



CENTRO DE INVESTIGACIÓN EN MATEMÁTICAS, A.C.

**Detección de Rostros en Imágenes Basado
en Multclasificación
con aplicación a Video Conferencia.**

TESIS

que para obtener el grado de

**Maestría en Ciencias con Especialidad en Computación y
Matemáticas Industriales**

presenta

Jorge Cruz Pérez

Director de Tesis

Dr. Rogelio Hasimoto Beltrán

Noviembre de 2004

Guanajuato, Gto, México



CIMAT
BIBLIOTECA

Detección de Rostros en Imágenes Basado en Multclasificación
con aplicación a Video Conferencia.

por

Jorge Cruz Pérez

I.C., Instituto Tecnológico de Oaxaca (2000)

Sometido al Área de Ciencias de la Computación
en el cumplimiento parcial de los requisitos para el grado de

Maestro en Ciencias de la Computación y Matemáticas Industriales

en el

CENTRO DE INVESTIGACIÓN EN MATEMÁTICAS A.C.

Noviembre 2004

© Centro de Investigación en Matemáticas A.C., 2004

El autor por este medio concede al Centro de Investigación en Matemáticas A.C.
permiso de reproducir y
distribuir copias de este documento de tesis en entero o en parte.

Firma del autor
Área de Ciencias de la Computación
Noviembre, 2004

Certificado por
Dr. Rogelio Hasimoto Beltrán
Dr. en Ingeniería Eléctrica y Computación, Grupo de Cómputo Matemático
Director de Tesis

Aceptado por
Dr. Salvador Botello Rionda
Coordinador de la Maestría en Ciencias de la Computación

C I M A T
B I B L I O T E C A

019182

Detección de Rostros en Imágenes Basado en Multclasificación con
aplicación a Video Conferencia.

por

Jorge Cruz Pérez

Sometido al Área de Ciencias de la Computación
en Noviembre, 2004, en el cumplimiento parcial de los
requisitos para el grado de
Maestro en Ciencias de la Computación y Matemáticas Industriales

**Detección de Rostros en Imágenes Basado en Multiclasificación con
aplicación a Video Conferencia.**

por
Jorge Cruz Pérez

Sometido al Área de Ciencias de la Computación
en Noviembre, 2004, en el cumplimiento parcial de los
requisitos para el grado de
Maestro en Ciencias de la Computación y Matemáticas Industriales

Resumen

En este trabajo de tesis se aborda el problema de la detección de rostros en imágenes o secuencias de video, formulando un conjunto de reglas que aseguran la decisión adecuada, en cada etapa del desarrollo de un sistema diseñado con estos fines. Para esto se utiliza la información de las componentes de cromacidad en los espacios de color RGB (Red, Green, Blue), HSV (Hue, Saturation, Value) y HMMD (Hue, Max, Min, Diff). Esta estrategia de segmentación de color tiene el propósito de reducir el espacio de búsqueda en la imagen a fin de emplear técnicas de análisis basadas en la imagen, y además es robusta a cambios de iluminación intensos. Se proponen y adaptan técnicas y algoritmos rápidos, sencillos y eficientes para el conteo, separación y localización de regiones conectadas, junto con técnicas de procesamiento basadas en la semántica de la imagen, en particular el Análisis de Componentes Principales (PCA), para la clasificación de la pertenencia de cada una de las regiones dentro de la categoría de rostro. Uno de los objetivos que se persigue es el de no imponer restricciones a requerimientos tales como: operación en tiempo real, independencia de la persona, cierta independencia en la posición, naturalidad de los gestos, fondos (backgrounds) complejos y dinámicos, iluminación variable, o el número de rostros contenidos en la imagen. Todo lo anterior representa un primer paso en el objetivo principal, que es la implementación de un sistema automático de codificación priorizada de regiones de interés (RoI), u otros sistemas o aplicaciones en que puede ser de gran utilidad.

Director de Tesis: Dr. Rogelio Hasimoto Beltrán
Título: Dr. en Ingeniería Eléctrica y Computación, Grupo de Cómputo Matemático

Agradecimientos.

Estoy muy agradecido con todas las personas e instituciones para lograr realizar este trabajo, por su conocimiento, aportación y motivación como son: Primeramente mi asesor de tesis, el Dr. Rogelio Hasimoto Beltrán, por su ayuda y orientación, así como el tiempo dedicado para la realización de este trabajo. A mis revisores de tesis, el Dr. Mariano Rivera Meraz y el Dr. Arturo Hernández Aguirre, por el tiempo dedicado a la revisión así como las observaciones y sugerencias hechas para la presentación y contenido de este trabajo.

Quiero agradecer también a los profesores del Área de Ciencias de la Computación, por los conocimientos y experiencias transmitidas en clase y fuera de clase, al CIMAT por el apoyo como institución y apoyo económico aportado así como el equipo proporcionado sin las cuales no se tendrían estos resultados.

Agradecer también al CONACYT y su programa de becas, por el apoyo proporcionado durante los dos años de estudio de maestría, ya que fue el soporte económico fundamental para mis estudios.

Agradecer especialmente a mis padres, Jorge Cruz y Gregoria Pérez, por la motivación y soporte emocional, así como a mis hermanas Magy y Paty, que me ayudaron a estar tranquilo, contento y motivado para seguir estudiando cuando estaba lejos de mi hogar. Y una mención especial a mi mascota, darkito, por la algarabía y lealtad con la que me recibía al visitar mi casa y que murió una semana antes de mi regreso final, faltándome esa algarabía a mi regreso.

Finalmente a mis compañeros de Maestría, a Ángel, Víctor y Omar, la comunidad de la T, a Moisés y Rosa, Rocky y Yolanda, a cada uno en un momento dado, por su tiempo, apoyo y disposición para resolver dudas y motivación para seguir. También gracias a mis compañeros de generaciones anteriores y posteriores por su ayuda.

Índice general

1. Introducción.	1
1.1. Motivación.	1
1.2. Descripción del problema y Trabajo Actual.	2
1.2.1. Trabajo actual (Estado del Arte).	3
1.2.2. Problemas relacionados con la detección de rostros.	8
1.3. Alcance y contribución de la tesis.	9
1.4. Organización de la tesis.	10
2. Estrategia de Multiclasificación para la Detección de Rostros.	11
2.1. Introducción al problema en el ámbito de Reconocimiento de Patrones.	11
2.1.1. Adquisición de datos.	11
2.1.2. Selección y extracción de características.	14
2.1.3. Módulo de clasificación.	16
2.2. Clasificación de Imágenes para la Detección.	20
2.3. Descripción general del sistema de multiclasificación.	21
3. Segmentación de color.	23
3.1. Conceptos básicos.	23
3.1.1. Reflexión de la piel.	24
3.2. Modelado del color de la piel.	24
3.2.1. Espacios de color.	25

3.2.2. Espacios de color empleados para el modelado y detección del color de la piel.	26
3.2.3. Modelado de la distribución del color de la piel	32
3.3. Esquema de segmentación basado en RGB, HSV y HMMD.	36
3.3.1. Clasificador.	36
3.3.2. Diseño del clasificador.	37
3.3.3. Esquema de segmentación de color.	51
3.3.4. Algoritmo de segmentación de color	51
3.4. Resultados preliminares de la segmentación de color.	53
4. Estrategias para la localización y separación de regiones.	55
4.1. Estrategias de diferenciación de regiones de piel.	57
4.1.1. El espacio de color HMMD como indicador de regiones de piel.	57
4.1.2. Morfología matemática para la mejora de subimágenes.	59
4.2. Esquema de localización y conteo de regiones.	63
4.2.1. Algunos conceptos básicos sobre regiones conectadas.	63
4.2.2. Etiquetado de regiones conectadas.	64
4.2.3. Estrategia de etiquetado de regiones conectadas.	64
4.3. Separación de regiones no convexas por proyección de vectores.	65
4.3.1. Algunos conceptos básicos sobre vectores en 2D.	66
4.3.2. Estrategia de división de regiones no convexas.	67
4.4. Algoritmo general de localización y separación de regiones.	71
5. Técnicas de Verificación de Rostros.	73
5.1. Técnicas rápidas.	73
5.1.1. Análisis de la forma de la región.	74
5.1.2. Aproximación de una elipse al contorno.	77
5.2. Detección de rostros usando PCA.	80
5.2.1. Introducción a los eigenrostros.	80

5.2.2. Cálculo de Eigenrostros.	82
5.2.3. Detección usando eigenrostros	84
5.3. Algoritmo de Verificación de Regiones para su clasificación en Rostro o no.	85
6. Implementación y Resultados Experimentales.	87
6.1. Implementación.	87
6.2. Resultados experimentales	87
6.2.1. Discusión.	91
7. Conclusiones y Trabajo Futuro.	93
7.1. Contribución de la tesis.	94
7.2. Conclusiones.	94
7.3. Trabajo futuro.	95
A. Herramienta de Software DERO	97
A.1. Detalles de implementación.	97
A.1.1. Introducción	97
A.1.2. Lenguaje de Programación	98
A.1.3. Descripción esquemática del sistema	98
A.1.4. Ventanas y botones	99

Índice de figuras

1.1. La ubicación y atributos de un rostro	4
1.2. Técnicas empleadas en la detección de rostros.	5
2.1. Etapas de un sistema de Reconocimiento de Patrones.	12
2.2. a) Selección de las 3 variables más significativas. b) Transformación de las variables originales.	15
2.3. Subconjunto de entrenamiento (a) Pixeles de piel, (b) Rostros	18
2.4. Diagrama de bloques del clasificador planteado.	19
2.5. Diferencia en las imágenes de prueba, (a) img. izq. Secuencia de Video, (b) img. der. Fotografía normal.	21
2.6. Esquema general de clasificación en la Detección de Rostros.	21
3.1. Espacio de color RGB	27
3.2. Espacio de color HSV	28
3.3. Diferentes representaciones del espacio HSV (a) Cono, (b) Cilindro	29
3.4. Espacio de color HMMD	30
3.5. Cuantización uniforme de color en el plano MMD (a), y el plano H (b).	31
3.6. (a) Cuantización escalar uniforme en el plano MMD, (b) Las líneas rectas en (a) corresponden a curvas en el plano SV.	32
3.7. (a) La componente R es mayor que las componentes G y B. (b) Límites superior e inferior en el plano GB	39

3.8. Modelo HSV (a) Desde el punto de vista adecuado se puede observar una región compacta en en plano HS.(b) Comportamiento semicircular de la región de piel en el plano SV.	40
3.9. Comportamiento de las componentes de piel en el plano SV, (a) aproximación por una línea semicircular, (b) aproximación por una línea recta.	41
3.10. Comportamiento de las componentes <i>hue</i> y <i>diff</i> del modelo <i>HMMD</i> , para la región de piel	42
3.11. Umbral de densidad de probabilidad para el rechazo.	44
3.12. (a) Función que puede considerarse un kernel al responder a las cuatro propiedades enumeradas, (b) Kernel Gaussiano en 2D.	47
3.13. Esquema general de Segmentación de Color	50
3.14. Resultados de la fase de segmentación de color (Ver Sección 3.4 para descripción). 53	
3.15. Resultados de la fase de segmentación de color (Ver Sección 3.4 para descripción). 54	
4.1. Esquema de procesamiento para la detección de regiones más probables	56
4.2. Imágenes de prueba	57
4.3. Resultados de la diferenciación de distintas regiones de piel utilizando descriptor de color con el modelo de color <i>HMMD</i>	58
4.4. Dilatación, (a) Imagen A, (b) resultado de la dilatación con un elemento estructural de 3×3	60
4.5. Erosión, (a) Imagen A, (b) resultado de la erosión con un elemento estructural de 3×3	60
4.6. Apertura (a) Imagen A, (b) resultado de la apertura con un elemento estructural de 3×3	61
4.7. Cerradura, (a) Imagen A, (b) resultado de la cerradura con un elemento estructural de 3×3	62
4.8. Operaciones morfológicas en niveles de gris (a)Imagen original, (b) erosión, (c) dilatación, (d) apertura, (e) cerradura, se utilizo un elemento estructural plano de 3×3	62
4.9. Resultados de la eliminación de detalles menores de la imagen usando morfología en niveles de gris en la diferenciación de regiones	63

4.10. Resultados del algoritmo de conteo y localización de regiones. (Los resultados se muestran en color o niveles de gris)	65
4.11. (a) un vector y sus componentes (A_x, A_y) , (b) suma de $A+B$, (c) resta de $A-B$, (d) el vector <i>comp_{BA}</i>	67
4.12. (a) Puntos del contorno de una región convexa y sus vectores de posición, (b) segmentos de línea dirigidas secantes al perímetro y las componentes del vector v_i a lo largo de v_{i-1}	68
4.13. (a) Comportamiento de los vectores secantes al perímetro para regiones no convexas, (b) el vector v_j que mejor se aproxima a la región convexa	69
4.14. Contornos de las imágenes de prueba utilizadas en esta sección y puntos de aproximación	69
4.15. Algunos resultados para las imagens mostradas, separación de regiones y los vectores perímetro o líneas secantes. En la primer columna, la cara que esta en primer plano junto con su sombra forman una sola región, las líneas muestran el resultado de la separación. En la segunda columna la cara del Sr. Presidente esta unida con un región sobre esta, el resultado de la separación se muestra con líneas continuas sobre la imagen. En la tercer columna el rostro de enmedio se ve afectado por una región detrás, las líneas continuas muestran la separación correspondiente.	70
5.1. Esquema de procesamiento en la etapa de verificación	74
5.2. Ajuste de una elipse, (a) imagen de un rostro, (b) la región de piel y el conjunto de puntos de ajuste, (c) elipse aproximada al conjunto de puntos.	80
5.3. (a),(b),(c),(d),(e),(f),(g) los siete primeros Eigenrostros del conjunto de análisis, (h) el rostro promedio de este conjunto.	81
6.1. Resultados del sistema de detección para imágenes de Video conferencia	89
6.2. Resultados del sistema de detección para imágenes de fotográficas y de tv. En esta serie de imágenes no se utilizó PCA.	90
A.1. DERO	98
A.2. Ventana principal de DERO	99

A.3. Ventana de diálogo, Abrir Imagen. 99
A.4. Ventana de configuración de parámetros 101

Índice de Tablas

6.1. Conjunto de prueba. 88
6.2. Tiempos registrados en las etapas del sistema 88
6.3. Porcentajes de detección. 90

Capítulo 1

Introducción.

1.1. Motivación.

El constante desarrollo tecnológico y el crecimiento considerable del poder computacional ha permitido el desarrollo de algoritmos complejos que tratan de imitar el comportamiento humano. Como por ejemplo los desarrollados en el área de Visión por Computadora. Tradicionalmente, los sistemas de visión por computadora han sido empleados en tareas tediosas y repetitivas, por mencionar un ejemplo en la inspección de líneas de ensamblaje, o tareas más específicas, como en la investigación médica, etc. Una línea de investigación que se ha desarrollado en los últimos años involucra tareas de visión más generales como reconocimiento de rostros o su aplicación en codificación de video.

De esta última, las técnicas de codificación de video, se desprende la motivación principal para el desarrollo del trabajo de esta tesis. El video o imágenes sin comprimir requieren una considerable capacidad de almacenamiento y un alto ancho de banda para su transferencia. Para lograr manejar grandes volúmenes de información eficientemente, se necesitan técnicas de compresión para reducir el tamaño del archivo. Últimamente el enfoque en el procesamiento de señales se está desviando hacia el procesamiento de información *basada en contenido*. Este *contenido* es subjetivo y el procesamiento de información basada en contenido desafía a una formalización completa aunado a técnicas para su representación y manipulación.

La mayor parte de las soluciones en codificación de video presentadas en la literatura se enfocan en incrementar el ancho de banda ya sea mejorando la red u otras técnicas para lograr mejor desempeño en la transmisión, tales como FEC (Forward Error Correction). Sin embargo, técnicas de compresión específica de imágenes han sido poco exploradas. Video Conferencia

es una de las aplicaciones más importantes en multimedia que presenta serios problemas para su transmisión; esto es, requiere un ancho de banda relativamente alto sobre un núcleo de transporte con recursos limitados. Este problema puede ser aminorado separando el fondo (background) de la información importante (foreground). El fondo en muchos de los casos no es importante para la mayoría de los participantes en la sesión, sin embargo, el primer plano (foreground) y más específicamente la parte de la cara, debe ser tan clara como sea posible. Así las técnicas de compresión deben tomar ventaja de los requerimientos establecidos para estas dos partes. Esto es background y foreground se pueden codificar con diferentes calidades de compresión (con mejor calidad a la región de interés), dependiendo del ancho de banda requerido para su transmisión.

La separación de la imagen en dos partes es el objetivo de este trabajo y deberá ser completamente automática; cumpliendo con requerimientos de color establecidos en estándares de codificación para imagen y video (JPEG, MPEGx, H.26x). Aunado a esto, los estándares de codificación como MPEG4 y MPEG7 permiten la composición de una escena con imágenes heterogéneas múltiples, permitiendo así realizar la labor de compresión de Regiones de Interés en particular.

Adicionalmente el problema de detección de rostros es sumamente importante en un contexto general, ya que es el primer paso en sistemas automáticos de reconocimiento de rostros, de acceso automático, de vigilancia o identificación criminal, sistemas de seguimiento o rastreo, interacción humano computadora, bases de datos de multimedia y como se apuntó anteriormente en video conferencia y multimedia. En general el problema de detección de rostros se considera como un problema general de reconocimiento de patrones.

1.2. Descripción del problema y Trabajo Actual.

La detección de rostros es una de las tareas visuales que nosotros como humanos podemos hacer sin mayor esfuerzo. Sin embargo, en términos de visión computacional, esta tarea no es fácil. Podemos definir el problema de la siguiente manera:

Dada una imagen digital fija o una secuencia de video, detectar y localizar un número desconocido (si existen) de rostros.

En la solución de esta tarea se emplean técnicas de segmentación, extracción de características y verificación de rostros que deben ser adecuadas al ámbito del problema, proponiéndose nuevas ideas, técnicas, estrategias o algoritmos en los casos en los que las herramientas existentes

no ofrezcan los mejores resultados.

En sistemas de video conferencia existe una restricción adicional relacionada con el tiempo de procesamiento, es decir, se desea obtener resultados en el menor tiempo posible para ser usados en tiempo real. Esta restricción es sumamente importante, que además nos limita en diversos aspectos, debido a que las técnicas a emplear además de la implementación deben ser muy bien analizadas y justificadas con el fin de cumplir este requerimiento. En la Figura 1.1 presentamos los atributos mínimos necesarios para ubicar un rostro en una imagen, estas son las coordenadas al origen de la imagen de una región que lo encierra, en este caso un rectángulo circunscrito.

Desde el ámbito de la psicología, el problema central en el reconocimiento de patrones es el estudio de los mecanismos por los que las señales externas estimulan los órganos sensoriales y se convierten en experiencias perceptuales significativas, o dicho de otra forma, como realizamos el etiquetado de estos estímulos tan complejos asignándoles un nombre. Estos procesos continúan siendo desconocidos en su mayor parte y no se ha encontrado un modelo concluyente sobre cómo nuestro sistema nervioso realiza este reconocimiento. No obstante, se admite que esta tarea debe realizarse siguiendo un esquema general como el que se detalla a continuación. Antes del reconocimiento, un patrón debe ser percibido por los órganos sensoriales. Además, el mismo patrón o alguno similar (de la misma clase) debe haberse percibido y recordado previamente. Finalmente, debe establecerse alguna correspondencia entre la percepción actual y lo recordado. En resumen, esta aproximación al reconocimiento de patrones se centra en el estudio del mecanismo de reconocimiento presente en los seres vivos.

El uso intensivo de ordenadores u otros dispositivos electrónicos en los últimos años ha impulsado el estudio y aplicación de técnicas de reconocimiento de patrones. La aproximación al reconocimiento de patrones que adoptaremos esta basada en las teorías y técnicas de reconocimiento implementables en un sistema informático.

1.2.1. Trabajo actual (Estado del Arte).

El problema de detección de rostros ha recibido bastante atención en la literatura, en esta se encuentran una gran variedad de métodos y algoritmos propuestos que resuelven el problema parcialmente o bajo ciertas restricciones, existiendo pocos trabajos en los que realmente no se este sujeto a restricciones en las condiciones de la imagen.

De los primeros esfuerzos en la detección de rostros, se usaron simples heurísticas y técnicas



Figura 1.1: La ubicación y atributos de un rostro

simples, bastante rígidas debido a las restricciones impuestas, además de que asumen la disponibilidad de un solo rostro en la imagen, Schmidt [27]. En la década pasada el interés se centró en proponer esquemas de segmentación más robustos, particularmente aquellos que usan análisis de movimiento, color o información generalizada, Kapur [18]. Numerosos sistemas diseñados con el propósito de encontrar personas o rostros en imágenes se han propuesto por variados grupos de investigación. Proyectos que se enfocan en imágenes a color en donde su tarea básica es la detección de una región facial, Kapur[18], otras propuestas incluyen la detección de características faciales adicionales como son los ojos y la boca, Hsu [14], etc. Para mayor referencia consultar el estudio sobre trabajos presentados en la literatura de Hjelmás [13] y Yang [41].

Muchos de los algoritmos existentes confían extensamente en información heurística tomada de varias imágenes modeladas bajo condiciones fijas. En tareas más generales de localización de rostros en varias posiciones con fondos complejos, muchos de estos algoritmos fallarán debido a su naturaleza rígida.

El uso de estadísticos y redes neuronales también se han adaptado para la solución de este problema. Algunos de los sistemas propuestos por Rowley [26] confían en esquemas de entrenamiento de redes neuronales y cálculo de medidas de distancia entre conjuntos de entrenamiento para detectar rostros, o detectores basados en la Regla de Decisión de Bayes, como el propuesto por Kanade [28]. En la Figura 1.2 presentamos gráficamente esta clasificación.

En general podemos clasificar las diferentes técnicas empleadas en la detección de rostros en las siguientes categorías:

- **Métodos basados en conocimiento de características.** Emplean el conocimiento de las características del rostro y siguen la metodología tradicional de detección donde

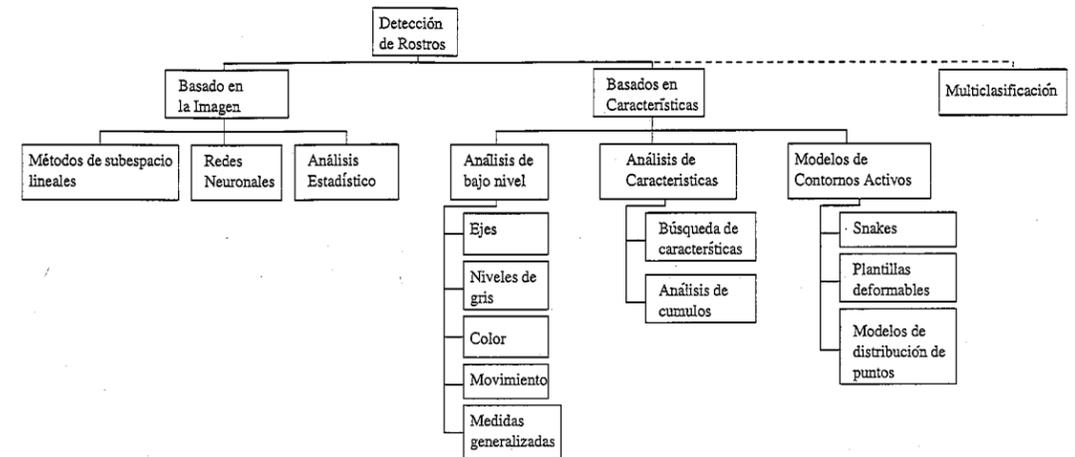


Figura 1.2: Técnicas empleadas en la detección de rostros.

las características de bajo nivel son derivadas del análisis del conocimiento *a priori*. Propiedades como el color de la piel, la geometría de la cara o movimiento son explotadas.

- **Métodos basados en la imagen.** Directamente clasifican la imagen como rostro, usando mapeos y esquemas de entrenamiento. Las redes neuronales son un ejemplo de estos métodos.
- **Métodos basados en multiclasificación.** Combinan los métodos anteriores para lograr resolver la tarea en condiciones de tiempo real.

Métodos basados en conocimiento de características.

Los métodos basados en conocimiento de características se pueden clasificar en tres grupos principales: (1) Análisis de características locales, (2) Análisis de características globales y (3) Modelos de Deformables.

Las técnicas basadas en el *análisis de características locales* utilizan las propiedades faciales tales como ejes o bordes, color y movimiento. Los *ejes* pueden proveer la base para muchas características faciales derivadas como contorno de los ojos, etc. La información de **color** ha sido usada para la detección de piel y es por supuesto muy útil en la detección de rostros. La información de **movimiento** provee un método efectivo de localización descartando fondos estáticos. La existencia de un par de ojos, la estimación de contornos de imágenes en movimiento,

y el flujo óptico son usados para resolver esta tarea basados en información de movimiento.

El conocimiento de la geometría de la cara también se ha empleado como *Características Globales* en la detección de rostros, tales como: la forma de la cara, el par de ojos, el contorno de la cabeza (en forma elipsoidal) y el cuerpo (abajo de la cabeza). Medidas generalizadas de bajo nivel como una operador de simetría basado en operaciones a nivel de ejes se han propuesto, Reisfeld [25]. La búsqueda de las características faciales adecuadas, aquellas más prominentes en una cara permiten el uso de características menos prominentes usando medidas antropométricas de la geometría de la cara. Por ejemplo, una pequeña área de una secuencia con un interlocutor, implica una escenario de "una cabeza sobre los hombros" y un par de ojos en el área de la cabeza incrementa la confianza de esta región.

Modelos Deformables. Contornos Activos (Snakes), Plantillas Deformables (DT), y Modelos de Distribución de Puntos (PDM), se han empleado en características complejas y no rígidas tales como el límite de la cabeza, la pupila de los ojos o el seguimiento de los labios. Las *Snakes* o *Contornos Activos* son comunmente usados para localizar el contorno de la cabeza. El modelo general consiste de la inicialización aproximada al contorno de la cabeza, cuando se encuentra con los ejes se va moviendo hasta que se ajusta. La evolución de la Snake se logra minimizando una función de energía, análoga a sistemas físicos. Presentan problemas cuando se ven atrapadas en características falsas, un bajo brillo en la imagen presenta problemas en la detección de ejes, además no son eficientes al extraer características no convexas debido a su tendencia de mantener una mínima curvatura.

Localizar el contorno de la cara no es una tarea fácil debido a que las características locales son difíciles de organizar en un modelo único usando contornos genéricos. Las *Plantillas Deformables* incorporan información global de las características adicionales para mejorar el proceso de extracción. Sin embargo la evolución de la plantilla es sensible a su posición inicial debido a su estrategia de ajuste prefijado, los requerimientos de tiempo computacional son altos debido a su implementación secuencial en el proceso de minimización, y otros factores que afectan su desempeño.

PDM son una descripción compacta parametrizada de la forma basada en estadística. La arquitectura y el proceso de ajuste es diferente de los otros modelos aquí señalados. El contorno de un PDM es discretizado en un conjunto de puntos etiquetados, las variaciones de estos puntos son primero parametrizados sobre un conjunto de entrenamiento que incluye objetos de diferentes tamaños y posiciones, Hjelmas [13].

Métodos basados en la imagen.

A diferencia de la categoría anterior, la representación basada en la imagen es directamente clasificada como clase prototipo de rostro o no, empleando un esquema de mapeo y entrenamiento, usando directamente la intensidad de los pixeles. También se categorizan en tres grupos: (1) Métodos de Subespacio Lineal, (2) Redes Neuronales y (3) Aproximaciones Estadísticas.

Los *Métodos de Subespacio Lineal* incluyen el Análisis de Componentes Principales (PCA), Análisis de Discriminación Lineal (LDA) y Análisis de Factores (FA). Las imágenes de rostros humanos recaen en un subespacio, de todo el espacio de imágenes posibles. Para representar este subespacio, podemos usar aproximaciones neuronales, como se describe más adelante, o métodos cercanos al análisis estadístico multivariado. En los 80s Sirovich and Kirby [32] desarrollan una técnica usando PCA para representar eficientemente rostros humanos. Dado un conjunto de diferentes imágenes, se encuentran los componentes principales de la distribución de los rostros, expresados en términos de eigenvectores. Cada rostro individual puede ser representado como una combinación lineal de los eigenvectores más grandes, conocidos comunmente como eigenfaces, con pesos apropiados. Recientemente Pentland [22] ha desarrollado aun más esta técnica con un esquema probabilístico.

Las *Redes Neuronales* han sido extensamente usadas en problemas de reconocimiento de patrones, incluyendo por supuesto la detección de rostros. Las Redes Neuronales Artificiales hoy en día son mucho más que estructuras MLP (Multiple Layer Perceptron), arquitecturas modulares, algoritmos de aprendizaje complejos, autoasociativos y redes de comprensión, entrenadas para detectar rostros en diferente posición y orientación, con algunos contras como búsquedas en una dimensionalidad muy alta. Una introducción a algunos métodos básicos con redes neuronales para detección de rostros se puede encontrar en Viennet [36]. Para obtener resultados adecuados se debe contar con un gran número de ejemplos positivos y negativos, además de algoritmos de aprendizaje robustos. Otro ejemplo muy rápido computacionalmente que está en este momento mostrando muy buenos resultados es el aprendizaje basado en Adaboost, Viola y Jones [15]. Propuestas con arquitecturas diferentes se han presentado recientemente.

Existen otras técnicas de *Aproximaciones Estadísticas* basadas en teoría de la información, detección por Máxima Verosimilitud (ML), Maquinas de Soporte Vectorial (SVM), Métodos Bayesianos (BM), etc. Estos métodos son técnicas muy robustas en el procesamiento de imágenes en escala de grises, ya que su búsqueda esta basada en la imagen. La mayoría de los algoritmos reportados se basan en escaneo por ventaneo para detectar rostros a diferentes escalas, que

resulta computacionalmente caro. Para lograr un desempeño en tiempo real, una combinación de métodos basados en la imagen con métodos basados en características son empleados.

Métodos basados en multclasificación.

Las técnicas de multclasificación utilizan características faciales junto con modelos holísticos para la clasificación de imágenes en la clase rostro. Primero características como color o movimiento se emplean para localizar regiones probables evitando así una búsqueda exhaustiva, posteriormente modelos basados en la imagen se emplean para verificar estas regiones y clasificarlas como rostro. Este es el esquema a seguir en este trabajo de tesis, analizando primeramente características locales tomando información de color para después utilizar métodos más robustos en la clasificación basados en la semántica de la imagen, siguiendo un esquema de solución basado en un problema general de reconocimiento de patrones.

1.2.2. Problemas relacionados con la detección de rostros.

Los retos asociados con la detección de rostros dependen de ciertos factores como:

- Posición. Las imágenes de un rostro pueden variar dependiendo de su posición relativa a la cámara (de frente, posición 3/4, de perfil, hacia arriba o hacia abajo).
- Presencia o ausencia de algunas características faciales (barba, bigote o lentes).
- Expresión facial. La apariencia del rostro se ve afectada directamente por la expresión facial.
- Oclusión. Los rostros pueden estar parcialmente ocultos ante la lente.
- Orientación de la imagen. La apariencia de un rostro varía dependiendo de la rotación sobre el eje óptico de la cámara.
- Condiciones de la imagen. Iluminación (espectro, distribución de la fuente e intensidad), las características de la cámara (la respuesta del sensor, control de ganancia, lentes) y la resolución.

El rostro de una persona en particular puede variar debido a las condiciones mencionadas. Aunque los rostros de diferentes individuos tienen las mismas características (ojos, nariz, boca), la forma de estas y la relación espacial entre estas difieren de persona a persona. Más aun estas

características tienen variación por sí mismas. El rango de patrones faciales permisibles es mayor aun cuando las condiciones de la imagen no se tienen bajo control.

Aunque la aplicabilidad de las técnicas resulta, a priori, muy amplia, no hay un método que sea la panacea. Diversas razones hacen que los sistemas de reconocimiento de formas o rostros operativos sean muy específicos del problema a resolver:

1. La naturaleza de los patrones: símbolos, dibujos, imágenes, objetos tridimensionales, etc.
2. Los requerimientos del sistema, especialmente el tiempo de respuesta hace que algunos métodos de reconocimiento, aún siendo superiores en éxito no sean aplicables en la práctica.
3. Factores económicos: un sistema equipado con diferentes sensores y equipos de procesamiento muy potentes pueden dar resultados muy satisfactorios pero no pueden ser asumidos por los usuarios.

Estos factores hacen que un sistema adecuado para un problema sea inaplicable para otro, lo que posibilita el estudio y desarrollo de nuevas técnicas. Debe considerarse, además, que el problema tratado o el reconocimiento de patrones no constituye un campo de estudio cerrado sino que las técnicas relacionadas con este campo pueden encontrarse en otras ramas de la Ciencia y de la Tecnología.

1.3. Alcance y contribución de la tesis.

La principal contribución de esta tesis pueden resumirse de la siguiente forma:

- Un método novedoso y computacionalmente rápido de clasificación de color de regiones con tono de piel, basado en información de las componentes de espacios de color RGB, HSV y HMMD, formando así un espacio híbrido, que además están soportados por estándares de codificación como JPEG-2000, MPEG4 y MPEG7.
- Una técnica de diferenciación de regiones de tonos de piel basada en un descriptor de color con la información del espacio de color HMMD (Hue, Max, Min, Diff), la cual auxilia ampliamente en la separación de regiones continuas pertenecientes a diferentes objetos clasificados como regiones de piel.
- Un algoritmo rápido y novedoso de partición de regiones no convexas.

- Además de modificaciones y planteamientos alternos a algoritmos de etiquetado de regiones conectadas , a fin de lograr reducir el tiempo computacional en la obtención de resultados.
- El uso conjunto de estas técnicas para llevar a cabo la tarea impuesta de detección de regiones en la categoría de rostros.

El desarrollo de un sistema de detección de rostros general es un reto arduo debido a que el sistema debe enfrentarse a un rango muy amplio de variación en la configuración de estos patrones. Como se mencionó en la Sección 1.2.2, cubrir este rango completo de variación de imágenes con uno o más rostros o ninguno es casi imposible. Hasta la fecha el problema de detección de rostros es un problema abierto.

En el análisis se deberá contar con imágenes a color con condiciones de iluminación normales en ambientes interiores o exteriores, pero no bajo el reflejo de una fuente de luz diferente a la luz blanca ya que afecta el rango de color de la piel, ya que esto provoca que el tono de la piel se encuentre en un rango diferente al considerado como color natural en términos generales. Por supuesto también la resolución de la imagen o región juega un papel importante; en nuestro análisis el rostro más pequeño deberá ser superior a los 18×20 pixeles.

1.4. Organización de la tesis.

Esta tesis consta de 7 capítulos incluyendo esta Introducción y un Apéndice, el resto del trabajo está organizado de la siguiente manera:

En el **Capítulo 2** presentamos el esquema general de segmentación propuesto para la solución del problema planteado siguiendo un enfoque de reconocimiento de patrones, en el **Capítulo 3** revisamos algunos conceptos básicos de color y sus espacios de representación, revisamos las técnicas de segmentación de color basadas en el tono de la piel y presentamos la propuesta de nuestro modelo híbrido basado en los espacios de color RGB, HSV y HMMD para la obtención de las regiones de piel, en el **Capítulo 4** presentamos el esquema de procesamiento y análisis realizado en las zonas etiquetadas como piel, para identificar, localizar y dividir el número de regiones probables, en el **Capítulo 5** revisamos las técnicas de verificación de rostros empleadas como último paso de nuestro sistema de detección, en el **Capítulo 6** presentamos los resultados obtenidos así como los detalles de implementación, finalmente en el **Capítulo 7** presentamos las conclusiones y el trabajo futuro a realizar. En el **Apéndice A** se describe la herramienta de software desarrollada en este trabajo de tesis.

Capítulo 2

Estrategia de Multclasificación para la Detección de Rostros.

2.1. Introducción al problema en el ámbito de Reconocimiento de Patrones.

El desarrollo del trabajo elaborado en esta tesis consistió en el estudio del conjunto apropiado de variables de representación del conjunto de datos, la variabilidad entre estos, las medidas de similitud entre patrones, formulando un conjunto de reglas que aseguran la decisión adecuada, en cada etapa del desarrollo de un sistema para resolver el problema de Detección de Rostros, siguiendo una aproximación estadística, en el ámbito de Reconocimiento de Patrones.

El esquema seguido para la construcción de nuestro sistema de detección puede verse desde un punto de vista funcional, donde la entrada es un patrón natural y el resultado es una etiqueta. No debe entenderse que todos los sistemas de reconocimiento de patrones deben incorporar todas estas unidades, Figura 2.1.

A continuación profundizaremos en los diferentes módulos funcionales de este sistema general.

2.1.1. Adquisición de datos.

La entrada al sistema de reconocimiento o clasificación es un vector numérico que contiene los valores muestreados y cuantificados (o binarizados) de una serie de señales naturales. De una manera más formal, suponiendo patrones n -dimensionales, un patrón X es una variable

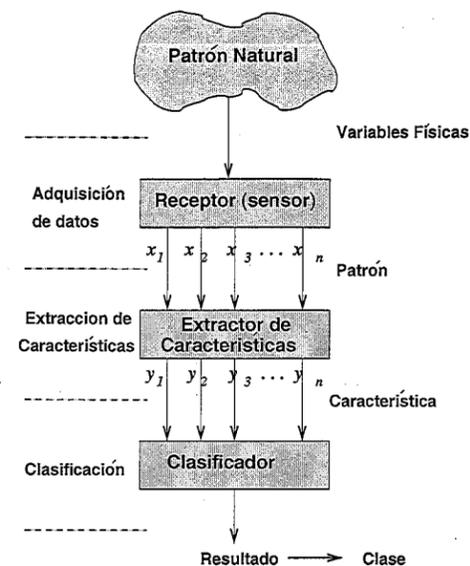


Figura 2.1: Etapas de un sistema de Reconocimiento de Patrones.

aleatoria n -dimensional compuesta por n componentes, x_1, x_2, \dots, x_n variables aleatorias, tales que $x_i \in G_i$ para $i = 1, 2, \dots, n$ y $G_i = G = \{0, 1, \dots, 255\}$ son los valores de representación de las variables.

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

En el caso de imágenes digitales a color, en el modelo de color RGB cada color aparece descompuesto en sus componentes espectrales primarias: rojo (R), verde (G) y azul (B). Por conveniencia se supone que todos los colores han sido normalizados y escalados, es decir se supone que todos los valores de rojo, verde y azul están en el rango $\{0 - 255\}$. Los puntos de la diagonal en el cubo RGB que va de $(0,0,0)$ a $(255,255,255)$ pueden representarse visualmente por 256 niveles de gris, escalados desde el negro (nivel de gris 0 o $(0,0,0)$ en RGB) al blanco (nivel de gris 255 o $(255, 255, 255)$ en RGB).

Espacio de representación.

Con esta aproximación un patrón no es más que un punto en el *espacio de representación de los patrones* que es un espacio de dimensionalidad determinada por el número de variables consideradas. Esta aproximación concluye que es razonable que los patrones pertenecientes a una misma clase estén cercanos en el espacio de representación mientras que aquellos que pertenezcan a clases diferentes deberían estar en diferentes regiones del espacio de representación. Un patrón se representa como un punto en el *espacio de patrones P*. El espacio de patrones P es un espacio de dimensionalidad determinada por el número de variables consideradas y se define como el conjunto de todos los valores posibles que puede tomar patrón X, esto es

$$P = \times_{i=1}^n G_i$$

donde \times denota el producto cartesiano. Los valores típicos de x_i están en el conjunto $G_i = \{0, 1, \dots, 255\}$.

En el caso de una imagen como un objeto en si, por ejemplo considerando un patrón de 19×19 píxeles que representa un rostro, la dimensión para la representación de los componentes del vector es de $256^{19 \cdot 19} = 2^{8 \cdot 361} = 2^{2888}$, el cual es un espacio de una dimensión gigantesca.

Los espacios de representación para imágenes a color utilizados son: RGB, HSV y HMMD, los cuales se describirán en el capítulo 3. En el capítulo 5 se hablará sobre el espacio de representación para rostros.

Similaridad entre patrones.

La tarea fundamental de un sistema de reconocimiento de patrones (clasificador) es la de asignar a cada patrón de entrada una etiqueta. Dos patrones diferentes deberían asignarse a una misma clase si son similares y a clases diferentes si no lo son. La cuestión que se plantea ahora es la definición de una medida de similaridad entre patrones. Aunque existen varias maneras de expresar esta similaridad, estas son muy dependientes del problema a resolver.

Supongamos un sistema de adquisición perfecto (sin ruido). Podemos asegurar que:

1. La adquisición repetida del mismo patrón debería proporcionar la misma representación en el espacio de patrones. Por ejemplo, una cámara debería proporcionar siempre la misma imagen de la misma escena si las condiciones externas no cambiaran.

2. Dos patrones diferentes deberían proporcionar dos representaciones diferentes.
3. Una ligera distorsión aplicada sobre un patrón debería proporcionar una pequeña distorsión de su representación.

Estas consideraciones sugieren que si las representaciones de dos patrones están muy cercanas en el espacio de representación, entonces los patrones deben tener un alto grado de similaridad.

Variabilidad entre patrones.

La suposición de un sistema de adquisición perfecto no deja de ser eso, un suposición. Los sistemas de adquisición introducen, por su naturaleza, cierta distorsión o ruido, lo que produce una variabilidad en la representación de los patrones. Aunque es posible controlar eficientemente en muchos casos esta distorsión mediante el calibrado de los sistemas de adquisición aparece otra fuente de variabilidad por la propia naturaleza de los patrones. Con mucha frecuencia, patrones de una misma clase difieren, incluso significativamente.

2.1.2. Selección y extracción de características.

El problema que se trata de resolver es el de extraer la información *relevante* para la clasificación entre la suministrada por los sensores (datos en bruto). De forma general este problema puede plantearse como sigue. Dado un conjunto de patrones n -dimensionales

$$X = [x_1, x_2, \dots, x_n]^T$$

se trata de obtener un nuevo conjunto (características) d -dimensionales

$$Y = [y_1, y_2, \dots, y_d]^T$$

donde $d \leq n$.

Este objetivo puede abordarse de dos formas:

1. **Reduciendo la dimensionalidad de los datos.** Si los patrones son de alta dimensionalidad, el costo computacional asociado a la clasificación puede ser muy alto. Como otra consideración computacional hay que considerar el espacio de almacenamiento adicional que supone guardar los valores de nuevas variables. Además, algunas de las variables pueden ser redundantes con otras y no aportar información adicional.

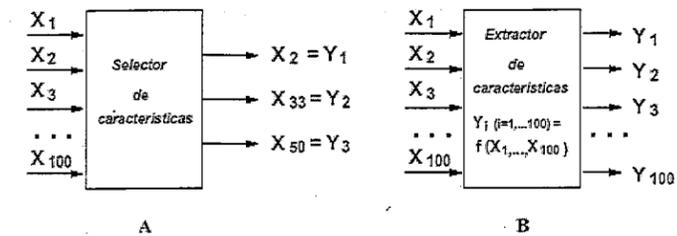


Figura 2.2: a) Selección de las 3 variables más significativas. b) Transformación de las variables originales.

Las técnicas dedicadas a seleccionar las variables más relevantes se dicen de *selección de características* y reducen la dimensionalidad de los patrones. Este proceso puede esquematizarse como se indica en la Figura 2.2 (a), en el que un módulo selector recibe patrones n -dimensionales (en el ejemplo, $n = 100$) y proporciona como resultado las d variables más significativas (en el ejemplo, $d = 3$) de acuerdo a algún criterio a optimizar. En este caso $d < n$ y el conjunto de las de variables seleccionadas es un subconjunto del conjunto original de variables.

2. **Cambiando el espacio de representación.** El objetivo es obtener una nueva representación de los patrones en la que los agrupamientos aparezcan bien separados si son de diferente clase y que haya un agrupamiento por clase. Esto puede conseguirse aplicando alguna transformación sobre los datos originales. Estas transformaciones suelen ser transformaciones lineales y el objetivo suele ser maximizar la varianza. Estas técnicas reciben el nombre de extracción de características y producen un nuevo conjunto de variables. Este proceso puede esquematizarse como se indica en la Figura 2.2. b), en el que un módulo extractor recibe patrones n -dimensionales (en el ejemplo, $n = 100$) y proporciona como resultado nuevos patrones n -dimensionales de acuerdo a algún criterio a optimizar. Es posible que las nuevas variables estén implícitamente ordenadas, por lo que proporcionan, adicionalmente un procedimiento de selección. En este caso $d = n$ y las variables seleccionadas no forman un subconjunto del conjunto original de variables. PCA es un ejemplo de esta categoría de la cual se hará uso en este trabajo en la etapa de clasificación de regiones de piel en la categoría de rostros. También el cambio de la información de color a otros espacios de representación más adecuados o que aportan mayor separación de regiones son algunas estrategias que se emplearán en este trabajo.

2.1.3. Módulo de clasificación.

El objetivo final de un sistema de Reconocimiento de Patrones es el etiquetar de forma automática patrones de los cuales desconocemos su clase. Suponemos que el sistema dispone de un módulo de adquisición de datos y que se han seleccionado previamente las variables más significativas.

El conjunto de clases.

En primer lugar debe establecerse claramente el objetivo final del sistema: que salidas debe proporcionar. Si suponemos que todos los patrones a reconocer son elementos potenciales de J clases distintas denotadas $w_j, j = 1, 2, \dots, J$, llamaremos

$$\Omega = \{w_1, w_2, \dots, w_J\}$$

al conjunto de las clases informacionales. Conviene tener en cuenta que una clase informacional es la denominación que se da a una clase conocida y con significado.

Como veremos, resulta conveniente ampliar el conjunto Ω incorporando una nueva clase, llamada la clase de rechazo. Así, se define la clase de rechazo (w_0) como una clase que se asigna a todos los patrones para los que no se tiene una certeza aceptable de ser clasificados correctamente en alguna de las clases de Ω . Se dice que

$$\Omega^* = \{w_1, w_2, \dots, w_J, w_0\}$$

es el conjunto extendido de clases informacionales.

Así definimos nuestros conjuntos de clases, para la clasificación de piel como:

$$\Omega^* = \{w_1 = \text{"piel"}, w_0 = \text{"nopiel"}\}$$

y la para clasificación de regiones:

$$\Omega^* = \{w_1 = \text{"Rostro"}, w_0 = \text{"noRostro"}\}$$

El clasificador.

Una vez establecido el conjunto de clases se procede a la construcción del clasificador. Como puede intuirse, la construcción del clasificador no es una tarea trivial ni directa e involucra una serie de etapas como:

1. La elección del modelo.
2. Aprendizaje (entrenamiento del clasificador).
3. Verificación de los resultados.

Un clasificador o regla de clasificación es una función $d : P \rightarrow \Omega^*$ definida sobre los patrones X tal que para todo patrón X , $d(X) \in \Omega^*$. En caso de imágenes digitales el resultado de la clasificación es "palpable": puede presentarse en forma de una imagen de etiquetas, asignando un color a cada una. En el caso de una región, señalando el contorno que la rodea.

Como se ha comentado anteriormente, los patrones de una misma clase presentan cierta variabilidad natural. No obstante, deben estar (relativamente) cercanos en el espacio de representación y lejanos (relativamente) respecto a los patrones de otras clases. Esta situación, en un caso ideal, hace que se distingan diferentes agrupamientos en el espacio de representación, uno por cada clase considerada y en estos casos es posible asociar regiones disjuntas en P a cada una de las clases representadas. El esquema general se presenta en la Sección 2.3.

Aprendizaje.

Se suele utilizar indistintamente los términos aprendizaje y entrenamiento para referirse al proceso de construcción del clasificador. El aprendizaje puede realizarse de dos maneras muy diferentes.

1. Aprendizaje supervisado.

Un aprendizaje supervisado requiere disponer de un conjunto de patrones de los cuales se conoce su clase verdadera. A este conjunto se le denomina *conjunto de entrenamiento*. Este tipo de entrenamiento se denomina *entrenamiento supervisado* y los clasificadores así obtenidos *clasificadores supervisados*. Disponer de un conjunto de entrenamiento supone que alguien se ha preocupado de etiquetar los patrones de ese conjunto. Esta tarea la suele realizar un experto en el campo en el que se va a realizar el reconocimiento y generalmente viene impuesto.

2. Aprendizaje no supervisado.

El aprendizaje no supervisado se realiza a partir de un conjunto de patrones del que no se conoce su clase verdadera. En ocasiones, ni siquiera se conoce el número de clases. Básicamente, se traduce en encontrar agrupamientos. El objetivo suele ser el de verificar

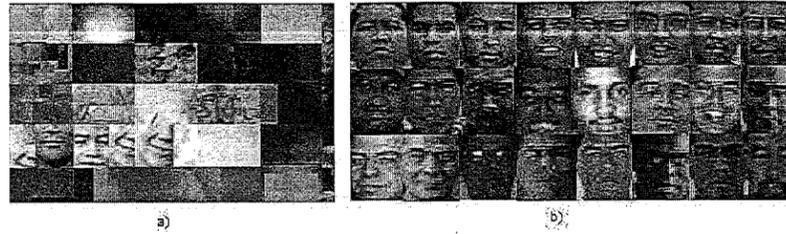


Figura 2.3: Subconjunto de entrenamiento (a) Píxeles de piel, (b) Rostros

la validez del conjunto de clases informacionales para una clasificación supervisada. Las técnicas utilizadas suelen denominarse métodos de agrupamiento o clustering.

En este trabajo disponemos de un conjunto de entrenamiento, por lo que obviamente seguiremos un esquema de aprendizaje supervisado.

Un conjunto de aprendizaje o conjunto de entrenamiento, T , consiste en un conjunto de parejas (X_i, c_i) donde $X_i \in P$ y $c_i \in \Omega$ a las que llamamos prototipos.

$$T = \{(X_1, c_1), (X_2, c_2), \dots, (X_N, c_N)\}$$

donde N es el número de prototipos. Mediante X_i nos referimos al patrón i -ésimo del conjunto de prototipos y mediante c_i a la clase cierta del patrón X_i . Evidentemente, $T \in P \times \Omega$

Si N es el número total de prototipos, mediante N_j nos referimos al número de prototipos de clase w_j , $j = 1, 2, \dots, J$. Así, $N = \sum_{j=1}^J N_j$.

En la Figura 2.3 presentamos solo un subconjunto del conjunto de entrenamiento, (a) los píxeles clasificados como piel, (b) imágenes clasificadas como rostros.

Aprendizaje supervisado paramétrico y no paramétrico.

Si consideramos que en un caso ideal cada agrupamiento representa a una clase y cada clase tiene asociado un agrupamiento bien diferenciado de los demás, un problema de clasificación supervisada puede plantearse como la búsqueda de las superficies que separan los diferentes agrupamientos. Estas superficies se denominan *superficies de decisión*. Las superficies de decisión determinan *regiones de decisión* de forma que cada clase tiene asociada una región en P y la decisión sobre la clase a asignar a un nuevo patrón se hará en base a la región en la que éste se encuentra en P .

La búsqueda de estas superficies (análogamente, regiones) de decisión se puede abordar de

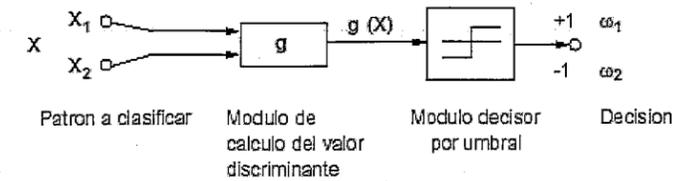


Figura 2.4: Diagrama de bloques del clasificador planteado.

dos maneras, dependiendo de si se conoce o supone un determinado modelo estadístico para las clases.

1. Si se supone un completo conocimiento a priori de la estructura estadística de las clases, el aprendizaje se reduce a la estimación de los parámetros que determinan las funciones de densidad de probabilidad de las clases. Las fronteras de decisión están definidas por las distribuciones de probabilidad de las clases. Los clasificadores construidos bajo esta suposición se conocen como *clasificadores paramétricos*.
2. Si no se supone un determinado modelo estadístico, bien por desconocimiento o por la imposibilidad de asumir un modelo paramétrico adecuado, el problema resulta más complejo y se puede abordar desde diferentes perspectivas. Las fronteras de decisión están definidas por los prototipos. Estos clasificadores se conocen como *clasificadores no paramétricos*.

En un problema de clasificación entre dos clases, dado un patrón X , la decisión sobre la clase que le corresponde puede formularse en base a la evaluación de una *función discriminante*, $g(X)$, de forma que, por ejemplo, si $g(X) > 0$, X se etiqueta de clase 1 y si $g(X) < 0$ de clase 2. Así, la frontera de decisión entre las dos clases es el lugar de los puntos que verifican la Ecuación $g(X) = 0$. La frontera de decisión divide el espacio de representación en dos regiones disjuntas: las regiones de decisión. Así, R_i , la región de decisión asociada a la clase i , ($i = 1, 2$), es la formada por aquellos puntos a los que se asignaría la clase i ,

$$R_1 = \{X; d(X) = w_1\} \text{ equivalentemente, } R_1 = \{X; g(X) > 0\}$$

$$R_2 = \{X; d(X) = w_2\} \text{ equivalentemente, } R_2 = \{X; g(X) < 0\}$$

Un esquema modular del clasificador adecuado para este problema se muestra en la Figura 2.4.

En consecuencia, para construir el clasificador en primer lugar debe estudiarse cómo se distribuyen los patrones en cada clase (la estructura estadística de las clases) y establecer

una función discriminante adecuada utilizando los prototipos disponibles. En esto consiste básicamente el **entrenamiento supervisado paramétrico**. En el caso paramétrico, el Reconocimiento de Patrones puede verse, desde un punto de vista muy general, como un problema de estimación de funciones de densidad de probabilidad en un espacio multidimensional para posteriormente dividir el espacio en regiones de decisión asociadas a clases. La herramienta fundamental para esta tarea es la estadística y particularmente la teoría de la decisión de Bayes. Cuando no puede suponerse un modelo paramétrico para las funciones de densidad de probabilidad asociadas a las clases se utilizan modelos no paramétricos para estimar las funciones de densidad de probabilidad. En estos casos la única información disponible para la clasificación es el conjunto de prototipos. El método más simple entre los no paramétricos es el del vecino más cercano, que consiste en etiquetar un patrón con la etiqueta del prototipo más cercano.

2.2. Clasificación de Imágenes para la Detección.

Después de un amplio análisis respecto a las imágenes de prueba con las que se contó y los resultados obtenidos para cada una de estas, se generalizó una división de imágenes en: imágenes de Video Conferencia o secuencias, que en general en su contenido aparece uno o más interlocutores en un ambiente de iluminación controlado y con un fondo (background) simple e Imágenes Fotográficas, que es cualquier imagen estática obtenida con alguna cámara digital de la que por su fuente no se tiene certeza del contenido, cumpliendo además cada una de las imágenes con los requisitos impuestos anteriormente, Sección 1.3. En este trabajo de tesis se plantean dos esquemas diferentes para la obtención de resultados para cada una de las categorías de imágenes mencionadas, ver Figura 2.5.

Secuencias o Imágenes de Video Conferencia. Estas imágenes por lo general se obtienen bajo un ambiente controlado, por esta razón se requiere de menor análisis. La diferencia con el otro esquema clasificación es que el paso de diferenciación de regiones de piel utilizando un descriptor en el espacio de color HMMD y la separación de regiones no convexas no se emplea, disminuyendo así el tiempo computacional.

Imágenes Fotográficas. Estas imágenes son mucho más complejas debido al completo desconocimiento y nulo control de éstas, para mejorar los resultados obtenidos se requirió de mayor análisis, como se describirá más adelante, sin embargo estando todavía cumpliendo los requerimientos de tiempo computacional (menores a décimas de segundo).



Figura 2.5: Diferencia en las imágenes de prueba, (a) img. izq. Secuencia de Video, (b) img. der. Fotografía normal.

2.3. Descripción general del sistema de multclasificación.

Como se mencionó en el Capítulo 1, el uso de múltiples técnicas para lograr un mejor desempeño resulta muy útil, las características faciales se usan para localizar las regiones más probables, posteriormente los métodos basados en la imagen se emplean para verificar las regiones resultantes. Al igual que en el trabajo de Wu[38] la arquitectura de nuestro sistema es jerárquica, en cada etapa un porcentaje de la imagen, considerada como la unión de un conjunto de regiones, es descartada, dejando solo aquellas regiones que son más probables y finalmente aquellas clasificadas como rostros. El sistema de Multclasificación consiste en tres diferentes etapas, partiendo un frame de una secuencia de video o una imagen estática se analiza en nuestro sistema, las regiones ya localizadas son la salida para una aplicación posterior, este sistema se muestra en la Figura 2.6:

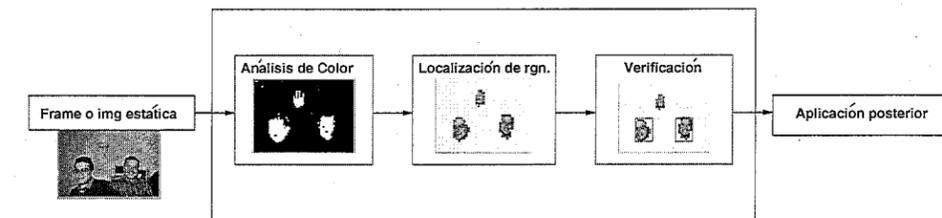


Figura 2.6: Esquema general de clasificación en la Detección de Rostros.

Análisis de color. Se propone un clasificador de color robusto y rápido como primer etapa, para extraer todas las regiones semejantes al tono piel, para posteriormente asignar un valor de probabilidad de pertenecer a la clase *piel* a aquellas regiones de la imagen en las que se tenga el mismo color de piel respecto al modelo planteado, basado en un espacio de

color híbrido que combina información de los espacio de color RGB, HSV y HMMD. El objetivo es aislar aquellos rasgos de la imagen que merecen evaluación con criterios más específicos de clasificación, reduciendo el espacio de búsqueda, evitando así búsquedas exhaustivas.

Localización de Regiones. En esta etapa se emplean algoritmos de diferenciación de regiones de piel en base al descriptor de color dominante en el espacio de color HMMD. Se utilizan operadores morfológicos en escala de grises para eliminar detalles menores en el resultado preservando la forma general de las regiones. Adicionalmente se emplean algoritmos, con ciertas modificaciones dando prioridad a regiones convexas, para el etiquetado de regiones conectadas, además de obtener los datos necesarios para el análisis de descriptores de la forma de la región. Finalmente se propone un algoritmo de separación y suavizamiento de regiones convexas basadas en el conocimiento de características a priori como el hecho de que la cabeza humana tiene forma elipsoidal.

Verificación. La última etapa consiste en la verificación de cada una de estas regiones resultantes a fin de determinar si cada una de ellas se clasifica como Rostro, esto analizando la semántica de la región empleando técnicas basadas en la imagen.

Cada una de estas etapas se describirá a detalle en los capítulos 3, 4 y 5.

Capítulo 3

Segmentación de color.

3.1. Conceptos básicos.

La luz es color, un pequeño experimento para comprobar esto consiste en colocar un cristal (prisma de base triangular) por el que ha de pasar un rayo de luz, lo que observaremos es la descomposición de la luz blanca en los seis colores del espectro, y de manera inversa se puede hacer un experimento similar. Así, la luz blanca, esta formada por luz de siete colores, comenzando en violeta, después azul, cian, verde, amarillo, naranja y finalmente rojo; que al incidir sobre algún cuerpo éste absorbe algunos de estos colores y refleja otros.

El color es una sensación subjetiva, esto es nadie puede asegurar a ciencia cierta que percibe los colores igual que otro. De cualquier forma los humanos vemos mas o menos igual y partiendo de esta premisa estudiamos el color. En la teoría del color, estudiada por Thomas Young, propone su teoría tricromática de colores primarios, confirmada en los 60s, esta teoría se basa en que cualquier color se compone de tres colores primarios (rojo, verde y azul), partiendo de estos se pueden obtener la infinidad de mezclas. Los procesos de iluminación, reflexión y su detección son aspectos importantes para modelar el comportamiento de la luz, comenzando por las fuentes de ésta, viajando a través de la escena, interactuando con diferentes objetos, y finalmente alcanzando el sensor de la cámara. Esto involucra la composición espectral y reflexión de las fuentes de luz y superficies, respectivamente, sus posiciones y orientaciones, la aspereza de la superficie en terminos ópticos, y las características de la cámara.

3.1.1. Reflexión de la piel.

El análisis óptico de la piel humana así como su modelado son importantes en áreas diversas como aplicaciones médicas, desarrollo cosmético, gráficos por computadora y visión por computadora. La piel es, a diferencia de otros órganos, directamente visible. El tono de piel es influenciado por las capacidades de filtrado de tres principales agentes: la melanina en la epidermis, el caroteno contenido en la dermis y en la grasa subcutánea, y los tubos capilares de la sangre a través de la dermis. La melanina es un pigmento café y el caroteno es un tinte orgánico, Störring [33].

El espectro final de la piel esta formado por al interacción entre la piel y la luz, pulsos de luz son transmitidos, absorbidos o reflejados a través de las capas de la piel. El espectro de la piel humana forma una serie homóloga continua debido a la caracterización causada por la absorción de melanina y la hemoglobina. Al igual que la mayoría de los objetos naturales, la piel tiene un espectro variable debido principalmente a la cantidad, densidad y distribución de melanina. Así, la piel puede ser clasificada como un material óptico inhomogéneo debido a que debajo de la superficie existen partículas colorantes que interactúan con la luz, produciendo dispersión y coloración. Un estudio más detallado de las características de la piel se encuentra en los trabajos de Störring [33] y Martinkauppi [21].

3.2. Modelado del color de la piel.

En la década pasada la detección basada en el color de la piel ha sido usada como indicador con mayor frecuencia en visión por computadora para detectar, segmentar, rastrear o seguir rostros o manos. En la construcción de un sistema basado en color, por lo general nos enfrentamos a tres principales problemas:

- La elección del espacio de color a utilizar.
- La manera de modelar la distribución del color de la piel.
- La forma en que debemos de procesar la información para resolver la tarea de Detección.

En este capítulo tratamos los dos primeros puntos, dejando la discusión del último punto para los capítulos 4 y 5.

En la literatura encontramos diversos estudios hechos sobre el comportamiento de la cromacidad de la piel en diferentes espacios de color, [21], [35], [33]. Muchos de ellos indican

que los tonos varían principalmente en las intensidades mientras que son más similares en las coordenadas (componentes) de cromacidad.

3.2.1. Espacios de color.

Primeramente se necesita una representación apropiada de las señales de color con metas a emplear esta información en áreas de visión por computadora, procesamiento de imágenes, gráficos por computadora, etc. esta representación se hace con los diferentes *espacios de color*. Un *modelo de color* es un modelo matemático abstracto que describe la forma en que los colores pueden ser representados como una secuencia finita (tuplas) de números, típicamente tres o cuatro valores o *componentes* de color. Sin embargo, un modelo de color sin ninguna función de mapeo asociada referenciada a un espacio de color es más o menos un sistema de color arbitrario sin mucha conexión a los requerimientos de una aplicación dada. Agregando una cierta función de mapeo entre el modelo de color y una cierta referencia al espacio de color, definen en su conjunto un nuevo *espacio de color*. Por ejemplo, AdobeRGB y sRGB son dos espacios de color diferentes, ambos basados en el modelo RGB.

La referencia estándar generalmente usada es el espacio de color *CIE Lab*, el cual fue específicamente diseñado para abarcar todos los colores que el humano promedio puede ver. Aunque es el espacio de color más preciso es demasiado complejo para un uso extensivo. Además de que un espacio de color es un término más específico para una cierta combinación de un *modelo de color* junto con una función de mapeo de color, el término *espacio de color* tiende a ser usado también para identificar modelos de color, ya que al identificar un espacio de color automáticamente se identifica su modelo de color asociado. Desafortunadamente esto resulta en una ambigüedad con respecto a los dos términos y son generalmente usados indistintamente. En un sentido general, un espacio de color puede ser definido sin usar de un modelo de color establecido.

Para los humanos, la representación de color más intuitiva, en el ámbito de gráficos por computadora, está en términos de la tonalidad, saturación y brillo. Si el color se visualizara en una pantalla la representación común está basada en los colores primarios rojo, azul y verde, el espacio RGB. Si deseáramos imprimir algún color es necesario transformarlos a los colores sustractivos como cian, amarillo y magenta. En el campo de procesamiento de imágenes y visión computacional los espacios de color podrían dividirse en:

- Espacios de color orientados a dispositivos, asociados con dispositivos de entrada, salida y procesamiento de señales.

- Espacios de color orientados a usuarios, que son utilizados como puente entre el usuario y el hardware utilizado para manipular la información de color.
- Espacios de color independientes del dispositivo, usados para especificar las señales de color independientemente de las características del dispositivo dado o aplicación.

La mayoría de las cámaras, fotográficas o de video, tienen alguna clase de salida en RGB, consecuentemente las imágenes se almacenan en coordenadas RGB. Como mencionamos anteriormente en el Capítulo 2, es conveniente cambiar el espacio de representación, debido a que en el modelo RGB las señales de intensidad y cromacidad están muy correlacionadas, esto se describirá más adelante.

Densidad del espacio de color.

El espacio RGB puede ser implementado en varios niveles específicos para su representación, dependiendo de la capacidad del sistema a usar. La forma más común de ésta es una implementación de 24 bits, 8 bits por cada canal. Esto significa que por cada uno de los tres canales se almacenan 256 niveles discretos de color. Cualquier espacio de color basado en el modelo RGB está limitado a $256 \times 256 \times 256 = 16.7$ millones de colores. Esto es lo que denominamos profundidad en bits.

3.2.2. Espacios de color empleados para el modelado y detección del color de la piel.

Espacio de color RGB.

El ojo humano posee tres tipos de células cónicas fotorreceptoras de luz, que son las responsables de la visión de color humana. Cada una de estas tres células tiene diferentes curvas de respuesta espectral, similar a las cámaras RGB. Debido a estos tres tipos de fotorreceptores, tres componentes numéricos son suficientes para describir un color para un sistema visual humano. RGB es un modelo de *color aditivo*, debido a que describe la cantidad y tipo de luz necesaria a emitir para reproducir un color dado. En RGB se almacenan los valores individuales para rojo, verde y azul.

En la literatura los espacios RGB están clasificados en lineales y no lineales. Un espacio RGB lineal significa que es lineal en intensidad, mientras tanto, un espacio RGB no lineal ($R'G'B'$) significa que es no lineal en su intensidad. Un ejemplo de RGB no lineal se

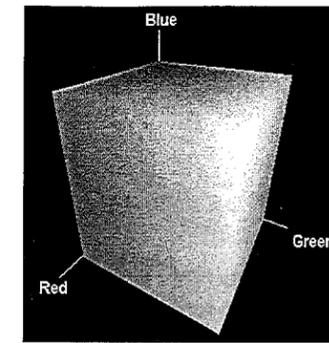


Figura 3.1: Espacio de color RGB

presenta cuando la intensidad tiene la *corrección-gama* para compensar la no linealidad de las pantallas CRT. Variedad de cámaras se construyen con la *corrección-gama* y varios formatos de almacenamiento almacenan los valores de RGB con esta corrección. El espacio RGB lineal es usado en gráficos por computadora y en visión computacional. En la Figura 3.1 se aprecia gráficamente este modelo en 3D.

Espacio de color HSV.

La representación basada en tonalidad-saturación-intensidad representa un espacio de color orientado al usuario, el cual es principalmente usado en gráficos por computadora. Los componentes están mejor enfocados para la interacción humana que RGB, debido a su relación con la percepción humana del color, saturación y luminancia, Figura 3.2. Existen varias versiones: HSI (Hue, Saturation, Intensity), HSB (Hue, Saturation, Brightness), HSL (Hue, Saturation, Lightness), HSV (Hue, Saturation, Value). Estos difieren principalmente en el cálculo del último componente, el cual puede ser de *Intensidad lineal* o *Intensidad no-lineal*. En su uso, *hue* es representada como una región circular, y una región triangular para representar *sat* y *val*. Otra visualización de este espacio es la de un cono. En esta representación, *hue* está representado en una formación cónica tridimensional del círculo de colores. La saturación es representada como la distancia del centro a una sección circular del cono y *val* es la distancia desde el pico del cono. Estas representaciones se presentan en la Figura 3.3 (a).

Conversión de RGB a HSV. La conversión de RGB a HSV es no-lineal, y se muestra a continuación:

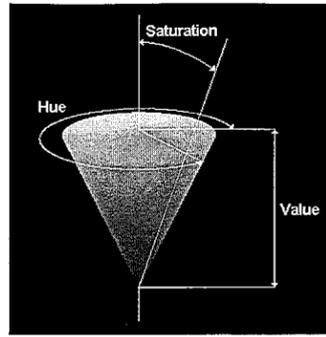


Figura 3.2: Espacio de color HSV

$$\begin{aligned} H &= \arccos\left(\frac{\frac{1}{2}[(R-G)+(R-B)]}{\sqrt{(R-G)^2+(R-B)(G-B)}}\right) \\ S &= \frac{\max(R,G,B) - \min(R,G,B)}{\max(R,G,B)} \\ V &= \frac{\max(R,G,B)}{255} \end{aligned} \quad (3.1)$$

El espacio *HSV* puede interpretarse de mejor manera si lo vemos como un cilindro, no un cono, donde *H* representa aproximadamente un ángulo de 0 a 360 grados, *S* es el radio del círculo ($0 \leq S \leq 1$), y *V* es la altura del cilindro ($0 \leq S \leq 1$), Figura 3.3 (b). Esta aproximación se puede ver como el modelo matemático mejor aproximado al espacio de color *HSV*. Sin embargo, en la práctica el número de niveles de *sat* y *hue* visualmente distintos decrece cuando el valor de *val* va disminuyendo hasta el negro. Este efecto se hará notar en comparación al modelo de color HMMD.

En términos prácticos, la componente *H* se puede calcular de la siguiente manera:

$$h = \begin{cases} \text{INDEFINIDO} & \text{si } (x = v) \\ \left(\frac{G-B}{\max-\min}\right) 60 & \text{si } (R = \max) \\ \left(2 + \frac{B-R}{\max-\min}\right) 60 & \text{si } (G = \max) \\ \left(4 + \frac{R-G}{\max-\min}\right) 60 & \text{si } (B = \max) \end{cases} \quad (3.2)$$

donde *INDEFINIDO* = 0, *min* = $\min(R, G, B)$ y *max* = $\max(R, G, B)$.

El sistema coordenado polar de los espacios de Hue-Sat, resulta en una naturaleza cíclica del espacio de color, haciéndolo inconveniente para modelos de piel paramétricos donde se necesita regiones compactas de colores de piel para su mejor desempeño. Una representación diferente de

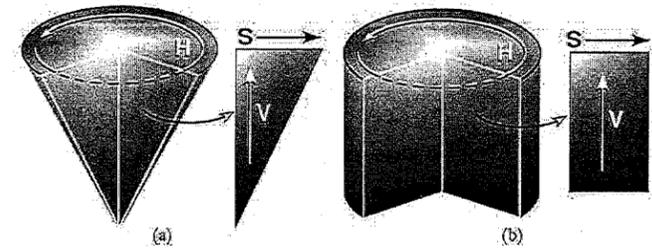


Figura 3.3: Diferentes representaciones del espacio HSV (a) Cono, (b) Cilindro

Hue-Sat usando coordenadas cartesianas puede usarse:

$$X = S \cos H, Y = S \sin H \quad (3.3)$$

Espacio de color HMMD.

Para obtener una cuantización de color "elegible" (cuantización de color escalar uniforme) de un espacio de color completo con una métrica de similitud simple entre dos colores, el espacio de color debe tener, necesariamente, la propiedad de ser perceptualmente uniforme. *Perceptualmente uniforme* es la propiedad que la similitud perceptual de dos colores es medida como la distancia entre dos puntos de color en el espacio de color, es decir, pequeñas distancias en el espacio de color corresponden con pequeñas diferencias en el color percibido. Aunque espacios de color como el CIELAB o CIELUV se consideran perceptualmente uniformes, sus transformaciones desde RGB demandan mayor computo y funciones de transformación más complejas. En este ámbito LG-ELITE (LG Corporate Institute of Technology - Information Technoloy Lab) ha desarrollado un modelo de color denominado HMMD y un esquema de cuantización de color conveniente.

Semántica de los parámetros de modelo HMMD. Existen cinco distintos parámetros en el modelo HMMD, sin embargo tres de ellos (*Hue, Max, Min* o *Hue, Diff, Sum*) son suficientes para definir el espacio de color. La semántica de los parámetros es como sigue:

- **Hue(h):** *Hue* puede ser representado como un ángulo (de 0 a 360 grados) en un círculo. Cuando en ángulo se incrementa, *hue* cambia de rojo (0°), amarillo (60°), verde (120°), azul (240°), a rojo (360° = 0°).
- **Max:** *Max* indica que tanto negro tiene, dando una sensación de matiz.

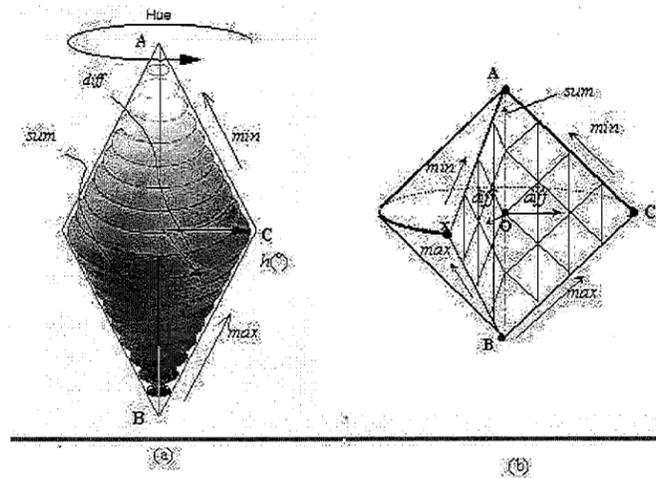


Figura 3.4: Espacio de color HMMD

- **Min:** *Min* indica que tanto blanco tiene, dando una sensación de tinte.
- **Diff:** *Diff* indica que tanto gris contiene y que tan cercano esta a un color puro, dando una sensación de tono.
- **Sum:** *Sum* simula el brillo del color.

Formación del espacio de color hmmd. *Hue* es el ángulo alrededor del eje vertical (AB en la Figura 3.4) perpendicular al eje *diff*. *max*, *min*, *diff* y *sum* estan definidos sobre el plano MMD que consiste en dos ejes (*min* y *max*) que tiene un valor constante de *hue*. Vale observar que *diff* y *sum* son parámetros auxiliares determinados por la diferencia y la suma de *max* y *min*. El plano MMD es un triángulo rectángulo (ABC o ABx en la Figura 3.4(b).) formado por el eje vertical AB, el eje *min* y el eje *max*. Los tres vértices representan blanco, color puro o negro, respectivamente, de arriba hacia abajo. El nivel de gris cambia a lo largo del eje vertical, *min* varia de negro y color puro a blanco, *max* varia de negro a color puro y blanco, y *diff* varía de blanco y negro a color puro.

Transformación RGB a HMMD. La transformación de RGB a HMMD es no-lineal pero sencilla como se muestra:

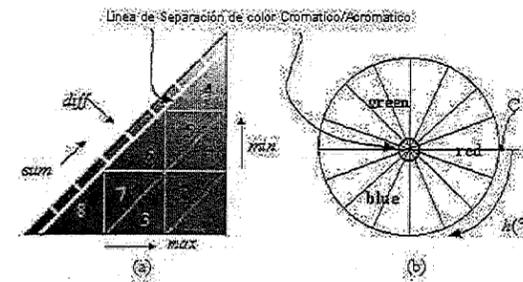


Figura 3.5: Cuantización uniforme de color en el plano MMD (a), y el plano H (b).

Cada rango de *max*, *min*, *diff* y *sum* esta entre 0-1 y *h* esta entre 0 y 360, entonces:

$$max = Max(r, g, b), \quad (3.4)$$

$$min = Min(r, g, b), \quad (3.5)$$

$$diff = max - min, \quad (3.6)$$

$$sum = (max + min)/2, \quad (3.7)$$

Si *diff* = 0, *h* esta indefinido (para colores acromáticos), de otra manera:

$$h = \begin{cases} \frac{g-b}{max-min} 60 & si (r = max \wedge (g - b) > 0) \\ \frac{g-b}{max-min} 60 + 360 & si (r = max \wedge (g - b) < 0) \\ \left(2.0 + \frac{b-r}{max-min}\right) 60 & si (g = max) \\ \left(4.0 + \frac{r-g}{max-min}\right) 60 & si (b = max) \end{cases}$$

Cuantización de color HMMD. El modelo HMMD esta uniformemente cuantizado por cada uno de sus parámetros *hue*, *max*, *min* y *diff* como se ilustra en la Figura 3.5. El espacio primero es dividido en regiones cromática y acromática, de manera que estas regiones son tratadas diferente. La región acromática es teóricamente la línea de *diff* = 0, pero esta región puede ser extendida un cierto rango tanto como el ojo humano lo permita. Las dos regiones se dividen por la línea de *diff* = *Umbralgris* de manera que la región cuyo componente *diff* es menor que *Umbralgris* es acromática y la región donde *diff* es mayor que *Umbralgris* es la región cromática. La línea de separación denominada *línea de separación cromática/acromática* en la Figura 3.5.

Comparado con la cuantización en el plano SV del modelo HSV, la cuantización en el plano MMD corresponde a líneas rectas a diferencia de las curvas del plano SV, la cuantización en el plano MMD es trivial que corresponde a complicados vectores de cuantización en el plano SV.

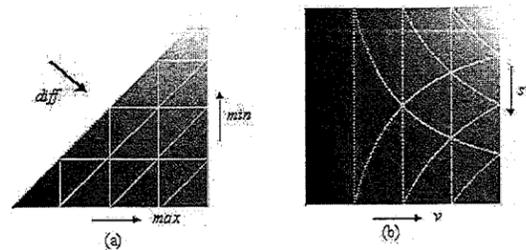


Figura 3.6: (a) Cuantización escalar uniforme en el plano MMD, (b) Las líneas rectas en (a) corresponden a curvas en el plano SV.

La relación entre los planos SV y MMD se ilustra en la Figura 3.6.

3.2.3. Modelado de la distribución del color de la piel

En la literatura encontramos una gran variedad de modelos de la distribución del color de la piel en diferentes espacios de color. En general en este trabajo solo mencionaremos los métodos existentes de una forma rápida a manera de información general y nos abocaremos a los métodos utilizados en este trabajo que fueron probados y tomados en cuenta de acuerdo a los resultados obtenidos, estos métodos son:

Regiones definidas explícitamente.

Consiste en definir regiones explícitas en el espacio de color adecuado, adoptamos este esquema como primer paso de nuestro sistema ya que se obtienen resultados muy rápidos, computacionalmente hablando, reduciendo así el espacio de búsqueda en la imagen a analizar. Peer[23], Gomez[11] y [12], Fleck[9].

Modelado no paramétrico.

La idea principal del modelado no paramétrico es estimar la distribución del color de la piel con los datos de entrenamiento sin derivar en un modelo explícito del color de la piel, entre estos se encuentran:

- Normalized lookup table.

Varios algoritmos de detección y seguimiento emplean un esquema de segmentación de píxeles de piel basados en histogramas. El espacio de color (generalmente solo la

información de crominancia), es cuantizada en un número de segmentos determinados, cada uno correspondiente a un particular rango de color, en pares, para el caso de dimensión dos, o triadas, para el caso de dimensión tres, referido comúnmente como *lookup table (LUT)*. Cada segmento o bin del histograma almacena el número de veces que un particular color ocurre en el conjunto de entrenamiento. Después, el histograma es normalizado, convirtiendo el histograma en una distribución de probabilidad discreta:

$$P_{piel}(c) = \frac{piel[c]}{Norm}$$

donde $piel(c)$ es el valor de histograma en el bin correspondiente, que corresponde al color c , y $Norm$ es el coeficiente de normalización, que puede ser la suma de todos los valores del histograma o el máximo valor del bin presente. Estos valores normalizados corresponden como veremos al valor de *verosimilitud* de que cada vector de color corresponda a piel.

- Clasificador de Bayes.

Dado un prototipo de cierta clase a analizar w_i , la **probabilidad a priori** de una clase es el conocimiento *a priori* acerca de ésta clase antes de realizar un experimento. La probabilidad a priori de una clase i se expresa como $P(w_i)$, en adelante usaremos una notación más compacta definida como π_i . Las probabilidades de I clases no pueden ser negativas y su suma debe ser la unidad:

- $\pi_i \geq 0, i=1,2,\dots, I$, y
- $\sum_{i=1}^I \pi_i = 1$

Para una variable aleatoria x cuya probabilidad depende de la clase considerada, $p(x | w_i)$ es la **función de densidad de probabilidad condicional** de x dada la clase w_i . Para cada clase w_i se verifica que $\int_x p(x | w_i) = 1$.

Si conocemos π_i y $p(x | w_i)$ de antemano. Supongamos también que disponemos de una medida x . La pregunta clave es: **¿ cómo influye este conocimiento sobre la decisión acerca de la clase a asignar ?**

Probabilidad a priori y Probabilidad Condicional (verosimilitud) se combinan en la regla de Bayes, que indica como el valor observado, x , modifica la decisión basada en π_i , a través de $p(x | w_i)$, introduciendo la **probabilidad a posteriori**, $P(w_i | x)$ que se interpreta como *la probabilidad de que la clase cierta sea w_i dado que el valor observado es x .*

$$P(w_i | x) = \frac{p(x | w_i)\pi_i}{p(x)} \text{ donde } p(x) = \sum_{j=1}^I p(x | w_j)\pi_j \quad (3.8)$$

Para un problema de dos clases con iguales a priori ($p_{i1} = p_{i0} = \pi$),

$$p(x) = p(x | w_1)\pi + p(x | w_0)\pi = \pi(p(x | w_1) + p(x | w_0))$$

por lo que las probabilidades a posteriori tienen la siguiente expresión:

$$P(w_1 | x) = \frac{p(x|w_1)}{p(x|w_1)+p(x|w_0)}$$

$$P(w_0 | x) = \frac{p(x|w_0)}{p(x|w_1)+p(x|w_0)}$$

En cualquier caso,

- Si $p(x | w_j) = 0$, $P(w_j | x) = 0$ y $P(w_i | x) = 1$, $j, i = 1, 0$ $j \neq i$
- Si $p(x | w_1) = p(x | w_0)$, $P(w_1 | x) = P(w_0 | x) = 0.5$
- Para todo x , $P(w_1 | x) + P(w_0 | x) = 1$

La regla de clasificación de Bayes es una extensión natural del cálculo de la probabilidad a posteriori: asignar a x la clase para la que su probabilidad a posteriori sea mayor.

Para justificar esta decisión, calculamos la probabilidad de cometer un error en la toma de una decisión. Dado un patrón x cualquiera,

$$P(error | x) = \begin{cases} P(w_1 | x) & \text{si decidimos } w_0 \\ P(w_0 | x) & \text{si decidimos } w_1 \end{cases}$$

Para cada x , minimizamos la probabilidad de error escogiendo w_1 si $P(w_1 | x) > P(w_0 | x)$ y w_0 si $P(w_0 | x) > P(w_1 | x)$. La probabilidad media de error viene dada por la integral

$$P(error) = \int_{-\infty}^{\infty} P(error, x)dx = \int_{-\infty}^{\infty} P(error | x)p(x)dx$$

Así, si para cada x $P(error | x)$ es el mínimo posible, esta integral debe ser la mínima posible, la consecuencia es que la regla de Bayes minimiza la probabilidad media de error. Para un problema de clasificación de dos clases se tendrá que:

Decidir w_1 si $P(w_1 | x) > P(w_0 | x)$ y w_0 en otro caso.

que se conoce como **Regla de Decisión de Bayes**.

Utilizando la expresión de probabilidad a posteriori en función de la probabilidad condicional y la probabilidad a priori (Ecuación 3.8) y teniendo en cuenta que $p(x)$ es

un factor de normalización que sirve para asegurar que $P(w_1 | x) + P(w_0 | x) = 1$, la siguiente regla es equivalente a la anterior:

Decidir w_1 si $P(x | w_1)\pi_1 > P(x | w_0)\pi_1$ y w_0 en otro caso.

Cuando $P(x | w_1) > P(x | w_0)$, x no es capaz de aportar información sobre su clase verdadera y la decisión se toma según los valores de π_1 y π_0 . Por otro lado, cuando $\pi_1 = \pi_0$ la decisión se toma en base a los valores de $P(x | w_i)$, y en estos casos se habla de la *verosimilitud* de x respecto a w_j y la regla de decisión correspondiente:

Decidir w_1 si $P(x | w_1) > P(x | w_0)$, y w_0 en otro caso.

se conoce como **regla de máxima verosimilitud**. En general, ambos factores (probabilidad a priori y probabilidad condicional) son importantes en la toma de decisiones y la Regla de Bayes los combina de forma que se garantiza que se alcanza la mínima probabilidad de error.

■ Self Organizing Map.

Self Organizing Map (SOM) es una de las más populares redes neuronales no supervisadas. En Brown [3] se propone un detector de esta categoría.

Modelado paramétrico.

La necesidad por una representación más compacta de la distribución de la piel en ciertas aplicaciones junto con la ventaja de generalizar e interpolar los datos de entrenamiento generó el desarrollo de modelos paramétricos, además de inconvenientes en eficiencia presentados por los esquemas no paramétricos, así estos métodos se clasifican en:

■ Gausiano Simple.

Es uno de los esquemas utilizados en el desarrollo de este trabajo, el cual se describirá a detalle en la siguiente sección.

■ Mezcla de Gaussianas.

Es un modelo más sofisticado, capaz de describir regiones de distribución complejas, la *función de distribución de probabilidad* es:

$$p(x | w_i) = \sum_{j=1}^k \Pi_j p_j(x | w_i)$$

donde k es el número de componentes, Π_j son los parámetros de la mezcla, que debe cumplir $\sum_{j=1}^k \Pi_j = 1$ y $p_j(x | w_i)$ son las funciones de distribución de probabilidad Gaussianas, cada una con su propio vector de medias y matriz de covarianza. Los detalles de este modelo se pueden encontrar en el trabajo de Yang [40].

- Multiple Gaussian Cluster.

Aproximaciones de regiones (clusters) de piel o variantes del algoritmo de k -medias son empleados. Sazonov [35] utiliza este modelo.

- Elíptico Boundary Model.

Es una variación al modelo Gaussiano simple, propuesta por Lee [19], el cual es, según Lee, igualmente rápido y simple de entrenar y evaluar, dando resultados superiores en la detección.

Modelos dinámicos.

El modelo utilizado puede ser menos general, más específico, entonado para una persona concreta, cámara o condiciones de luz. La fase de inicialización es posible cuando la región del rostro es fácilmente discriminable o se selecciona manualmente. Es óptima para las condiciones dadas. El modelo debe ser capaz de adaptarse a las condiciones de iluminación, cuando la distribución del color varía en el tiempo. Este modelo debe ser rápido en la fase de entrenamiento y clasificación y capaz de adaptarse por sí mismo cuando las condiciones cambien.

Una referencia más detallada de los métodos de modelado de la distribución del color de la piel existentes y los espacios de color utilizados se presenta en el trabajo de Sazonov [35].

3.3. Esquema de segmentación basado en RGB, HSV y HM-MD.

3.3.1. Clasificador.

Una función discriminante para la clase i , definida como $g_i(X)$, tiene la propiedad de que $g_i(X)$ alcanza un valor mayor que cualquier otra función discriminante $g_j(X)$, $i \neq j$, si X es una variable que pertenece a una región de decisión asociada a la clase w_i . Un clasificador basado en la Regla de Bayes puede expresarse fácilmente de esta manera, para el caso del mínimo error

se tiene el conjunto de funciones discriminantes:

$$g_i(X) = P(X | w_i)$$

de forma que el mayor valor discriminante se alcanza para la clase con mayor probabilidad a posteriori.

La elección de un conjunto de funciones discriminantes no es única. Considerar la siguiente propiedad: Si f es una función monótona y creciente y $g_i(X)$ es un conjunto de funciones discriminantes, sustituir $g_i(X)$ por $f(g_i(X))$ proporciona el mismo resultado. Así, los siguientes conjuntos de funciones discriminantes son equivalentes:

$$g_i(X) = P(w_i | X) \quad (3.9)$$

$$g_i(X) = \frac{P(X | w_i)\pi_i}{\sum_{j=1}^J P(X | w_j)\pi_j} \quad (3.10)$$

$$g_i(X) = P(X | w_i)\pi_i \quad (3.11)$$

$$g_i(X) = \log P(X | w_i) + \log \pi_i \quad (3.12)$$

la elección de uno en particular se hará en base a criterios computacionales.

Esta última ecuación, Ecuación 3.12, es la que se tomará en cuenta en el análisis posterior, ya que es barata computacionalmente hablando.

Si $p(X | w_i) \sim N(\mu_i, \Sigma_i)$, la Ecuación 3.12 puede evaluarse fácilmente sustituyendo en esta ecuación la expresión de $p(X | w_i)$ así:

$$g_i(X) = -\frac{1}{2}(X - \mu_i)^T \Sigma_i^{-1}(X - \mu_i) - \frac{d}{2} \log 2\pi - \frac{1}{2} \log |\Sigma_i| + \log \pi_i \quad (3.13)$$

que es la expresión de la forma general de las funciones discriminantes cuando se asume funciones de densidad de probabilidad normal. El término constante $-\frac{d}{2} \log 2\pi$ lo podemos descartar. Este esquema es importante debido a las metas impuestas que incluyen construir un sistema eficiente y rápido computacionalmente, auxiliar en la tarea de detección de rostros.

3.3.2. Diseño del clasificador.

Por supuesto al optar por un clasificador que resuelva nuestro problema debemos actuar con cautela. Disponemos de un conjunto de prototipos y adoptamos un esquema de aprendizaje

parámetro. Así, el diseño de un clasificador paramétrico requiere, de tres etapas:

1. Análisis del conjunto de aprendizaje.
2. Aprendizaje.
3. Clasificación.

Análisis del conjunto de aprendizaje.

Disponemos una muestra inicial de 480000 píxeles clasificados como piel, sabemos que la meta final en la regla de decisión, es discriminar adecuadamente entre los píxeles de piel y no-piel. Una forma rápida es definir regiones de piel explícitamente, lo cual resulta muy rápido computacionalmente. La principal dificultad es lograr altos porcentajes de detección, sin antes también haber escogido un espacio de color adecuado y reglas de decisión adecuadas, existen varios autores que proponen regiones explícitas, Peer 2003 [23] y Fleck 2002 [9]. Recientemente se han propuesto métodos que usan algoritmos de aprendizaje complejos que encuentran reglas de decisión simples con altos porcentajes de detección, Gomez y Morales 2002 [11]. Sin embargo estos métodos requieren de una mayor análisis y a veces resultan algo engorrosos.

Después de un amplio análisis y observaciones minuciosas en este trabajo se definieron y utilizaron regiones explícitas en el modelo de color RGB y un posterior análisis que asigna una probabilidad de ser piel a un píxel, denominado *mapa de veracidad*, utilizando un modelo paramétrico. El primer paso de este esquema es:

- Clasificar un píxel como probable si el componente R, del modelo RGB, es mayor que los componentes G y B, como podemos ver en la Figura 3.7 (a), estos límites son dos planos, ambos que pasan por el origen, uno con vector normal $(1, -1, 0)$, y el otro con vector normal $(1, 0, -1)$, donde si el píxel está de un lado u otro del plano en el espacio, se considera probable o no.
- Después se definen dos límites superior e inferior en el plano GB de este modelo de color, el límite superior está definido por la línea con pendiente 1.10 que pasa por el punto $(0.0, 0.3)$, de este plano y el límite inferior por la línea con pendiente 1.20 que pasa por el punto $(0.0, 0.24)$, Figura 3.7 (b). Este esquema de selección es mucho más rápido computacionalmente, comparado a los resultados de Gomez [12], dando muy buenos resultados dentro de los métodos de regiones explícitas.

Esto implica realizar un ajuste de la mejor línea recta a los puntos en este plano, regresión lineal, usando la técnica de ajuste lineal por mínimos cuadrados. Teniendo esta recta, empíricamente movemos la coordenada y de esta línea de ajuste, de manera que cuando cierto porcentaje de puntos del conjunto estén por arriba o por abajo de la región en el plano GB se definan el límite superior e inferior respectivamente. El porcentaje de puntos arriba o abajo de esta línea fue de 99%. La pendiente también se modificó ligeramente para lograr el porcentaje de puntos mencionado.

- El siguiente paso es utilizar un modelo paramétrico para calcular el mapa de veracidad, al final conservamos regiones grandes de píxeles llenando huecos pequeños si los hay y eliminar píxeles aislados mal clasificados utilizando un filtro de mediana en este mapa umbralizado.

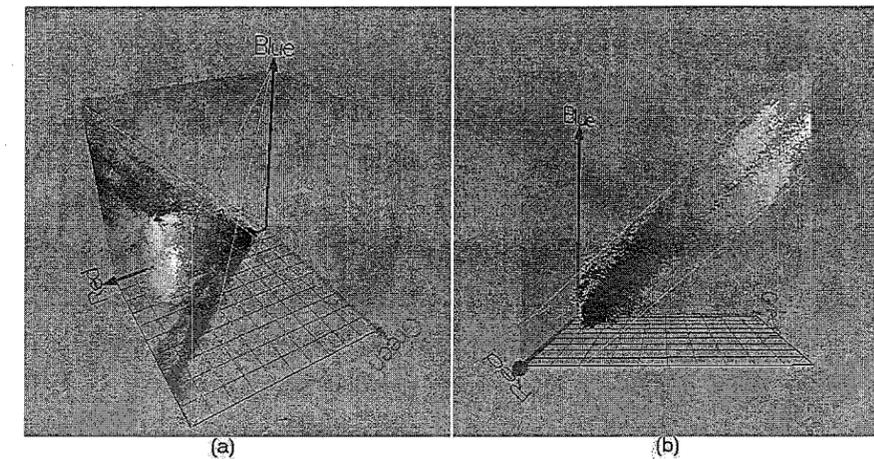


Figura 3.7: (a) La componente R es mayor que las componentes G y B. (b) Límites superior e inferior en el plano GB

Aprendizaje.

Desde un punto de vista paramétrico, habiendo analizado el conjunto de aprendizaje, escogemos un espacio de representación más adecuado, la elección del modelo HSV, se basa en el análisis y observación hechas en el conjunto de aprendizaje, además de su amplio uso en trabajos similares como Zarit [42], Sigal[30], Fleck [9], etc, seleccionado el espacio de representación hacemos la estimación de los parámetros que determinan la función de densidad y de las probabilidades a priori.

Para la estimación de los vectores medios μ_i ($c_i = piel$) los cuales determinamos con los siguientes estimadores:

$$\hat{\mu}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} X_j \quad (3.14)$$

para el vector medio, y

$$\hat{\Sigma}_i = \frac{1}{N_i - 1} \sum_{j=1}^{N_i} (X_j - \mu_i)(X_j - \mu_i)^T \quad (3.15)$$

para la matriz de covarianza, donde N_i es el número de prototipos de la clase i y X_j es el j -ésimo prototipo de esa clase.

Los parámetros obtenidos de este análisis fueron:

$$\mu_{piel} = \begin{pmatrix} hue = 20^\circ \\ sat = 0.45 \end{pmatrix} \quad (3.16)$$

$$\Sigma_{piel} = \begin{pmatrix} 25 & 0.5 \\ 0.5 & 0.25 \end{pmatrix} \quad (3.17)$$

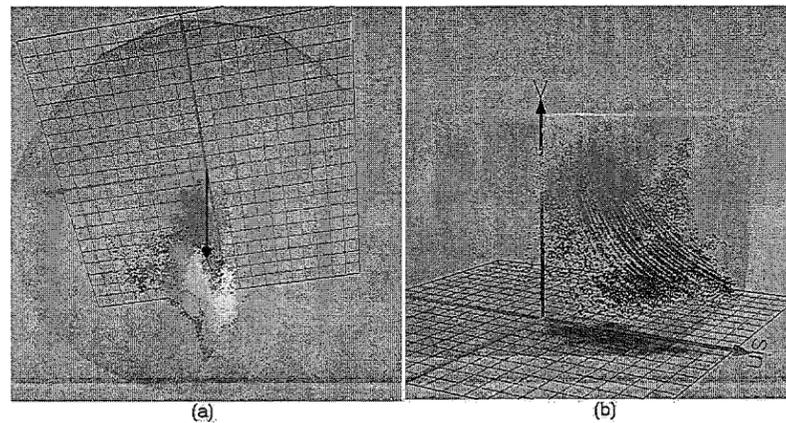


Figura 3.8: Modelo HSV (a) Desde el punto de vista adecuado se puede observar una región compacta en el plano HS. (b) Comportamiento semicircular de la región de piel en el plano SV.

Sin embargo, como se observa en la Figura 3.8, se presenta un comportamiento regular, parametrizable, en el plano SV del modelo HVS, es decir, a mayor valor de la componente Val la componente Sat va disminuyendo, así podemos aproximar un eje promedio mediante un semicírculo cuyo centro son las coordenadas (1, 1) y un radio que establecerá el valor promedio

de la componente Sat o aproximar una línea recta con pendiente (-2.5) que pasa por el punto (0.3, 1.0) de este plano, esto mediante regresión lineal, Figura 3.9. Esto lo podemos formular de la siguiente manera:

$$\mu_{Sat} = \begin{cases} 1 - \sqrt{radio^2 - (1 - Val)^2} & \text{si } Val < radio \\ 0.85 & \text{si } Val = 0.85 \end{cases} \quad (3.18)$$

con $radio = 1 - Sat$ con una aproximación de una línea semicircular, y

$$\mu_{Sat} = \begin{cases} 1 - \sqrt{radio^2 - (1 - Val)^2} & \text{si } Val < radio \\ 0.85 & \text{si } Val = 0.85 \end{cases} \quad (3.19)$$

para una aproximación con una línea recta.

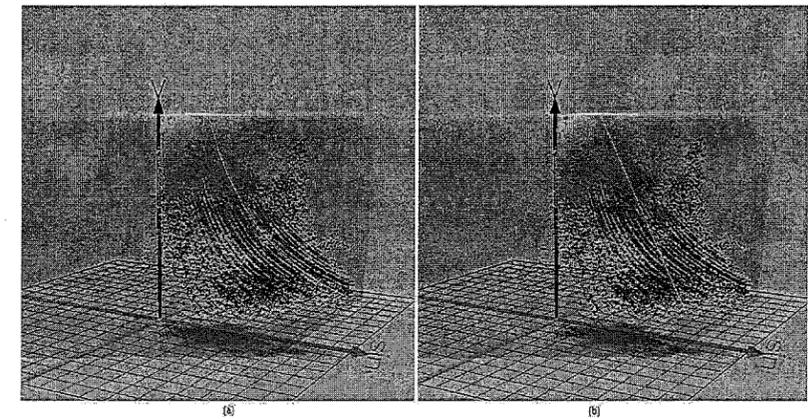


Figura 3.9: Comportamiento de las componentes de piel en el plano SV, (a) aproximación por una línea semicircular, (b) aproximación por una línea recta.

Un esquema alternativo, ahora en el modelo HMMD, es utilizar las componentes de hue y $diff$, ya que como se vio en la Sección 3.2.2, Figura 3.6, este comportamiento es lineal en el plano MMD, es decir la variación de la componente $diff$ equivalente a la saturación de color, se encuentra más concentrada a lo largo de la línea sum que es una región más compacta, Figura 3.10, los parámetros obtenidos de este análisis fueron:

$$\mu_1 = \begin{pmatrix} hue = 20^\circ \\ diff = 0.40 \end{pmatrix} \quad (3.20)$$

$$\Sigma_1 = \begin{pmatrix} 25 & 0.15 \\ 0.15 & 0.30 \end{pmatrix} \quad (3.21)$$

en este trabajo se utilizaron los dos esquemas de los espacios de color mencionados.

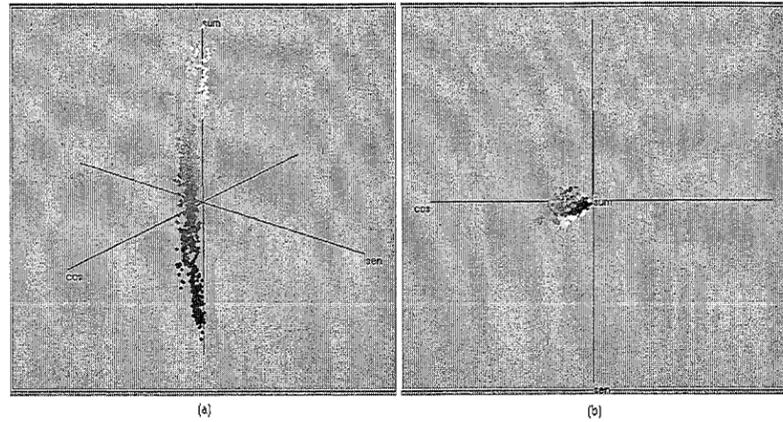


Figura 3.10: Comportamiento de las componentes *hue* y *diff* del modelo *HMMD*, para la región de piel

Las probabilidades a priori se consideran iguales, un pixel con el tono de piel puede ocurrir con la misma probabilidad que un pixel con un tono diferente, así $\pi_{piel} = \pi_{nopiel}$.

Llegar a esta conclusión en el uso de un espacio de color adecuado y un modelado de la distribución de los pixeles piel, requirió de considerable tiempo de investigación, análisis y observación de los resultados obtenidos, además del estudio de varios modelos de color y su comportamiento, de la utilización, análisis e implementación de métodos propuestos en la literatura y su comparación, de la propuesta de esquemas sencillos que auxiliaran en la reducción del tiempo computacional, esto en términos generales, consumió la mayor parte del tiempo dedicado a este trabajo de investigación, donde se logró finalmente mejorar el desempeño del clasificador.

Clasificación.

Una vez estimados los parámetros, la clasificación de nuevos patrones, independientemente de los utilizados para la estimación se hará en base a la función o conjunto de funciones discriminantes impuesto por el modelo adoptado. En cualquier caso, el esquema es similar. Para clasificar un pixel X ,

1. Calcular $g_i(X)$ para $i = 1$ la clase *piel*, la clase 0 -*nopiel*-, la clase de rechazo, es el conjunto de puntos que no están en la región de piel.
2. $d(X) = w_{piel}$ si $g_c(X) > umbral$.

Clasificación paramétrica. Clasificador de mínima distancia de Mahalanobis.

Si el conjunto de patrones sigue una distribución normal, tiende a representarse formando un único agrupamiento de manera que el centro de este agrupamiento está definido por el vector medio y la forma por la matriz de covarianza. La distancia de Mahalanobis es una métrica que considera la distinta dispersión de las variables en el espacio, lo cual es adecuado y barato computacionalmente en su adopción, así asumiendo que las probabilidades a priori (π_i) son iguales, no son significativas para $g_i(X)$, y dado que solo analizamos una sola clase podemos excluir el término $\frac{1}{2} \log |\Sigma|$, finalmente se obtiene la siguiente función discriminante:

$$g_i(X) = -(X - \mu_i)^T \Sigma_i^{-1} (X - \mu_i), \quad (3.22)$$

donde $(X - \mu_i)^T \Sigma_i^{-1} (X - \mu_i)$ es la distancia de Mahalanobis (al cuadrado) de X a μ_i . La regla de decisión óptima se formula de la siguiente manera:

Para clasificar un pixel X , calcular $(X - \mu_i)^T \Sigma_i^{-1} (X - \mu_i)$, la distancia de Mahalanobis (al cuadrado) de X a la clase (i) asignándole la etiqueta si es menor a un umbral predefinido.

Dicho de otra forma, si $\delta_{M^2}(X, \mu_i)$, la distancia de Mahalanobis (al cuadrado) de X al centro de la clase *piel*,

$$d(X) = w_{piel} \text{ si } \delta_{M^2}(X, \mu_i) < umbral, \quad (3.23)$$

donde el umbral se define a continuación y $\delta_{M^2}(X, \mu_i) = (X - \mu_i)^T \Sigma_i^{-1} (X - \mu_i)$, hablando de una sola clase. Note que minimizar (3.22) equivale a minimizar (3.13) con respecto a X para Σ_i y μ_i fijas, cambiándole el signo.

Asignando a la clase de rechazo.

Usando las reglas de clasificación basadas en funciones discriminantes

$$d(X) = w_c \text{ si } g_c(X) = \max_{i=1, \dots, K} \{g_i(X)\}$$

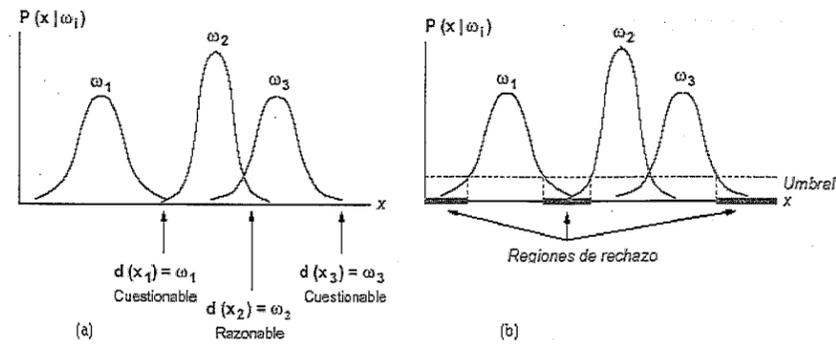


Figura 3.11: Umbral de densidad de probabilidad para el rechazo.

conducen inevitablemente, a que cada uno de los patrones se etiquetan con una de las K clases disponibles. Sin embargo, es deseable que algunos patrones deban ser descartados, o mejor dicho, asignados a la *clase de rechazo* w_{nopei} . Por ejemplo para un problema de clasificación unidimensional de tres clases distribuidas normalmente con iguales probabilidades a priori. La clasificación se hace en base a la clase para la que se obtiene el máximo valor de densidad de probabilidad, Figura 3.11 (a). En este ejemplo, etiquetar x_1 y x_3 resulta arriesgado ya que la máxima densidad de probabilidad es demasiado baja. A diferencia, x_2 puede etiquetarse con cierta garantía ya que la densidad de probabilidad máxima es suficientemente alta.

Los puntos dudosos, aquellos en los que la densidad de probabilidad no es lo suficientemente alta, se deben asignar a la clase de rechazo, w_0 , mediante una técnica de *umbralización*. En la Figura 3.11 (b), se indica como se establece, gráficamente, el umbral de probabilidad de corte. En este caso, el valor umbral es común para todas las clases de manera que si U es este valor umbral, la regla de clasificación puede formularse como:

Sea w_c tal que $P(x | w_c) = \max_{i=1, \dots, K} \{P(x | w_i)\}$

$$d(X) = \begin{cases} w_c & \text{si } P(x | w_c) > U \\ w_0 & \text{si } P(x | w_c) \leq U \end{cases} \quad (3.24)$$

Esta regla requiere establecer el umbral U , pero este valor no puede fijarse *ad-hoc*. Además, en la práctica, los umbrales se aplican a funciones discriminantes más que a densidades de probabilidad.

La función discriminante que proporciona el mínimo error de clasificación cuando las clases siguen una distribución normal viene dado por la Ecuación 3.13 y la regla de decisión asociada,

con la clase de rechazo, puede escribirse:

Sea w_c tal que $g_c(X) = \max_{i=1, \dots, K} \{g_i(X)\}$

$$d(X) = \begin{cases} w_c & \text{si } g_c(X) > U_c \\ w_0 & \text{si } g_c(X) \leq U_c \end{cases} \quad (3.25)$$

donde U_c es el umbral para asignar a la clase $w_{c=piel}$. La estimación de U_c puede hacerse como sigue.

Una clasificación es aceptable ($d(X) = w_c$) si

$$-\frac{1}{2}(X - \mu_c)^T \Sigma_c^{-1}(X - \mu_c) - \frac{1}{2} \log |\Sigma_c| > U_c$$

o lo que es lo mismo, si

$$(X - \mu_c)^T \Sigma_c^{-1}(X - \mu_c) < -2U_c - \log |\Sigma_c| \quad (3.26)$$

La parte izquierda de la Ecuación 3.26 sigue una distribución χ^2 con d grados de libertad, donde el número de variables $d=2$, cuando X esta normalmente distribuida.

El procedimiento a seguir es consultar la tabla χ^2 para determinar el valor de la forma cuadrática $(X - \mu_c)^T \Sigma_c^{-1}(X - \mu_c)$ por debajo del cual hay un determinado porcentaje de puntos.

Clasificación no paramétrica. Estimadores basados en kernel.

Considerando que el clasificador de Bayes es el clasificador que proporciona el mínimo error, la formulación dada por:

$$P(w_i | x) = \frac{p(x | w_i)\pi_i}{p(x)} \quad \text{donde } p(x) = \sum_{j=1}^I p(x | w_j)\pi_j \quad (3.27)$$

el numerador de esta ecuación puede usarse para construir un conjunto de funciones discriminantes para cada clase.

El objetivo final es etiquetar un patrón X utilizando el siguiente conjunto de funciones discriminantes:

$$g_i(X) = P(X | w_i)\pi_i$$

donde no se supone nada acerca de la forma funcional de $P(X | w_i)$. Tan solo disponemos de un conjunto de prototipos, T , y a partir de él debemos estimar tanto el valor de $P(X | w_i)$ como el de π_i .

El cálculo de las probabilidades a priori es relativamente sencillo: basta con utilizar las frecuencias relativas de cada clase en el conjunto de prototipos o, simplemente, no considerarlas, asumiendo que todas son iguales. El cálculo de la densidad de probabilidad es más complejo, y puede abordarse desde un punto de vista geométrico. La cuestión es: ¿De qué forma puede utilizarse la información proporcionada por el conjunto de prototipos para inferir, a partir de éste, el valor de la función de densidad de probabilidad mediante una interpretación geométrica?. Si fijamos un volumen v en P y consideramos varias regiones en este espacio, resulta evidente que

- En una región en la que $P(X | w_i)$ tiene un valor bajo, la probabilidad de encontrar un patrón de clase w_i es pequeña.
- A la inversa, hay una alta probabilidad de encontrar un patrón de clase w_i en una región que tiene asociada un alto valor de la función de densidad.

De manera informal se podría decir que dado un conjunto de patrones, hay una alta probabilidad de encontrar un patrón en una región densamente poblada y baja en regiones poco pobladas en las que las observaciones están más dispersas.

Los estimadores de Parzen generan información acerca de $P(X | w_i)$ proporcionada por cada prototipo, individualmente, del conjunto de prototipos. Antes de nada, introduciremos una nueva notación: mediante Z_i^m nos referiremos al m -ésimo prototipo de la clase w_i . Así, dado un prototipo de clase w_i , Z_i^m , podemos afirmar:

- $P(Z_i^m | w_i) > 0$
- Bajo la suposición de continuidad de $P(X | w_i)$, ésta tomará valores positivos y distintos de cero en una inmediata vecindad de Z_i^m .
- Cuanto más nos alejemos de Z_i^m menos puede afirmarse sobre $P(X | w_i)$ basándonos únicamente en Z_i^m .

Con estas afirmaciones, fácilmente asumibles, podemos concluir que la información acerca de $P(X | w_i)$ obtenida a partir de Z_i^m puede representarse mediante una *función kernel* (o

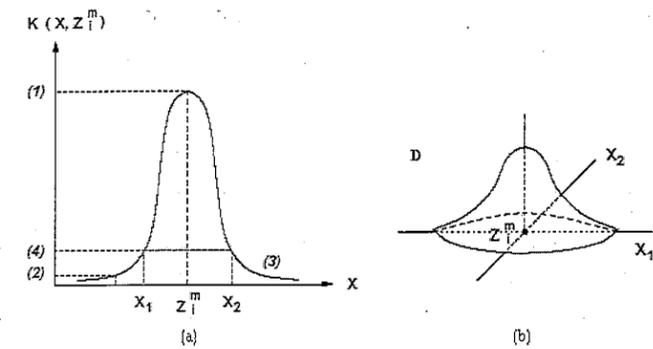


Figura 3.12: (a) Función que puede considerarse un kernel al responder a las cuatro propiedades enumeradas, (b) Kernel Gaussiano en 2D.

simplemente *kernel*) definida como: Un *kernel*, al que notamos por $K(X, Z_i^m)$, es una función centrada en el punto Z_i^m , que alcanza un máximo en él y que decrece monótonamente conforme la distancia $\delta(X, Z_i^m)$ se incrementa. Por concretar aun más, las características deseables de K deberán ser las siguientes:

- $K(X, Z)$ debería alcanzar el máximo para $X = Z$.
- $K(X, Z)$ debería ser aproximadamente cero para valores de X distantes de Z .
- $K(X, Z)$ debería ser una función suave (y continua) y decrecer monótonamente conforme aumenta la distancia $\delta(X, Z)$.
- Si $K(X_1, Z) = K(X_2, Z)$, X_1 y X_2 deberían tener el mismo grado de similitud con Z .

Una vez establecida la aportación de cada prototipo, hay que considerar cómo puede conjugarse la información que proporcionan todos los prototipos de una clase. Un *estimador* de $P(X | w_i)$ basado en kernel es:

$$\hat{P}(X | w_i) = \frac{1}{N_i} \sum_{m=1}^{N_i} K(X, Z_i^m) \quad (3.28)$$

o lo que es lo mismo, el estimador de la densidad de probabilidad de X supuesto que es de clase w_i es la media de las aportaciones individuales obtenida al evaluar la función kernel sobre los N_i prototipos de esa clase.

Un aspecto muy importante a considerar es que la contribución en un punto dado al

estimador $\hat{P}(X | w_i)$ depende del *rango de influencia del kernel* (también llamado *ancho del kernel*). Esto se traduce en que el estimador es muy dependiente de los datos disponibles:

- Cuando las muestras están muy dispersas, el rango del kernel debería ser relativamente grande.
- Cuando están agrupadas, el rango del kernel debería ser menor para considerar tan solo la inmediata vecindad de los prototipos.

Forma general de una función kernel.

Antes de describir los kernel usados habitualmente, debemos estudiar cómo debería ser la forma funcional de una función kernel. Devijver y Kittler [6] proponen que una función kernel debe ser de la siguiente manera:

$$K(X, Z_i^m) = \frac{1}{\rho^d} h \left[\frac{\delta(X, Z_i^m)}{\rho} \right] \quad (3.29)$$

donde:

- ρ es un parámetro del estimador, estrictamente positivo, que satisface:

$$\lim_{N_i \rightarrow \infty} \rho^d(N_i) = 0. \quad (3.30)$$

Esta condición sugiere que el ancho del kernel depende, en última instancia, del tamaño del conjunto de aprendizaje. Cuanto mayor sea el número de prototipos, menor será el ancho del kernel.

- $\delta(X, Z_i^m)$ es una métrica definida sobre P . Como veremos, la métrica puede estar determinada por el kernel que se vaya a emplear.
- $h[\cdot]$ es una función que alcanza un máximo cuando $\delta(X, Z_i^m) = 0$ y es monótona decreciente conforme $\delta(X, Z_i^m)$ aumenta.

Si se exige que $h[\cdot]$ sea no negativa, la única que se impone es que

$$\int K(X, Z_i^m) dx = 1. \quad (3.31)$$

Puede demostrarse, Devijver y Kittler, que las condiciones impuestas por 3.30 y 3.31 garantizan

que $\hat{P}(X | w_i)$ dado por

$$\hat{P}(X | w_i) = \frac{1}{N_i} \sum_{j=1}^{N_i} K(X, Z_i^j)$$

es una *función de densidad de probabilidad* y proporciona una estimación *consistente e insesgada* de $P(X | w_i)$.

Las funciones kernel más empleadas, habitualmente, son:

- Kernel Gaussiano.
- Kernel Hipercúbico.
- Kernel Hiperesférico.

Kernel Gaussiano.

La forma funcional de una función Kernel Gaussiano es la siguiente:

$$K(X, Z_i^m) = \frac{1}{(2\pi)^{d/2} |\sigma|^{1/2}} \exp \left(-\frac{1}{2} (X - Z_i^m)^T \Sigma^{-1} (X - Z_i^m) \right) \quad (3.32)$$

en esta ecuación el *ancho* está determinado por la matriz de covarianza Σ .

En su forma más general es posible tener en cada dimensión un ancho diferente y considerar núcleos en los que se contemple la correlación entre las variables. Por esta razón se toma la distancia de Mahalanobis entre el patrón y el prototipo considerado.

Para reducir el costo computacional suele tomarse una matriz de covarianza diagonal. Esto implica la suposición de correlación nula y por lo tanto, una simplificación importante en el cálculo de la distancia, que se transforma en una distancia Euclídeana. La Ecuación 3.32 se transforma en

$$K(X, Z_i^m) = \frac{1}{(2\pi)^{d/2} |\sigma|^{1/2}} \exp \left(-\frac{1}{2} (X - Z_i^m)^T \Sigma^{-1} (X - Z_i^m) \right) \quad (3.33)$$

en la que Z_i^m es la componente j -ésima del prototipo m -ésimo de la clase w_i .

Ancho del kernel.

El parámetro ρ que venimos utilizando indica el *ancho* o *rango de influencia* del kernel. Para resolver cuál es el ancho del kernel apropiado para un problema determinado, Devijver y Kittler, proponen un método que depende del tamaño del conjunto de prototipos. Un valor de

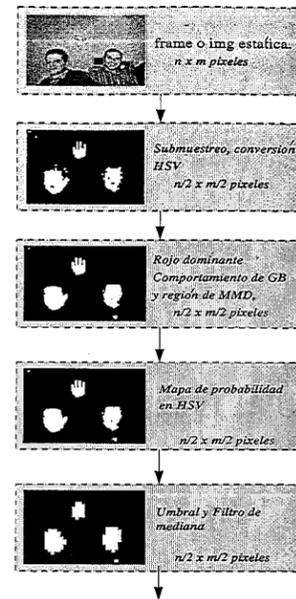


Figura 3.13: Esquema general de Segmentación de Color

ρ es adecuado si satisface la condición:

$$\lim_{N_i \rightarrow \infty} \rho^d(N_i) = 0$$

En particular, se puede tomar

$$\rho^d(N_i) = N_i^{-\frac{\eta}{d}}$$

donde η es un valor del intervalo (0, 1).

Aunque pareciera que el tipo de kernel a adoptar es el factor determinante de la bondad de la estimación, diversos autores (véase, por ejemplo [6]) reconocen que el valor del ancho del kernel es mucho más importante. Este valor puede ser común a todas las clases o diferente, siendo la última la mejor opción.

Resultados utilizando funciones kernel Gaussianas.

Los resultados fueron muy similares comparados con la segmentación o clasificación paramétrica, el problema principal al que nos enfrentamos fue obtener los prototipos de la clase adecuados para obtener resultados admisibles con el compromiso de no elevar el número de estos prototipos para no aumentar sobremedida o exagerar el tiempo computacional empleado.

La estrategia utilizada para obtener mejores prototipos, partiendo de una imagen de $m \times n$ pixeles donde todos los pixeles están clasificados como piel, es decir, nuestro conjunto de aprendizaje, tomar una media y matriz de covarianza por subimagen tomadas a cada k intervalos. El número de prototipos se calcula de la siguiente manera: $N_{pr} = \frac{m}{k} \times \frac{n}{k}$, para una imagen de 256×256 pixeles y $k=8$, el número de prototipos es $\frac{256}{8} \times \frac{256}{8} = 32 \times 32 = 1024$ prototipos. El número de prototipos dependerá del parámetro k , mientras más grande k menor número de prototipos y viceversa. El ancho del kernel se tomó en función de las varianzas.

Para 1024 prototipos el tiempo computacional en la clasificación de pixeles de piel fue mayor a un segundo para una imagen de 320×240 pixeles. De esto concluimos que el número de prototipos deberá ser menor a fin de reducir el tiempo computacional, nuestro experimento final consistió de 256 prototipos de la clase piel. Debido a que los resultados obtenidos fueron muy similares, se continuó empleando una clasificación paramétrica.

3.3.3. Esquema de segmentación de color.

El esquema seguido se describe a continuación: dada una imagen el primer paso es submuestrear esta, tomando un pixel si y uno no, como resultado tenemos una imagen de 1/4 de tamaño de la imagen original, posteriormente se realiza la conversión al espacio de color HSV y HMMD. En la siguiente etapa se analiza el comportamiento de las componentes de color en RGB y HMMD como se describió anteriormente, después, para cada pixel en la región descrita, se evalúa el clasificador para decidir si el pixel pertenece a la clase piel, finalmente se aplica un filtro de mediana a la imagen binaria resultante para eliminar pixeles aislados o llenar huecos pequeños. Este esquema se muestra en la Figura 3.13.

3.3.4. Algoritmo de segmentación de color

El algoritmo de clasificación lo podemos resumir de la siguiente manera:

1. Inicialización.

$I_{rgb} \leftarrow$ Imagen a color

$Chsv(k)(i, j) \leftarrow 0$, para $k=0,1,2$, arreglo de conv. a HSV

$Chmmd(k)(i, j) \leftarrow 0$, para $k=0,1,2$, arreglo de conv. a HMMD

$M_{eti}(i, j) \leftarrow 0$, arreglo bidimensional de etiquetas por cada pixel de la imagen

$M_{prob}(i, j) \leftarrow 0$, arreglo bidimensional de etiquetas

$filas \leftarrow Irgb.Height$ no. de filas de la imagen
 $cols \leftarrow Irgb.Width$ no. de columnas de la imagen
 $hue \leftarrow 0.45$, valor promedio de comp hue
 sat , valor promedio de comp sat
 $\mu_1 \leftarrow [15, 0.40]$, vector, media de la clase piel, componentes hue y sat
 $M_{cvar} \leftarrow [25, 0.5; 0.5, 0.25]$, matriz de covarianza, componentes hue y sat

2. Cambio de espacio de color

Por cada pixel impar ($i \bmod 2 = 1, j \bmod 2 = 1$) de la imagen

$Chsv \leftarrow conversion.HSV(Irgb)$

$Chmmd \leftarrow conversion.HMMD(Irgb)$

3. Selección de pixeles más probables.

Por cada pixel (i, j) de la imagen

Analizar comportamiento en el plano GB si $R > G$ y $R > B$

Analizar comportamiento en la comp. diff, si $diff > 0.03$ y $diff < 0.58$

$M_{etiq} \leftarrow 1$ si las componentes GB estan entre el límite inferior y superior ya definidos.

4. Mapa de veracidad

Por cada pixel (i, j) de la imagen, si $M_{etiq} = 1$

Calcular el valor medio de sat de acuerdo a (3.18)

$\mu_1 \leftarrow (hue, sat)$ la componente sat obtenida del paso anterior

$M_{prob}(i, j) = \delta_{M^2}(X, \mu_1)$

si $M_{prob}(i, j) < umbral$

$M_{etiq} \leftarrow 0$

5. Aplicar filtro de mediana al arreglo binario M_{etiq}

6. Salida

$Chsv$ imagen en HSV

$Chmmd(i)$ imagen en HMMD

M_{etiq} imagen de etiquetas

M_{prob} mapa de veracidad

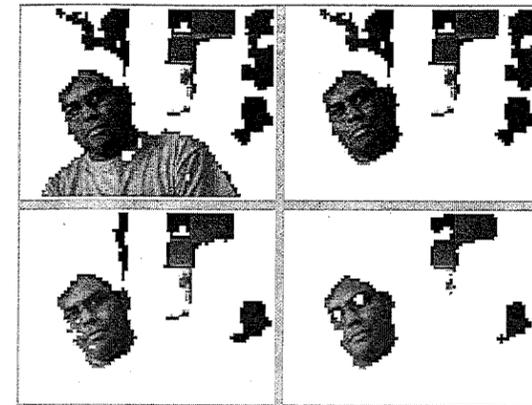


Figura 3.14: Resultados de la fase de segmentación de color (Ver Sección 3.4 para descripción).

3.4. Resultados preliminares de la segmentación de color.

En esta sección presentamos algunos resultados preliminares para algunas imágenes de prueba de la segmentación en la etapa de color, Figuras 3.14 y 3.15, lo que podemos observar en la secuencia de imágenes, de arriba hacia abajo y de izquierda a derecha, es la reducción del espacio de búsqueda cuando se toman aquellos pixeles donde la componente $R > G$ y $R > B$, quedando solo los pixeles con colores en el rango del rojo, imagen superior izquierda, la reducción subsecuente al tomar solo los pixeles que estan dentro de la región definida en el plano GB, imagen superior derecha, reducción final al tomar aquellos pixeles dentro de la región definida en la componente $diff$ del modelo HMMD, imagen inferior izquierda, y la segmentación de color final al umbralizar el mapa de veracidad, imagen inferior derecha.

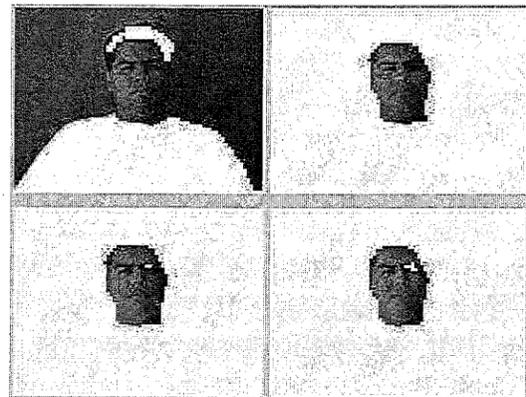


Figura 3.15: Resultados de la fase de segmentación de color (Ver Sección 3.4 para descripción).

Capítulo 4

Estrategias para la localización y separación de regiones.

La etapa subsecuente una vez hecha la segmentación de color es la localización o ubicación de éstas regiones dentro de la imagen y la separación que en cada caso corresponda, en el caso de que alguna de estas regiones representa o contiene dos o más regiones que puedan ser discriminables entre si, esto debido a que la representación de color es una característica de bajo nivel, ya que hasta este punto solo hemos analizado cada uno de los pixeles por separado, lo cual no es suficiente para hacer la discriminación o clasificación como rostros. Esta segmentación únicamente captura aquellos pixeles que están más cercanos al color de la piel propuesto, pero no su semántica, como resultado existen ambigüedades en los pixeles que conforma las regiones de piel, de hecho hasta este punto no se puede hablar ciertamente de regiones en sí, si no de pixeles etiquetados como piel, es el paso siguiente el que determinará si estos pixels conforman un región y si la región una vez delimitada puede clasificarse dentro de la clase de regiones catalogadas como Rostros.

El esquema de procesamiento de esta etapa es el siguiente, Figura 4.1:

Descriptor de color. Como primer paso, utilizando un descriptor de color, la componente *diff* del espacio de color HMMD, cuantizamos en un número de regiones predefinido de manera que regiones de color que hayan quedado juntas pero que representen a dos o más regiones distintas pueda ser diferenciables. La elección de este descriptor y de la manera en que es utilizado, Sección 4.1, se basa en la restricción de tiempo impuesta al sistema, de manera que con los resultados preliminares obtenidos, este esquema resulto el más rápido

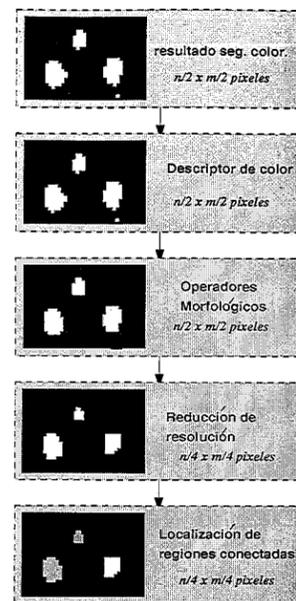


Figura 4.1: Esquema de procesamiento para la detección de regiones más probables

y apropiado para las condiciones dadas. En el caso de imágenes de video conferencia, que como se comentó anteriormente se tienen controlados ciertos factores como la iluminación, etc. este paso no será necesario, evitando con esto un proceso adicional que nos reduce el tiempo de cómputo.

Operadores morfológicos. El segundo paso es limpiar u ordenar en un sentido más general las regiones obtenidas del paso anterior, es decir, pixeles o regiones muy pequeñas de pixeles que hayan sido clasificados dentro de una región particular o mal clasificados y que se encuentran aislados deben ser eliminados o corregidos, utilizando operadores morfológicos en niveles de gris, este procedimiento se describirá en la Sección 2 de este capítulo. Este paso no será necesario para imágenes de video conferencia.

Reducción de la resolución de la imagen. Este paso se realiza con el fin de acelerar la búsqueda de regiones conectadas, la resolución de la imagen sera $m/2 \times n/2$ renglones y columnas, respectivamente, de la imagen de $m \times n$ pixeles del tamaño anterior.

Localización de regiones conectadas. El siguiente paso es uno de los más importantes en esta etapa, es el conteo y localización de las regiones presentes, la estrategia adoptada consiste en un llenado de regiones (floodfill) adaptado a las condiciones de las regiones

de la imagen dando prioridad a regiones convexas, esta estrategia es sumamente rápida además de una reducción adicional en el tamaño del arreglo de etiquetas que nos auxilia a reducir el tiempo computacional. Se aplica un paso final en la búsqueda de regiones, para las imágenes clasificadas como fotográficas, este consiste en aplicar un algoritmo de separación de aquellas regiones para el cual el descriptor de color sea el mismo, no produzca los resultados esperados o cuando simplemente estas regiones sean semejantes pero tengan cada una un significado diferente, este esquema esta basado en un algoritmo de proyección de vectores priorizando aquellas características convexas de una región particular.

En la Figura 4.2 presentamos tres imágenes de prueba, que dentro de la gama de imágenes existentes en nuestro conjunto de prueba son casos complicados, en la imagen a la izquierda se encuentran tres rostros que no estan de frente, a diferentes escalas cada uno de ellos, con elementos en la imagen de fondo con colores muy parecidos al tono de la piel como la puerta de madera, pantalones de vestir y paredes de fondo, en la imagen del centro también encontramos muchos elementos en el fondo de ésta con diversos tonos muy parecidos al color de la piel además de regiones que son piel de otras partes del cuerpo, en la imagen de la derecha presentamos una imagen menos complicada, un rostro de frente con un fondo con tono rosado claramente diferenciable a simple vista pero que también presenta retos interesantes en su solución.



Figura 4.2: Imágenes de prueba

4.1. Estrategias de diferenciación de regiones de piel.

4.1.1. El espacio de color HMMD como indicador de regiones de piel.

En los experimentos base del estandar MPEG-7 para búsqueda y obtención de imágenes en bases de datos multimedia, el espacio HMMD es muy efectivo y favorable comparado con el espacio HSV, Manjunath [20]. Dentro de una gama de descriptores de color soportados por



Figura 4.3: Resultados de la diferenciación de distintas regiones de piel utilizando descriptor de color con el modelo de color *HMMD*

MPEG-7 como descriptores de histograma, de color dominante y descriptor de disposición. El *descriptor de color dominante DCD* nos da la distribución del color dominante en una imagen. A diferencia de la cuantización por segmentos en un histograma, la especificación de color en un descriptor de color dominante está limitado solo por la cuantización del espacio de color. El propósito es proveer una representación efectiva, compacta e intuitiva de colores presente en una región de interés.

Este descriptor nos auxilia a describir globalmente así como localmente la distribución espacial del color en imágenes para búsquedas muy rápidas. A diferencia de histogramas de color, este descriptor proporciona una representación mucho más compacta, con ciertas desventajas como bajo desempeño en algunas aplicaciones. Los colores de una región dada están agrupados en un número pequeño de colores representativos. El descriptor consiste en los colores representativos, sus porcentajes en una región, la coherencia espacial del color, y la variación del color. Esta es más conveniente para representar un objeto o región de la imagen donde un número pequeño de colores es suficiente para caracterizar la información de color de una región de interés, aunque también es aplicable a la imagen completa. La cuantización de color es usada para extraer un número pequeño de colores representativos en cada región/imagen.

Como vimos en el capítulo anterior el modelo de color *HMMD* es muy útil debido a las características en cada uno de sus componentes, un esquema rápido de segmentación de diferentes regiones de piel obtenidas del paso anterior de segmentación es la utilización del modelo de color *HMMD* y en específico de la componente *diff*, ya que obviamente el color dominante en este caso es el color de la piel, así que es necesario una segunda componente que nos ayude a diferenciar entre regiones para resolver el problema. En la Figura 4.3 se muestran los resultados intermedios para las imágenes de prueba de la Figura 4.2.

4.1.2. Morfología matemática para la mejora de subimágenes.

Conceptos básicos sobre morfología matemática.

El lenguaje de la morfología matemática binaria es el de la teoría de conjuntos. Los conjuntos en morfología matemática representan las formas presentes en imágenes binarias o de niveles de gris. El conjunto de todos los píxeles blancos en una imagen en blanco y negro (binaria) constituye una descripción completa de la imagen.

Los puntos en un conjunto sobre los que se aplica la transformación son el conjunto de puntos seleccionado y el complementario el no seleccionado. En las imágenes binarias los puntos seleccionados son los que no pertenecen al fondo.

Las operaciones primarias morfológicas son la erosión y la dilatación. A partir de ellas podemos componer las operaciones de apertura y cerradura. Son estas dos operaciones las que tienen mucha relación con la representación de formas, la descomposición y la extracción de características. Antes de mencionar estos conceptos, presentaremos algunas notaciones que se usaran en adelante, el punto 0 es un punto arbitrario en \mathbb{R}^2 al que nos referiremos como el *origen*. Con cualquier punto x en \mathbb{R}^2 , asociamos el vector \vec{Ox} . Sea B un conjunto arbitrario. Denotamos \check{B} el conjunto transpuesto:

$$\check{B} = \{-x \mid x \in B\}$$

Además, para cualquier a en \mathbb{R}^2 , denotamos B_a la traslación del conjunto B por un vector a :

$$B_a = \{a + x \mid x \in B\}$$

Dilatación binaria.

Consideremos un conjunto $X \subseteq \mathbb{R}^2$. La dilatación de X por B es el conjunto de todos los puntos x en \mathbb{R}^2 tales que la intersección entre X y B_x es no vacía. El resultado es el conjunto denotado por $X \oplus \check{B}$:

$$X \oplus \check{B} = \{x \in \mathbb{R}^2, X \cap B_x \neq \emptyset\} \quad (4.1)$$

en Figura 4.4 mostramos un ejemplo de dilatación para una imagen binaria.

Erosión binaria.

La erosión es la operación morfológica dual, de la dilatación. La erosión se define de manera

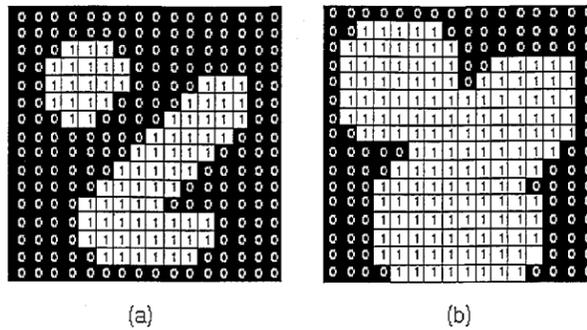


Figura 4.4: Dilatación, (a) Imagen A, (b) resultado de la dilatación con un elemento estructural de 3×3

similar: X erosionado por B se denota $X \ominus \tilde{B}$ es el conjunto x en \mathbb{R}^2 tales que B_x esta totalmente incluido en X :

$$X \ominus \tilde{B} = \{x \in \mathbb{R}^2 \mid B_x \in X\} \quad (4.2)$$

un ejemplo de erosión se encuentra en la Figura 4.5.

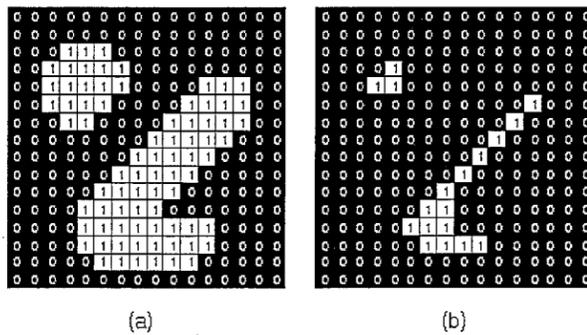


Figura 4.5: Erosión, (a) Imagen A, (b) resultado de la erosión con un elemento estructural de 3×3

El conjunto B , el cual juega un rol particular en estas transformaciones, es llamado el *elemento estructural*. La dilatación y erosión son operadores duales, esto significa que la dilatación de un conjunto X es idéntica al complemento de la erosión del conjunto X^c

Apertura y cerradura.

La erosión y la dilatación usualmente se emplean por pares, bien la dilatación seguida por la erosión o al revés. En cualquier caso, el resultado de esta aplicación sucesiva de erosiones y

dilataciones es una eliminación de detalles menores que el elemento estructural, sin distorsionar la forma global del objeto.

Un aspecto fundamental de esta aplicación sucesiva es el hecho de que la aplicación continuas e intermitentes de erosiones y dilataciones es idempotente. La apertura y la cerradura proporcionan los medios por los cuales seleccionar subformas y superformas de una forma compleja.

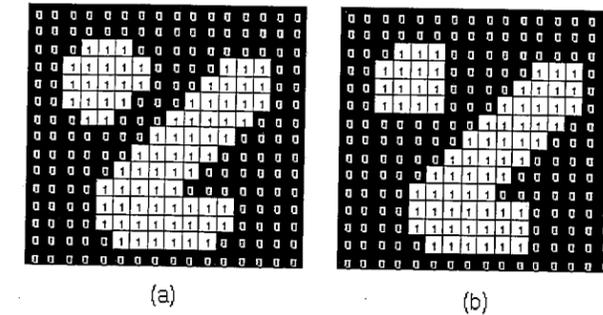


Figura 4.6: Apertura (a) Imagen A, (b) resultado de la apertura con un elemento estructural de 3×3

La apertura de X por un elemento estructural B se denota $X \circ B$, se define como:

$$X \circ B = (X \ominus \tilde{B}) \oplus B$$

que en palabras establece que la apertura de X por B es simplemente la erosión de X por B , seguido de la dilatación del resultado por el conjunto transpuesto \tilde{B} de B . Un ejemplo binario se muestra en la Figura 4.6.

La cerradura de X por un elemento estructural B se denota $X \bullet B$, se define como:

$$X \bullet B = (X \oplus \tilde{B}) \ominus B.$$

En la Figura 4.7 se muestra el ejemplo para una imagen binaria.

Morfología de niveles de gris.

Estos operadores morfológicos pueden fácilmente extenderse a imágenes de niveles de gris, el objetivo es el uso de la morfología de niveles de gris para extraer componentes de la imagen que son útiles en la representación y descripción de las formas. Por motivos de simplicidad, morfologistas comunmente se restringen al uso de un elemento estructural plano. En este caso

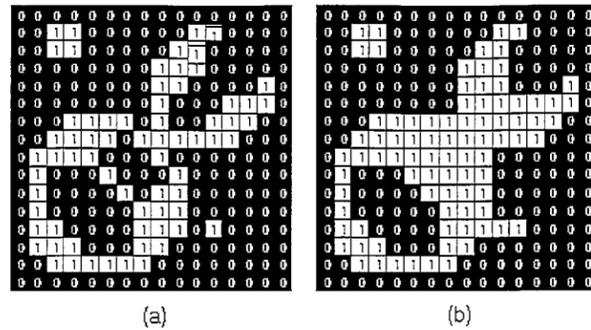


Figura 4.7: Cerradura, (a) Imagen A, (b) resultado de la cerradura con un elemento estructural de 3×3

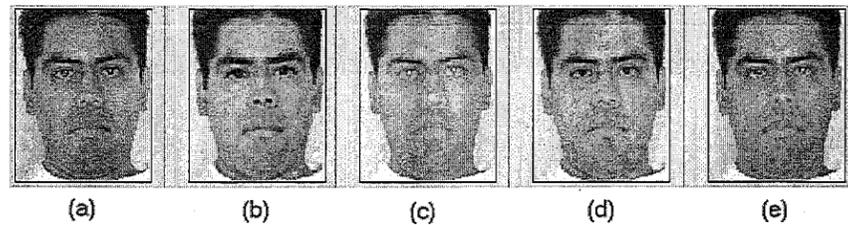


Figura 4.8: Operaciones morfológicas en niveles de gris (a) Imagen original, (b) erosión, (c) dilatación, (d) apertura, (e) cerradura, se utilizó un elemento estructural plano de 3×3 .

particular, la dilatación y erosión de una función f por un elemento estructural plano B es:

$$\forall x \in \mathbb{R}^2 \begin{cases} (f \oplus \tilde{B})(x) = \sup_{y \in B_x} f(y) \\ (f \ominus \tilde{B})(x) = \inf_{y \in B_x} f(y) \end{cases} \quad (4.3)$$

En particular, la dilatación ensancha picos y llena valles mientras que la erosión tiene el efecto opuesto. Las operación de apertura y cerradura también pueden extenderse a niveles de gris. Un efecto remarcable es que la apertura tiende a suprimir los picos, mientras que la cerradura tiende a llenar valles. Estos operadores son muy útiles en la segmentación de regiones, ya que nos auxilian en la eliminación de detalles menores o ruidosos, sin distorsionar la forma global del objeto. Debido a estas ventajas su uso es aplicable en el proceso de segmentación de regiones de piel que claramente son diferenciables. Además de emplear estos operadores morfológicos con el fin descrito anteriormente, también se planea su uso para dar énfasis a las regiones oscuras del rostro como son las cejas y los ojos y así poder determinar su ubicación más fácilmente.

Estos operadores se utilizan para mejorar los resultados de esta etapa, en específico el operador de cerradura en niveles de gris, los resultados para las imágenes de prueba de la Figura 4.2 se muestran en la Figura 4.9.



Figura 4.9: Resultados de la eliminación de detalles menores de la imagen usando morfología en niveles de gris en la diferenciación de regiones

4.2. Esquema de localización y conteo de regiones.

4.2.1. Algunos conceptos básicos sobre regiones conectadas.

Para considerar dos pixeles como conectados, sus valores correspondientes debe ser ambos del mismo conjunto de valores V . En una imagen de grises, V puede ser cualquier rango de niveles de gris, para una imagen binaria simplemente tenemos que $V = \{1\}$. Para formular un criterio de adyacencia para la conectividad, introducimos la notación de *vecindad*. Para un pixel p con coordenadas (x, y) el conjunto de pixeles dado por:

$$N_4(p) = \{(x+1, y), (x-1, y), (x, y+1), (x, y-1)\} \quad (4.4)$$

es llamado *4-vecinos*. La vecindad de *8-vecinos* se define como:

$$N_8(p) = N_4(p) \cup \{(x+1, y+1), (x-1, y+1), (x-1, y-1), (x+1, y-1)\} \quad (4.5)$$

De las definiciones 4.4 y 4.5 podemos inferir la definición de *conectividad-4* o *conectividad-8*.

Dos pixeles p y q , ambos con valores del conjunto V tiene conectividad-4 si $q \in N_4(p)$ y conectividad-8 si $q \in N_8(p)$.

Hablando en términos de conectividad -4, un pixel p esta conectado a un pixel q si p tiene conectividad-4 con q o si p esta conectado un tercer pixel el cual esta conectado a si mismo con q , en otras palabras, dos pixeles p y q estan conectados si existen un camino de p a q en el

cual cada pixel tiene conectividad-4 con el siguiente.

4.2.2. Etiquetado de regiones conectadas.

Un conjunto de pixeles en una imagen que están conectados unos con otros es denominada *componente conectada* o *región conectada*. Encontrar todas las componentes conectadas en una imagen y etiquetarlas se denomina *etiquetado de regiones conectadas*. El etiquetado de regiones o componentes conectados consiste en recorrer la imagen y agrupar pixeles en regiones o componentes basados en su conectividad.

La manera en que se realiza el etiquetado de regiones conectadas es recorrer la imagen, pixel a pixel (de arriba hacia abajo y de izquierda a derecha), para identificar regiones de pixeles conectadas. La implementación puede hacerse para imágenes binarias o imágenes en escala de grises con diferentes medidas de conectividad admisibles. Adoptando un esquema de conectividad-8. El operador de etiquetado de regiones conectadas recorre la imagen moviéndose sobre una fila de pixeles hasta que algún punto p (donde p denota un pixel a ser etiquetado en cualquier etapa del proceso de recorrido) para el cual $V = \{valor_p\}$. Cuando esto es cierto, se examinan los vecinos de p que ya se han encontrado en el recorrido, los pixeles vecinos arriba de p y vecinos a la izquierda de p . Basados en esta información, la etiqueta de p se realiza de la siguiente manera:

- Si los vecinos encontrados son todos 0, asignamos una nueva etiqueta $valor_{p+1}$ a p , si no,
- si solo un vecino tiene $V = valor_p$, asignamos esta etiqueta a p , si no,
- si uno o más vecinos tiene $V = valor_p$, asignamos una de las etiquetas a p y hacemos nota de la equivalencia.

Después de completar el recorrido, los pares de etiquetas equivalentes son clasificadas en clases de equivalencia asignándoles una única etiqueta a cada clase. Como paso final, se realiza un segundo recorrido sobre la imagen, en el cual cada etiqueta es reemplazada por la etiqueta asignada por su clase de equivalencia.

4.2.3. Estrategia de etiquetado de regiones conectadas.

La estrategia empleada en la localización y conteo se basó en algunas modificaciones en la implementación del algoritmo de etiquetado de regiones conectadas, estas modificaciones se realizaron con el objetivo de extraer características adicionales de cada una de las regiones

conectadas como son: *Área* definida como en número total de pixeles contenidos en la región conectada, *Perímetro* definido como el número de pixeles que tienen $n-1$ vecinos para una conectividad n y el rectángulo vertical envolvente de la región o las coordenadas inferior y superior a la izquierda y derecha. Dando además prioridad a regiones convexas, esto es, un círculo por ejemplo puede llenarse en un solo recorrido con una única etiqueta a diferencia de una región no convexa, permitiendo así asignar una sola etiqueta a este tipo de regiones en una sola pasada. En la Figura 4.10 mostramos el etiquetado de regiones conectadas en color o niveles de gris para las imágenes de prueba de la Figura 4.2.

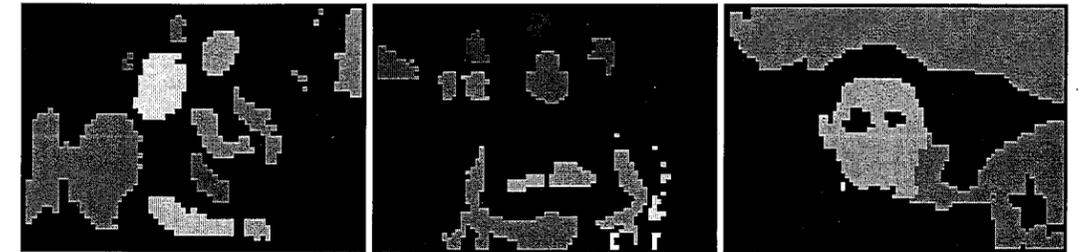


Figura 4.10: Resultados del algoritmo de conteo y localización de regiones. (Los resultados se muestran en color o niveles de gris)

4.3. Separación de regiones no convexas por proyección de vectores.

Después del resultado anterior en el que cada una de las regiones candidatas fueron localizadas, existe la siguiente interrogante ¿Estas regiones están o no unidas entre si?, suponiendo que puedan estar dos regiones diferentes unidas entre si y no exista en la región alguna característica que así lo haga ver, excepto el contorno de esta región, es decir, la forma elipsoidal y convexa, cuando esto no se cumple se puede analizar la región tomando este conocimiento a priori para obtener regiones elipsoidales con ligeras deformaciones pero consideradas en general como convexas. Esta técnica es denominada *particionamiento* por Wei [37]. A diferencia de Wei donde realizan un particionamiento aleatorio basado en la razón del eje mayor y menor de una región r_i , nuestra propuesta de basa en el análisis del contorno de la región, particionando r_i si se presenta alguna discontinuidad o un cambio brusco de dirección en el contorno.

Algunos ejemplos donde se presenta esta unión de regiones se muestran en la Figura 4.15,

en la fila superior se observa que para estas imágenes el descriptor de color dominante no logró separar estas regiones adecuadamente, por lo cual fue necesario proponer una técnica de particionamiento de regiones basado en la forma de la región conservando la convexidad dado que sabemos que un rostro tiene forma elipsoidal.

4.3.1. Algunos conceptos básicos sobre vectores en 2D.

El espacio vectorial V_2 de dimensión 2, es el conjunto de todos los pares ordenados (A_x, A_y) de números reales, llamados **vectores**. Los números A_x y A_y en (A_x, A_y) son los **componentes** del vector. La magnitud del vector A es $\|A\| = \sqrt{A_x^2 + A_y^2}$, el vector perpendicular a A es el vector A^\perp , este es el vector tal que A y A^\perp forman un ángulo recto, aunque existen dos vectores perpendiculares a un vector dado, uno rotado 90° en el sentido contrario a las manecillas del reloj y el otro a 90° en el sentido de las manecillas, definimos el vector A^\perp perpendicular a

$$A = \begin{bmatrix} A_x \\ A_y \end{bmatrix}$$

como:

$$A^\perp = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} A = \begin{bmatrix} -A_y \\ A_x \end{bmatrix}$$

Si $P_1 = (x_1, y_1)$ y $P_2 = (x_2, y_2)$ son dos puntos, entonces el vector A en V_2 que corresponde a $\vec{P_1P_2}$ es $A = (x_2 - x_1, y_2 - y_1)$

Si se tienen dos vectores $A = (A_x, A_y)$ y $B = (B_x, B_y)$, la suma de A y B es $A + B = (A_x + B_x, A_y + B_y)$. El **producto escalar** $A \cdot B$ de $A = (A_x, A_y)$ y $B = (B_x, B_y)$ es $A \cdot B = A_x B_x + A_y B_y$

Sean A y B dos vectores en V_2 con $B \neq 0$. La **componente de A a lo largo de B** se denota por $comp_B A$ y se define como:

$$comp_B A = A \cdot \frac{1}{\|B\|} B \quad (4.6)$$

Obsérvese que $comp_B A$ es positivo si $0 \leq \theta < \pi/2$, o negativo si $\pi/2 < \theta \leq \pi$. Cuando $\theta = \pi/2$, la componente es 0. Escribiéndolo de otra forma, la componente de A a lo largo de B es igual al producto escalar de A y un vector unitario que tiene la misma dirección que B , Figura 4.11 (d).

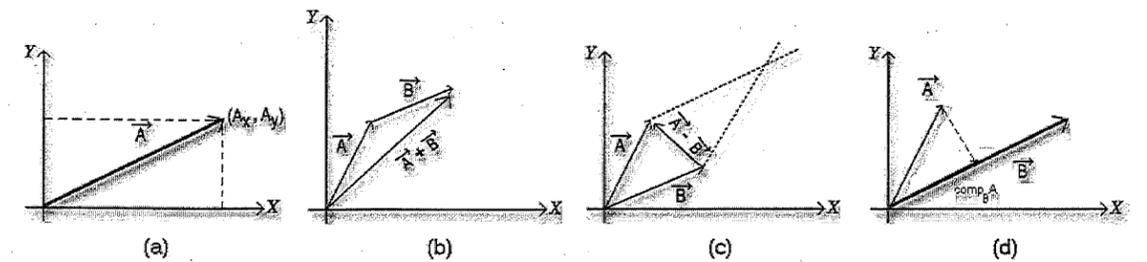


Figura 4.11: (a) un vector y sus componentes (A_x, A_y) , (b) suma de $A+B$, (c) resta de $A-B$, (d) el vector $comp_B A$

4.3.2. Estrategia de división de regiones no convexas.

La estrategia de división de regiones se basa en la proyección de vectores para regiones convexas de la siguiente manera, supongamos el conjunto de n puntos que conforman una aproximación al contorno real de la región a analizar $P = \{(x_i, y_i) \mid (x_i, y_i) \in \text{Perímetro}\}$, Figura 4.12.(a), de hecho cada punto pertenece al contorno real, pero son tomados tal que son una submuestra de este conjunto, esto de acuerdo a un parámetro de manera que tengan una separación entre pixeles uniforme o aleatoriamente. El conjunto de vectores v_{n-1} se obtiene de la resta de los vectores de coordenadas $\vec{p_i p_{i-1}} = (x_i - x_{i-1}, y_i - y_{i-1})$, es el conjunto denominado **vectores perímetro** v_n , es decir, el conjunto de líneas dirigidas secantes al perímetro que pasan por los puntos $(p_i - p_{i-1})$, la característica principal, en este ámbito de los vectores perímetro, es que la componente del vector v_i a lo largo de v_{i-1} , $comp_{v_i} v_{i-1}$ es positiva, y la componente v_i a lo largo del vector perpendicular a v_{i-1} , $comp_{v_i} v_{i-1}^\perp$ también es positiva, Figura 4.12.(b).

Cuando cada una de estas componentes de los vectores a lo largo de su vector anterior son positivos, el algoritmo propuesto mantiene esta región como una región única. Para una región no convexa, Figura 4.13, en algún momento la componente de v_i a lo largo de v_{i-1}^\perp no es positiva, significando esto que la región es cóncava, cuando esto sucede hacemos una anotación de este vector y el punto correspondiente que señala una ubicación o punto probable de separación de regiones, si las componentes del vector siguiente a v_{i+1} mas el vector v_i , es decir, las componentes de $v_i + v_{i+1}$ a lo largo de v_{i-1} y v_{i-1}^\perp , son ambas positivas se continua evaluando de la misma forma, esto evita considerar una región con pequeños cambios de dirección en el contorno de una región como no convexa, con esto evitamos separar una región en un punto inadecuado, lo cual denominamos una discontinuidad hacia afuera, Figura 4.13 (a). Sin embargo se puede dar el caso de que ambas componentes del vector v_i sean positivas pero en el siguiente paso, la componente del vector v_{i+1} a lo largo de v_i^\perp no es positiva y alguna de las componentes de

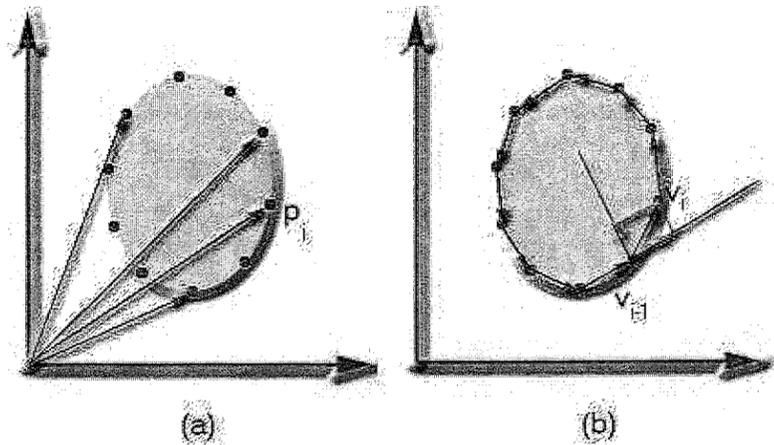


Figura 4.12: (a) Puntos del contorno de una región convexa y sus vectores de posición, (b) segmentos de línea dirigidas secantes al perímetro y las componentes del vector v_i a lo largo de v_{i-1} .

$v_{i+1} + v_{i+2}$ a lo largo de v_i no sean positivas, lo que denominamos una discontinuidad hacia adentro, en este caso evaluamos las componentes de $v_{i+1} + v_{i+2}$ a lo largo de v_i , si ambas son positivas continuamos evaluando en el siguiente vector, Figura 4.13 (a).

Cuando ninguno de los casos anteriores se cumple nos encontramos con una discontinuidad en el contorno de la region de análisis suficiente que amerita, bajo nuestro estudio de las regiones de imagenes de prueba, en el que este tipo de cambios de dirección corresponden a regiones que deben ser separadas, por lo cual actuamos de la siguiente manera:

Calculamos las componentes del vector de coordenadas que va del punto p_i a p_j , $j = i, i + 1, \dots, n$, el vector $\vec{p_i p_j} = (x_{p_j} - x_{p_i}, y_{p_j} - y_{p_i})$ a lo largo de v_i , Figura 4.13 (b), cuando ambas componentes son positivas se empieza a tomar un criterio de la mínima magnitud del vector $\vec{p_i p_j}$, aquel vector $\vec{p_i p_j}$ cuya magnitud sea menor, $\min(\|\vec{p_i p_j}\|)$ y cuyas componentes sean positivas se considera como el vector secante correspondiente, y los puntos p_j como p_i son los puntos que define la línea que separa estas regiones.

Algunos resultados experimentales se muestran en la Figura 4.15, en la fila superior se muestra la imagen y líneas secantes al perímetro y como se separaron estas regiones, en la fila inferior se muestran el conjunto de puntos de aproximación al contorno antes de la separación, en esta figura solo se muestran para visualización, en el sistema desarrollado el manejo es interno.

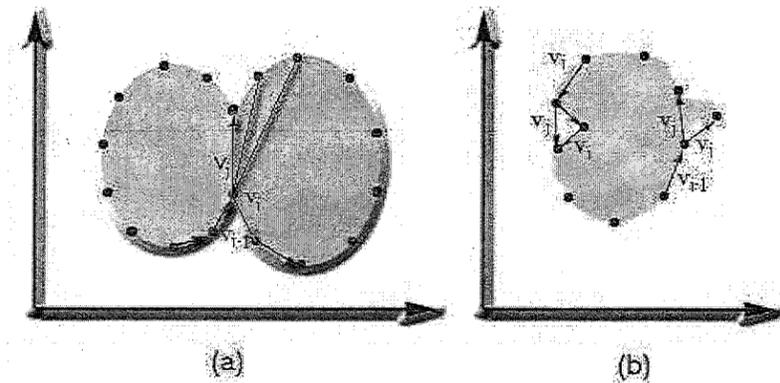


Figura 4.13: (a) Comportamiento de los vectores secantes al perímetro para regiones no convexas, (b) el vector v_j que mejor se aproxima a la región convexa



Figura 4.14: Contornos de las imágenes de prueba utilizadas en esta sección y puntos de aproximación

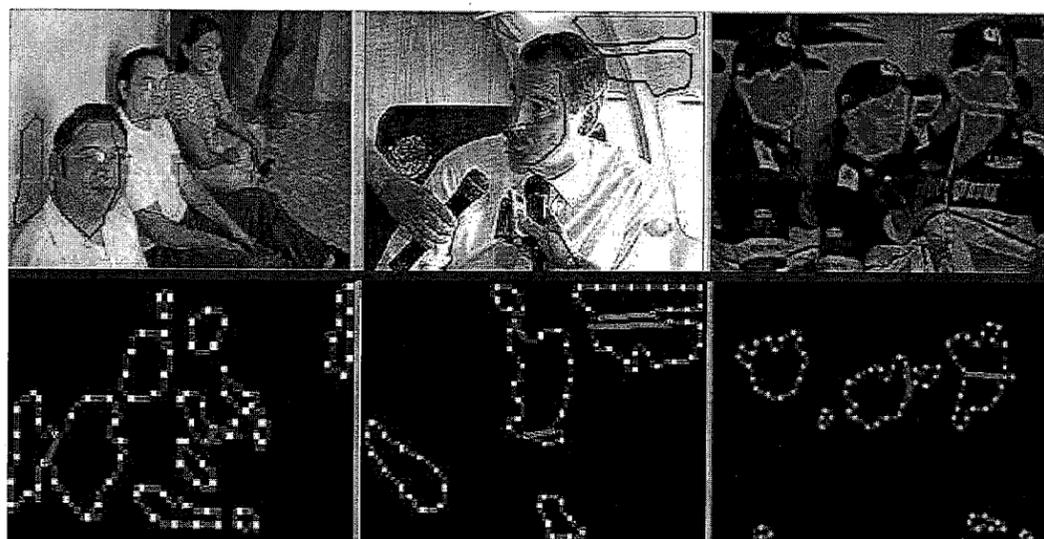


Figura 4.15: Algunos resultados para las imagens mostradas, separación de regiones y los vectores perímetro o líneas secantes. En la primer columna, la cara que esta en primer plano junto con su sombra forman una sola región, las líneas muestran el resultado de la separación. En la segunda columna la cara del Sr. Presidente esta unida con un región sobre esta, el resultado de la separación se muestra con líneas continuas sobre la imagen. En la tercer columna el rostro de enmedio se ve afectado por una región detrás, las líneas continuas muestran la separación correspondiente.

4.4. Algoritmo general de localización y separación de regiones.

El algoritmo general para la localización y separación de regiones probables es el siguiente:

1. Inicialización.

$Chmmd(diff) \leftarrow 0$, arreglo bidimensional con la componente *diff* del modelo
HMMD de toda la imagen

Min_{etiq} , arreglo bidimensional de etiquetas del paso anterior

$Mout_{etiq} \leftarrow 0$, arreglo bidimensional de etiquetas

$M_{reg} \leftarrow 0$, arreglo bidimensional de regiones en la imagen

$Arr_{pos} \leftarrow 0$, arreglo bidimensional de coordenadas y descriptores de las regiones a localizar

$Num_{obj} \leftarrow 0$, número de objetos en la imagen

2. Descriptor de color para las regiones de piel

Si $Min_{etiq} = 1$

- Calcular el histograma de la componente *diff*, con 4 bins
- Observamos el comportamiento del histograma, si el valor de cada bin es muy diferente se separan las regiones
- Si se separan las regiones, se asigna el valor del bin correspondiente $Mout_{etiq}$.

3. Operador morfológico para mejorar la imagen de etiquetas

Aplicar a $Mout_{etiq}$ el operador de cerradura en niveles de gris

4. Reducción de resolución

Reducir la imagen $Mout_{etiq}$ a la mitad de acuerdo a la etiqueta dominante

5. Localización de regiones.

Ejecutar el algoritmo de Etiquetado de regiones conectadas, asignar a M_{reg} por cada pixel la etiqueta correspondiente y una etiqueta diferente por cada región, en cada cada región encontrar los siguientes descriptores:

Asignar en la posición del arreglo Arr_{pos} correspondiente, las coordenadas del punto sup.izq., sup.der., inf.izq. y inf.der. del min. rectángulo circunscrito, área de la región y coordenadas del punto medio

Asignar a Num_{Obj} el número de regiones localizadas

6. Contornos y separación de regiones en si amerita.

- Por cada número de regiones localizadas, detectar sus contornos.
- Asignar a cada m puntos del contorno como punto de aprox., con $m=4$.
- Ejecutar el algoritmo de separación de regiones por proyección de vectores.

Si hubo regiones a dividir encontrar los descriptores mencionados para estas regiones, actualizar los de las regiones divididas, y sumar el número de regiones adicionales a Num_{Obj}

7. Salida.

Num_{Obj} Numero de regiones localizadas más las divididas.

M_{reg} Arreglo de las regiones etiquetadas.

Arr_{pos} Coordenadas y descriptores de las regiones.

Capítulo 5

Técnicas de Verificación de Rostros.

El último paso de nuestro sistema de detección de rostros es la clasificación de las regiones consideradas como piel en la clase rostro, esta tarea es la más difícil, en el ámbito de reconocimiento de patrones, debido a una gran variedad de factores ya descritos, el desarrollo de métodos para detectar rostros puede basarse en un simple análisis de la forma de la región hasta esquemas basados en algoritmos complejos como son los métodos basados en la semántica de la imagen. En este capítulo presentamos algunas técnicas rápidas analizando la región y la técnica de PCA para la verificación de estas regiones. En la Figura 5.1 presentamos gráficamente el procesamiento a desarrollar en la imagen.

5.1. Técnicas rápidas.

El análisis de la forma consiste en aplicar algunas reglas heurísticas simples basadas en descriptores de las regiones de piel, dado un conocimiento apriori sobre la forma que debe tener cada región, para así determinar que la región en análisis es un rostro o no, lo cual es muy rápido en términos computacionales. Como tal se emplean algunos de estos descriptores en este trabajo, pero asignando los resultados a un vector de rasgos que conjuntamente con métodos basados en la imagen nos proporcionan mejores resultados.

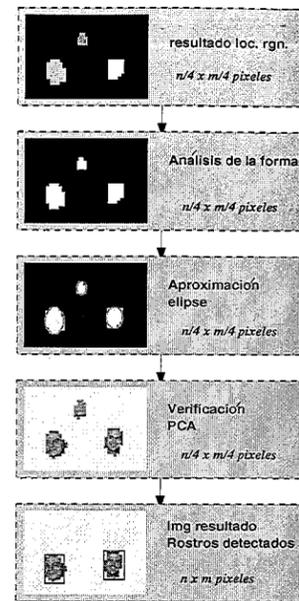


Figura 5.1: Esquema de procesamiento en la etapa de verificación

5.1.1. Análisis de la forma de la región.

Algunos descriptores simples de regiones.

En general, los descriptores son algún conjunto de números generados para describir una forma o región dada, el uso de forma o región se usará indistintamente. Esta forma puede no ser enteramente reconstruida con estos descriptores, pero los descriptores de diferentes formas pueden ser suficientes de manera que estas formas puedan ser discriminadas. Algunos descriptores simples son:

Área. El número de píxeles en la forma o región.

Perímetro. El número de píxeles en la frontera de la forma o región. En la medición del perímetro, se toma en consideración la conectividad definida entre píxeles, y algunas consideraciones adicionales, por ejemplo, en la distancia, podemos contar la distancia igual a 1 para píxeles con conectividad-4 y $\sqrt{2}$ para conectividad-8.

(no)Compacidad o (no)Circularidad. Que tan cerrada-compacta es (o no es) la forma o región: se define como $perimetro^2/area$. La forma compacta más común es el círculo (4π). Todas las demás formas tienen una compacidad mayor a 4π .

Excentricidad. Es el cociente entre la cuerda de longitud más corta de la forma o región a la cuerda de longitud más larga perpendicular a esta. Esta es una manera de definirla aunque existen otras.

Elongación. El cociente de la altura entre el ancho del mínimo rectángulo circunscrito en cualquier dirección. En otras palabras, podemos rotar un rectángulo de manera que el mínimo de estos se ajuste a la forma, entonces comparamos el largo contra el ancho.

Rectangularidad. Que tan rectangular es una forma o región (que tanto llena el mínimo rectángulo circunscrito): $área\ del\ objeto / área\ del\ rectángulo\ circunscrito$.

Orientación. La dirección general de la forma o región.

Descriptores topológicos. Si deformáramos una región, por ejemplo, si tuviéramos una lámina de material hecho de goma, habría algunas formas que se podrían hacer y otras que no. La *topología* se refiere a la propiedad de la forma de no cambiar, siempre y cuando no se rompan o unan partes de ésta. Un descriptor topológico útil es el *Número de Euler E*: el número de componentes conectados C menos el número de huecos H :

$$E = C - H$$

Aunque el número de Euler es un descriptor simple, puede resultar útil para separar regiones simples.

Contorno (poligonal) Convexo. Aunque no es estrictamente una propiedad topológica, podemos describir las propiedades de la forma midiendo el número o tamaño de las concavidades en la forma. Primeramente encontramos la poligonal convexa de la forma y sustraemos la forma misma. El resultado son huecos ("islas") o concavidades ("bahías"). Esto resulta útil, por ejemplo, cuando tratamos de distinguir la letra O de la C .

Puntos extremos. Es otra forma de describir la forma. La manera más simple es el rectángulo circunscrito, el rectángulo más pequeño que contiene completamente al objeto. Otra manera más es encontrar ocho puntos definidos como: arriba izquierda, arriba derecha, izquierda superior, derecha superior, abajo izquierda, abajo derecha, izquierda inferior, derecha inferior, (por supuesto algunos de estos puntos pueden ser los mismos). Conectando los pares opuestos de los puntos extremos podemos crear cuatro ejes que describen la forma. Estos ejes pueden usarse

por si mismos como descriptores o podemos usar una combiación de ellos. Por ejemplo, el eje mayor y su opuesto (no necesariamente ortogonal) pueden ser etiquetados como el eje mayor y eje menor, respectivamente. El cociente de estos ejes puede usarse para definir la excentricidad. La dirección del eje mayor puede usarse (como aproximación) para definir la orientación del objeto.

Perfiles. El *perfil* o *proyección* es una característica muy útil basada en la región. El *perfil vertical* es el número de pixeles de la región en cada columna. El *perfil horizontal* es el número de pixeles en la región en cada fila. También podemos definir un *perfil diagonal*, que es contar los pixeles por cada diagonal definida.

Por ejemplo estos han sido usados en aplicaciones de reconocimiento de caracteres, donde una L y una T tienen perfiles muy diferentes.

Momentos. Otra manera de describir la forma de un objeto es usando propiedades estadísticas llamadas *momentos*. Para el caso 2-dimensional, supongase que tenemos una función discreta sobre dos variables. Podemos definir momentos sobre un punto arbitrario, generalmente sobre cero o sobre la media. El *i*-ésimo momento sobre cero es:

$$m_{ij} = \frac{\sum_{x=1}^N x^i y^j f(x, y)}{\sum_{y=1}^M \sum_{x=1}^N f(x, y)}$$

Así, $m_{00} = 1$. m_{10} es el x componente μ_x de la media, y m_{01} es el y componente μ_y de la media.

Definimos los momentos centrales como:

$$\mu_{ij} = \frac{\sum_{x=1}^N (x - \mu_x)^i (y - \mu_y)^j f(x, y)}{\sum_{y=1}^M \sum_{x=1}^N f(x, y)} \quad (5.1)$$

Respecto a 0, $\mu_{10} = \mu_{01} = 0$

Se pueden usar los momentos para generar un descriptor de formas muy adecuado. Supongase una imagen binaria. Los momentos μ_{20} y μ_{02} son las varianzas de x y y respectivamente. El momento μ_{11} es la covarianza entre x y y . Podemos ver que la covarianza

determinan la orientación de la forma. La matriz de covarianza C es

$$C = \begin{bmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{bmatrix}$$

Encontrando los eigenvalores y eigenvectores de C , podemos determinar la excentricidad de la forma (que tan elongada es esta) como el cociente entre los eigenvalores, y también determinar la dirección de elongación usando la dirección del eigenvector que corresponde al eigenvalor más grande en valor absoluto. La orientación se puede determinar también de esta forma:

$$\theta = \frac{1}{2} \tan^{-1} \frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \quad (5.2)$$

Existen muchas más formas útiles de combinar estos momentos para describir una forma. Aun más, si se cuentan con los suficientes momentos puede reconstruirse completamente la región. También, se puede aproximar un círculo que tiene la misma media de la forma o una elipse que tiene la misma media y covarianza.

5.1.2. Aproximación de una elipse al contorno.

Este método trabaja sobre un conjunto de puntos pertenecientes al contorno de la región, siendo un método no iterativo de ajuste específico de una elipse.

Una elipse es un caso especial de una cónica general que puede ser descrita por un polinomio implícito de segundo orden

$$F(x, y) = ax^2 + bxy + cy^2 + dx + ey + f = 0$$

con la restricción para ser una elipse de:

$$b^2 - 4ac < 0$$

donde a, b, c, d, e, f son los coeficientes de la elipse y (x, y) son las coordenadas de los puntos que están sobre ésta. El polinomio $F(x, y)$ es llamado la *distancia algebraica* del punto (x, y) a una cónica dada. Introduciendo los vectores

$$\mathbf{a} = [a, b, c, d, e, f]^T$$

$$\mathbf{x} = [x^2, xy, y^2, x, y, 1]$$

podemos reescribir en forma vectorial

$$F_a(x) = \mathbf{x} \cdot \mathbf{a} \quad (5.3)$$

El ajuste de una cónica general a un conjunto de puntos $(x_i, y_i), i = 1, \dots, N$ puede aproximarse minimizando la suma al cuadrado de las distancias algebraicas de los puntos a la cónica representada por los coeficientes \mathbf{a} :

$$\min_a \sum_{i=1}^N F(x_i, y_i)^2 = \min_a \sum_{i=1}^N (F_a(\mathbf{x}_i))^2 = \min_a \sum_{i=1}^N F(\mathbf{x}_i \cdot \mathbf{a})^2$$

Esto puede resolverse directamente usando una aproximación de mínimos cuadrados estándar, pero el resultado es una cónica general, y se necesita cumplir la restricción de que sea una elipse. Para asegurar que sea una elipse específica, el discriminante $b^2 - 4ac$ debe ser considerado. Sin embargo, esta restricción es difícil de cumplir [8]. Aunque la imposición de esta desigualdad es difícil en general, se tiene la libertad de arbitrariamente escalar los parámetros, así se incorpora este escalamiento en la restricción e imponiendo una restricción de igualdad, $4ac - b^2 = 1$. Esta es una restricción cuadrática que puede ser expresada en forma matricial

$$b^2 - 4ac = \mathbf{a}^T \begin{bmatrix} 0 & 0 & -2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ -2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \mathbf{a} = \mathbf{a}^T \mathbf{C} \mathbf{a} = 1$$

así el problema de ajuste de una elipse puede reformularse como:

$$\min_a \|\mathbf{D}\mathbf{a}\|^2 \text{ sujeto a } \mathbf{a}^T \mathbf{D}\mathbf{a} = 1 \quad (5.4)$$

donde la *matriz de diseño* \mathbf{D} de tamaño $N \times 6$,

$$\begin{bmatrix} x_1^2 & x_1 y_1 & y_1^2 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_i^2 & x_i y_i & y_i^2 & x_i & y_i & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_N^2 & x_N y_N & y_N^2 & x_N & y_N & 1 \end{bmatrix}$$

representa la minimización por mínimos cuadrados y la *matriz restricción* \mathbf{C} de tamaño 6×6 expresa la restricción. Resolviendo por mínimos cuadrados, primero, aplicando multiplicadores de Lagrange, se obtienen las condiciones para una solución óptima de \mathbf{a}

$$\mathbf{S}\mathbf{a} = \lambda \mathbf{C}\mathbf{a} \quad (5.5)$$

$$\mathbf{a}^T \mathbf{C}\mathbf{a} = 1$$

donde \mathbf{S} es la *matriz de dispersión* de tamaño 6×6 ,

$$\mathbf{S} = \mathbf{D}^T \mathbf{D} = \begin{bmatrix} S_{x^4} & S_{x^3 y} & S_{x^2 y^2} & S_{x^3} & S_{x^2 y} & S_{x^2} \\ S_{x^3 y} & S_{x^2 y^2} & S_{x y^3} & S_{x^2 y} & S_{x y^2} & S_{x y} \\ S_{x^2 y^2} & S_{x y^3} & S_{y^4} & S_{x y^2} & S_{y^3} & S_{y^2} \\ S_{x^3} & S_{x^2 y} & S_{x y^2} & S_{x^4} & S_{x y} & S_x \\ S_{x^2 y} & S_{x y^2} & S_{y^3} & S_{x y} & S_{y^4} & S_y \\ S_{x^2} & S_{x y} & S_{y^2} & S_x & S_y & S_1 \end{bmatrix}$$

donde el operador \mathbf{S} denota la suma

$$S_{x^a y^b} = \sum_{i=1}^N x_i^a y_i^b$$

Después, el Sistema de ecuaciones 5.5 se resuelve usando descomposición de eigenvectores generalizados. Existen seis soluciones reales $(\lambda_j, \mathbf{a}_j)$, pero debido a

$$\|\mathbf{D}\mathbf{a}\|^2 = \mathbf{a}^T \mathbf{D}^T \mathbf{D} \mathbf{a} = \mathbf{a}^T \mathbf{S} \mathbf{a} = \lambda \mathbf{a}^T \mathbf{C} \mathbf{a} = \lambda$$

lo que se busca es el eigenvector \mathbf{a}_k correspondiente al mínimo eigenvalor positivo λ_k . Después

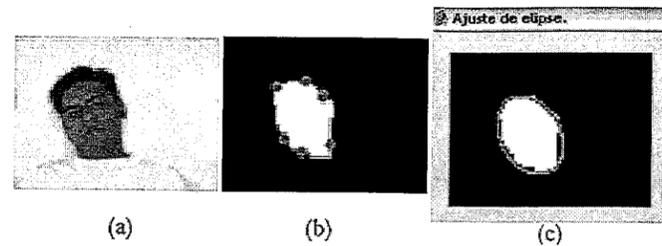


Figura 5.2: Ajuste de una elipse, (a) imagen de un rostro, (b) la región de piel y el conjunto de puntos de ajuste, (c) elipse aproximada al conjunto de puntos.

del escalamiento asegurando $\mathbf{a}_k^T \mathbf{C} \mathbf{a}_k = 1$, se obtiene la solución que minimiza la Ecuación 5.4 que representa el mejor ajuste de una elipse para un conjunto de puntos dado. Esta aproximación fue propuesta por Fitzgibbon [8]. En la Figura 5.2 se muestra un ejemplo del ajuste de una elipse al conjunto de puntos señalado.

5.2. Detección de rostros usando PCA.

Como hemos visto en las secciones previas, el problema de detección de rostros con un modelo explícito de características faciales llega en algunas ocasiones a tener algunos problemas debido al carácter impredecible de la apariencia del rostro y las condiciones ambientales circundantes al adquirir los datos, en muchas de las imágenes empleadas también existen regiones catalogadas como piel, que podrían no serlo. Como resultado existen la necesidad de técnicas con un desempeño aun mayor para rostros en cualquier posición y en escenarios con fondos muy complejos. Formulando el problema como uno de aprendizaje para reconocer patrones de rostros desde ejemplos, la aplicación de un conocimiento a priori del rostro es evitado. La idea básica es reconocer patrones de caras mediante un procedimiento de entrenamiento el cual clasifica ejemplos en clases prototipos de *Rostro* y *noRostro*. La comparación entre estas clases y un arreglo de intensidades en 2D (de aquí el nombre de basadas en la imagen) extraídas de una imagen de entrada nos permite que la decisión de la existencia de un rostro sea tomada.

5.2.1. Introducción a los eigenrostros.

Las imágenes de rostros existen en un subespacio del espacio de todas las imágenes. Para representar este subespacio, podemos usar aproximaciones con redes neuronales y técnicas similares, sin embargo también existen varios métodos más relacionados al análisis estadístico



Figura 5.3: (a),(b),(c),(d),(e),(f),(g) los siete primeros Eigenrostros del conjunto de análisis, (h) el rostro promedio de este conjunto.

multivariado que pueden ser aplicados. Una de estas técnicas es la denominada **Análisis de Componentes Principales** o **PCA** que se encuentra dentro de la categoría de los métodos de subespacio lineal.

En el lenguaje de la teoría de la información, lo que queremos es extraer la información relevante de una imagen de un rostro, codificarla de la mejor forma posible, y compararlo con un conjunto de rostros codificados de la misma forma. Una aproximación simple para extraer la información contenida en una imagen de un rostro es de alguna manera capturar la variación en una colección de rostros y usar esta información para codificar y comparar imágenes individuales.

En términos matemáticos, deseamos encontrar los componentes principales de la distribución de rostros, o los eigenvectores de la matriz de covarianza del conjunto de rostros, considerando la imagen como un punto (o vector) en un espacio de dimensionalidad muy alta. Los eigenvectores son ordenados, cada uno de acuerdo a la cantidad de variación respecto a las imágenes de rostros. Estos eigenvectores pueden verse como un conjunto de características las cuales juntas caracterizan la variación entre rostros. Cada configuración de la imagen contribuye de mayor o menor manera a cada eigenvector, el cual podemos mostrar como una imagen de una silueta fantasma la cual se denomina *eigenrostro*, Figura 5.3.

Cada rostro individual puede ser representado exactamente en términos de una combinación lineal de eigenrostros. Así mismo cada rostro puede ser aproximado usando solamente los mejores eigenrostros, aquellos que corresponden a los eigenvalores más grandes, los cuales indican mayor variación dentro del conjunto de imágenes de rostros. Los mejores K eigenrostros atraviesan un espacio K -dimensional, el *espacio de rostros*, de todas las imágenes posibles.

La idea de usar eigenrostros fue motivada por la técnica desarrollada por Sirovich y Kirby [32], para representar eficientemente rostros usando PCA. Comenzando con un conjunto de rostros, calculan el mejor sistema de coordenadas para la compresión de la imagen, donde cada coordenada es lo que denominaron *eigenimagen*. Ellos argumentan, al menos en principio, que cualquier colección de rostros puede ser reconstruida aproximadamente encontrando un conjunto de pesos adecuados por cada imagen y un conjunto pequeño de imágenes estándar. Los pesos que describen cada rostro se encuentran proyectando el vector de la imagen del rostro a lo largo de cada eigenimagen.

La técnica para la detección de rostros, dada la imagen que en su conjunto puede ser un rostro o no, involucra las siguientes operaciones de inicialización:

1. Adquirir un conjunto inicial de imágenes de rostros (el conjunto de entrenamiento).
2. Calcular los eigenrostros del conjunto de entrenamiento, manteniendo solo las K imágenes que corresponden a los mayores eigenvalores. Estas K imágenes definen el espacio de rostros.

Estas operaciones se realizan antes de la realizar la tarea de detección de rostros, es decir, las operaciones no se realizan en tiempo real.

Teniendo inicializado el sistema, los siguientes pasos son necesarios para la tarea de detección:

1. Calcular el conjunto de pesos basado en la imagen de entrada con los K eigenrostros más significativos, proyectando la imagen de entrada sobre cada eigenrostro.
2. Determinar si la imagen es un rostro en sí, si la imagen esta lo suficientemente cerca al espacio de rostros. Ésta es la etapa de clasificación.

5.2.2. Cálculo de Eigenrostros.

Sea $I(x, y)$ una imagen representada por arreglo de dimensión dos de $N \times N$ con 256 niveles de intensidad. Esta imagen también puede considerarse como un vector de dimensión N^2 . La idea principal en el análisis de componentes principales, también conocida como expansión Kahunen-Loeve, es encontrar la mejor representación para la distribución de los rostros dentro del espacio total de imágenes.

El procedimiento básico para calcular el espacio de rostros y la distancia desde el espacio de rostros (dfs) es como sigue: Tenemos un conjunto de M imágenes de rostros, $\Gamma_1, \Gamma_2, \dots, \Gamma_M$,

cada imagen I_i se representa como un vector Γ_i . El rostro promedio se define como:

$$\Psi = \frac{1}{M} \sum_{i=1}^M \Gamma_i \quad (5.6)$$

Cada rostro difiere del rostro promedio, el vector:

$$\Phi_i = \Gamma_i - \Psi \quad (5.7)$$

es la diferencia.

Este conjunto de vectores grandes es sujeto al análisis de componentes principales, donde la búsqueda se realiza en un conjunto de M vectores ortonormales, \mathbf{u}_n , que mejor describe la distribución de los datos. El k -ésimo vector, \mathbf{u}_k , se escoge de tal manera que:

$$\lambda_k = \frac{1}{M} \sum_{n=1}^M (\mathbf{u}_k^T \Phi_n)^2 \quad (5.8)$$

es un máximo, sujeto a:

$$\mathbf{u}_l^T \mathbf{u}_k = \delta_{lk} = \begin{cases} 1, & \text{si } l = k \\ 0, & \text{de otra manera} \end{cases} \quad (5.9)$$

Los vectores \mathbf{u}_k y escalares λ_k son los eigenvectores y eigenvalores, respectivamente, de la matriz de covarianza:

$$C = \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^T = AA^T \quad (5.10)$$

donde la matriz $A = [\Phi_1, \Phi_2, \dots, \Phi_M]$. La matriz C , sin embargo, es de N^2 por N^2 , y determinar los N^2 eigenvectores y eigenvalores es una tarea intratable para imágenes de tamaños regulares. Así se necesita un método implementable computacionalmente para encontrar estos eigenvectores.

Si el número de puntos en el espacio es menor que la dimensión del espacio ($M < N^2$), entonces habrá solamente $M - 1$, en vez de N^2 , eigenvectores significativos, los restantes tendrán asociados eigenvalores igual a cero. Afortunadamente podemos resolver para los eigenvectores de dimensión N^2 , primero resolviendo para los eigenvectores de una matriz de $M \times M$, y después tomando una combinación lineal apropiada de las imágenes de rostros Φ_i .

Considere los eigenvectores v_i de $A^T A$ tales que

$$A^T A v_i = \mu_i v_i \quad (5.11)$$

Multiplicando ambos lados por A , tenemos

$$A A^T A v_i = \mu_i A v_i \quad (5.12)$$

donde podemos ver que $A v_i$ son los eigenvectores de $C = A A^T$.

Continuando con este análisis, construimos una matriz de $M \times M$, $L = A^T A$, donde $L_{mn} = \Phi_m^T \Phi_n$, y encontramos los M eigenvectores, v_l de L . Estos vectores determinan la combinación lineal de los M rostros del conjunto de entrenamiento para formar los eigenrostros u_l ,

$$u_l = \sum_{k=1}^M v_{lk} \Phi_k, \quad l = 1, \dots, M \quad (5.13)$$

Con este análisis los cálculos son reducidos lo suficiente, de un orden del número de píxeles de la imagen N^2 a un orden igual al número de imágenes en el conjunto de entrenamiento (M).

5.2.3. Detección usando eigenrostros

Este procedimiento se aplica por cada región que queremos clasificar como rostro o no, así dada una imagen desconocida Γ .

1. Calcular la imagen diferencia, la imagen desconocida menos el rostro promedio.

$$\Phi = \Gamma - \Psi$$

2. Calcular los pesos como la proyección del rostro sobre cada eigenvector, para obtener una aproximación lineal con los K eigenrostros más significativos:

$$\hat{\Phi} = \sum_{i=1}^K w_i u_i$$

3. Calcular la diferencia (difs),

$$e_d = \|\Phi - \hat{\Phi}\|$$

4. si $e_d < T_d$, entonces Γ es un rostro.

La distancia e_d es llamada *distancia desde el espacio de rostros (difs)*.

5.3. Algoritmo de Verificación de Regiones para su clasificación en Rostro o no.

El algoritmo general de esta etapa se describe a continuación:

1. Inicialización.

Arr_{Loc} , las coordenadas y descriptores de las regiones detectadas en la etapa de localización.

$M_{rgn} \leftarrow 0$, arreglo bidimensional de regiones en la imagen

2. Verificación de la forma de la región

Por cada región localizada en la imagen

- Calcular su compacidad.
- Calcular los momentos m_{20}, m_{11}, m_{02} y orientación de la región, Ecuación 5.1 y 5.2
- Aproximar la elipse correspondiente.
- Si $compacidad < umbr_{comp}$ y el área de la región entre el área de la elipse $> umbr_{elip}$ evaluar con PCA

3. Verificación con PCA.

Evaluar con la técnica de PCA

Si $difs < umbr_{difs}$ se clasifica como Rostro

4. Salida.

N Número de rostros detectados en la imagen.

salida.txt Archivo de texto con las coordenadas de las regiones clasificadas como rostros.

Capítulo 6

Implementación y Resultados Experimentales.

En este capítulo presentamos los resultados de una serie de experimentos con el fin de evaluar el desempeño de nuestro sistema de detección de rostros basado en un esquema de multclasificación en la solución del problema planteado.

6.1. Implementación.

Se trabajo en dos tipos de computadoras, desktop y laptop, las características del equipo de cómputo utilizado son las siguientes: computadora de escritorio con procesador Pentium IV a 1.8 Ghz, 256 MB de memoria RAM y una computadora portátil con procesador Pentium IVM a 1.8 Ghz, 256 MB de memoria RAM, los resultados fueron muy parecidos en ambos equipos, en cuanto a tiempo de procesamiento, las tablas de tiempo que se muestran son los resultados obtenidos con el equipo de escritorio. El lenguaje de programación con el que se implementó el sistema fue c++ con la herramienta de desarrollo Builder5, el sistema se denominó DERO (DEtección de ROstros).

6.2. Resultados experimentales

Contamos con un conjunto de secuencias de video con un lapso de un segundo a 16 frames x seg, un conjunto de imágenes tomadas de la transmisión de programas de televisión, y un conjunto de imágenes fotográficas tomadas con una cámara digital y recolectadas de internet. En

estos últimos conjuntos, imágenes de tv e imágenes fotográficas, no se tiene controlados ningún tipo de parámetro y en general son imágenes más complicadas. El total es de 40 imágenes con diferente número de rostros en estas, los resultados obtenidos se muestran en la tabla 6.3. La evaluación del error es de la siguiente forma: se contó visualmente el número de rostros por cada imagen, lo que forma nuestro groundtruth (información que es correcta), esto se muestra en la Tabla 6.1, se evaluó cada imagen con DERO, el porcentaje de detecciones correctas fue el número de rostros correctamente identificados respecto al total de estos, también se enlistan el número de falsas detecciones, Tabla 6.3.

Imágenes o secuencias de imágenes	Número total de imágenes en el conjunto	Número de rostros por conjunto de imágenes
Imágenes fotográficas	10	63
Imágenes VideoConf.	22	29
Imágenes TV	8	9

Tabla 6.1: Conjunto de prueba.

El tiempo de procesamiento es muy pequeño, ya que en todos los casos aun para imágenes grandes, del orden de 800×600 pixeles, el tiempo esta por debajo de las décimas de segundo. Tabla 6.2, lo cual se considera como ejecución en tiempo real.

Etapa del sistema	Tiempos de procesamiento, en milisegundos imagen de 320×240	Tiempos de procesamiento, en milisegundos imagen de 640×480
Transformación de color	004	006
Mapa de veracidad	005	005
Loc. de regiones	menor a 1	menor a 1
Verificación	menor a 10 por rgn.	menor a 10 por rgn.
Total	0.020	0.022

Tabla 6.2: Tiempos registrados en las etapas del sistema

Algunos de los resultados se han mostrado a lo largo de este trabajo, a continuación se muestran resultados finales del programa para diversas imágenes, Figura 6.1 y Figura 6.2.

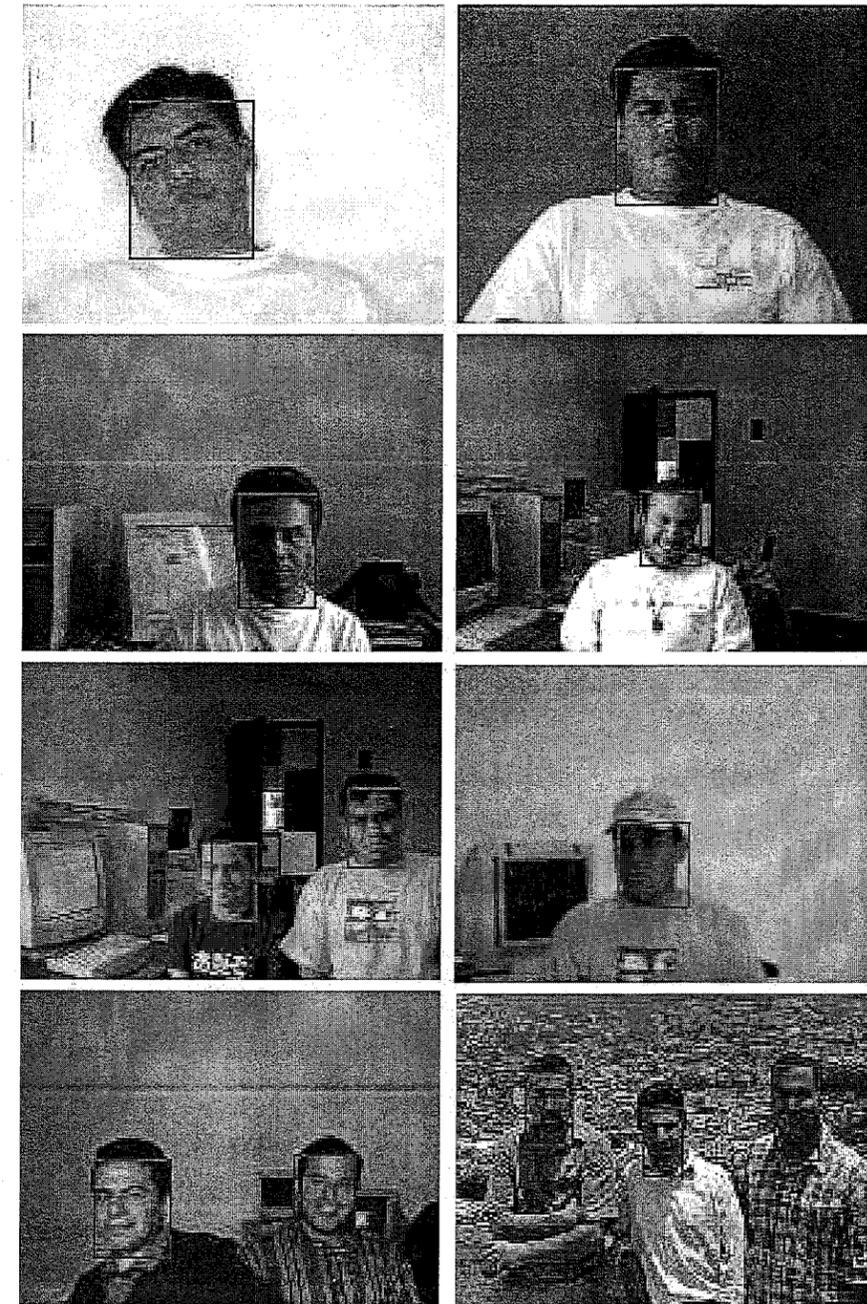


Figura 6.1: Resultados del sistema de detección para imágenes de Video conferencia

Imágenes o secuencias de imágenes	Porcentaje de Detección Correcta	Número de Rgn. Falsas
Imágenes fotográficas	79 %	7
Imágenes VideoConf	93 %	5
Imágenes TV	88 %	1

Tabla 6.3: Porcentajes de detección.

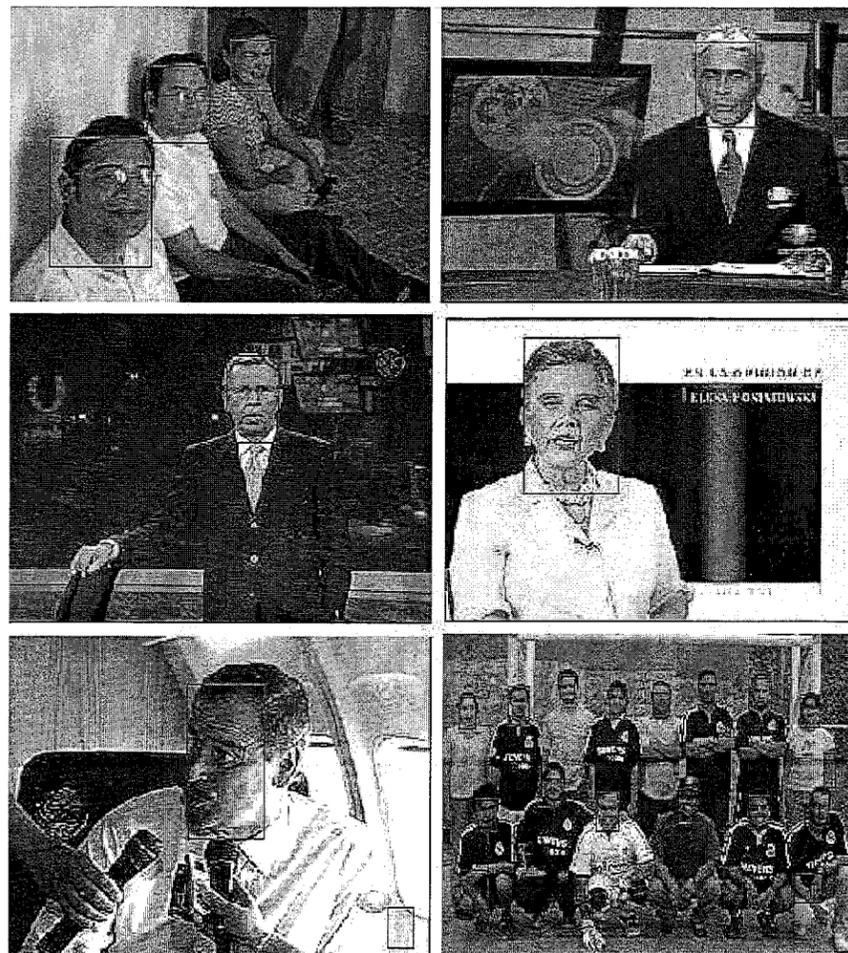


Figura 6.2: Resultados del sistema de detección para imágenes de fotográficas y de tv. En esta serie de imágenes no se utilizó PCA.

6.2.1. Discusión.

El sistema presentado funciona muy bien en la localización de regiones más probables, esto es en el tono de piel. Evidentemente es necesario emplear técnicas más robustas en el análisis de la semántica de la región, ya que basarse en descriptores simples no siempre es robusto y además técnicas como PCA son muy rígidas, debido a que el rostro debe quedar centrado exactamente a los contornos de la cara y cualquier región adicional como el cuello, etc. generan resultados pobres. Adicionalmente en el análisis con PCA solo se usaron rostros de frente, lo cual implica que un rostro fuera de esta posición no podrá ser clasificado como rostro a menos que se hiciera el análisis con PCA para diferentes posiciones de este.

Con el objeto de llevar a cabo una comparación con otros métodos, los valores del porcentaje de detección, tiempo de procesamiento y falsas detecciones deben ser evaluados y analizados en igualdad de condiciones. Los métodos propuestos por Kanade [28], Viola and Jones [16], Rowley [26], son en este momento los puntos de comparación. Estos métodos se han propuesto en los últimos años y son los que logran mejores resultados de los presentes en la literatura. Sin embargo cada autor tiene una configuración diferente, esto es: el conjunto de prueba, la arquitectura de implementación y hardware, y el número y tamaño de las imágenes varían. En general podemos decir que los porcentajes de detección son muy buenos, arriba del 90 %, aunque el tiempo de procesamiento es relativamente alto, que en la mayoría de los casos no se ejecutan en tiempo real, excepto el método de Viola y Jones [16]. Aunque nuestro esquema propuesto tiene porcentajes de detección ligeramente menores a los métodos mencionados y altos números de regiones falsamente detectadas, logramos detectar las regiones de color de la piel muy bien, en un tiempo mucho más rápido que los esquemas propuestos basados en color, nuestra meta como trabajo futuro es continuar con esta línea de investigación utilizando técnicas de procesamiento basadas en la imagen como son el uso algoritmos de aprendizaje, como redes neuronales, etc., para mejorar la etapa de clasificación de estas regiones además de no limitar la posición de la cara. Creemos que con un mayor estudio de estas técnicas y su implementación lograremos mejores porcentajes de detección y reducir las regiones mal clasificadas, logrando un desempeño igualable a los métodos hasta ahora propuestos con un tiempo de procesamiento mucho menor.

Capítulo 7

Conclusiones y Trabajo Futuro.

En un futuro no muy lejano, el reconocimiento de rostros y gestos, será una interface humano computadora muy común, dada su potencialidad más intuitiva y amigable al tradicional teclado y ratón. La forma más adecuada de realizar estas interfaces es empleando métodos basados en visión por computadora. Seis requerimientos para sistemas de reconocimiento de gestos han sido identificados en este ámbito [33]: *operación en tiempo real, independencia de la persona, cierta independencia en la posición, naturalidad de los gestos, fondos (backgrounds) complejos y dinámicos e iluminación variable*, la mayoría de estos requerimientos son también válidos para la detección o reconocimiento de rostros. El uso de información múltiple como color, movimiento o forma, combinado con métodos basados en la imagen, como redes neuronales, etc. es una buena aproximación que ha sido empleada.

El uso de color tiene las ventajas de ser invariante a rotación y tamaño (Sección 3), dando un dimensión extra en comparación a métodos de escala de grises, además de que es rápida de procesar. En años recientes, ha habido un avance significativo en el área de detección, Viola [16], o reconocimiento de rostros, sin embargo por la complejidad del problema, ninguno de los métodos propuestos podrán abarcar la variedad de situaciones que podrían presentarse en la práctica, solo se resolverá parcialmente estos problemas. En esta ámbito, algunos requerimientos que nos atañen como operación en tiempo real, fondos complejos e iluminación variable pueden ser mejorados o solucionados vistos desde diferentes puntos de vista o planteando estrategias alternativas.

7.1. Contribución de la tesis.

En este trabajo de tesis hemos estudiado el uso de diversas estrategias basadas en características de bajo nivel conjuntamente con el uso de una técnica basada en la imagen como es PCA. Los resultados más importantes son el planteamiento de un esquema de multclasificación en el problema de detección, un enfoque no tomado hasta ahora en la selección de componentes de color piel e ideas alternativas en las etapas intermedias que se consideran comunes, como son etiquetado de regiones conectadas, pero no por eso menos importantes.

Las principales contribuciones de estas tesis se presentan a continuación:

- Se presentó un esquema general de procesamiento en tiempo real para detección de rostros basado en clasificación de bajo nivel tomando información de color y técnicas basadas en regiones de la imagen para verificar la existencia de rostros en esta.
- Al aplicar reglas de decisión sencillas y eficientes en la segmentación de color y conteo de regiones, como las propuestas, se reduce el tiempo de procesamiento.
- Con el uso de técnicas de segmentación de color robustas de la piel se logró un buen porcentaje de detección mientras se evitan búsquedas exhaustivas.
- No se tienen restricciones especiales en la imagen, deberán ser escenas del mundo real cotidianas, aunque aun no clasificamos rostros que no estén de frente, lo cual es una meta a cumplir en el trabajo futuro.
- Respecto a otros métodos propuestos el número de rostros contenidos en la imagen no está limitado.

7.2. Conclusiones.

El color es un atributo muy útil para la identificación, detección y localización de un objeto. Debido a que es una característica de bajo nivel, este es computacionalmente adecuado y por lo tanto conveniente para aplicaciones en tiempo real, lo cual se ha demostrado con el presente trabajo, además es robusto a rotaciones u oclusiones. Algunas de las desventajas de tomar la información de color, como la dependencia en la iluminación y/o reflexión, lo limita para algunas aplicaciones en tareas más generales, es decir sistemas bajo ninguna restricción, sin embargo, se puede cubrir una gama de colores semejantes al tono de la piel a manera de evitar tales situaciones.

La meta de este trabajo fue la construcción de un sistema efectivo en tiempo real que automáticamente detecte aquellas regiones de la imagen que correspondan a rostros humanos, si estos existen en la imagen, y que pueda ser utilizado como primer módulo de un sistema general en una posterior aplicación ya sea Video Conferencia u otra aplicación. La idea general fue el empleo de técnicas sencillas, pero computacionalmente rápidas y efectivas.

Concluyendo, tenemos un buen método de segmentación de color, que además es sumamente rápido, para los requerimientos de tiempo real impuestos, que nos permite el empleo de técnicas de procesamiento más complejas, las cuales se planean usar y se mencionan en el trabajo futuro. Esto para lograr un mejor porcentaje en las detecciones correctas y evitar falsas detecciones impuestas por los esquemas rígidos empleados.

7.3. Trabajo futuro.

- Implementar estos resultados en un sistema de Video Conferencia para la transmisión de Regiones de Interés.
- Trabajar con bases de datos de imágenes de rostros que se mencionan en la literatura, las cuales estén disponibles o solicitar estas por los medios adecuados, para contar con un conjunto de entrenamiento suficiente, conjuntamente con la implementación y experimentación de métodos basados en la semántica de la imagen más sofisticados, como son Redes Neuronales, Máquinas de Soporte Vectorial, etc. Incluyendo también la experimentación con el uso de información temporal de la imagen, como es la información generada de frames anteriores para mejorar resultados o eliminar algunas restricciones impuestas.
- No limitar la posición de la cara, en los casos en los que aun no se puede determinar la localización de un rostro, además de identificar las características faciales presentes en el rostro como los ojos y la boca.
- Es necesario un análisis exhaustivo de las condiciones de iluminación y características de la cámara.
- Mejorar aun más, en lo posible, la detección de píxeles de piel, respetando la restricción de ejecución en tiempo real, utilizando además información espacial de la imagen. Utilizar la información del mapa de veracidad en la segmentación de color, conjuntamente con alguna combinación estratégica de componentes de color o intensidad.

Apéndice A

Herramienta de Software DERO

A.1. Detalles de implementación.

A.1.1. Introducción

Se desarrolló una herramienta de software, denominada DERO (DEtección de ROstros), bajo el sistema operativo Windows, para detectar el número de rostros presentes en una imagen, si los hay. Este sistema se puede considerar con el primer paso para una aplicación posterior en la localización automática de regiones de interés (RoI) en sistemas de videoconferencia, o en otras aplicaciones como las mencionadas en el Capítulo 1. El sistema en esta primer versión acepta como **entrada** una imagen en formato BMP (Windows bitmap, .bmp) o una imagen en formato JPEG (.jpeg o .jpg). Si existen algún rostro el resultado del sistema se mostrará gráficamente, además generará un archivo de texto con las coordenadas de las regiones de interés como **salida**.

Suposiciones de entrada

- La imagen será a color, es decir, una imagen de 24 bits de profundidad.
- Los rostros deberán tener la suficiente iluminación para ser distinguibles.
- Los rostros no deberán estar girados completamente hacia atrás, ni tampoco parcialmente ocultos.

A.1.2. Lenguaje de Programación

El lenguaje de programación empleado fue C++. La herramienta de desarrollo utilizada fue Borland Builder 5, esto debido a las estructuras y funciones, y ciertas bondades provistas para la manipulación de imágenes. Además de un ambiente amigable en la creación de interfaces gráficas de usuarios.

A.1.3. Descripción esquemática del sistema

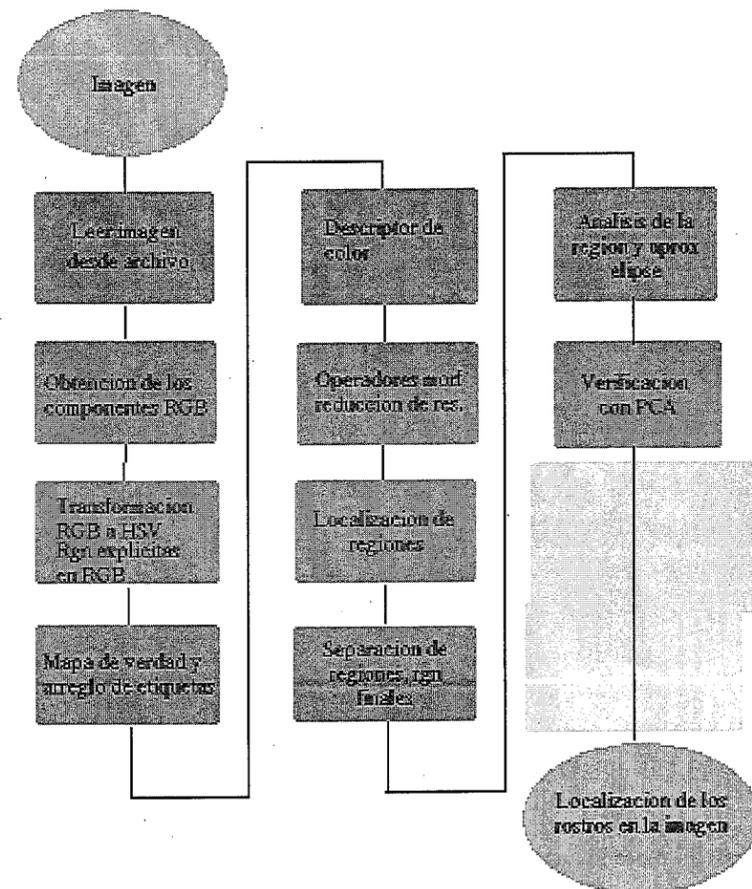


Figura A.1: DERO

A.1.4. Ventanas y botones

La ventana principal del programa DERO, Figura A.2, y los botones de manipulación, que se describirán a continuación:

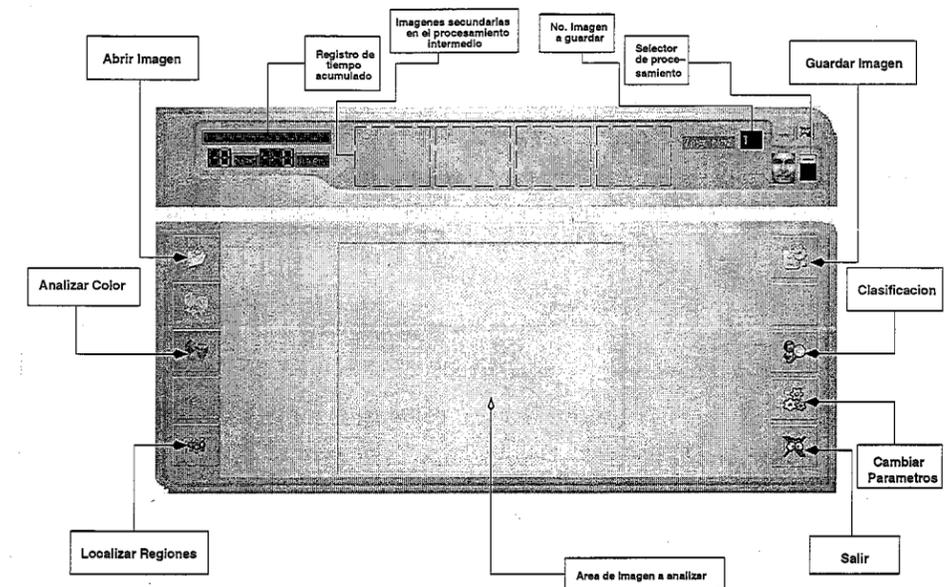


Figura A.2: Ventana principal de DERO

Abrir Imagen.

El botón de **abrir imagen** muestra una ventana de diálogo, la cual permite al usuario abrir una imagen para su análisis, hacer la detección de rostros en la imagen, si los hay, esta imagen se mostrará en la parte central de la ventana, en el área de imagen a analizar. Figura A.3.

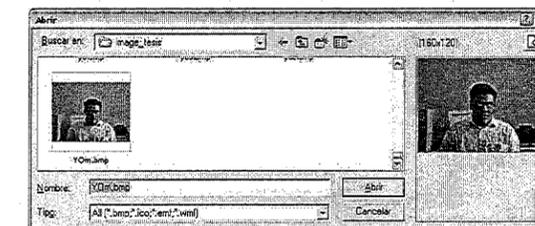


Figura A.3: Ventana de diálogo, Abrir Imagen.

Selector de procesamiento.

Este selector de procesamiento de imagen determina el procesamiento a realizar a la imagen, este está ubicado en la parte superior derecha de la ventana principal, Figura A.2. La configuración es de la siguiente manera: posición superior, se trata de una imagen de video conferencia, donde se tiene un ambiente controlado; posición intermedia o inferior, para procesar una imagen compleja, la diferencia entre estas dos posiciones radica en diferenciar regiones usando el descriptor de color dominante. El valor por omisión es la posición superior. Este selector se puede cambiar en cualquier etapa para una imagen abierta, comenzando nuevamente en la segmentación de color.

Analizar Color.

Ejecuta el algoritmo de segmentación de color descrito en la Sección 3, los resultados parciales se muestran en la barra superior de imágenes.

Localizar regiones.

Ejecuta el algoritmo de detección de regiones descrito en la Sección 4, los resultados parciales se muestran en la barra superior de imágenes.

Clasificación.

Por cada región de la imagen, la etapa de clasificación verifica que cada una de estas regiones pertenezca a la clase *Rostro*, cada región clasificada como *Rostro* se enmarcaba en un rectángulo que lo encierra. El técnica utilizada se describió en la Sección 5.

Cambiar parámetros.

Este botón abre la ventana de configuración de parámetros de la segmentación de color, estos son los vectores medios y matriz de covarianza en los espacios de color HSV y HMMD. Figura A.4.

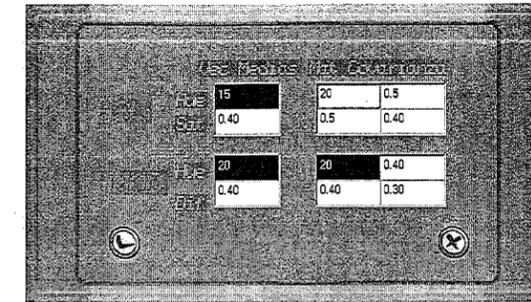


Figura A.4: Ventana de configuración de parámetros

Guardar Imagen.

El botón de guardar imagen muestra una caja de edición, en la parte superior derecha de la ventana principal, donde se selecciona el número de la imagen a guardar, numeradas de izquierda a derecha comenzando en 1. Posteriormente muestra una ventana de diálogo, la cual permite al usuario guardar la imagen seleccionada en formato BMP (Windows bitmap).

Salir.

Cierra la ventana y sale del programa.

Bibliografía

- [1] G. Bradski, *Computer Vision Face Tracking For Use in a Perceptual User Interface*, Intel Technology Journal 1998.
- [2] J. Brand, J. Mason, *A comparative assessment of three approaches to pixel level human skin-detection*. In Proc. of the International Conference on Pattern Recognition, vol. 1, 10561059.2000.
- [3] D. Brown, I. Craw, J. Lewthwaite, *A som based approach to skin detection with application in real time systems*. In Proc. of the British Machine Vision Conference, 2001.
- [4] C. Chistopoulos, D. Berg, A. Skodras, *The colour in the upcoming mpeg-7 standard*, in Proc. X European Signal Processing Conference (EUSIPCO-2000), Finland, 2000.
- [5] P. Colantoni, *ColorSpace Visualization*, a real time 3D visualizer tool of space colors. <http://www.couleur.org>
- [6] P. Devijver, J. Kittler, *Pattern Recognition. A statistical Approach*, Prentice Hall, 1982.
- [7] J. Ding, J. Furgesin, E. Sha, *Application Specific Image Compression for Virtual Conferencing in The International Conference On Information Technology: Coding and Computing (ITCC00)*.
- [8] A. Fitzgibbon, M. Pilu., R. Fisher, *Direct least squares fitting of ellipses*, in Proc. of the 13th international Conference on Patter Recognition, 253-257, Vienna, 1996.
- [9] M. Fleck, D. Forsyth, C. Bregler, *Finding naked people*, in Proc. of the ECCV, vol 2, 595-602, 2002.
- [10] R. Frischholz, *The Face Detector Homepage* <http://home.t-online.de/home/Robert.Frischholz/index.html>, <http://www.facedetection.de>

- [11] G. Gomez, E. Morales, *Automatic feature construction and a simple rule induction algorithm for skin detection*. In Proc. of the ICML Workshop on Machine Learning in Computer Vision, 3138.2002.
- [12] G. Gomez, *On selecting colour components for skin detection*. In Proc. of the ICPR, vol. 2, 961964. 2000.
- [13] E. Hjelma, B.L. Low, *Face Detection: A Survey* in Computer Vision and Image Understanding 83, 236-274, 2001.
- [14] R. Hsu, M. Abdel-Mpftabeb, *Face Detection in Color Images*, in IEEE. Trans. Pattern Analysis and Machine Intelligence, vol 24, no.5, pp 696-706.
- [15] M. Jones, P. Viola, *Rapid Object Detection Using a Boosted Cascade of Simple Features*, IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 511-518, dec. 2001.
- [16] M. Jones, P. Viola, *Robust Real-time Object Detection*, Technical report, 2001.
- [17] M. Jones, P. Viola, *Fast Multi-view Face Detection*, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2003.
- [18] J. Kapur, *Face Detection in Color Images*, EE499 Design Project. University of Washington.
- [19] J. Y. Lee, S. I. Yoo, *An elliptical boundary model for skin color detection*. In Proc. of the 2002 International Conference on Imaging Science, Systems, and Technology. 2002.
- [20] B.S. Manjunath, J.R. Ohm, V. Vasudevan, A. Yamada, *Color and Texture Descriptors*, in IEEE Trans. on Circuits and Systems for Video Technology, Vol 11, no. 6, June, 2001.
- [21] B. Martinkauppi, *Face color under varying illumination - analysis and applications*, Academis Dissertation, Dep. of Electrical and Information engineering and Infotech Oulu, University of Oulu, 2002. ISBN 951-42-6788-5.
- [22] B. Moghaddam, A. Pentland, *Probabilistic visual learning for object representation*, in IEEE Trans. on Pattern Anal. Mach. Intell. 19, 1997.
- [23] P. Peer, J. Kovac, F. Solina, *Human skin colour clustering for face detection*. Submitted to EUROCON 2003 - International Conference on Computer as a Tool.
- [24] R. Qian, I. Sezan, *Face Tracking Using Robust Statistical Estimation*, in Proc. of Multimodal Conference On Multimodal Interfaces, 1998.

- [25] D. Reisfeld, Y. Yeshurum, *Robust Detection of facial features by generalised symmetry*, in Proc. of 11th Int. Conf. on Pattern Recognition, The Hague, the Netherlands, August 1992.
- [26] H.A. Rowley, S. Baluja, T. Kanade, *Neural Network Face Detection*, IEEE. Trans. Pattern Analysis and Machine Intelligence, 20, 1, 1998, pp 23-38.
- [27] R. Schmidt, E. de Campos, *Detection and Tracking of Facial Features in Video Sequences* in Lecture Notes in Artificial Intelligence, vol 1793, pp. 197-206, 2000.
- [28] H. Schneiderman, T. Kanade, *A statistical model for 3D object detection applied to faces and cars* in IEEE Conference in Computer Vision and Patter Recognition , 2000.
- [29] K. Schwerdt, J. Crowley, *Robust Face Tracking Using Color*, Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pp.90-95 Marzo 2000.
- [30] L. Sigal, S. Sclaroff, V. Athitsos, *Estimation and prediction of evolving color distributions for skin segmentation under varying illumination*. In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, vol. 2, 152159.2000.
- [31] T. Sikora, *The MPEG-7 Visual Standard for Content Description - An Overview*, in IEEE Trans. on Circuits and Systems for Video Technology, Vol 11, no. 6, June, 2001.
- [32] L. Sirovich, M. Kirby, *Low-dimensional procedure for the characterization of human faces*, Journal Opt. Soc. Amer. 4. 1987, 5219-524
- [33] M. Störring, *Computer Vision and Human Skin Colour*, Ph.D. Dissertation, Faculty of Engineering and Science, Aalborg University. 2004
- [34] M. Turk, A. Pentland, *Eigenfaces for recognition*, Journal of Cognitive Neuroscience. 3, 7186, 1991.
- [35] V. Vezhnevets, V. Sazonov, A. Andreeva, *A Survey on Pixel-Based Skin Color Detection Techniques*, Proc. Graphicon, pp. 85-92, Russia, 2003.
- [36] E. Viennet, F. Fogelman, *Connectionist methods for human face processing* in Face Recognition: From Theory to Application. Springer-Verlag, Berlin/New York, 1998.
- [37] G. Wei, I. Sethi, *Face detection for image annotation*, Elsevier Pattern Recognition Letters, 20, 1313-1321, 1999.

- [38] H. Wu, J. Zeleck, *A Multi-classifier Based Real-Time Face Detection System* in Journal of IEEE Transactions on Robotics and Automation, Submitted.
- [39] L. Yang, M. Robertson, *Multiple-face tracking system for general region-of-interest video coding*, IEEE Proc. International Conference on Image Processing, pp. 1347-1350, 2000.
- [40] M. Yang, N. Ahuja, *Gaussian mixture model for human skin color and its application in image and video databases*. In Proc. of the SPIE: Conf. on Storage and Retrieval for Image and Video Databases (SPIE 99), vol. 3656, 458466. 1999.
- [41] M. Yang, D. Kriegman, N. Ahuja, *Detecting Faces in Images: A Survey* IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 24, No. 1, Enero 2002.
- [42] B. D. Zarit, B. J. Super, F. K. H. Quek, *Comparison of five color models in skin pixel classification*. In ICCV99 Int. Workshop on recognition, analysis and tracking of faces and gestures in Real-Time systems, 5863. 1999.

Índice alfabético

- Ajuste de elipses, 77
 A una región por Momentos, 77
 Al contorno de la región, 77
- Análisis de Componentes Principales
 Distancia desde el espacio de rostros, 84
 Cálculo de Eigenrostros, 82
 Eigenrostros, 80
- Aprendizaje, 17
 No supervisado, 17
 Supervisado, 17
- Arquitectura del sistema de clasificación, 21
- Clases informacionales, 16
- Clasificación
 No paramétrica, 45
 Paramétrica, 43
- Clasificador, 16
 Aprendizaje, 17
 Conjunto de aprendizaje, 18
- Descriptor de color dominante, DCD, 58
- Descriptores de regiones, 74
 Descriptores simples, 74
 Poligonal convexa, 75
 Puntos extremos, 75
 Topológicos, 75
 Momentos, 76
 Perfiles, 76
- Espacio de representación, 13
- Espacios de color, 25
 Densidad, 26
 espacio RGB lineal, 26
 espacio RGB no-lineal, 26
 HMMD, 29
 HSV, 27
 perceptualmente uniforme, 29
- Etiquetado de regiones conectadas, 63
 Conectividad, 63
 Vecindad, 63
- función de densidad de probabilidad condicional, 33
- Función kernel, 47
 Ancho del kernel, 49
 Kernel Gaussiano, 49
- Modelo de distribución del Color de la Piel, 32
 Modelado no paramétrico, 32
 Modelos dinámicos, 36
 Modelado paramétrico, 35
 Regiones explícitas, 32
- Morfología en niveles de gris, 61
 Apertura y cerradura, 62
 Dilatación, 62
 Erosión, 62
- Morfología matemática, 59
 Apertura y cerradura, 60
 Dilatación binaria, 59

Erosión binaria, 59

Patrones

Similaridad entre patrones, 13

Variabilidad entre patrones, 14

Probabilidad a priori, 33

Problemas relacionados con la detección de rostros., 8

Proyección de vectores, 65

Reflexión de la piel, 24

Regla de clasificación de Bayes, 34

verosimilitud, 35

Selección y extracción de características., 14

Técnicas de detección de rostros, 4

Basados conocimiento de características, 5

Basados en la imagen, 5

Basados en multclasificación, 5

Umbralización, 44