtener los pixeles más cercanos al conjunto y en la distancia de Mahalanobis simplemente se calcula la distancia de que un color este en el centro (Vector esperanza) asumiendo que tiene una distribución normal. Y finalmente se muestra la correlación de varianzas esta parte es la relación  $r, g$  (RGB Normalizado), y nos indica si se tiene alguna dependencia entre estos valores.

En el capítulo cinco se describen detalladamente el proceso de los algoritmos detectores de piel, se realiza una optimización a este proceso para mejorar su tiempo y en el capítulo seis se presentan los resultados obtenidos, se calcula el proceso de detección de rostros con los distintos espacios de color y muestra el modelo elegido y optimizado.

# Capítulo 2

## ANTECEDENTES

### 2.1 Introducción

En este capítulo se dará una revisión de los trabajos realizados anteriormente en las técnicas de reconocimiento de caras. Se muestra el estado general del problema de reconocimiento facial (Figura 2.1): obtención de una imagen de una secuencia de video o una imagen, identificar una persona en la escena. La solución del problema involucra segmentación de caras desde la escena, extracción de rasgos desde una región y procesado de las FACs (facial action coding) o reconocimiento facial.

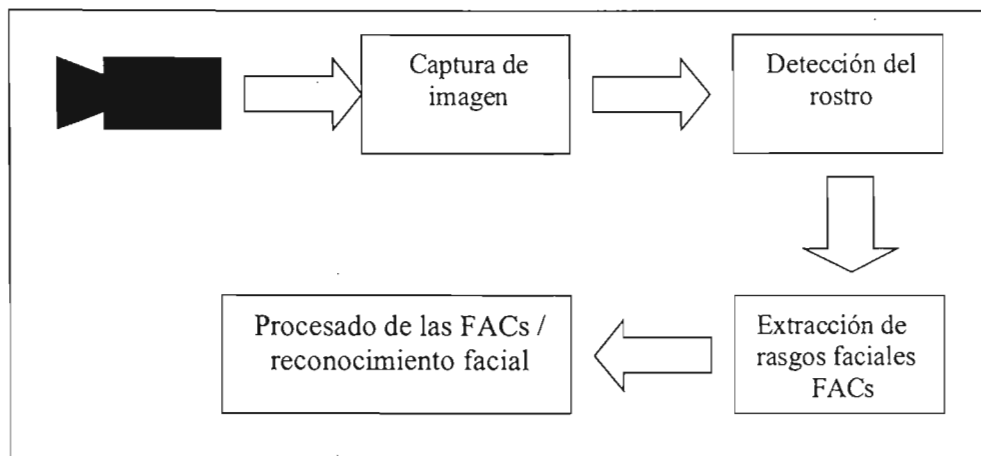


Figura 2.1 Diagrama general sobre el procesamiento de FACs y reconocimiento de facial.

### 2.2 Segmentación de caras

El primer paso hacia el reconocimiento automático de caras es la detección de estas en las imágenes o videos y separarlas del fondo.

Este es un paso esencial para hacer un sistema totalmente automático. Sin embargo es el más difícil. Por lo tanto este problema es más explorado que el de reconocer los segmentos de la cara.

Para localizar la “cara” se utilizan técnicas de segmentación de imágenes. Dichas técnicas asignan a cada pixel una región, dentro de un número finito de regiones, caracterizadas por tener propiedades comunes entre los elementos de la región y diferentes con las de otras regiones. Dentro de las técnicas de segmentación destacan las orientadas a regiones, uso de umbrales analizando histogramas en blanco y negro (b/n) o

en color, unión de bordes de fronteras, análisis de texturas y uso de movimiento. Las técnicas más recientes que se basan en reconocimiento de patrones, obtienen la información de manera implícita mediante el aprendizaje desde ejemplos o patrones.

Un enfoque que resulta muy interesante en los métodos de segmentación es el empleo de técnicas estadísticas, mediante las cuales las clases son caracterizadas por medias estadísticas de bajo orden como: media, varianza, correlación de funciones o densidad espectral de potencia. De esta forma, el problema de segmentación de una imagen se convierte en un problema de optimización estadística, lo que produce una mayor precisión en la caracterización de las imágenes.

Las técnicas de segmentación de imágenes basadas en modelos estocásticos pueden ser supervisadas o no supervisadas. En las supervisadas, los parámetros del modelo se obtienen de un conjunto de datos de entrenamiento, mientras que en las no supervisadas los parámetros del modelo se estiman directamente de la imagen a analizar. El diseño de un sistema de segmentación autónomo implica el empleo de técnicas no supervisadas; sin embargo, la baja fiabilidad que ofrecen algunos métodos a la alta complejidad de otros hace que no sean muy utilizados en la práctica, siendo un tema actual de investigación.

Los sistemas de localización de caras que usan eigenfaces [6], son obtenidas principalmente por métodos de análisis de componentes de la cara, como son los ojos, la boca, la nariz, oídos, etc. Estas técnicas que generalmente son mucho más precisas y obtienen mas información sobre la imagen analizada requieren que la persona se encuentre de frente en la imagen ya que de lo contrario puede llegar a desaparecer alguna de las características antes mencionadas. Lo que implica no tener buenos resultados. También algunos de estos algoritmos utilizan clusters de computadoras para desarrollar esta tarea para hacer una detección en secuencias de video ya que requieren gran consumo de tiempo de procesamiento.

Otro de los métodos de localización de caras es el espacio de cromático normalizado, en este espacio, lo que se intenta es separar el color de la piel de los demás colores (vease [7-15]). El color de la piel (sea cual sea el grupo étnico al que pertenezca la cara a detectar) puede ser modelado mediante una distribución gaussiana sencilla. Las características faciales de mayor resolución pueden ser buscadas de una forma secuencial (comenzando con los ojos), o por grupos mayores de características. La mayoría de las técnicas anteriores se utilizan sólo para detección de cabezas, hombros y en escenas de frente.

### ***2.2.1 Segmentación de caras desde imágenes fijas***

Esta técnica se utiliza para algunas aplicaciones como reconocimiento de rostros en donde se tienen escenas con vistas frontales, tamaños similares y fondo uniforme (por ejemplo en identificaciones o pasaportes). La segmentación de las imágenes mediante conexiones heurísticas o detección de bordes, para poder utilizar escenas más generales y omitir algunas de las condiciones anteriores se usan redes neuronales en pequeñas ventanas de la imagen.

Las redes neuronales son entrenadas con ejemplos positivos (imágenes de rostros) y con ejemplos negativos (imágenes que no contengan rostros) y con éstas se determina si existe un rostro o no en la imagen. El tiempo de procesamiento es alto ya que requiere



hacer una búsqueda en imágenes de distintas resoluciones, y es imposible usar este método para detectar rostros en tiempo real con una computadora estándar.

### ***2.2.2 Segmentación de caras desde video***

En el caso anterior sólo se usa información de la intensidad de la imagen para detectar el rostro. En el caso del tiempo real se puede usar mas información como es el movimiento.

Por ejemplo Turk y Pentland [1], usan diferencia entre imágenes para detectar el movimiento de personas y el fondo que es estático. El primer paso para realizar la diferencia de imágenes es transformar la imagen (en tonos de gris o color) a una imagen binaria, esto mediante una función de umbral. Entonces la imagen binaria consiste de algunas “manchas de movimiento”, que son analizadas heurísticamente. Se utilizan pocas reglas como son: “el movimiento de la cabeza debe ser pequeño y continuo” y “la cabeza es la pequeña mancha superior sobre la gran mancha (el cuerpo)”. Incluso se determina el tamaño de la cabeza y el movimiento de ésta asumiendo que la cabeza se mueve.

Aparte de la intensidad y el movimiento, el color puede ser otra información útil. Se puede tomar el color de la piel como otra característica. Los sistemas que utilizan el color, son conocidos como sistemas de detección de la piel. La variación de los diferentes sistemas de detección de piel ocurre en las siguientes áreas: diferentes espacios de color, el adiestramiento en línea y fuera de línea, técnicas para ordenar e identificar los colores correspondientes a la piel. La mayoría de los algoritmos de detección de piel usan la segmentación del histograma basados en el color o una variación de este método. Quek y Mysliwiec [2] [3], emplean una tabla de vistas (índice de colores) específica para el usuario con un costo de entrenamiento en tiempo de ejecución. De la imagen muestra, se calcula la tabla de vistas ordenada, permitiendo una segmentación más fácil. Para cada celda de la tabla corresponde un valor en RGB, el detector de piel cuenta cuantos pixeles caen en cada índice. Se normaliza el histograma obteniendo valores probabilísticos. El algoritmo construye la probabilidad de encontrar piel. Ésta es encontrada tomando el valor de cada pixel en la imagen y con la tabla de vistas se encuentra la probabilidad de cada pixel. Solo estos pixeles de alta probabilidad son pasados a la imagen resultante.

### ***2.3 Utilidad de los detectores de piel***

Los sistemas detectores de rostros y los que siguen los rostros utilizan en su mayoría algún algoritmo de detección de la piel. Esto se utiliza frecuentemente en el reconocimiento facial, extracción de rasgos faciales donde los algoritmos detectores de piel son el paso inicial. Algunas de las investigaciones al respecto se presentan en la siguiente tabla:

Autor	Año	Método	Rasgos usados	Posición de la cabeza
Féraud et al.	2001	Redes neuronales	Movimiento; Color; textura	Frontal y perfil
Maio et al.	2000	Plantillas faciales; transformación de Hough	Textura; Imágenes dirigidas	Frontal
García et al.	1999	Análisis estadístico wavelet	Color; Coeficiente wavelet	Frontal y próximo de frente.
Wu et al.	1999	Modelo difuso de color; Coincidencia de plantillas.	Color	Frontal y perfil
Rowley et al.	1998	Redes neuronales	Textura	Frontal
Sung et al.	1998	Aprendizaje	Textura	Frontal
Yang et al.	1998	Modelo de color; Segmentación multiescala.	Color de piel; textura	Frontal
Colmenarez et al.	1997	Aprendizaje	Proceso de Markov	Frontal
Yow et al.	1997	Rasgos y redes de Belief.	Rasgos de geometría facial.	Frontal y perfil
Lew et al.	1996	Markov Random Field		Frontal

*Tabla 2.1 Investigaciones relacionadas con los detectores del piel. [24]*

Los algoritmos de reconocimiento y análisis de expresiones faciales requieren el analizar la imagen para detectar las expresiones faciales y determinar que unidades de acción se encuentran presentes en la imagen. Este proceso puede llegar a ser muy pesado de tal manera que no puede llevarse a cabo, si no están normalizadas las imágenes. Por esta razón se tiene que crear un proceso de detección de rostros para recortar las imágenes.

La ventaja de detectar la piel, es que en los detectores de rostros y seguimientos de caras el proceso es mucho más rápido, debido a que al realizar una segmentación inicial de la imagen, es mucho más barato (tiempo de procesador) que analizar toda la imagen y por esta razón es un paso inicial en los algoritmos de reconocimiento, análisis de expresiones faciales y rastreo de la cara; Esto puede ser muy útil como apoyo visual en la forma de indicar la dirección a sillas de ruedas “inteligentes” para personas discapacitadas [22].

# Capítulo 3

## TEORIA Y ESPACIOS DEL COLOR

### 3.1 Introducción

El color es el resultado de la percepción de la luz, que es parte del espectro electromagnético en una región entre 400nm a 700nm. Un histograma es una distribución del color en un espacio de color y a sido usado para la visión por computadora para el entendimiento de imágenes. Por ejemplo el análisis del histograma del color es la llave para aplicar modelos físicos basados en visión por computadora. Se ha mostrado que el color no aparece aleatoriamente en un plano, sino que aparece de una forma altamente homogénea en grupos específicos de colores [18 - 19]. El histograma del color de la piel humana coincide con esta observación como se mostrará en capítulos posteriores.

### 3.2 Teoría del Color

Las relaciones entre la luz y la materia, juega un papel crucial en la apariencia visual. La luz forma parte de las radiaciones electromagnéticas, que, en general (por ejemplo las ondas radioeléctricas) no estimulan el sistema visual humano. Se reserva el nombre de radiación luminosa o de luz a las que tienen un efecto visual, y que no representan más que una ínfima parte del dominio de las radiaciones y deben poseer una longitud de onda comprendida entre unos límites, mal definidos, fijados entre los 400 y 780 nm.



Figura 3.1 Longitud de onda (nm)  
Posición de los colores en el espectro de la luz natural.

Hacia el año (1802) Young emite la hipótesis de la trivariancia visual, idea según la cual no deben existir más de tres tipos de receptores en el ojo para ver los colores ya que sabía que LeBlon había impreso correctamente los colores, no con las siete tintas tradicionales, sino con una cuidadosa elección de tres únicas. Young imaginó entonces que hay tres tipos de receptores en el sistema visual, uno sensible al rojo, un segundo al amarillo (cambiado poco después al verde) y un tercero al azul. Poco antes Grassmann había establecido las leyes fundamentales que ponen las bases de la estructura matemática para la medida del color, y que todavía hoy perduran. Los resultados de Maxwell conducen a representar los colores por tres números y por consiguiente poder figurar como un punto en un espacio geométrico.

Es esta calibración la que, juntamente con las leyes de Grassmann, permite predecir el resultado de cualquier combinación de colores.

### **3.3 Bases de la colorimetría tricromática**

Un principio enunciado por Newton (1672), reafirmado por Young (1802) y verificado en la práctica, afirma que cualquier color puede ser igualado por la suma de tres colores primarios convenientemente elegidos. Este principio fundamental es la base sobre la que se sustenta toda la colorimetría. Una técnica experimental simple para llevar a cabo la duplicación de un color es la siguiente: el observador mira en un instrumento óptico cuyo campo visual está dividido en dos partes, la luz cuyo color debe igualarse se introduce en una mitad y la otra se ilumina con las luces de las tres fuentes primarias. Variando las cantidades de estas luces puede encontrarse una combinación que produce una igualación de color exacta de las dos partes del campo visual, y sólo una combinación dada produce una igualación de color. El color desconocido puede especificarse por esas cantidades llamadas **valores triestímulo**. Los valores triestímulo así obtenidos constituyen una especificación del color para el observador y las condiciones que determinan el estímulo equivalente.

### **3.4 Leyes de Grassman**

Los diversos experimentos de hace ya muchos años sobre mezclas aditivas de colores fueron estudiados por Hermann Günter Grassman y se conocen como leyes de Grassman. Se entiende por mezcla aditiva de colores la combinación de dos estímulos que actúan de forma que entran en el ojo simultáneamente e inciden en la misma zona de la retina. Estas leyes dan lugar a otras propiedades importantes en la igualación del color:

1. Para especificar una mezcla de color son necesarias y suficientes tres variables independientes. De aquí el nombre de "triestímulo" y de "colorimetría tricromática" y el principio según el cual todos los tonos de color pueden igualarse con una adecuada mezcla de tres triestímulos diferentes, siendo la única condición a la hora de elegirlos que ninguno de ellos pueda obtenerse (igualarse) como mezcla de otros dos. Matemáticamente se puede expresar:

$$c(C) = r(R) + g(G) + b(B)$$

donde r, g, b son la luminancia, en lúmenes, de tres fuentes de referencia R, G, B (roja, verde, azul) necesarios para igualar c lúmenes de un color dado C.

Lo esencial de esta ley es que existe una y sólo una combinación de tres primarios que igualan cualquier color percibido.

2. Los estímulos que evocan la misma apariencia de color producen resultados idénticos en mezclas aditivas de colores.
3. En una ecuación de color la igualdad no varía si ambos miembros se multiplican por una constante K (ley de proporcionalidad). Esto supone que una vez igualado un estímulo las proporciones de cada componente de la mezcla aditiva no cambia cuando lo hace la luminancia del estímulo de color.

4. Si se suman dos colores cualesquiera, evaluados con los mismos estímulos de referencia, la suma algebraica de ambos es equivalente a la suma algebraica de las cantidades de los estímulos de referencia necesarios para igualar los colores en cuestión (ley de aditividad)

### ***3.5 Síntesis tricromática o modelo RGB***

De la trivariancia visual se puede deducir que es necesario mezclar tres colores para reproducir cualquier otro. Naturalmente, la mezcla debe hacerse en una proporción adecuada de los tres colores, y los tres números que expresan la proporción de mezcla deberán servir para caracterizar el color reproducido. A partir de esta simple idea, desarrollaremos un sistema de coordenadas en el que cada color real (tanto monocromático como de espectro continuo) vendrá definido por tres números.

$$C = R + G + B$$

Donde es un valor de intensidad para R, G, B.

Por consiguiente, los representaremos por R, G y B. En el sistema CIE (1931) se toman como primarios el 435.8 nm (azul), el 546.1nm (verde) y el 700nm (rojo).

Esta representación es muy utilizada en el procesamiento de imágenes y se utilizan tres planos de bytes para representar las intensidades de R, G, B.

#### ***3.5.1 Representación cúbica - cubo de Maxwell***

Con estos tres colores primarios se puede representar un color mediante un cubo llamado de Maxwell. En este cubo se define un sistema de ejes coordenados donde cada eje representa cada color primario R, G, B. En este se representa la relación entre el componente tricromático R, G y B.

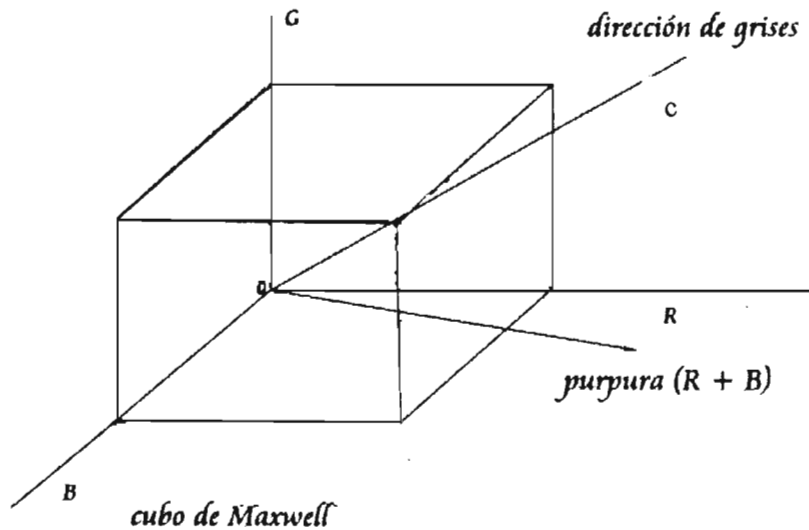


Figura 3.2 Cubo de Maxwell

Del cubo se obtienen las siguientes conclusiones:

- La dirección (orientación) espacial del vector OC definida por el punto C (color C) define el matiz del color.
- La longitud del vector define la luminancia del mismo.
- En la diagonal del plano G-R se tiene el color mezcla entre rojo y verde, o sea amarillo.
- Idéntica situación pasa en los planos B-G donde la diagonal corresponde al cian y en el plano B-R donde la diagonal indica el púrpura.
- Para la obtención del blanco se mezclan R, G y B en proporciones iguales, con sus valores máximos de R, G, B.

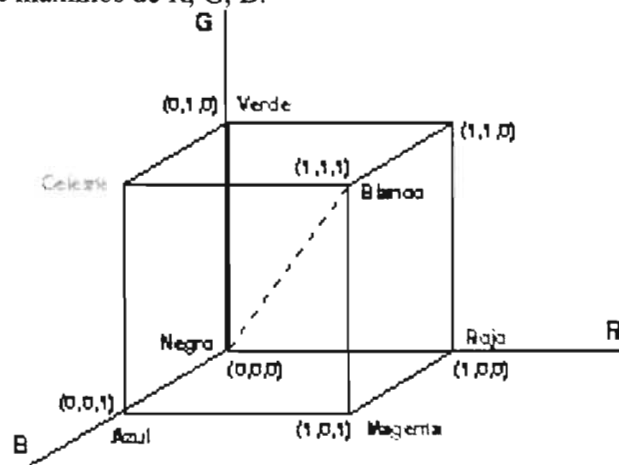


Figura 3.3 Cubo de color RGB. La diagonal principal indica los niveles de grises, desde el negro en el origen (0, 0, 0) al blanco en el punto (1,1,1)

En la figura 3.4 se muestra una imagen en color de verduras y frutas junto con sus tres bandas o planos de color rojo, verde y azul. Obsérvese que el mayor nivel de gris en cada banda corresponde con el color predominante en cada objeto (fruta o verdura).

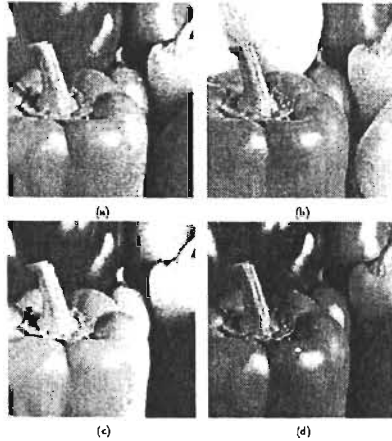


Figura 3.4 A) Imagen original en color. B) Plano de rojo. C) Plano de verde y (D) plano de azul

En una imagen digital de solo una banda (B, C y D) un pixel tiene asociado un único valor (nivel de gris) mientras que en una imagen multibanda un pixel se representa mediante un vector. Si suponemos que disponemos de los planos de color individualmente y cada patrón es un pixel de la imagen, cada patrón puede representarse por un vector de dimensión 3 ( $n = 3$ ),

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \text{ donde } \begin{cases} x_1 \text{ es el nivel de gris en el plano R} \\ x_2 \text{ es el nivel de gris en el plano G} \\ x_3 \text{ es el nivel de gris en el plano B} \end{cases}$$

En este caso, por la notación introducida anteriormente,  $G_i = \{0, 1, \dots, 255\}$  para  $i = 1, 2, 3$ . Así, un patrón de rojo intenso podría tener asociado el vector  $[250, 5, 2]^T$  mientras que otro de verde intenso podría tener asociado el vector  $[0, 255, 2]^T$ .

### 3.6 Sistema Munsell

Albert Munsell fue un profesor de pintura cuya afición, poco frecuente entre los artistas, de expresar científicamente los parámetros de los elementos que utilizaba para su trabajo y sobre todo para enseñar a sus alumnos, le hizo desarrollar un sistema que hoy sigue utilizándose sobre todo en Estados Unidos.

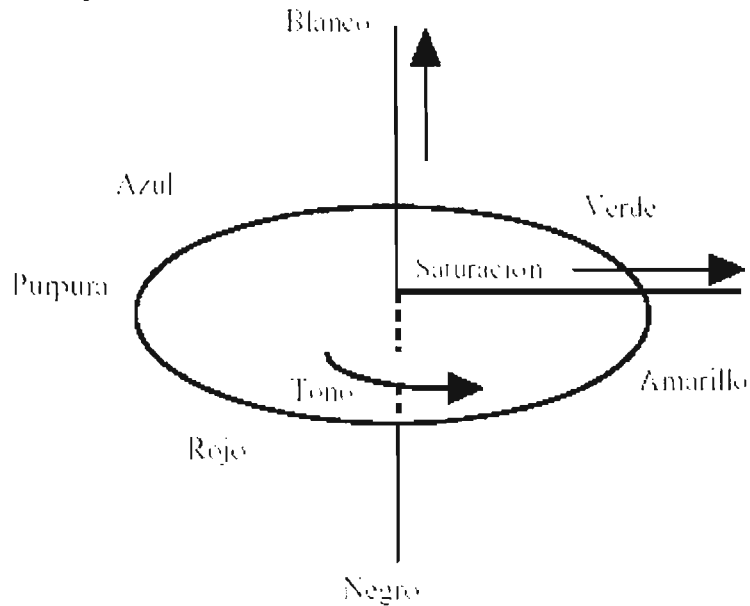
En 1905 publicó un artículo donde afirmaba que el color puede ser especificado mediante un sistema apropiado, basado sobre el *tono*, *saturación* y *luminosidad* (Hue, Chroma y Intensity) de nuestras sensaciones, en lugar de procurar describirlos mediante los infinitos y variantes colores de los objetos naturales.

Proponía un sistema que “estableciera las tres dimensiones del color y midiera cada una de ellas mediante una escala apropiada”. Las ventajas de su sistema fueron descritas por él mismo como se detallan:

- Se reemplazan definiciones vagas y abstractas del color por una notación definida.
- Cada nombre de un color, autodefine su grado de tono, claridad y croma.
- Cada color se puede registrar y comunicar mediante un código.
- Se puede escribir la especificación de un color y verificarla mediante pruebas físicas.

- e) Los colores nuevos no perturban la clasificación ordenada, ya que les está reservado un lugar.
- f) El decoloramiento se puede definir y representar gráficamente a ciertos intervalos poniéndose, así, de manifiesto su progreso en términos de tono, claridad y croma.

El sólido de colores propuesto tiene un eje principal que es el lugar donde se representan los colores que van desde el blanco hasta el negro (extremo superior e inferior respectivamente), su simetría polar en el eje horizontal indica, mediante el módulo del radio vector la saturación (croma) del color, cuyo tono está dado por el ángulo de ese radio vector con el eje de coordenadas.



*Figura 3.5 Representación del espacio de Munsell*

El tono (H) de un color indica su posición en una escala espaciada de 100 tonos. Esta escala está compuesta por los 10 tonos fundamentales (5 principales y 5 intermedios) igualmente espaciados.

La notación de los colores está compuesta por los nombres de los colores más usados: rojo, amarillo-rojizo, amarillo, amarillo-verdoso, verde, verde-azulado, azul, violeta (púrpura-azulado), púrpura y rojo-púrpura.



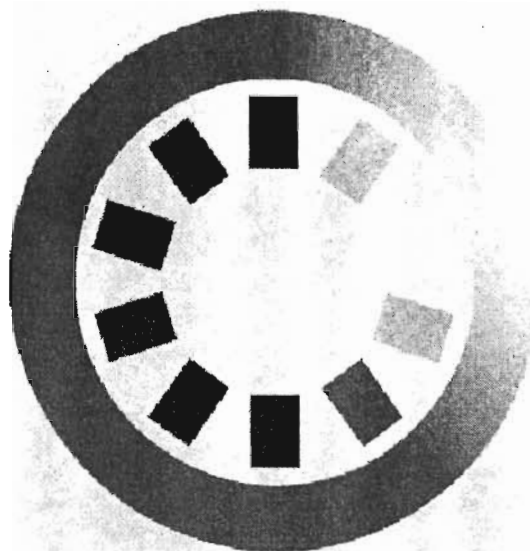


Figura 3.6 Distribución de los colores más usados en el sistema de Munsell.

Las iniciales del color unidas a la especificación numérica, determinan el color de una forma muy descriptiva. Así, el 0 es el RP-R (rojo-púrpura rojo), pasando por el 5, R (rojo); el 25, Y (amarillo); el 45, G (verde); el 65, B (azul); y el 85, P (púrpura) y los intermedios YR en 15 (amarillo-rojo); el 35, GY (verde-amarillo); el 55 BG (azul-verde); el 75 PB (púrpura-azul) y el 95 RP (rojo-púrpura).

La luminosidad es la magnitud que corresponde a la escala de grises, e indica lo luminoso u oscuro que es un color respecto a una escala de grises que se extiende desde el negro absoluto (0) al blanco absoluto (10). Así el símbolo 5 representa un gris medio.

La saturación indica el grado de separación entre un tono determinado y un gris de la misma claridad. La escala de la saturación se extiende desde 0, para un gris, hasta 10, 12, 14 o más, dependiendo de lo saturado que sea el color que se va a evaluar.

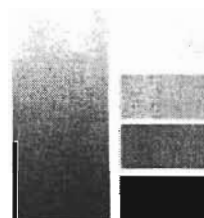


Figura 3.7 En la parte superior un ejemplo de claridad y abajo un ejemplo de croma.

### 3.7 Modelo HSI

Se denomina longitud de Onda Dominante a la luz que se ve y que corresponde a una determinada cantidad de "Tono".

La saturación corresponde a la pureza del color, asociada a la proporción de luz pura de la longitud de onda dominante y la luz blanca necesaria para definir el color. Un color 100% puro tiene una saturación del 100% y no contiene luz blanca. Colores que se alejen del color puro contienen una mezcla entre luz blanca y el color puro comprendida entre 0% y 100% de saturación. La luz blanca y los grises tienen un 0% de saturación ya que no contienen color ni ninguna longitud de onda dominante. La luminosidad es la cantidad de luz.

Concepto	Colorimetría
Tono ( <b>H</b> ue)	Longitud de onda dominante
Saturación ( <b>S</b> aturation)	Pureza
Luminosidad ( <b>L</b> ightness, <b>I</b> ntensity, <b>C</b> roma, <b>B</b> rightness, <b>V</b> alue)	Luminosidad

Tabla 3.1 Modelo HSI  
Variaciones del modelo HSI son HSL, HSV, HSB.

Este modelo está definido por el tono, la luminosidad y la saturación. El espacio se define en un espacio cilíndrico donde se define una doble pirámide hexagonal. H es el ángulo alrededor del eje vertical, siendo el rojo a 0°. Los colores tienen el mismo orden que en el diagrama CIE siguiéndolos en sentido antihorario. El color complementario está a 180° del color indicado. La saturación se mide radialmente variando desde 0 a 1. La luminosidad es 0 para el negro (en vértice inferior de la doble pirámide) y vale 1 para el blanco (en el vértice superior de la doble pirámide)

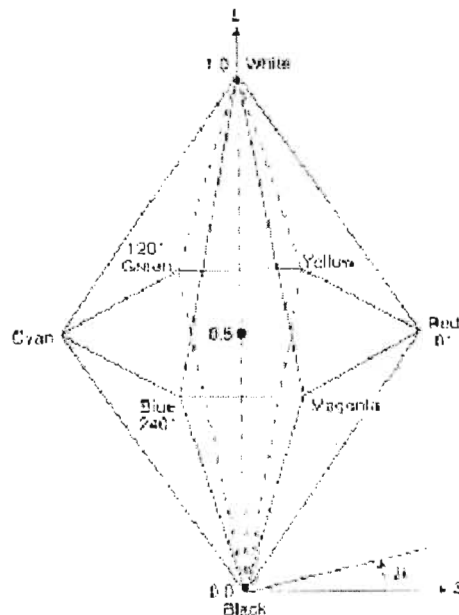
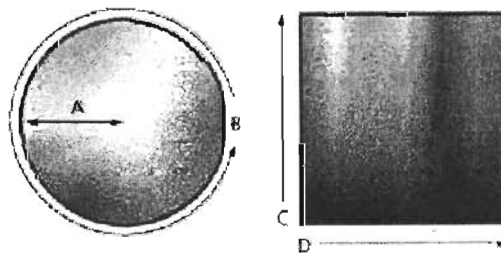


Figura 3.8 Modelo HSI

En este modelo los grises tienen todos el valor  $S=0$  y los tonos más saturados se encuentran a  $S=1$  y  $L=0.5$ .



A. Saturación B. Tono C. Brillo D. Todos los tonos

Figura 3.9 De lado izq. una paleta de colores basada en la Saturación y el Tono. De lado der. una basada en el brillo y el Tono.

Una variación de este modelo es el modelo HSV que fue creado en 1978. También se denomina HSB, siendo B el brillo. El sistema coordenado es cilíndrico, y el subconjunto de este espacio donde se define el color es una pirámide de base hexagonal.

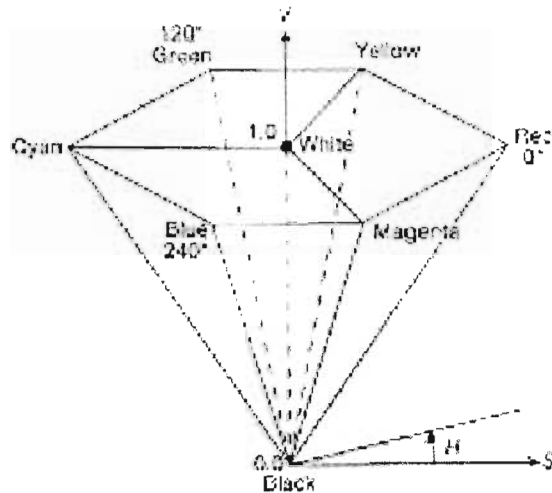


Figura 3.10 El área hexagonal corresponde a un valor de  $V=1$ , conteniendo los colores brillantes. Sin embargo hay que tener en cuenta que los colores con  $V=1$  no se perciben con el mismo brillo por el observador.

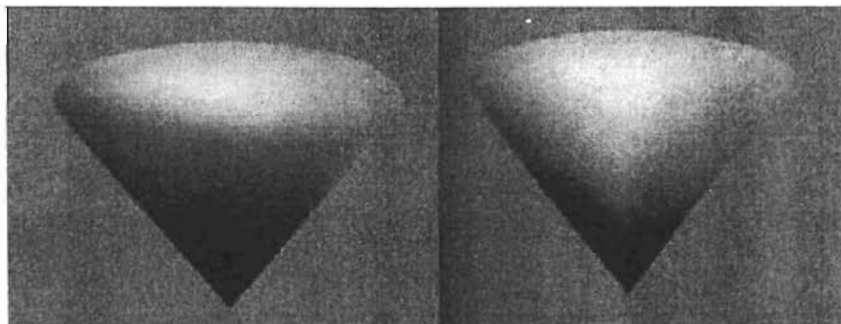


Figura 3.11 El tono se mide como el ángulo alrededor del eje S. El rojo se sitúa a  $0^\circ$ , el verde a los  $120^\circ$ , etc. Los colores complementarios son aquellos que se encuentren a  $180^\circ$  del señalado. El valor de S corresponde al valor desde 0 (coincidiendo con el eje de la pirámide) hasta 1, coincidiendo con el final del área hexagonal de la pirámide.

La pirámide tiene de altura la unidad. El vértice corresponde al negro con coordenadas  $S=0$  y  $V=0$ . El blanco corresponde a  $S=0$  y  $V=1$ . Los valores coincidentes con la altura y por lo tanto con  $V$  variando de 0 a 1, son los grises. Cuando  $S=0$  el valor de  $H=0$  no es importante y se dice que está indefinido. Cuando  $S > 0$ ; el valor de  $H$  empieza a tener importancia. Por ejemplo, el rojo puro se sitúa a  $H=0$ ,  $S=1$ ,  $V=1$ . Si se añade blanco  $S$  disminuye pero sin cambiar  $V$ . Las sombras se crean manteniendo  $S=1$  y disminuyendo  $V$ . Variando  $H$  corresponde a variar el color base de la mezcla.

### 3.8 Medida normalizada del color

El cálculo de la luminancia puede simplificarse al máximo en el caso de que los valores del modelo del color que se están tomando estén en un rango de 0 a 1. Esto quiere decir que el modelo de color este normalizado.

### 3.9 RGB Normalizado

El RGB normalizado corresponde a una transformación no lineal a partir de los componentes RGB, su representación se hace en coordenadas rectangulares y su cálculo es sencillo. Este espacio inicialmente llamado diagrama cromático corresponde al triángulo equilátero unitario conocido también como triángulo de Maxwell.

Definamos ahora las siguientes cantidades, que denominaremos coordenadas cromáticas:

$$r = \frac{R}{R+G+B}, \quad g = \frac{G}{R+G+B}, \quad b = \frac{B}{R+G+B}$$

Si representamos en el plano las coordenadas  $r$  y  $g$  nos referiremos a ella como lugar del espectro o *locus espectral*.

#### 3.9.1 Representación triangular de los colores

Haciendo cortar el plano  $r+g+b=1$  con el cubo de Maxwell se obtiene el triángulo equilátero de Maxwell, donde se puede representar la crominancia, manteniéndose constante la luminancia, independizándose de esta. Esto se representa en el triángulo de coordenadas  $(1,0,0)$ ,  $(0,1,0)$  y  $(0,0,1)$  o triángulo de Maxwell.

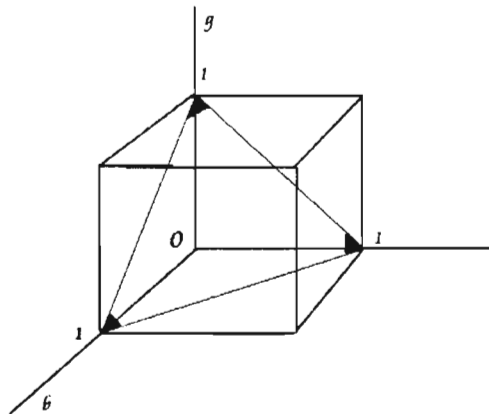


Figura 3.12 Triángulo equilátero de Maxwell

De este triángulo se obtienen las siguientes conclusiones:

- Los vértices del triángulo corresponden a  $b=1 \ g=0 \ r=0$ ;  $g=1 \ b=0 \ r=0$  y  $r=1 \ b=0 \ g=0$ .
- El centro del triángulo contiene al punto blanco de igual energía.
- El lado derecho representa a los colores del plano  $gr$ , formados por las diversas proporciones de los primarios R y G, en cuyo centro se encuentra el amarillo de  $r=0.5$  y  $g=0.5$ .

- El lado izquierdo del triángulo representa a los colores del plano  $gb$  formados por las diversas proporciones de los primarios G y B. En su centro esta situado el cian de  $g=0.5$  y  $b=0.5$ .
- En la base se tienen los colores del plano  $br$ , formado por las diversas proporciones de los primarios B y R, en cuyo centro se sitúa el púrpura de  $b=0.5$  y  $r=0.5$ .
- Los colores correspondientes a puntos en el interior del triángulo, son colores obtenidos a partir de un primario mezclado con algo de blanco, por lo tanto son colores no saturados.

Para incluir el concepto de luminancia es preciso trabajar con planos triangulares paralelos, a distancias distintas del origen.

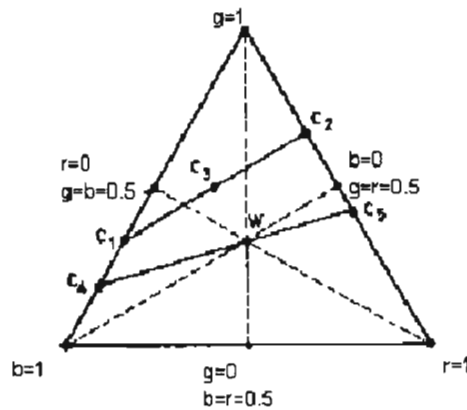


Figura 3.13 Obtención de la mezcla de colores en el triángulo de Maxwell

Actualmente el diagrama cromático corresponde a un triángulo rectángulo (al lado derecho) donde  $r$  y  $g$  son perpendiculares. Lo que nos da un función que realiza un mapeo de  $R^3 \rightarrow R^2$ .

$$r = \frac{R}{R+G+B}, \quad g = \frac{G}{R+G+B}$$

En ocasiones a este también se les llega a llamar colores puros ya que representan el color y minimizando la influencia de el brillo del color.

### 3.10 Espacio $YC_bCr$

CCIR 601 es el viejo nombre de un estándar publicado por el CCIR para decodificar las señales entrelazadas análogas de video en una forma digital. Incluye métodos para codificar 525-líneas 60 hertzios y 625-líneas las señales de 50 hertzios, con 720 muestras de la luminiscencia y 360 muestras del chromancia por línea. El sistema de codificación del color se conoce como YUV.

YUV, también conocido como  $YC_bCr$  y  $YP_bPr$ , es un espacio de color en el cual la  $Y$  es la componente de luminiscencia (el brillo) y  $U$ ,  $V$  son componentes del chromancia (color). Se utiliza comúnmente en los usos video, donde también se refiere como vídeo componente. Esta señal se crea de una fuente original del RGB ( rojo, verde y azul). Los

valores cargados de R, de G y de B se agregan juntos para producir una sola señal de Y, representando el brillo total, o la luminiscencia, de ese punto. La señal de U entonces es creada restando la Y de la señal azul del RGB original, y V restando la Y del rojo.

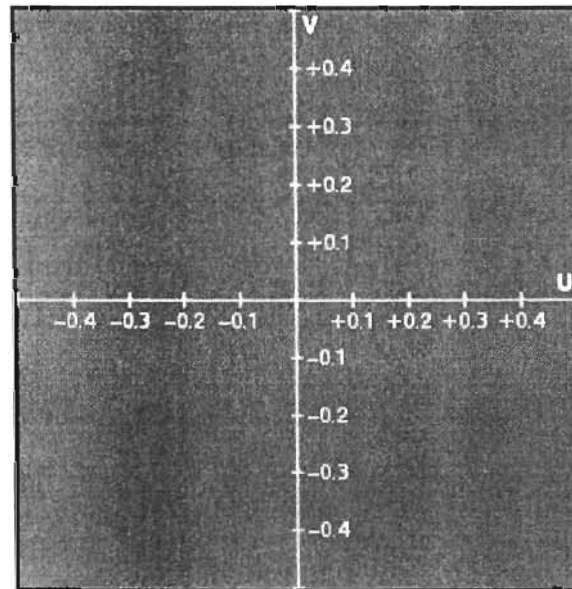


Figura 3.14 Ejemplo del plano u-v del color, valor de  $Y = 0,5$

De esta forma en la representación  $YC_bC_r$ , Y representa la luminancia, es decir la intensidad del pixel, mientras que  $C_b$  es el componente de azul y  $C_r$  es el componente de rojo, puesto que el verde se codifica ya en la señal de Y, las señales de  $C_b$  que resultan y de  $C_r$  son substancialmente más pequeñas que serían de otra manera si las señales originales del RGB. Nótese que la luminancia es el brillo producido por los tres colores primarios, lo que significa que el parámetro Y incluye información cromática.

Los valores que toman Y se encuentran en el rango de 16-235 (blanco y negro),  $C_b$  y  $C_r$  están definidos en el dominio 16-240 donde el 128 corresponde al cero.

Para realizar transformaciones entre el modelo  $YC_bC_r$  y el modelo RGB se utilizan las siguientes ecuaciones:

De  $YC_bC_r$  a RGB.

$$\begin{aligned} Y &= (77/256)R + (150/256)G + (29/256)B \\ C_b &= -(44/256)R - (87/256)G + (131/256)B + 128 \\ C_r &= (131/256)R - (110/256)G - (21/256)B + 128 \end{aligned}$$

RGB a  $YC_bC_r$

$$\begin{aligned} R &= Y + 1.371(C_r - 128) \\ G &= Y - 0.698(C_r - 128) - 0.336(C_b - 128) \\ B &= Y + 1.732(C_b - 128) \end{aligned}$$

# Capítulo 4

## FUNDAMENTOS DE PROBABILIDAD PARA EL ANÁLISIS DE IMÁGENES

### 4.1 Función de densidad de probabilidad normal unidimensional

Si tenemos  $n$  mediciones de un mismo fenómeno  $x_1, x_2, \dots, x_n$ , deberíamos tener  $x_1 = x_2 = \dots = x_n$  si no tuviéramos errores. En su anexo sobre el método de los mínimos cuadrados, “Nuevos métodos para la determinación de las órbitas de los cometas”, Legendre (1805) propone determinar el valor único  $z$  de la medición de manera que una función de los errores sea mínima:

$$\min_z \sum_{i=1}^n (x_i - z)^2$$

La solución es el promedio de las mediciones. Entre las funciones de densidad de probabilidad, la **función de densidad normal**, es desde entonces, un buen estimador para gran parte de fenómenos físicos y es muy tratada en la literatura por su fácil manejo analítico y sus interesantes propiedades. Como son:

1. **Parámetros que especifican la distribución.** La función de densidad normal queda completamente especificada por pocos parámetros. En el caso unidimensional, bastan únicamente dos parámetros: la media y la varianza. En el caso multidimensional, el vector medio y la matriz de covarianza.
2. **Sin correlación e independencia.** Dado un conjunto de patrones que siguen una distribución normal, si las variables asociadas están sin correlación, entonces son independientes.
3. La función de densidad normal es acertada en situaciones en las que un conjunto de patrones de una determinada clase toman valores en un rango *continuo* y alrededor de un *patrón promedio*. Esto es, *considera que los patrones de clases diferentes tienen distintos valores pero los valores de los patrones de una clase son lo más parecido posible*.
4. **La varianza frente a transformaciones lineales.** La distribución que sigue cualquier combinación lineal de una variable aleatoria normal es también normal (con diferentes parámetros). Además, siempre es posible encontrar una transformación *lineal y no singular* que hace que la nueva matriz de covarianza sea *diagonal*, esto es, *siempre puede encontrarse, para una distribución normal, un nuevo conjunto de ejes tal que las nuevas variables son independientes en este nuevo sistema*.

Esta propiedad es particularmente interesante cuando se aplican transformaciones lineales a los datos, con objeto de *resaltar algunas características* que se ponen de manifiesto con estas transformaciones.

Además, desde un punto de vista práctico la relación calidad-costo de la clasificación es mucho mejor que con otros modelos más complejos y los clasificadores diseñados bajo esta suposición son clasificadores **robustos**.

La función de densidad de probabilidad normal para una variable es la siguiente:

$$p(x | \omega_i) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}}$$

que indica la probabilidad de que, asumiendo que una cierta clase sea  $\omega_i$ , el patrón observado tenga el valor  $x$ . En la anterior ecuación.

$\mu_i = E [ x | \omega_i ]$  es la **media** de la clase  $i$ .

$\sigma_i^2 = E [ (x - \mu_i)^2 | \omega_i ]$  es la **varianza** de la clase  $i$ .

La función de densidad de probabilidad normal (unidimensional) está *completamente* especificada por dos parámetros:  $\mu_i$  y  $\sigma_i^2$ , y se suele abreviar por  $P(x | \omega_i) \rightsquigarrow N(\mu_i, \sigma_i^2)$ .

A continuación se representan tres funciones de densidad de probabilidad normales de media 0 y varianzas: 0.15, 1 y 2. Observe la forma simétrica y de “campana” que caracteriza a estas funciones. Recordar que el área bajo cada campana es 1 por lo que, informalmente hablando, las campanas bajas serán anchas mientras que las campanas estrechas serán altas. La “anchura” de las campanas está en relación inversa con el valor de la varianza: a menor varianza, los datos estarán más concentrados alrededor de la media y por lo tanto, la probabilidad de encontrar un valor cercano a la media aumenta la altura de la campana mayor.

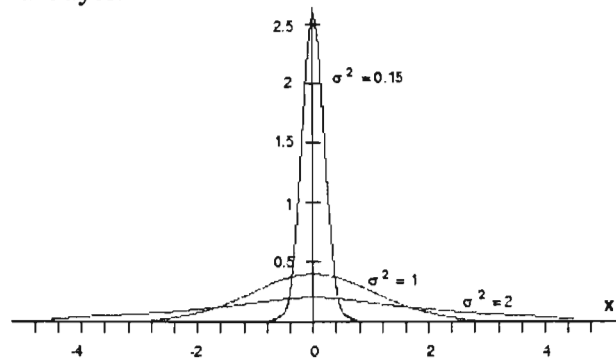


Figura 4.1 Ejemplo de distribución normal

Una propiedad interesante y útil de la función de densidad normal es la siguiente: *el área bajo la curva de la función de densidad de probabilidad normal puede calcularse de forma precisa según el número de desviaciones típicas.*



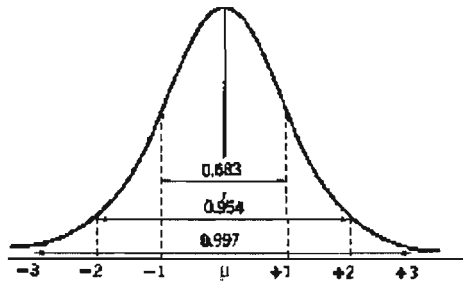


Figura 4.2 Áreas bajo la curva de la función de probabilidad gaussiana en función del número de desviaciones típicas.

Este valor indica la proporción de la población que se encuentra en determinados intervalos centrados en la media. Así, si  $\mu$  es el valor medio y  $\sigma$  es la desviación típica,

El 68.3% de las observaciones están en el intervalo  $[\mu - \sigma, \mu + \sigma]$ .

El 95.4% de las observaciones están en el intervalo  $[\mu - 2\sigma, \mu + 2\sigma]$ .

El 99.7% de las observaciones están en el intervalo  $[\mu - 3\sigma, \mu + 3\sigma]$ .

En la práctica,  $\mu_i$  y  $\sigma_i^2$  son desconocidos y deben *estimarse* a partir de los prototipos de la clase  $\omega_i$ .

$$\hat{\mu}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_j, \quad \hat{\sigma}_i^2 = \frac{1}{N_i - 1} \sum_{j=1}^{N_i} (x_j - \hat{\mu}_i)^2$$

donde:

$N_i$  es el número de prototipos de la clase  $i$ .

$x_j$  es el  $j$ -ésimo prototipo de la clase  $i$ .

## 4.2 Función densidad de probabilidad normal multidimensional

La forma de la función de densidad de probabilidad normal para  $d$  variables es una extensión directa de la unidimensional:

$$p(x | \omega_i) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_i|}} e^{-\frac{1}{2}(x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i)}$$

Ecuación general de probabilidad multivariada

donde:

$\mu_i = E[x | \omega_i]$  es el **vector medio** de la clase  $i$ ,

$\Sigma_i = E[(x - \mu_i)(x - \mu_i)^T | \omega_i]$  es la **matriz de covarianza** de la clase  $i$ ,

$|\Sigma_i|$  es el determinante de  $\Sigma_i$ ,

$\Sigma_i^{-1}$  es la matriz inversa de  $\Sigma_i$ ,

$(x - \mu_i)^T$  es el vector traspuesto de  $(x - \mu_i)$ .

La función de densidad de probabilidad normal multivariante está *completamente* especificada por los parámetros recogidos en  $\mu_i$  y  $\Sigma_i$ . Se suele abreviar como  $P(X|\omega_i) \rightsquigarrow N(\mu_i, \Sigma_i)$ .

A continuación se presenta una función de densidad de probabilidad normal para un conjunto de patrones bidimensionales

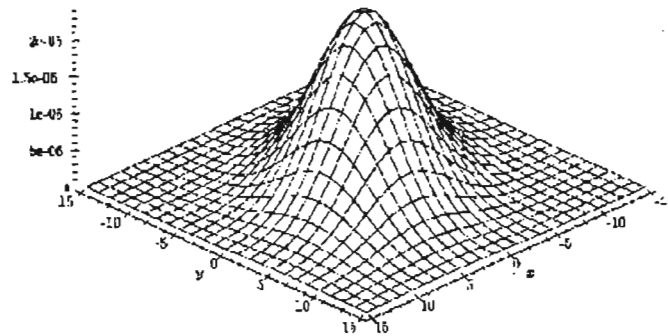


Figura 4.3 Representación de una función de probabilidad normal bidimensional

En la práctica, los parámetros que definen la distribución son desconocidos y deben estimarse a partir del conjunto de prototipos.

$$\mu_i = \begin{bmatrix} \mu_{i1} \\ \mu_{i2} \\ \vdots \\ \mu_{id} \end{bmatrix}, \Sigma_i = \begin{bmatrix} \sigma_{i11} & \sigma_{i12} & \cdots & \sigma_{i1d} \\ \sigma_{i21} & \sigma_{i22} & \cdots & \sigma_{i2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{id1} & \sigma_{id2} & \cdots & \sigma_{idd} \end{bmatrix}$$

Se calcula  $\mu_i$  y  $\Sigma_i$  como:

$$\hat{\mu}_i = \frac{1}{N_i} \sum_{l=1}^{N_i} x_l$$

$$\hat{\Sigma}_i = \frac{1}{N_i - 1} \sum_{l=1}^{N_i} (x_l - \hat{\mu}_i)(x_l - \hat{\mu}_i)^T$$

donde  $N_i$  es el número de prototipos de la clase  $i$  y  $x_l$  es el  $l$ -ésimo prototipo de esa clase.

### 4.3 Distancia de Mahalanobis

Si un conjunto de patrones siguen una distribución normal, tienden a representarse formando un único agrupamiento de manera que el centro de este agrupamiento está determinado por el vector medio y la “forma” por la matriz de covarianza.

De la ecuación general de probabilidad multivariada vista anteriormente se deduce que los puntos para los cuales el valor de la densidad de probabilidad constante están situados en hiperelipsoides para las que se verifica que la forma cuadrática  $(x - \mu)^T \Sigma_i^{-1} (x - \mu)$  es

constante. El valor de esta expresión es la **distancia de Mahalanobis** (al cuadrado) de  $x$  a  $\mu$ . En el caso bidimensional, estas elipses pueden verse claramente como se muestra.

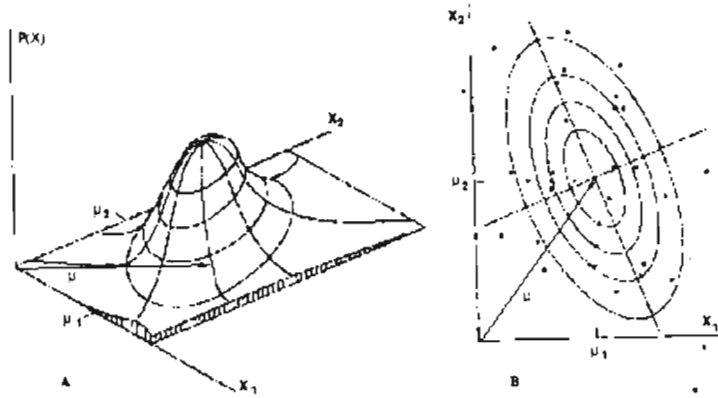


Figura 4.4 A) Densidad de probabilidad. B) Diagrama de dispersión

De lado izquierdo se muestra la gráfica de la densidad de probabilidad y de lado derecho representamos la proyección sobre el plano definido por las variables  $x_1$  y  $x_2$ .

En otras palabras, *los contornos de igual densidad de probabilidad son hiperelipsoides con una distancia de Mahalanobis a  $\mu$  constante*. Las direcciones de los ejes principales de estos hiperelipsoides están determinadas por los auto vectores de  $\Sigma$  y sus longitudes por los autovalores correspondientes.

Es muy importante tener en cuenta que esta distancia está ponderada por  $\Sigma$ , por lo que esta métrica considera la distinta dispersión de las variables en el espacio. Se debe tener en cuenta que el concepto de distancia, con una métrica de este tipo, es muy distinto al concepto de distancia en nuestro mundo Euclídeano. En la figura 4.5 se muestra un ejemplo de esta distancia, se representan los contornos de equiprobabilidad asociados a dos distribuciones normales con igual media y diferente matriz de covarianza. En azul se representa una distribución en la que las variables presentan una alta correlación (alto valor -relativo- de covarianza) y en rojo, una distribución en la que las variables están incorrelacionadas. Los contornos dibujados representan los mismos valores de probabilidad: el más interno de los rojos (el de mayor probabilidad), representa el mismo valor de probabilidad que el más interno de los azules; el segundo de los rojos, el mismo valor de probabilidad que el segundo de los azules, y así sucesivamente.

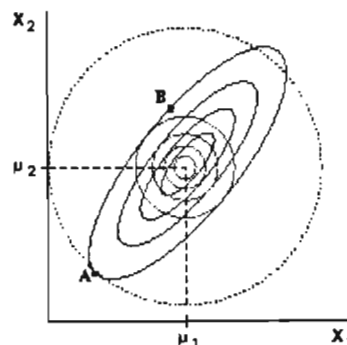


Figura 4.5 Dos distribuciones normales con igual media y diferente matriz de covarianza

Si consideramos la distribución azul, todos los puntos situados en la elipse más externa, por ejemplo, están a la misma distancia (de Mahalanobis) del centro. Así, los puntos indicados con A y B en la figura verifican que:

$$(A - \mu)^T \Sigma^{-1} (A - \mu) = (B - \mu)^T \Sigma^{-1} (B - \mu)$$

Considerando una métrica de Mahalanobis, esta afirmación es del todo correcta, aunque desde una perspectiva clásica (Euclídeana) esta afirmación es sorprendente. En la figura 4.5 se indica con un círculo de trazo punteado los puntos cuya distancia al centro es constante e igual a la distancia del punto A, al centro desde un punto de vista Euclídeano.

#### 4.4 Correlación de variables

Si tenemos una distribución normal, la matriz de covarianza determina la “forma” del agrupamiento. Dependiendo de la forma se deriva una medida normalizada, *el factor de correlación entre dos variables*, que indicará el grado de dependencia entre dos variables cualesquiera.

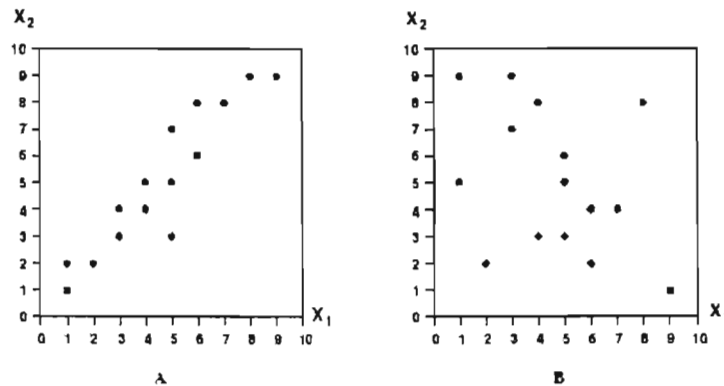


Figura 4.6 A) Alta covarianza. B) Baja covarianza

En la figura 4.6 se muestran dos variables que están estrechamente relacionadas y que presentan una alta covarianza en A ( $\Sigma_{12} = 5.9$ ). Las varianzas de cada variable son:  $\sigma_1^2 = 5.7$  y  $\sigma_2^2 = 7.1$ . Y en B se tienen dos variables con las mismas varianzas que en el caso anterior, aunque son independientes y tienen una baja covarianza ( $\Sigma_{12} = -2.3$ ).

La interpretación de los valores de las covarianzas tiene un serio problema: los valores concretos dependen de las unidades de medida, por lo que la interpretación de valores individuales carece de sentido. Si se desea conocer el grado de interrelación entre dos variables de una manera independiente de las unidades de medida se requiere algún tipo de *normalización*. Una medida normalizada del grado de relación entre variables la proporciona el *coeficiente de correlación*, definido como:

$$\rho_{ij} = \frac{\Sigma_{ij}}{\sigma_i \sigma_j}$$

donde  $\rho_{ij}$  es el coeficiente de correlación entre las variables  $i$  y  $j$ . Se calcula en función de las desviaciones típicas de cada una de las variables y de su covarianza.

Este coeficiente verifica que  $|\rho_{ij}| \leq 1$ . Por una parte, el valor máximo de la covarianza entre dos variables no puede ser mayor que el producto de sus desviaciones típicas. Por otra parte, dado que  $\sigma_i \geq 0$  y  $\sigma_j \geq 0$  el signo del valor de correlación depende del signo de  $\Sigma_{ij}$ . Así, forzosamente  $-1 \leq \rho_{ij} \leq 1$ .

Podemos expresar la matriz de covarianza,  $\Sigma$ , como el producto de tres matrices:  
 $\Sigma = \Gamma R \Gamma$

$$\Gamma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_d \end{bmatrix}, \quad R = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1d} \\ \rho_{21} & 1 & \cdots & \rho_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{d1} & \rho_{d2} & \cdots & 1 \end{bmatrix}$$

$\Gamma$  es una matriz diagonal que contiene las desviaciones típicas de las variables y  $R$  es una matriz (simétrica) que contiene los coeficientes de correlación entre las variables.

Así,

$$\Gamma R = \begin{bmatrix} \sigma_1 & \sigma_1 \rho_{12} & \cdots & \sigma_1 \rho_{1d} \\ \sigma_2 \rho_{21} & \sigma_2 & \cdots & \sigma_2 \rho_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_d \rho_{d1} & \sigma_d \rho_{d2} & \cdots & \sigma_d \end{bmatrix}$$

$$(\Gamma R) \Gamma = \begin{bmatrix} \sigma_1^2 & \sigma_2 \sigma_1 \rho_{12} & \cdots & \sigma_d \sigma_1 \rho_{1d} \\ \sigma_1 \sigma_2 \rho_{21} & \sigma_2^2 & \cdots & \sigma_d \sigma_2 \rho_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_1 \sigma_d \rho_{d1} & \sigma_d \rho_{d2} & \cdots & \sigma_d^2 \end{bmatrix}$$

Por un lado, como  $\rho_{ij} = \frac{\Sigma_{ij}}{\sigma_i \sigma_j}$ , entonces  $\Sigma_{ij} = \sigma_j \sigma_i \rho_{ij}$ .

Además, como  $\Sigma_{ij} = \Sigma_{ji}$ , entonces

$$\rho_{ij} = \frac{\Sigma_{ij}}{\sigma_i \sigma_j} = \frac{\Sigma_{ji}}{\sigma_i \sigma_j} = \rho_{ji}$$

Con esto queda demostrado que  $\Sigma = \Gamma R \Gamma$  para los elementos que no están en la diagonal.

La correlación entre una variable y ella misma es 1 ya que

$$\rho_{ii} = \frac{\Sigma_{ii}}{\sigma_i \sigma_i} = \frac{\sigma_i^2}{\sigma_i \sigma_i} = 1$$

Ya con esto tenemos las condiciones para especificar lo que entendemos por “forma” de una distribución. Si se proyecta una nube de puntos que siguen una distribución normal sobre un plano definido por los ejes  $X_i$  (abscisas) y  $X_j$  (ordenadas),

- la *superficie* está determinada por  $\Gamma$ , la matriz de desviaciones típicas.
- la *forma* está determinada por  $R$  (matriz de coeficientes de correlación).

Recordemos que  $|\rho_{ij}| \leq 1$ , o lo que es lo mismo,  $-1 \leq \rho_{ij} \leq 1$

1. Si  $\rho_{ij} = 0$ , la correlación es nula (las variables son *independientes*): los puntos se disponen aleatoriamente en un círculo ( $\sigma_1 = \sigma_2$ ) o en una elipse ( $\sigma_1 \neq \sigma_2$ ) centrado en  $(\mu_i, \mu_j)$ . Una correlación con valor 0 indica que no existe relación lineal en absoluto. Como mostramos en la siguiente figura.

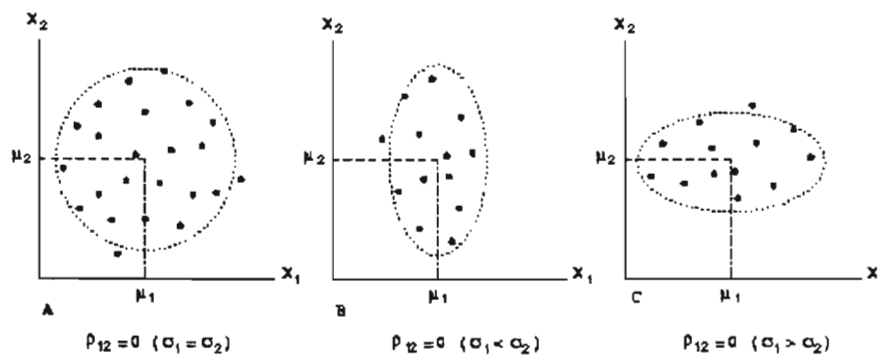


Figura 4.7 En la imagen se muestran unos ejemplos de correlaciones nulas.

2. Si  $0 \leq \rho_{ij} \leq 1$  los puntos se disponen en una elipse centrada en  $(\mu_i, \mu_j)$ . El eje principal tiene una *pendiente positiva* y una forma más o menos circular dependiendo de si el valor de  $\rho_{ij}$  está más o menos cercano a 0, lo que se traduce en una menor o mayor dependencia entre las variables. Una correlación positiva indica una relación lineal directa: cuando una variable crece, la otra también.

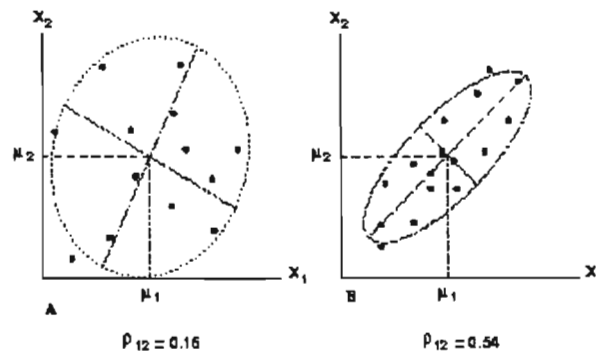


Figura 4.8 Ejemplos de correlación positiva

3. Si  $\rho_{ij} = 1$ , la correlación es *lineal y perfecta* ya que una variable depende linealmente de la otra. Los puntos se disponen a lo largo de una línea recta con pendiente positiva. A continuación se muestran dos ejemplos de correlación

lineal. El primero muestra una correlación lineal perfecta mientras que el segundo no lo es, aunque en la práctica se considere que la es.

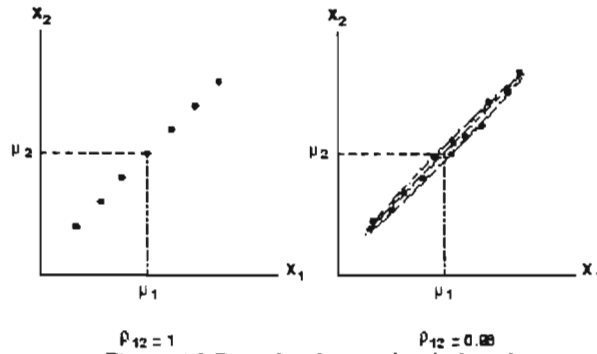


Figura 4.9 Ejemplos de correlación lineal

- Para valores de correlación negativos, la discusión es similar a la de los casos 2 y 3 salvo que la relación entre las variables es inversa: cuando una variable crece, la otra decrece.

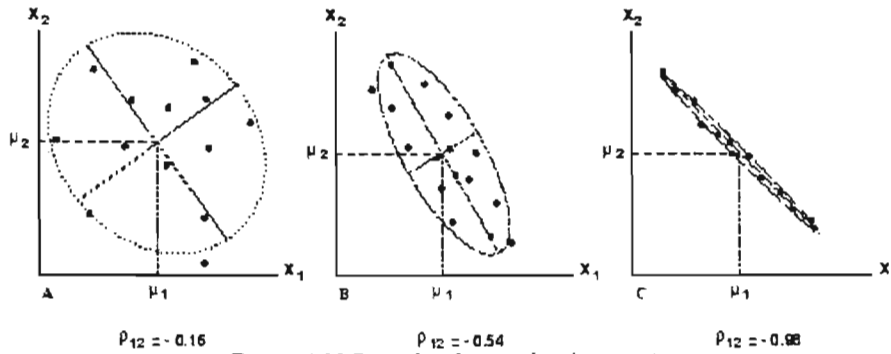


Figura 4.10 Ejemplos de correlación negativa

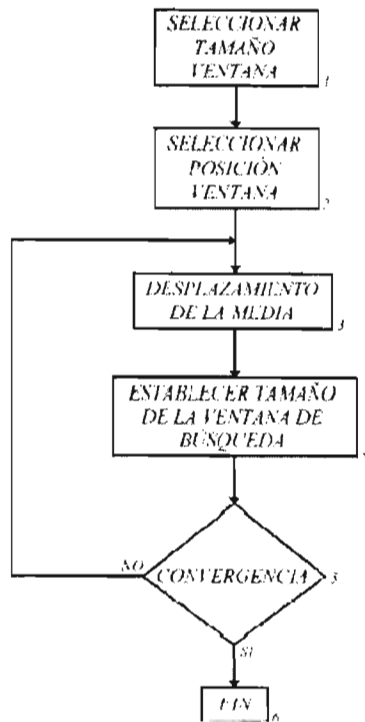
En la figura anterior se muestra en los casos A y B dos ejemplos de correlación en los que  $-1 \leq \rho_{ij} \leq 0$ , mientras que el caso C es un ejemplo de correlación lineal inversa casi perfecta. En todos los casos, la pendiente del eje principal es negativa.

En cualquier caso, la orientación y longitud de los ejes de las elipses que caracterizan las distribuciones, se deducen de los autovectores y los autovalores, respectivamente, de la matriz de covarianza.

#### 4.5 Algoritmo de Camshift

Este algoritmo se basa en estadística robusta, es decir, se tiende a desechar los datos atípicos de la muestra o los que se alejan demasiado del grupo, con lo que se tiende a eliminar el ruido.

El algoritmo trabaja como se muestra en el diagrama que podemos ver a continuación:



1. Elección del tamaño de la ventana de búsqueda.
2. Elección de la posición inicial de la ventana de búsqueda.
3. Desplazamiento de la media (una o varias veces) y actualización del momento de orden 0. El algoritmo de desplazamiento de la media se aplica a una imagen en la escala de grises que indica la probabilidad de que cada píxel pertenezca a la clase color piel.

Donde el momento de orden 0 es:

$$M_0 = \sum_i \sum_j I(i, j)$$

donde  $I(i, j)$  es el valor (probabilidad) del píxel en la posición  $(i, j)$  de la imagen, e  $(i, j)$  varían sobre toda la ventana de búsqueda.

4. Elección del tamaño de la ventana de búsqueda igual a una función del momento de orden 0 encontrado en el paso 3.
5. Evaluación de la convergencia. Si se cumple se finaliza el algoritmo, si no, se regresa al paso 3.

Figura 4.11 De lado derecho se muestra el diagrama de flujo y de lado izquierdo se muestra el algoritmo.

En otras palabras, el algoritmo de *Camshift* [25] trabaja con una imagen de la distribución de la probabilidad del color, de forma que calcula el centroide de la distribución de probabilidad del color dentro de la ventana de búsqueda, vuelve a centrar la ventana y calcula el área para obtener el siguiente tamaño de ventana. De esta forma no necesitamos volver a calcular la distribución de probabilidad del color en la imagen completa, sino que basta con restringir el cálculo de la distribución a una pequeña región de la imagen que envuelve a la ventana de búsqueda actual de *Camshift*.



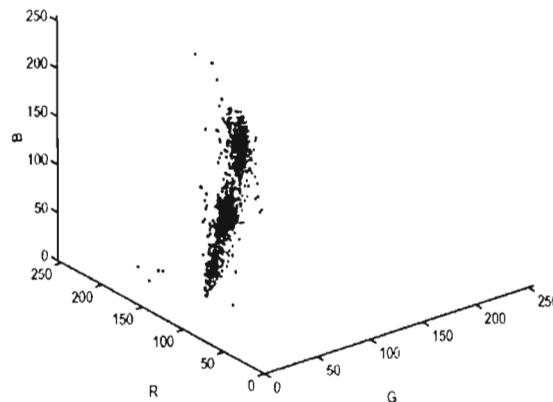
# Capítulo 5

## SEGMENTACIÓN DEL COLOR DE LA PIEL

### 5.1 Introducción

Se ha demostrado que los histogramas del color son representaciones estables del objeto, que no son afectadas por la obstrucción y pueden ser utilizados para distinguir entre una gran cantidad de objetos. A mediados de los años ochenta fue reconocido que el histograma del color para una sola superficie no homogénea con toques de luz tendrá una distribución planar en espacio de color. Se ha demostrado desde entonces que los colores no bajan aleatoriamente en un plano, pero forman racimos en los puntos específicos. Los histogramas del color de la piel humana coinciden con estas observaciones.

La figura 5.1 se muestran las ocurrencias del piel-color en el espacio de color del RGB (255x255x255). El color de la piel se agrupa en un área pequeña del espacio de color del RGB, es decir, sólo algunos de todos los colores posibles ocurren realmente en una cara humana.



*Figura 5.1 Los puntos representan los distintos tonos color de la piel. Se ve claramente que se agrupan en una pequeña área.*

Otro punto importante es que uno de los principales problemas que se tienen al trabajar con las imágenes a color es el tipo y color de la fuente de iluminación.

### 5.2 Compensación de luz

Con el fin de minimizar este problema se utilizó una técnica para compensar la luz ambiental.

La compensación de luz se va utilizar como paso inicial para el proceso de detección del color de la piel. La idea principal es tomar los colores más próximos al blanco y ajustar toda la imagen acoplando estos tonos al color blanco.

Para poder utilizar esta técnica de compensación de luz se tiene como primer paso encontrar los píxeles con los valores más altos de luminiscencia (nonlinear gamma-corrected luminance). Se tomó el 5% de los píxeles con los valores más altos como referencia. Estos valores de luminiscencia corresponden a la intensidad en las imágenes en niveles de gris. Por lo que se generó una copia de la imagen en niveles de gris.

Denotemos el promedio de 5% de los valores de luminiscencia como  $(R_i, G_i, B_i)$ . Los componentes RGB de la imagen son ajustados para igualar todos los valores RGB con luminiscencia mayor a  $(R_i, G_i, B_i)$  y se reescalan para que tomen el valor blanco  $(255, 255, 255)$  (equalizar la imagen basándonos en este valor obtenido).

Al utilizar esta técnica de compensación de luz no siempre realiza cambios notorios en las imágenes ya que en ocasiones el valor obtenido de luminiscencia es el blanco  $(255, 255, 255)$ .

A continuación se presenta el algoritmo de compensación de luz utilizado:

1. Transformar la imagen a niveles de grises *ImgGris*.
2. Localizar los píxeles con valores más altos de luminiscencia en la imagen (5% más altos en *ImgGris*).
3. Localizar el píxel promedio de luminiscencia en *ImgGris* y tomar las mismas coordenada  $(x, y)$  de la imagen RGB.
4. Reescalar (equalizar) los colores con respecto a los valores de ese píxel.

Para poder reescalar la imagen se utiliza la siguiente ecuación:

$$\left( \frac{255}{R_i}, \frac{255}{G_i}, \frac{255}{B_i} \right)$$

Se realiza el producto entre el valor del color anteriormente obtenido y los valores de la imagen a transformar.

$$\begin{aligned} R_j * \frac{255}{R_i} &= NR_j \\ G_j * \frac{255}{G_i} &= NG_j \\ B_j * \frac{255}{B_i} &= NB_j \end{aligned}$$

Donde  $j$  es el índice sobre el barrido de imagen a transformar y  $NR_j$ ,  $NG_j$ ,  $NB_j$  son los valores de la nueva imagen. Los resultados de esta técnica se muestran en la figura 5.2.



*5.2 Resultado después de la compensación de luz. De lado izquierdo se muestran las imágenes originales (sin aplicar la compensación) y de lado derecho las imágenes resultantes después de aplicar la técnica de compensación de luz.*

### 5.3 Representación del color de la piel

Una variedad de distribuciones espectrales de la luz, puede producir que al percibir dos colores (distintos), sean indistinguibles. La retina humana tiene tres tipos diferentes de células de cono fotorreceptoras del color con diferentes curvas del espectro, basado en el sistema humano de percepción del color, tres componentes numéricos son necesarios y suficientes para describir un color, a condición de que se utilicen las funciones espectrales apropiadas. Teóricamente, las coordenadas del color pueden ser definidas como integrales del producto del espectro  $U(\nu)$  del estímulo con tres funciones lineales del color que empareja independiente,  $r(\nu)$ ,  $g(\nu)$ ,  $b(\nu)$ .

$$R = \int_{\nu_1}^{\nu_2} \bar{r}(\nu)U(\nu)d\nu, \quad G = \int_{\nu_1}^{\nu_2} \bar{g}(\nu)U(\nu)d\nu, \quad B = \int_{\nu_1}^{\nu_2} \bar{b}(\nu)U(\nu)d\nu$$

Donde  $\nu$  es el suministro de luz.

Sabemos de antemano que distintas personas comúnmente presentan varios aspectos del color de la piel. Incluso para la misma persona, su aspecto del color de la piel será diferente si cambia de ropa o varían las condiciones de iluminación. Es decir muchos factores contribuyen al aspecto humano del color de la piel. Este espacio en 3D como lo es el RGB presenta estos inconvenientes ya que representa tanto el color y la intensidad de la iluminación.

El sistema visual humano se adapta a diversos brillos y a varias fuentes de iluminación. De tal forma que un mismo color se mantiene dentro de una amplia gama de las condiciones ambientales de la iluminación.

Por lo tanto es posible que eliminemos brillo de la representación del color de la piel y preservar información exacta del color. Si tomamos dos colores en RGB  $[r, g, b]$ , como  $[r_1, g_1, b_1]$  y  $[r_2, g_2, b_2]$ , decimos que son proporcionales si,

$$\frac{r_1}{r_2} = \frac{g_1}{g_2} = \frac{b_1}{b_2}$$

es decir tienen el mismo color pero distinto brillo. El brillo puede ser removido desde un espacio de color (en este caso RGB), mediante la normalización. Se tomará el RGB normalizado visto anteriormente. Este espacio lo llamaremos espacio cromático el cual toma los factores  $(r, g)$  [4], del RGB normalizado y son conocidos como colores puros ya que carecen del brillo, y son calculados como sigue.

$$r = \frac{R}{R+G+B}, \quad g = \frac{G}{R+G+B}$$

Lo que nos define una función de  $R^3 \rightarrow R^2$ , como se muestra en la figura 5.3. El factor azul es redundante ya que después de la normalización tenemos  $r+g+b=1$ . Las variaciones entre los diferentes tonos de color de la piel es reducida considerablemente lo que nos da que este modelo de espacio cromático es mejor que el RGB, otros de los beneficios que, se tienen es la reducción de parámetros al calcular el modelo del color de

la piel haciendo menos complejo los cálculos de este. Incluso se disminuyen las variaciones provocadas por el brillo de la fuente de iluminación.

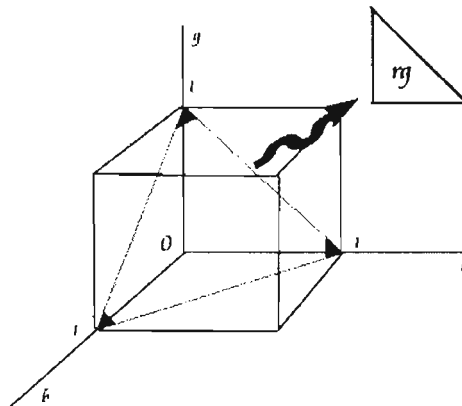


Figura 5.3 Representación del modelo cromático de los colores puros.

Otro punto importante que hay que tomar en cuenta en el color de la piel es la distribución que tiene en el espacio cromático. Como ya había comentado los distintos tonos de la piel humana se agrupa en áreas compactas del espacio RGB Ver figura 5.1. Y como resultado de la normalización la variación entre las distintas tonalidades es menor en el espacio cromático. En otras palabras el color de la piel se encuentra agrupados en un área pequeña del espacio cromático.

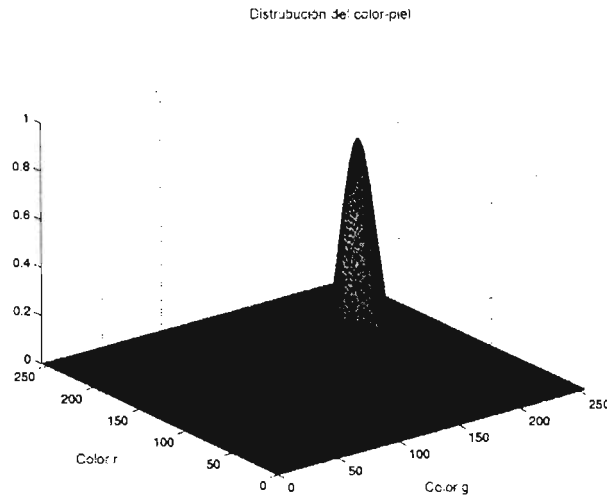


Figura 5.4 Distribución del color de la piel en el espacio normalizado.

Dependiendo del tipo de fuente de iluminación la distribución se desplaza en el espacio cromático. Por ejemplo la luz del sol desplazará el histogramas hacia el color azul porque contiene más azul que la iluminación fluorescente. Se muestra en la figura 5.5 los desplazamientos dependiendo del tipo de fuente de iluminación.

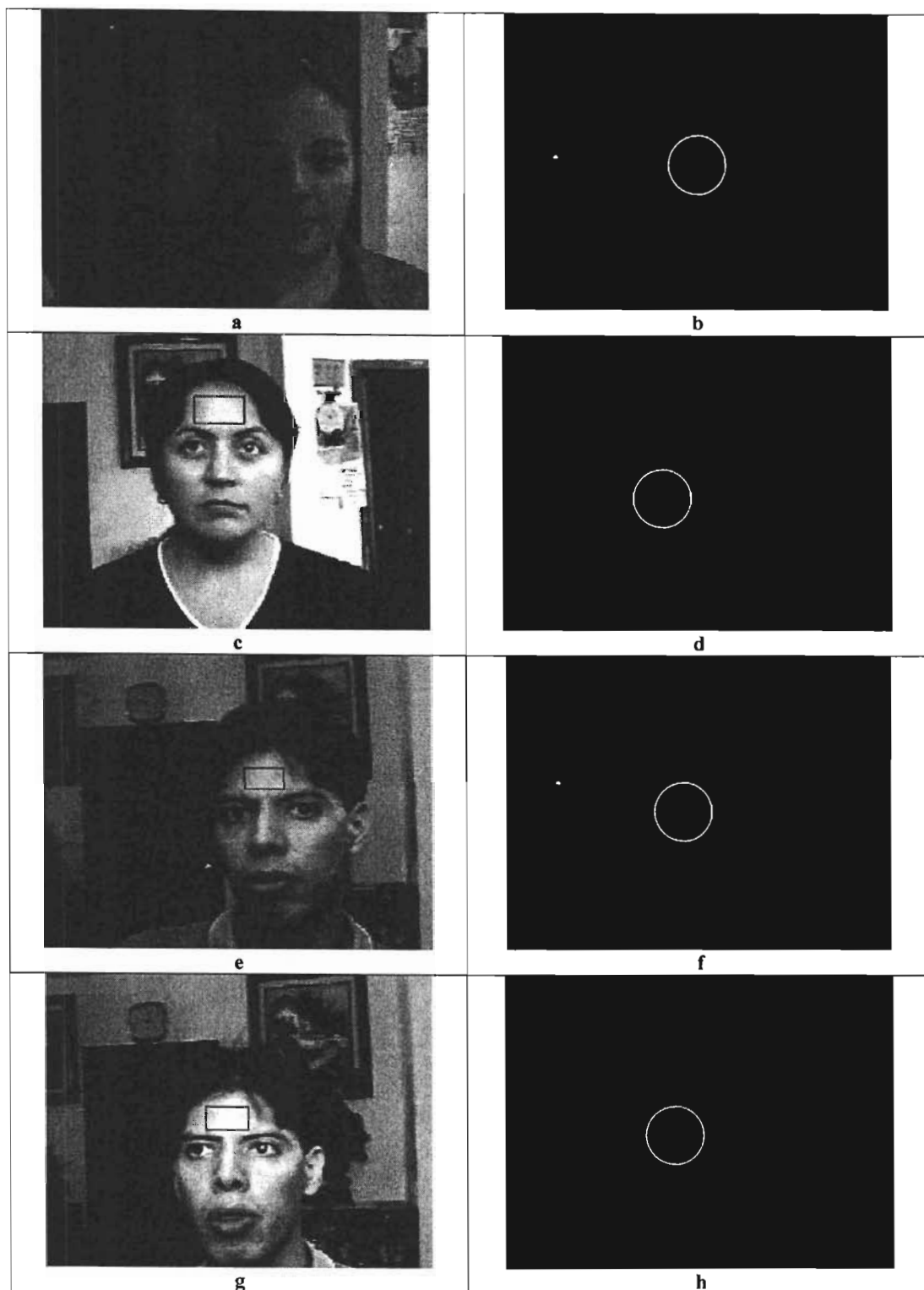


Figura 5.5 Se muestran de lado izquierdo (a,c,e,g) las imágenes con diferentes tipos de iluminación en las cuales se captura una muestra de color de la piel y del lado derecho (b,d,f,h) se graficaron los colores puros (encerrados en un círculo para mejor localización). En la imagen a la iluminación es incandescente (luz anaranjada), en la c la iluminación es fluorescente (luz blanca), en e la iluminación es mixta (luz fluorescente y luz incandescente) y en la g la iluminación es con luz solar. Como se muestra al variar las condiciones de iluminación (como son la intensidad y el color de luz), el color de la habitación. Los valores del color de la piel varían considerablemente.

Por lo tanto, es imposible caracterizar todas las distribuciones del color de la piel usando un modelo fijo. Por otra parte, aunque los colores de la piel de algunas personas parecen variar sobre una amplia gama, es posible modelar la distribución del color de la piel de cada uno bajo cierta condición de la iluminación. Puesto que la distribución del color de la piel tiene solamente dos variables en el espacio de color normalizado. Se tiene que la distribución del color de la piel del rostro en el espacio normalizado posee una distribución normal bivalente, lo que nos indica que es posible utilizar una distribución normal bivalente para caracterizar las distribuciones del color de la piel.

Como ya se menciona, la representación RGB se tiene dada por ternas de 8 bits con valores de 256 que nos da un total de  $256^3$  posibles valores. Al tomar las componentes de los colores por rojo, verde y azul. Se ve que se agrupan los distintos tonos del color de la piel por estas componentes como se muestra en la figura 5.1.

Al calcular las medias y varianzas de varios ejemplos de la figura 5.5 encontramos que las varianzas son similares lo que nos corrobora la distribución normal y como se vio también se encuentra ésta en espacios compactos. También sabemos por el trabajo de Jie Yang [5]. Que la distribución del color de la piel es normal donde los valores de  $r$  y  $g$ , y usando la matriz de covarianzas sabemos la forma del agrupamiento y obtenemos el factor de correlación entre dos variables que nos indica que tiene baja covarianza (ver sección 4.4). Lo que nos dice que la relación es nula y que no se pueden poner los valores en función del otro.

Considerando que posee una distribución normal  $N(\mu, \Sigma)$ , se trata de calcular la probabilidad que tienen los pixeles a ser del grupo color de la piel. Para esto se tiene que calcular el vector esperanza  $\mu = (\bar{r}, \bar{g})$  a un conjunto de las distintas muestras recolectadas. Utilizando las siguientes ecuaciones de

$$\bar{r} = \frac{1}{N} \sum_{i=1}^N r_i$$

$$\bar{g} = \frac{1}{N} \sum_{i=1}^N g_i$$

*Ecuaciones para obtener el vector esperanza  $\bar{\mu}$*

donde  $g_i$  y  $r_i$  son los valores de las distintas muestras de color de la piel.

Otro punto importante, para el cálculo de la probabilidad de que un pixel sea de la clase color de la piel, es el cálculo de varianza de la muestra de color de la piel.

Esto mediante:

$$\Sigma = \begin{bmatrix} \sigma_{rr} & \sigma_{rg} \\ \sigma_{gr} & \sigma_{gg} \end{bmatrix}$$

*Ecuación para obtener la matriz de varianzas*

Al calcular la matriz inversa ( $\Sigma^{-1}$ ) de  $\Sigma$  tenemos que  $\Sigma^{-1}$  queda de la siguiente forma.

$$\Sigma^{-1} = \frac{1}{\sigma_{rr} \cdot \sigma_{gg} - \sigma_{rg} \cdot \sigma_{gr}} \begin{pmatrix} \sigma_{rr} & \sigma_{rg} \\ \sigma_{gr} & \sigma_{gg} \end{pmatrix}$$

Como ya sabemos el vector esperanza y la matriz de varianza, son los únicos datos que requerimos para poder calcular la probabilidad. Entonces usando la fórmula de probabilidad multivariada

$$p(x | \omega_i) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_i|}} e^{-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i)}$$

podemos calcular la probabilidad de que un pixel tenga el color de la piel y tomar los de mayor probabilidad como pixeles del grupo de color de la piel.

Como ya se había mencionado una de las aplicaciones directas para los algoritmos detectores de los modelos el color de la piel es la localización de los rostros en imágenes. Para esto hay que aplicar este modelo a las imágenes de entrada y localizar el agrupamiento del color de la piel. Cada pixel de la imagen de entrada es transformado al espacio cromático normalizado antes visto y se compara su distribución con la distribución de nuestro modelo. Este proceso tiene la ventaja de ser muy rápido, lo que lo hace excelente candidato para la detección en tiempo real. En la figura 5.6 se muestran los resultados de utilizar este método y solo presentar los pixeles con mayor probabilidad.

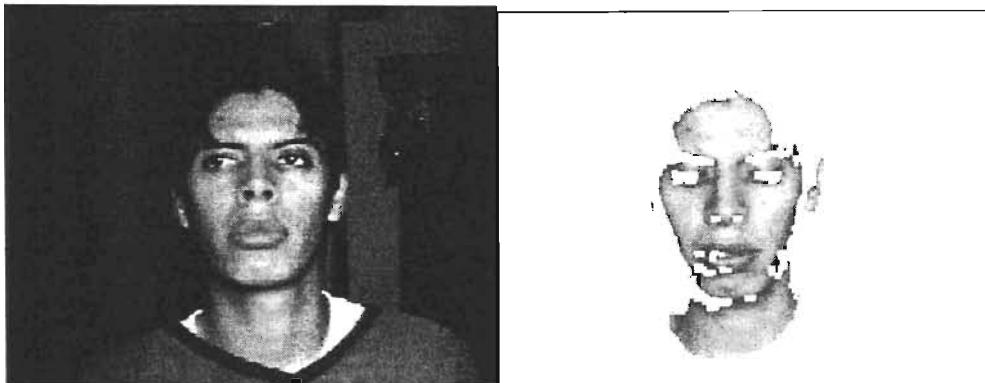


Figura 5.6 . Resultados al seleccionar el color de la piel con esta técnica

A partir de las imágenes obtenidas de este proceso (lado derecho figura 5.6), se utiliza el algoritmo de Camshift (véase sección 4.5) para obtener el marco del rostro o ventana de búsqueda.

Se va a tomar como entrada la imagen de probabilidades que pertenezca al conjunto de color de la piel y después de varias iteraciones en las que cambia el centro de masa y el tamaño de la ventana de búsqueda nos regresa la posición más probable en la que está el rostro.



## 5.4 Optimización del detector del color de la piel

Otro detector del color de piel utiliza un proceso equivalente al anteriormente descrito que utiliza el cálculo de la normal  $N(\mu, \Sigma)$ . Salvo que este utiliza la distancia de Mahalanobis (Véase sección 4.3) en lugar del cálculo de la normal. Esta distancia nos indica que tan lejos caen los valores del centro del agrupamiento de grupo color de la pixel.

Realizar el cálculo de esta distancia involucra menos operaciones que calcular la probabilidad de los pixeles como se muestra en la figura 5.7

$p(x   \omega_i) = \frac{1}{\sqrt{(2\pi)^d  \Sigma_i }} e^{-\frac{1}{2} (x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i)}$	$(x - \mu)^T \Sigma_i^{-1} (x - \mu)$
---	---------------------------------------

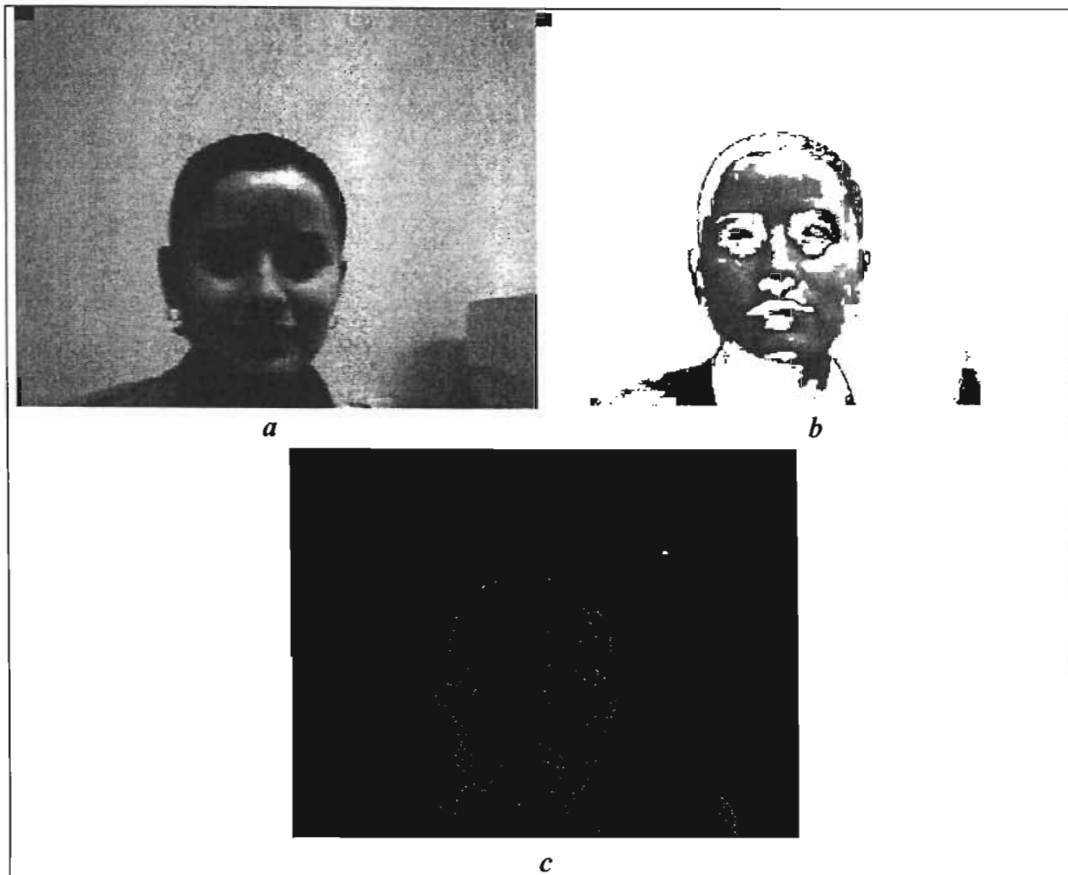
Figura 5.7 De lado izquierdo se muestra la función para calcular la probabilidad de un pixel, y de lado derecho se muestra la fórmula para calcular la distancia de Mahalanobis.

Como se puede notar el cálculo de la probabilidad contiene el cálculo de la distancia de Mahalanobis, calcular esta distancia nos implica una reducción importante de tiempo de procesamiento, ya que las operaciones extras son hechas para cada pixel en la imagen. Tomando en cuenta que las imágenes tomadas desde una cámara son en promedio de 320x200 pixeles tenemos 64 mil operaciones menos por imagen utilizando este método.

Ya que el algoritmo de Camshift busca el centro de gravedad de las imágenes y define una ventana de búsqueda que se ajusta para contener la mayoría de los pixeles con la probabilidad más alta.

Lo que llevo a tomar de nuevo la distribución de los pixeles, ésta también posee una distribución normal, ya que donde hay más pixeles en la parte central de la imagen, van disminuyendo conforme se alejan del centro. Así que, se realizaron dos proyecciones de las cantidades que ocurrían los pixeles del conjunto color de la piel con respecto al eje X y al eje Y. Tomando que posee una distribución normal, se procede a calcular la esperanza como ya se mostró anteriormente y a calcular la varianza para cada uno de los ejes. Al obtener la esperanza de X y Y obtenemos las coordenadas del equivalente al centro de masa en el algoritmo de Camshift y al obtener las varianzas es el equivalente de la ventana de búsqueda.

El resultado de este proceso se muestra en la figura 5.8.



*Figura 5.8. Resultado del proceso de detección de piel. La imagen con etiqueta a es tal como se captura de la cámara, la imagen con etiqueta b es el resultado la detección del color de la piel y al imagen con etiqueta c es el resultado final de la obtención de la ventana de búsqueda.*

# Capítulo 6

## IMPLEMENTACIÓN DE LA DETECCIÓN DE PIEL APLICADO A ROSTROS

### *6.1 Introducción*

Como se mostró anteriormente, la utilidad de los detectores de piel es muy grande ya que tienen la gran ventaja de que son muy rápidos y esto tiene la virtud de poderlos utilizar en dar solución a problemas en tiempo real, como lo son el reconocimiento facial, seguimiento de rostros, entre otros. En este capítulo se mostrarán los resultados obtenidos, también se describe la implantación de dicho detector y su aplicación en la detección de rostros en tiempo real.

### *6.2 Descripción del problema*

Se propone un conjunto de imágenes las cuales contengan una subimagen que contenga un rostro. Para ello se requiere que las imágenes que se utilicen estén a color, que el usuario este posicionado frente a la cámara y viendo a ésta.

Para lograr lo anterior se utilizó una cámara conectada a la PC la cual nos proporciona una secuencia de video en tiempo real. El software utilizado fue Visual C++ como lenguaje de programación, para manejo de cámara se utilizó Vision SDK y para el desarrollo de la interfaz se uso Qt.

El equipo en el que fue probado fue un pentium II a 400Mz, con sistema operativo Windows 2000 donde se procesan 15 cuadros por segundo con una resolución de 320x200. En este equipo se logró tener una buena fluidez en los resultados obtenidos. Aunque se trabajó también con equipos superiores obteniendo resultados ligeramente más rápidos.

Se utilizaron distintos espacios de color, para poder compararlos y poder decidir cual es el que mejor se adapta a nuestra solución. Para esto se desarrollaron dos versiones estables de la aplicación, una de ellas utiliza los modelos de color aunque más lenta que la otra que sólo utiliza el RGB normalizado por el que nos decidimos.

### 6.3 Etapas del desarrollo

Para poder solucionar el problema planteado, se implementó, el detector de piel con las características descritas en los capítulos anteriores. En el desarrollo se siguieron cinco etapas principales que se enumeran a continuación. (Véase figura 6.1)

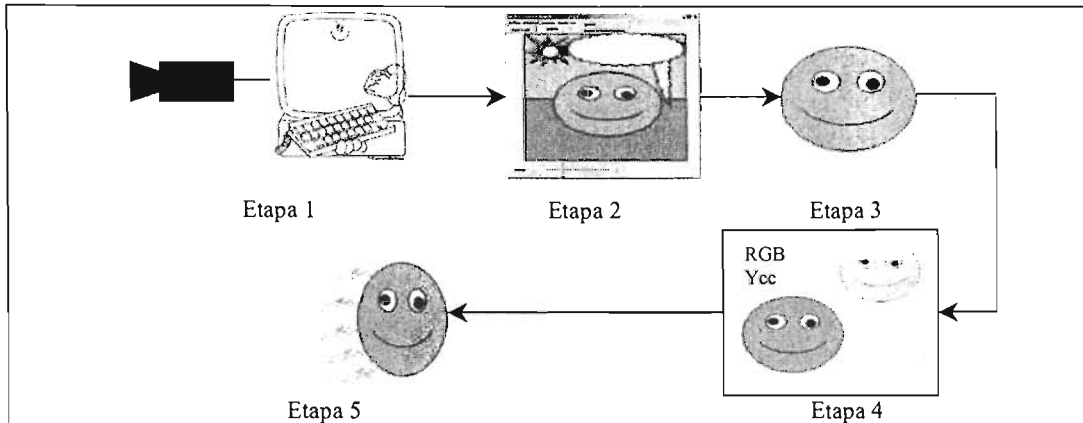


Figura 6.1 Diagrama de etapas del desarrollo.

1 captura de imágenes, 2 desarrollo de la interfaz, 3 detector de piel, 4 comparaciones, 5 optimización

#### 1.- Pruebas sobre la captura de imágenes mediante la cámara utilizando Vision SDK.

Se trabajó explorando las bibliotecas que ofrece Vision SDK, se probaron varios tipos de cámaras entre las cuales figuran webcam Alaska EZ-305, PC-camara DV2000 Musteck y DV4000 Musteck. En todas estas cámaras se manejó una resolución de 320x200.

#### 2.- Creación de la interfaz.

Se buscaron las maneras de manejar las imágenes obtenidas en la etapa anterior de una manera eficiente y se realizó la interfaz a la aplicación utilizando Qt.

#### 3.- Implementación de los detectores de piel.

Se implementó el detector de piel basado en el color mostrado con anterioridad.

#### 4.- Comparativa sobre los espacios de color utilizados.

Se realizaron comparaciones entre los distintos modelos de color (RGB normalizado,  $Y_{c}c_b$ , HSI), tomando como criterios de comparación los resultados obtenidos y el tiempo de ejecución.

#### 5.- Optimización del detector de piel.

Se eliminó de la aplicación todas las herramientas usadas en las comparativas, se optimizó el código y se implementaron los cambios ya mencionados al final del capítulo cinco.

### 6.4 Descripción de la interfaz

En la aplicación recibe de la cámara las secuencias de imágenes y las procesa en tiempo real. En la primera versión de pruebas y entrenamiento se tiene una barra de tareas que tiene las opciones que se muestran en la figura 6.2.

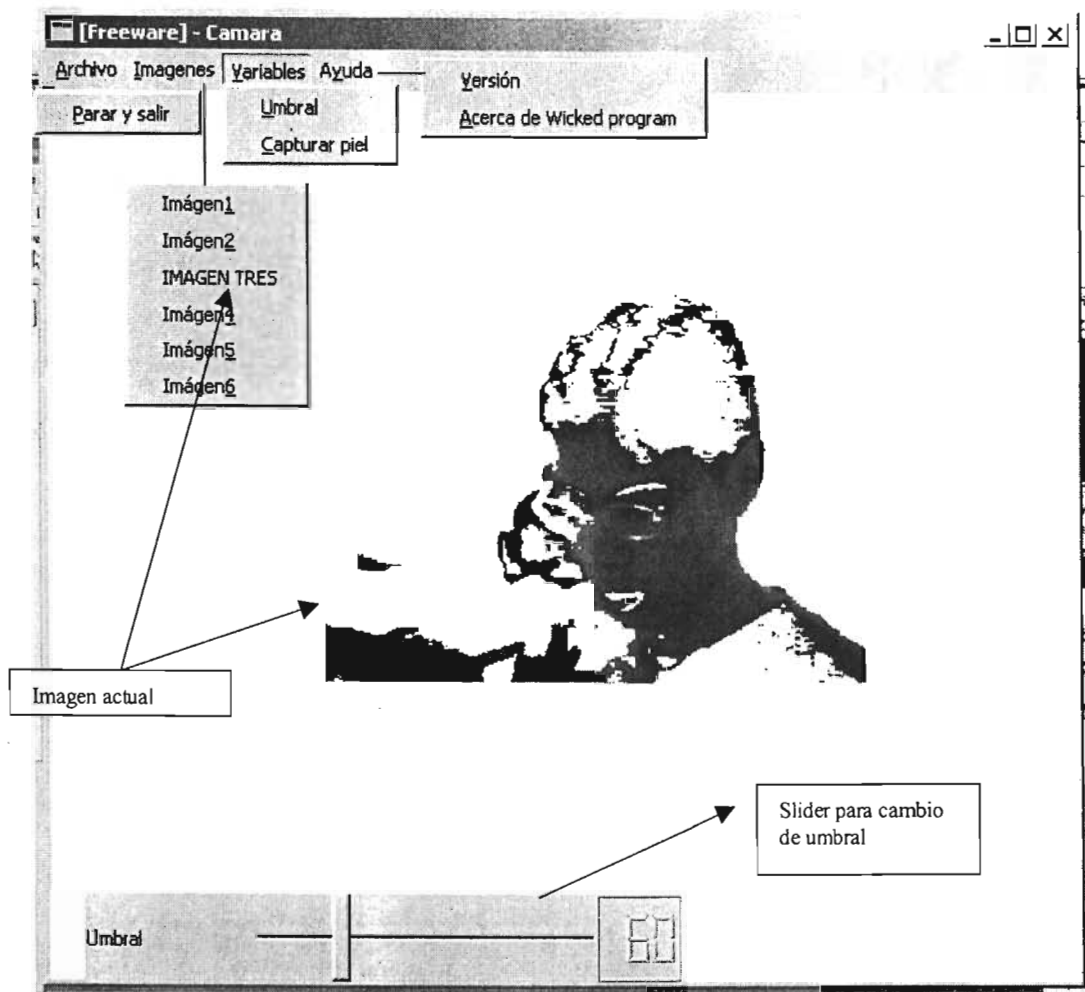


Figura 6.2 Aplicación desarrollada en esta Tesis.

*Archivo.*

*Parar y Salir.* Termina la aplicación

*Imágenes*

*Imagen 1* Imagen original.

*Imagen 2* Resultado de aplicar la compensación de luz

*Imagen 3* Resultado del algoritmo de detección de piel. RGB normalizado

*Imagen 4* Obtención de la ventana de búsqueda. RGB normalizado

*Imagen 5* Resultado del algoritmo de detección de piel. YCC y HSI

*Imagen 6* Resultado del algoritmo de detección de piel. YCC y HSI.

*Variables*

*Umbral* Se fija el umbral de tolerancia del umbral

*Captura piel* Captura despliega una muestra de color de la piel, calcula el vector esperanza, la matriz de varianzas toma estos nuevos valores y los almacena en un archivo. (Figura 6.3)

*Ayuda*

*Versión* Despliega la versión

Dependiendo de la comparación entre los espacios de color (YCC y HSI), los elementos *imagen 5* e *imagen 6* realizan la misma tarea que los elementos *imagen 3* e *imagen 4* pero con sus espacio correspondiente.



*Figura 6.3. El rectángulo negro es una muestra de piel tomada para el proceso de muestreo.*

### **6.5 Comparativa de modelos de color**

En esta etapa se realizó una comparativa entre los resultados obtenidos con el modelo RGB normalizado y los resultados obtenidos con los modelos  $Y C_{\tau} C_b$  y HSI [11, 16-17]. Los criterios tomados para calificar fueron:

1. la calidad de los resultados y
2. tolerancia a cambios de color de fuente de iluminación
3. el tiempo de respuesta.

Los detectores de grupo color de la piel que utilizan los modelos  $Y C_{\tau} C_b$  y HSI utilizan el mismo proceso anteriormente descrito en el modelo RGB normalizado, salvo que en el  $Y C_{\tau} C_b$  se toman las componentes  $C_{\tau}$ ,  $C_b$  y en el HSI tomamos las componentes H, S ya que éstas representan el color en sus respectivos modelos.

Se tomó primero el modelo  $Y C_{\tau} C_b$  el cual no generó resultados de muy buena calidad como se muestra en la figura 6.4 ya que acepta más colores que no sean del grupo color de la piel y por lo tanto se descato este modelo y no se prosiguió a comparar con los demás criterios.



Figura 6.4. De lado derecho el resultado al utilizar el espacio RGB normalizado y de lado izquierdo el modelo  $YC,C_b$ .

Con el modelo HSI los resultados son muy parecidos para clasificar las imágenes pero al calcular las distancias de Mahalanobis encontramos que los conjuntos son más compactos aunque de entrada esto puede parecer bueno dio el problema que tiene una menor tolerancia al realizar variaciones con la fuente de iluminación.

El proceso que se desarrolla en las comparativas es exactamente el mismo que con el RGB normalizado. Tomamos para empezar el HSI en donde sólo tomamos los valores  $H,S$ . Esto debido a que el término  $I$  representa la luminosidad de la imagen y los otros términos representan el color. Y sustituimos el  $H,S$  por los términos  $r,g$  del RGB normalizado.

Por estas razones se decidió quedarnos con el modelo RGB normalizado.

## 6.6 Obtención de ventana de búsqueda

Como ya se mencionó anteriormente, en el primer acercamiento para obtener el recorte de las imágenes se utilizó el algoritmo de Camshift (Sección 4.5), los resultados que da este algoritmo son buenos (Véase figura 6.5) aunque considerablemente lentos para ser tomados en tiempo real. Toma como entrada la imagen de probabilidades de que pertenezca al conjunto color de la piel y después de varias iteraciones en las que cambia el centro de masa y el tamaño de la ventana de búsqueda nos regresa la posición más probable en la que está el rostro.

En el proceso de optimización se dejó de utilizar la función normal multivariada para utilizar la distancia de Mahalanobis lo que implicó el dejar de tener la imagen de probabilidades. Entonces se consideró otra técnica para sustituir el algoritmo de Camshift y no pagar el costo de utilizar la imagen de probabilidades. Entonces se implementó la técnica descrita en la sección 5.4.

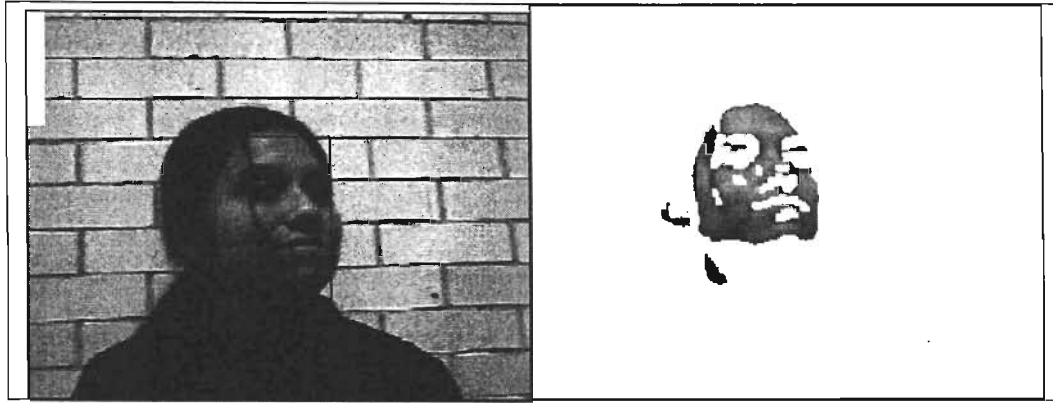


Figura 6.5. Obtención de la ventana de búsqueda utilizando el algoritmo Camshift.

## 6.7 Resultados

Se muestran en la tabla 6.1 una comparativa basada en los cuadros por segundo entre los resultados obtenidos antes y después de realizar la optimización del detector de piel.

Antes de la optimización	Después de la optimización
1.5 a 2.1 segundos por cuadros	0.9 al segundo por cuadros

Tabla 6.1 Comparativa de tiempos de respuesta en cuadros por segundo.

A continuación se muestra una serie de resultados obtenido mediante la última versión del sistema desarrollado en esta tesis (Véase figuras 6.7 a 6.15). Se encontró una mejora de 1.55965909 veces en promedio. (Véase figura 6.6)

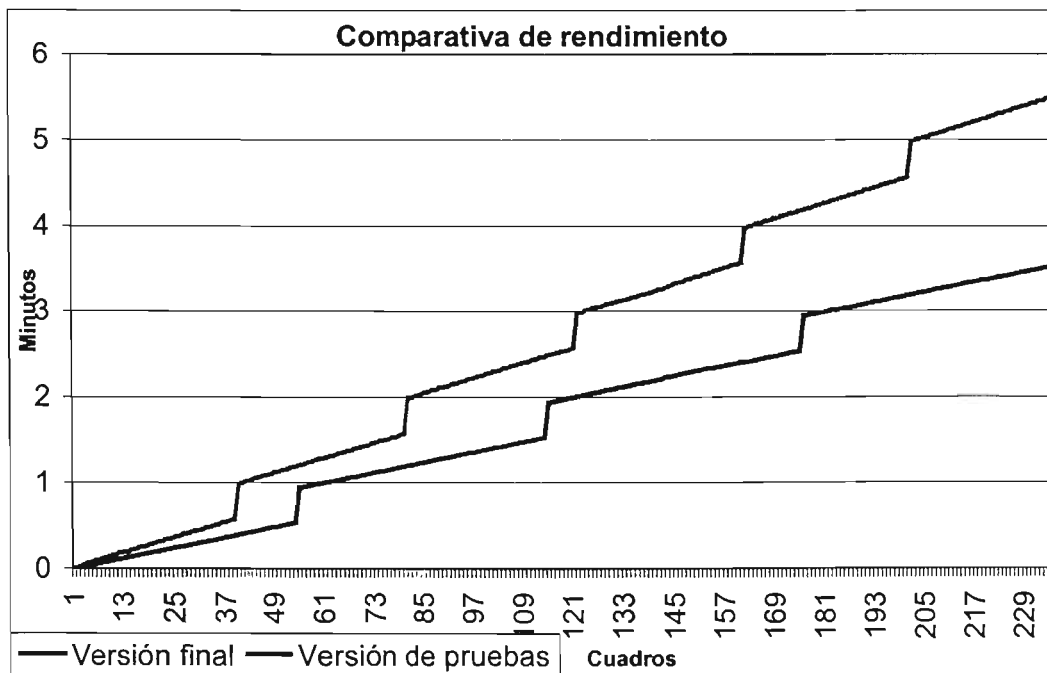


Figura 6.6. Gráfica promedio de las corridas entre la versión final y la versión antes de optimizar. Se compara el número de cuadros procesados contra el tiempo en de ejecución.





a



b



c

Figura 6.7

Resultados de la aplicación, las imágenes con etiqueta **a** son tal como se captura de la cámara, las imágenes con etiqueta **b** son los resultados la detección del color de la piel y las de etiqueta **c** son el resultado final de la obtención de la ventana de búsqueda. A continuación se muestran el las Figuras 6.8 a la 6.15 otros resultado en el mismo formato.

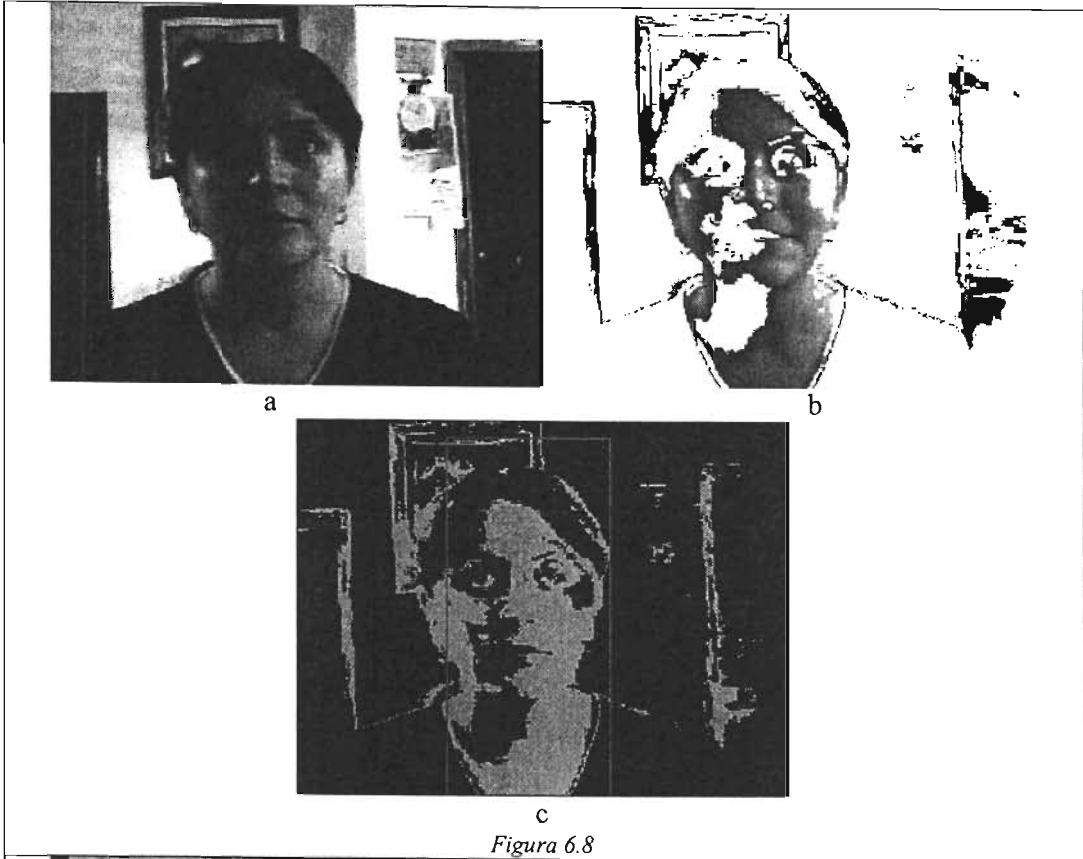


Figura 6.8

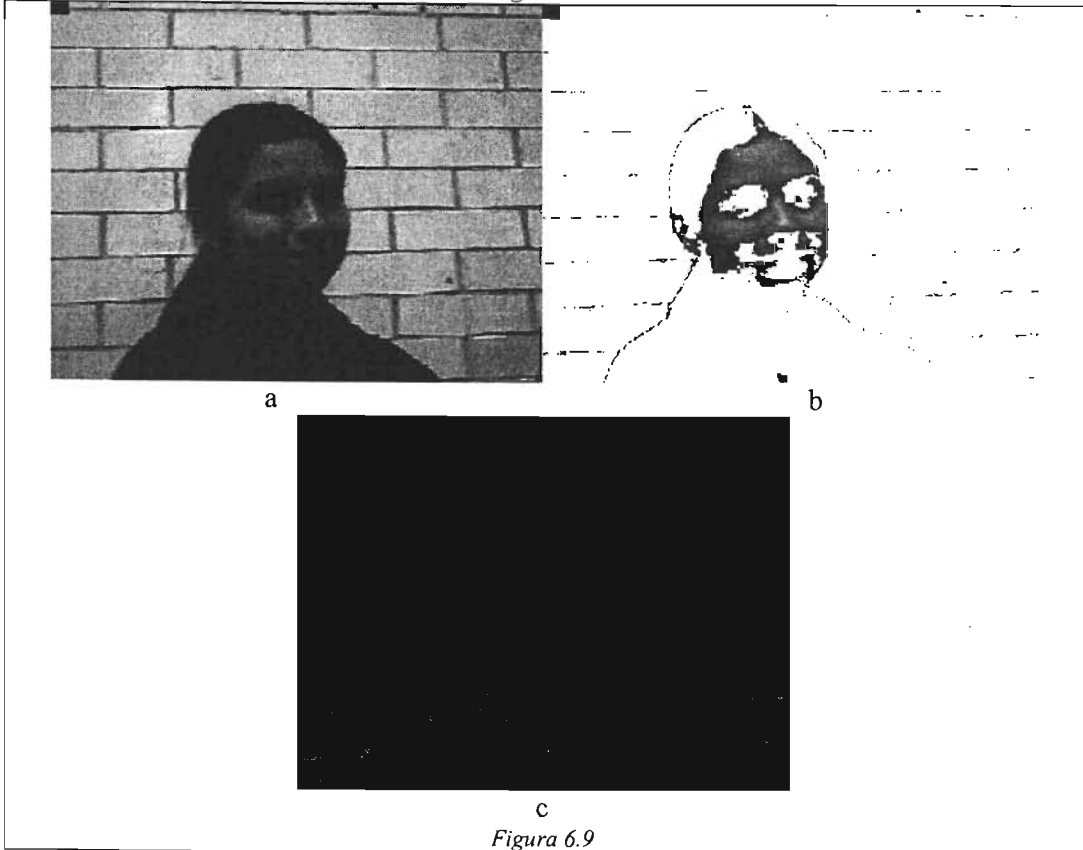


Figura 6.9



b



a



c

Figura 6.10



b



a



c

Figura 6.11

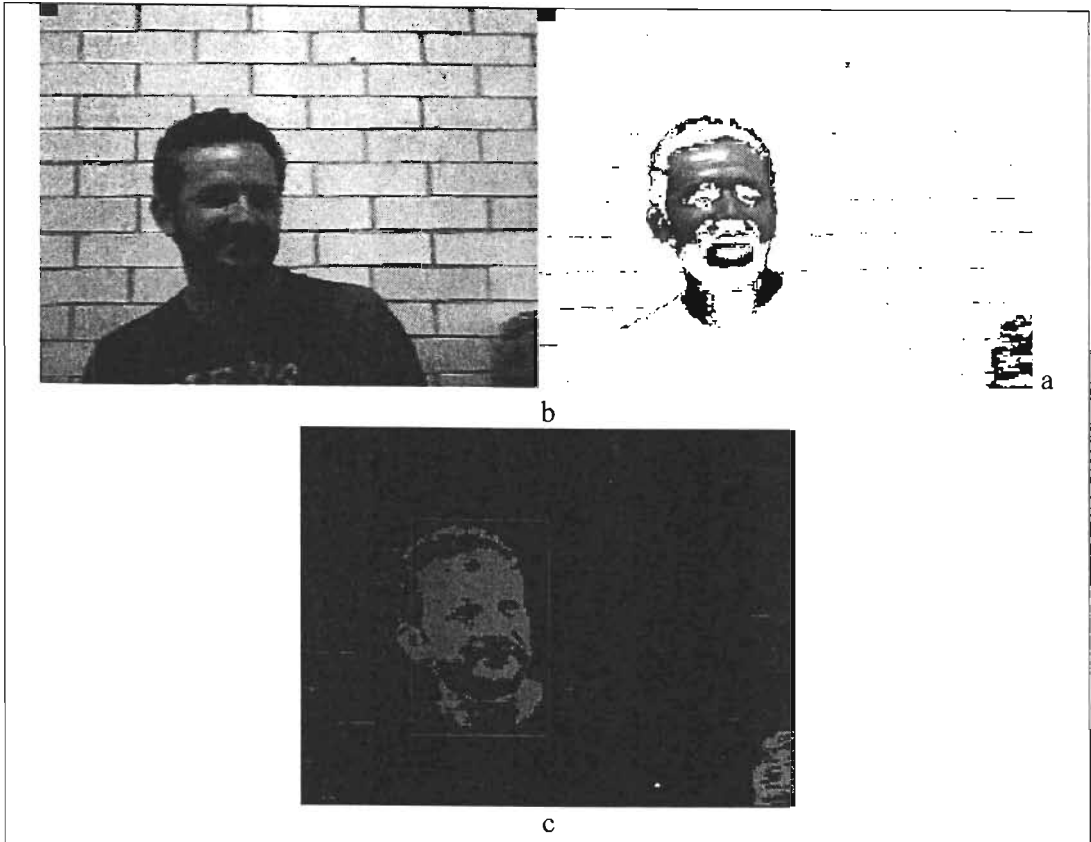


Figura 6.12

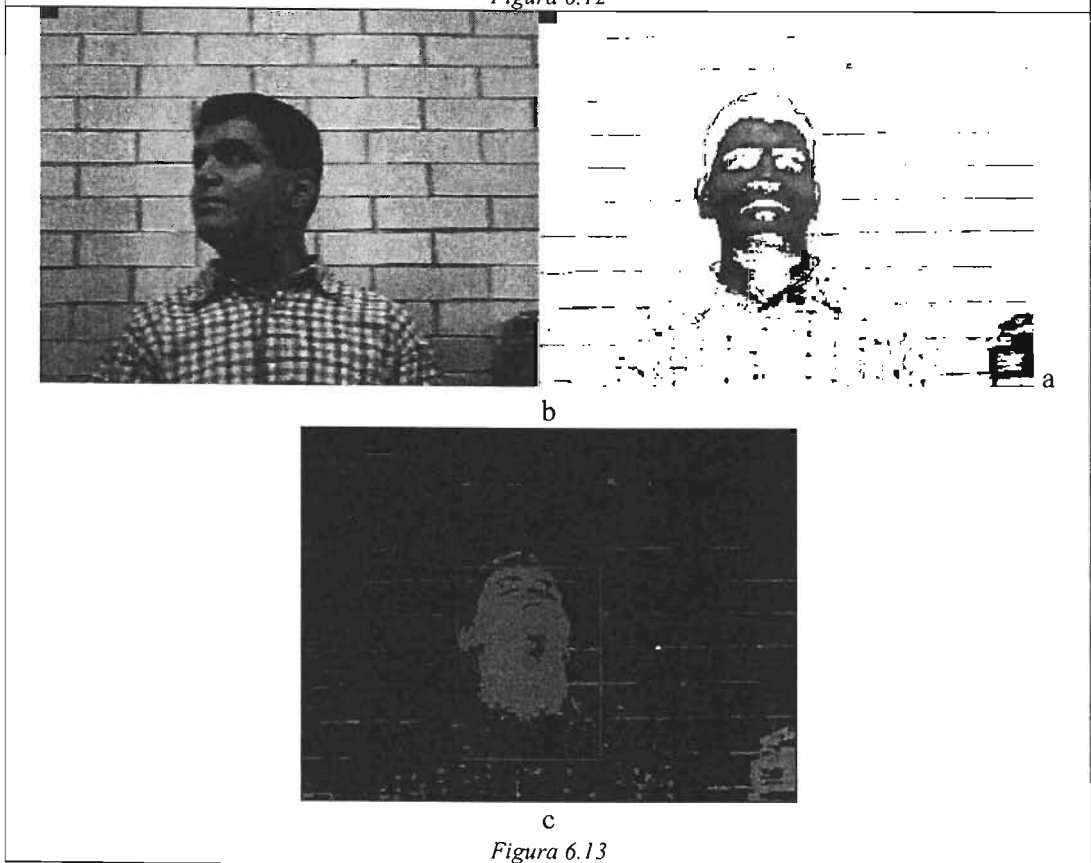


Figura 6.13



b



a

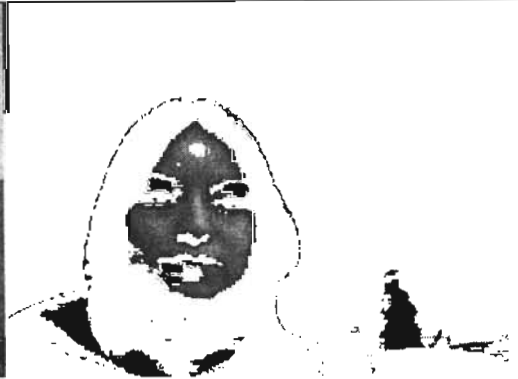


c

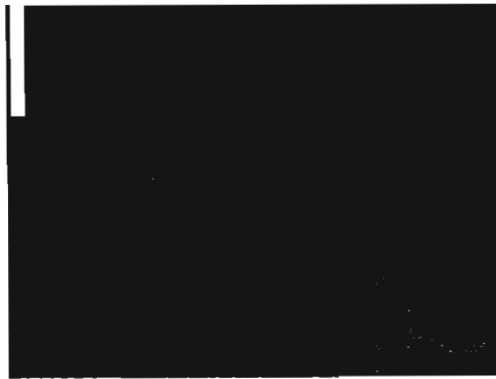
Figura 6.14



b



a



c

Figura 6.15

## CONCLUSIONES

En esta tesis se desarrolló un sistema para la detección de rostros, basado en la detección del color de la piel. Cuando se captura una imagen para detectar el color de la piel, se debe permitir cierta tolerancia a los cambios de iluminación, esto genera un cierto error en la detección.

El modelo RGB normalizado fue el modelo que más se adaptó a los cambios de iluminación entre los modelos comparados. Otra de las ventajas de este modelo, es que el cálculo de las componentes  $(r, g)$ , requiere un menor consumo de tiempo del procesador.

El uso de métodos probabilísticos y en particular el manejo de la normal dio lugar a realizar gran parte de este trabajo, sobretodo facilitó y agilizó la clasificación del color de la piel de una forma automática lo que permite hacer este proceso en tiempo real.

Aunque hay métodos que en algunos casos son mejores o sirven para mejorar la calidad de los resultados, como es el mapa de probabilidades de la imagen o el filtro gaussiano, etc., se eligió buscar la utilización de alternativas que fuesen menos pesadas, para hacer el proceso en tiempo real y poder realizar la detección lo más rápido posible.

### ***Trabajos a futuro***

Este trabajo es la parte inicial sobre un sistema de tutoriales. En este sistema se tiene que supervisar al usuario mediante una cámara conectada a la computadora. Con esta información se debe saber si el usuario está poniendo atención, conocer el estado de animo y basándose en esta información, el sistema debe incrementar o disminuir el ritmo del curso. Este trabajo es el paso inicial del sistema antes mencionado, ya que la detección de los rostros en la escena es crucial en la detección de expresiones faciales.

Para el uso de avatar<sup>2</sup>. Ya que se tiene realizar un levantamiento en 3D y obtención de la textura de la cara. Este trabajo al igual que en el caso anterior es el paso inicial ya que se debe hacer un recorte en la escena para obtener las caras de las cuales se va a realizar el levantamiento.

---

<sup>2</sup> Un avatar es un facsímil gráfico que se puede utilizar en las habitaciones de discusión virtual o chat rooms.

# ANEXO

## A.1 Filtro Gaussiano

El tema de filtros Gaussianos es muy amplio, aquí se expone sin detalles sobre este tema.

Los filtros Gaussianos son una clase de filtros lineales con pesos escogidos de acuerdo a la forma de la función Gaussiana. EL filtro Gaussiano es muy bueno para remover ruido que tiene la forma de una distribución normal. El hecho de que los pesos del filtro son escogidos de una distribución Gaussiana y que el ruido es también distribuido como una Gaussiana es sólo coincidencia.

La función Gaussiana en una dimensión es:

$$g(x) = e^{-\frac{x^2}{2\sigma^2}}$$

donde la Gaussiana cubre el parámetro  $\sigma$  que determina el ancho de la Gaussiana.

Para el procesamiento de imágenes, la función Gaussiana discreta es:

$$g[i, j] = e^{-\frac{i^2 + j^2}{2\sigma^2}}$$

donde  $i$  y  $j$  son las coordenadas de los pixeles. Esta función es usada como un filtro.

Las funciones Gaussianas tienen propiedades que hacen particularmente útil el procesamiento de visión. Estas propiedades indican que los filtros Gaussianos son efectivos para filtros de bajo flujo. Su implementación es eficiente, y pueden ser usados efectivamente en aplicaciones de visión práctica.

Las propiedades son las siguientes:

- En dos dimensiones, las funciones Gaussianas son simétricas rotacional mente. La propiedad de simetría rotacional implica que un filtro Gaussiano no se inclina en el eje de detección, en una dirección en particular.
- La función Gaussiana tiene la forma de una campana. No es conveniente que un filtrado Gaussiano reemplace cada pixel de la imagen con el valor promedio de sus vecinos, tal que el ancho dado a un vecino se decremente monótonamente con la distancia de un pixel central. Esta propiedad es importante porque una operación de filtrado que otorga más significado a pixeles más lejanos, distorsionará las características de la imagen.

- El ancho, y por lo tanto el grado de filtrado, de un filtro Gaussiano es parametrizado por  $\sigma$ , y la relación entre  $\sigma$  y el grado de filtrado es muy simple. Una  $\sigma$  grande implica un filtro Gaussiano grande y un filtrado extenso. El grado de filtrado se puede ajustar.
- Filtros Gaussianos grandes pueden ser implementados muy eficientemente porque las funciones Gaussianas son separables, es decir, filtros Gaussianos grandes pueden ser implementados por repeticiones aplicadas a filtros Gaussianos pequeños. Una función Gaussiana en dos dimensiones puede ser desarrollada por una Gaussiana en una dimensión, entonces se obtiene el resultado con el mismo filtro con orientación en dirección ortogonal a la Gaussiana, usada en la primera etapa. El orden de computabilidad de la implementación de filtros Gaussianos en dos dimensiones crece linealmente en el ancho del filtro en lugar de crecer cuadráticamente.

Una aproximación en el diseño de filtros Gaussianos es calcular la plantilla ancha directamente de la distribución Gaussiana discreta:

$$g[i, j] = ce^{-\frac{i^2+j^2}{2\sigma^2}}$$

donde  $c$  es una constante normalizada. Al reescribir ésta se tiene:

$$\frac{g[i, j]}{c} = e^{-\frac{i^2+j^2}{2\sigma^2}}$$

y cambiando un valor para  $\sigma^2$ , se puede evaluar sobre una plantilla de  $n \times n$ , para la cual el valor en  $[0,0]$  es igual a 1. Sin embargo, se diseña el filtro ancho para valores enteros, para facilitar el cálculo. Por lo tanto, se toma el valor en una de las esquinas del arreglo, y se escoge  $k$  tal que este valor se convierte en 1.



## BIBLIOGRAFÍA

- [1] Matthew Turk and Alex Pentland, *Eigenfaces for Recognition*, Journal of Cognitive Neuroscience, Vol 3 No. 1 Marzo 1991.
- [2] Quek, F. K. H., Mysliwicz, T., and Zhao, M.: *Fongermouse: A freehand pointing interface*. Proceedings of the International Workshop on Automatic Face and Gesture Recognition, Zurich, Switzerland, pages 372-377, June 1995
- [3] Mysliwicz, T.A.: *Fingermouse: A freehand computer point interface*. Technical report, University of Illinois at Chicago, 1994.
- [4] G. Wyszecki and W.S. Styles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*. Second Edition, John Wiley & Sons, 1982.
- [5] Jie Yang, Weier Lu, Alex Waibel. *Skin-Color Modeling and Adaptation*. In Proceedings of the 3rd Asian Conference on Computer Vision, volume 2, pages 687-694, 1998.
- [6] M.A. Turk and A. Pentland, *Face recognition using eigenfaces*, Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pag. 586-591, Maui HI.UA, 1991.
- [7] Cai, J., Goshtasby, A., and Yu, C.: *Detecting human faces in color images*. International Conference on Image Processing (ICIP '98) 3-Volume Set-Volume 1, 1998.
- [8] Chen, Q., Wu H., and Yachida, M.: *Face detection by fuzzy pattern matching*. IEEE 5<sup>th</sup> International Conference on Computer Vision, pages 591-596, 1995.
- [9] Collobert, M., Feraud, R., Tourneur, G. L., Bernier, D., Viallet, J. E., Mahieux, Y., and Collobert, D.: *Listen: A system for locating and tracking individual speakers*. International Conference on Automatic Face and Gesture Recognition, pages 283-288, October 1996.
- [10] Hunke, M. and Waibel, A.: *Face locating and tracking for human-computer interaction*. IEEE Computer, pages 1277-1281, November 1994.
- [11] Sobottka, K. and Pitas, I.: *Segmentation and tracking of face in color images*. Second International Conference on Automatic Face and Gesture Recognition, pages 236-241, October 1996.
- [12] Wu, H., Yokoyama, T., Pramadihanto, D., and Yacida, M.: *Face and facial feature extraction from color image*. International Conference in Automatic Face and Gesture Recognition, pages 345-350, October 1996.
- [13] Yang, J., Lu, W., and Waibel, A.: *Skin-color modeling and adaptation*. Technical report, Carnegie Mellon University, 1997. CMU-CS-97-146.

- [14] Yang, J., Lu, W., and Waibel, A.: *Skin-color modelling and adaptation*. Proceedings of ACCU'98, 2:687-694, 1998.
- [15] Yang, J., Lu, W., and Waibel, A.: *Tracking human faces in real-time*. **Technical report**, Carnegie Mellon University, 1995. CMU-CS-95-210.
- [16] Kjeldsen, R. and Kender, J.: *Finding skin color images*. International Conference on Automatic Face and Gesture Recognition, pages 312-317, October 1996.
- [17] Saxe, D. and Foulds, R.: *Toward robust skin identification in video images*. 2<sup>nd</sup> International Face and Gesture Recognition Conference, September 1996.
- [18] G.J. Klinker, S.A. Shafer, and T. Kanade\_ *Using a color reflection model to separate highlights from object color*, Proc. ICCV, Pp. 145-150, 1987.
- [19] R. Gershon. *The Use of Color in Computational Vision*. Ph. D thesis. University of Toronto.1987.
- [20] Solis, AL, Rios, H: *Interaction for Virtual Enviroments based on Facial Expresión Recognition*, Proceedings of workshop at IEEE Virtual Reality 2001, Yokohama, Japan.
- [21] GMD Report 138, GMD Forschungszentrum Informationstechnik GmbH, *The Future of VR and AR Interfaces*, Proceedings of workshop at IEEE Virtual Reality 2001, Yokohama, Japan.
- [22] L. M. Bergasa Pascual *Seguimiento facial, mediante visión Artificial, orientado a la ayuda a la Movilidad*. Tesis. Universidad de Alcalá Escuela politécnica. 1999
- [23] Solis, AL, Rios, H: *Facial Expression Recognition and Modelling for virtual Intelligent Tutoring System*. Lectures Notes in Artificial Intelligence 1793, MICAI 2000: Advances in Artificial Intelligence.
- [24] Rein-Lien Hsu, *Face Detection in Color Images*, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 696-706, May 2002
- [25] Gary R. Bradski, Microcomputer Research Lab, *Computer Vision Face Tracking For Use in a Perceptual User*
- [26] Eric Guerrino, Mike Kahn, and Ellen Kapito. *User authentication and encryption overview*, 1997.