



UNIVERSIDAD
NACIONAL
DE LA PLATA

Trabajo Final presentado para obtener el grado de Especialista
en Computación Gráfica, Imágenes y Visión por Computadora

ALGORITMOS DE IDENTIFICACIÓN DE PIEL HUMANA Y SU
RELACIÓN CON LOS SISTEMAS DE COLOR. SU APLICACIÓN A LA
SEGMENTACIÓN DE PIEL BASADA EN PÍXELES

Facultad de Informática - Universidad Nacional de La Plata

Autor:

Lic. César Estrebou

Directora:

Dra. Laura Lanzarini

Abril de 2020

ÍNDICE GENERAL

Índice general	1
1 Introducción	5
1.1. Motivación	5
1.1.1. Sistemas de Reconocimiento de Gestos	5
1.1.2. Segmentación en Sistemas de Reconocimiento de Gestos	6
1.2. Objetivos	6
1.3. Organización del Documento	7
2 Sistemas de Representación del Color	9
2.1. El Color	9
2.1.1. Espectro de Color	11
2.1.2. Círculo Cromático	12
2.1.3. Síntesis Aditiva y Sustractiva	12
2.1.4. Propiedades del color	13
2.1.5. Modelo de Color	14
2.2. Modelos de Color Básicos	15
2.2.1. Modelo de Color RGB	15
2.2.2. Modelo de Color RGB Normalizado	17
2.3. Espacios de Color Perceptivos	18
2.3.1. Modelo de Color HSV	18
2.3.2. Modelo de Color HSL	20
2.3.3. Modelo de Color HSI	22
2.4. Espacios de Color Ortogonales	23
2.4.1. Modelo de Color YUV	23
2.4.2. Modelo de Color YIQ	24
2.4.3. Modelo de Color YCbCr	26
2.4.4. Modelo de Color YCgCr	27
2.5. Espacios de Color Perceptiblemente Uniformes	28
2.5.1. Modelo de Color CIE-LAB	28
3 Algoritmos para Segmentación	31
3.1. Modelos No Paramétricos	31
3.1.1. Basados en Umbrales	31
3.1.2. Cuantificación Vectorial (VQ)	33

3.1.3.	K Vecinos Cercanos (KNN)	35
3.1.4.	K Medias	36
3.1.5.	Máquinas de Vectores de Soporte (SVM)	37
3.2.	Modelos Paramétricos	39
3.2.1.	Modelo de Mezclas Gaussianas (GMM)	39
3.2.2.	Modelos Ocultos de Markov (HMM)	42
3.3.	Clasificador de Bayes Ingenuo	43
3.4.	Modelos de Redes Neuronales Artificiales (ANN)	44
3.4.1.	Redes Neuronales Artificiales (ANN)	44
3.4.2.	Red de Perceptrones Multicapa	44
3.4.3.	Mapas Auto-organizativos	47
3.4.4.	Redes Neuronales de Energía de Coulomb Restringida (RCE)	50
4	La Piel en los Sistemas de Color	55
4.1.	Introducción	55
4.1.1.	Base de Datos de Piel	56
4.1.2.	Gráficos para Analizar	56
4.1.3.	Normalización de Escalas	56
4.2.	Distribución de la Piel en Sistemas de Color	56
4.2.1.	Distribución de la Piel en RGB	57
4.2.2.	Distribución de la Piel en RGB Normalizado	57
4.2.3.	Distribución de la Piel en HSV, HSL y HSI	58
4.2.4.	Distribución de la Piel en YUV e YIQ	60
4.2.5.	Distribución de la Piel en YCbCr e YCgCr	62
4.2.6.	Distribución de la Piel en LAB	63
4.3.	Análisis de la Distribución de la Piel	64
5	Juegos de Datos para Pruebas	67
5.1.	Características de un Buen Juego de Datos	67
5.2.	Juegos de Datos Utilizados por Artículos Estudiados	71
5.2.1.	IITK Face Database	71
5.2.2.	Pratheepan Dataset	71
5.2.3.	Aberdeen Pilot Dataset	72
5.2.4.	Computer Vision Laboratory Face Database	72
5.2.5.	Labeled Faces in the Wild Dataset	72
5.2.6.	Face Detection Data Set and Benchmark	73
5.2.7.	Edith Cowan University Dataset	73
5.2.8.	Hand Gesture Recognition Dataset	74
5.2.9.	BioID Face Database	74
5.2.10.	SFA Database	75
6	Discusión y Análisis de los Artículos de Estudio	77
6.1.	Sobre los Artículos Estudiados	78
6.2.	Algoritmos de Segmentación de los Artículos de Estudio	79

<i>ÍNDICE GENERAL</i>	3
6.3. Modelos de Color de los Artículos de Estudio	81
6.3.1. Uso de los Modelos de Color	81
6.3.2. Distribución de la Piel en los Modelos de Color	84
6.3.3. Un Experimento sobre Redes Neuronales de Energía de Coloumb	87
6.4. El Mejor Modelo de Color para Segmentar Piel	89
7 Conclusiones y Trabajos futuros	93
7.1. Conclusiones	93
7.2. Trabajos Futuros	94
Índice de figuras	96
Índice de cuadros	98
Lista de Abreviaturas y Símbolos	99
Bibliografía	101

INTRODUCCIÓN

1.1. Motivación

1.1.1. Sistemas de Reconocimiento de Gestos

Los sistemas de reconocimiento de gestos (SRG) han tenido un auge importante desde hace algunos años debido a el avance tecnológico y al gran potencial que tienen en todo tipo de aplicación.

Como cualquier sistema de reconocimiento de patrones (SRP), un sistema de reconocimiento de gestos se compone de cuatro etapas o procesos secuenciales donde la salida de una etapa es la entrada de la siguiente. La primera etapa de un SRP consiste en la adquisición cruda de algún tipo de dato a ser observado. Normalmente, un SRG captura el movimiento y/o configuración de las manos con algún tipo de sensor como una cámara de video, una cámara de tiempo de vuelo, guantes con sensores de posición o alguna combinación de todos estos. La segunda etapa es la de segmentación donde se remueve el ruido de la información y se aíslan los patrones de interés. En un SRG esta etapa consiste en recortar o extraer la mano. Frecuentemente suele utilizarse algún tipo de marcador con el objetivo de simplificar esta tarea. En la tercera etapa se realiza la extracción de las características representativas del patrón para obtener una nueva representación que puede ser clasificable por una computadora. En un SRG esta etapa se encarga de obtener una representación de la configuración de la mano y del movimiento de forma que pueda ser clasificada. Finalmente, en la última etapa se toman las características representativas extraídas y se realiza un esquema de clasificación para reconocer los patrones observados. En un SRG se combina la representación del gesto de la mano junto con el movimiento que lo acompaña y se realiza la clasificación para determinar su significado.

1.1.2. Segmentación en Sistemas de Reconocimiento de Gestos

En el grupo de investigación del que formo parte, hemos trabajado en la etapa de extracción de características y de clasificación de gestos [RQLE15, RQEL16, RQE⁺ 16b, RQE⁺ 16a] desde hace algún tiempo. Recientemente hemos comenzado a explorar diferentes alternativas que utilizan una cámara de video convencional en la etapa de segmentación sin la utilización de marcadores para recortar las manos[GEL17]. En la literatura sobre el estado del arte, muchos trabajos presentan variadas alternativas, pero frecuentemente utilizan como solución parcial o total la segmentación basada en píxeles del color de la piel para extraer las áreas correspondientes a las manos y al rostro.

Muchos autores publican soluciones que realizan transformaciones en una variedad de sistemas de color y aplican algoritmos de distinta naturaleza para determinar si un píxel de una imagen es piel o no. En particular he comenzado la investigación sobre los sistemas de color utilizados en distintos trabajos tratando de buscar una respuesta que eche al menos un poco de luz a la pregunta ¿Cuál es el modelo de color más apropiado para segmentar la piel por su color? Después de haber leído y analizado una buena cantidad de publicaciones con soluciones específicas y publicaciones que realizan revisiones sobre el estado del arte, no he conseguido una respuesta contundente. Es más, muchos autores en sus trabajos no justifican el uso de diferentes sistemas de color en sus aplicaciones de segmentación de piel, y cuando lo hacen lo hacen de manera vaga y sin fundamentos suficientes o sustentables.

1.2. Objetivos

El objetivo principal de este documento es determinar cuál es el modelo de representación de color más conveniente para realizar la segmentación basada en píxel de piel humana en imágenes. Para llevar adelante esto se requiere del estudio de la influencia de los modelos de color en los distintos algoritmos de segmentación basada en píxel generalmente usados en artículos publicados.

Como objetivo secundario se plantea documentar y describir en detalle los modelos de representación del color, los algoritmos de segmentación basada en píxeles y los juegos de datos utilizados para realizar pruebas y obtener los resultados. Esta información es de fundamental importancia para replicar y complementar los resultados obtenidos por los autores de los artículos para realizar futuros experimentos.

Para cumplir con estos objetivos, en este trabajo se describe y estudia en profundidad los sistemas de representación de color, se realiza un análisis de la distribución de la piel en cada uno de estos y se incluye más de treinta artículos publicados en diversos medios para su revisión y análisis. En particular se hace hincapié en la realización de un relevamiento lo más amplio posible de manera de incluir trabajos diversos con la mayor cantidad de modelos de representación de color y algoritmos de segmentación.

1.3. Organización del Documento

En este capítulo se realiza una introducción al tema y se describe la motivación, los objetivos y la forma en que se organiza el documento.

El segundo capítulo trata sobre el color y sus propiedades. Se describen los distintos sistemas o modelos de representación del color comúnmente usados y sus fórmulas conversión a partir del modelo de color RGB.

El tercer capítulo se describe en detalle los algoritmos de segmentación utilizados en los artículos de estudios.

En el capítulo cuarto se muestra la distribución de la piel en los distintos espacios de representación del color y se realiza un breve análisis sobre estas.

En el quinto capítulo se incluye el relevamiento y análisis de los juegos de datos utilizados en las pruebas de los artículos. También se incluye una sección con las características que éstos deben poseer, sobre todo para garantizar pruebas que permitan obtener resultados que puedan compararse de manera justa. Adicionalmente se incluyen juegos de datos disponibles de manera libre que cumplen con dichas características y no se encuentran mencionados en los artículos.

En el sexto capítulo se realiza un análisis general sobre los artículos. Primeramente, se realiza una crítica sobre los resultados y la manera de obtenerlos. Se hace un descripción, análisis y uso de los algoritmos de segmentación y de los modelos de color. Se presenta también una experiencia realizada para segmentación de piel para el reconocimiento de gestos y algunas conclusiones. Finalmente se analiza y discute las posibilidades de cuál es el modelo de color más conveniente para la segmentación de piel.

En el capítulo séptimo se realiza un resumen del aporte realizado por este trabajo y se presentan futuras posibles líneas de investigación y trabajos para continuar.

Finalmente, se listan las referencias a las citas de los artículos de estudio, libros y demás referencias bibliográficas del estado del arte. También se incluyen los índices de figuras, cuadros y glosario de términos.

SISTEMAS DE REPRESENTACIÓN DEL COLOR

Resumen

En este capítulo se introducen conceptos relacionados al color como el espectro visible de la luz, el círculo cromático, la formación del color por medio de la síntesis aditiva y sustractiva y propiedades del color. También se describen los diferentes sistemas de color que comúnmente se utilizan en el procesamiento de imágenes y que se encuentran en los distintos artículos de estudio.

A manera de guía, un buen resumen sobre los sistemas o espacios de colores utilizados para representar el color de la piel humana se puede encontrar en los artículos que analizan y recopilan revisiones tanto de sistemas de color como los algoritmos de clasificación/segmentación utilizados [VSA03, XZ06, KMB07, CGVRGPSP10].

2.1. El Color

La percepción del color está determinada por la sensibilidad del ojo humano a las longitudes de onda del espectro electromagnético que van aproximadamente desde unos 380 nm hasta uno 780 nm. Esta sección del espectro abarca los mismos colores que el arcoíris y comienza en la franja de las longitudes de onda bajas donde se perciben colores en la gama del violeta y finaliza en la franja de longitudes de onda altas donde se perciben colores en la gama del rojo.

El complejo sistema biológico ojo-cerebro transforma la radiación electromagnética en una serie de reacciones químicas procesadas por el cerebro para producir sensaciones visuales. Este proceso de transformación de la luz en señales biológicas es responsabilidad de los conos y de los bastones del ojo, que son células foto-sensitivas que trabajan en condiciones contrapuestas. La capacidad para ver de noche o en condiciones de escasa luz (visión escotópica) depende de unos 100 millones

de células denominadas bastones. Estos sintetizan un foto-pigmento llamado rodopsina que les permiten detectar el número de fotones que componen la luz, independientemente de la longitud de onda. Por este motivo, este tipo de célula presenta gran sensibilidad en condiciones de luz baja, pero satura rápidamente con luz alta, perdiendo la habilidad para emitir señales. Debido a la falta de sensibilidad a las longitudes de onda, estas células no tienen la capacidad de distinguir colores por lo que la visión en condiciones de luz pobre sólo distingue en tonos de grises.

En condiciones de luz diurna o buena luz (visión fotópica), unas 7 millones de células, denominadas conos, rigen la percepción del contraste de color y de la nitidez. El ojo humano posee 3 tipos de conos denominados rojo, verde y azul. La proporción de estos se distribuye aproximadamente alrededor del 60% son rojos, el 30% verdes y sólo el 10% azules. Cada tipo de cono tiene la capacidad de detectar cierto rango de colores o responder a cierto intervalo de longitudes de onda, presentando una zona de sensibilidad máxima a una longitud de onda específica (ver figura 2.1). La superposición entre los rangos de sensibilidad de los conos junto al proceso del cerebro que integra las señales que éstos envían es lo que permite la percepción de millones de colores. Para ejemplificarlo de manera concreta, la percepción del color amarillo es el resultado de la estimulación de los conos verdes y rojos y la falta de estimulación de los conos azules.

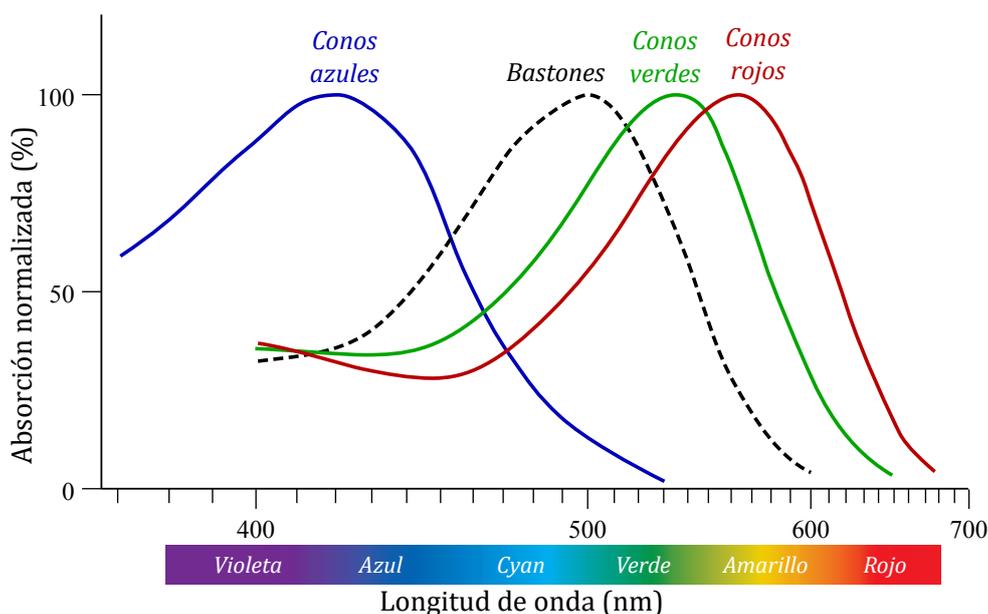


Figura 2.1: Sensibilidad normalizada de las células retinales (bastones y conos) del ojo humano

Un aspecto importante para resaltar sobre la percepción del color es que el proceso de reconocimiento de los colores depende del complejo sistema ojo-cerebro de cada persona en particular. Si bien es posible medir con precisión la longitud de onda de un color específico, el concepto del color percibido a partir de ella es completamente relativo, dependiendo de la persona en sí. Por este motivo, diferentes personas pueden interpretar un mismo color de forma diferente.

2.1.1. Espectro de Color

En la actualidad se acepta que la luz está compuesta de fotones cuyo comportamiento tiene propiedades tanto de onda como de partícula. La pequeña región del espectro electromagnético que el ojo humano es capaz de percibir recibe el nombre de espectro visible o espectro óptico de la luz. La radiación electromecánica que abarca el rango de longitudes de onda que las personas pueden percibir se denomina comúnmente luz visible o simplemente luz. Si bien no se puede establecer con precisión, la fisiología del ojo humano normalmente permite percibir longitudes de onda entre 400 nm y 700 nm, aunque algunas personas tienen la capacidad de percibir el espectro desde longitudes de onda que van desde 380 nm hasta 780 nm. La mayor parte del espectro de ondas no puede ser percibido por los humanos. Por debajo del espectro visible se encuentran las ondas correspondientes a la luz ultravioleta, rayos-x, rayos-gama mientras que las ondas por encima de las visibles corresponden a la luz infrarroja, microondas y ondas de radio.

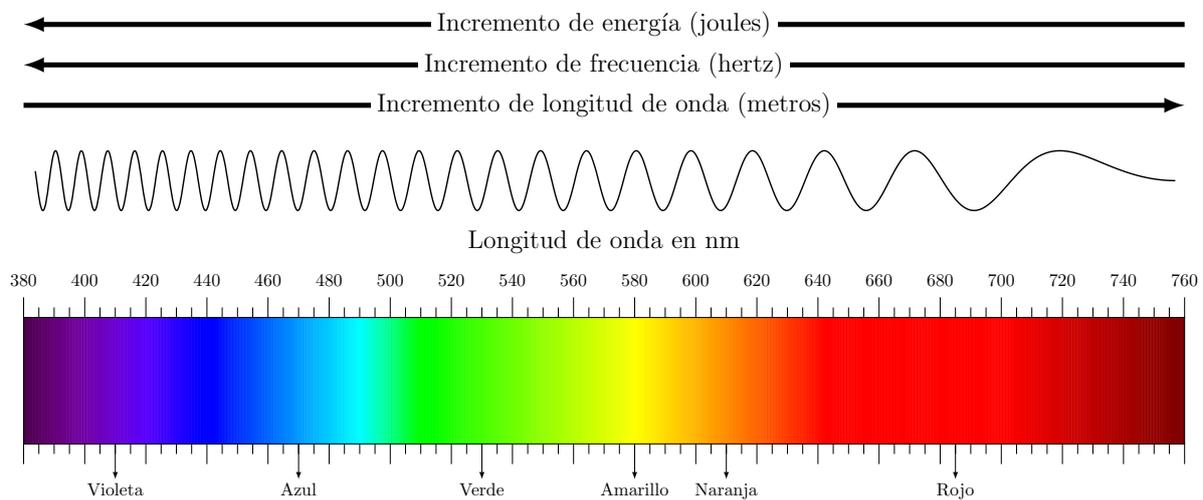


Figura 2.2: Espectro de luz visible.

En el arcoíris están incluidos todos los colores puros o monocromáticos (violeta, azul, celeste, verde, amarillo, naranja y rojo). Estos colores forman parte del grupo de colores espectrales por ser producidos por la luz visible de una sola longitud de onda.

Algo importante para destacar es que el espectro visible no contiene todos los colores que el ojo humano tiene la capacidad de distinguir. Colores sin saturar como el rosa, marrón o variaciones del púrpura como el magenta, no pueden reproducirse mediante una sola longitud de onda sino mediante la combinación de dos o más luces monocromáticas. Estos colores forman parte del grupo de colores no espectrales.

La luz, además de medirse por su longitud de onda, también puede medirse por su frecuencia y por su energía. La frecuencia del espectro de luz visible cubre el rango que va desde los 400 thz para el color rojo hasta los 750 thz para el color violeta. La cantidad de energía contenida en cada onda está relacionada proporcionalmente a la frecuencia. Cuanto mayor es la energía de una onda, mayor es su frecuencia y viceversa.

2.1.2. Círculo Cromático

El círculo cromático, también llamado rueda de colores o círculo de colores, es una representación visual abstracta y ordenada que muestra la relación cromática (matiz o tono) de los colores dispuesta en forma de círculo. Los colores se ordenan en la misma secuencia que los colores del arcoíris y el espacio que queda entre el rojo y púrpura se cubre con transiciones de magenta. Los colores primarios, secundarios y terciarios se ubican dentro del círculo de manera equidistante y los colores complementarios se ubican en el extremo opuesto de cada color primario.

El uso del círculo cromático es compatible tanto con los modelos sustractivos de color (artísticos, pictóricos), como con los modelos aditivos (lumínicos).

Los círculos cromáticos pueden representarse escalonados o continuos (ver figura 2.3). Las representaciones escalonadas pueden tener una cantidad de colores arbitraria, por lo general suelen tener 6, 12, 24, 48 o incluso más escalones.

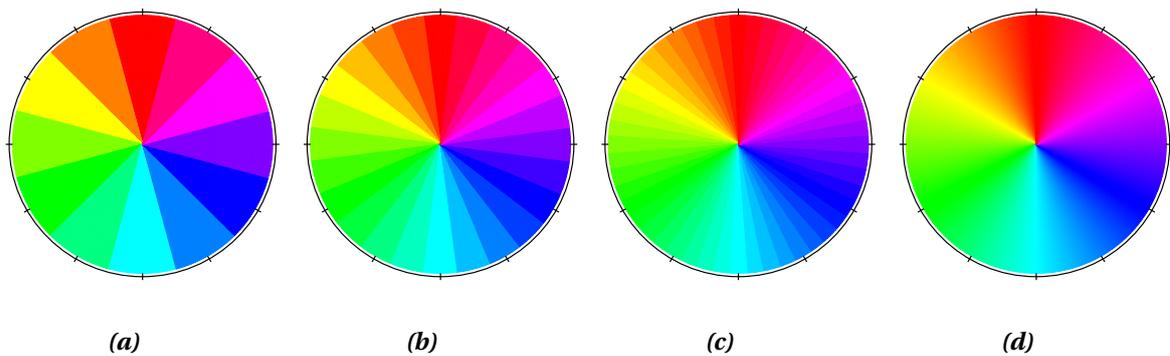


Figura 2.3: a) Círculo cromático escalonado de 12 colores. b) Círculo cromático escalonado de 24 colores. c) Círculo cromático escalonado de 48 colores. d) Círculo cromático continuo de colores

2.1.3. Síntesis Aditiva y Sustractiva

El proceso de formación o reproducción del color se realiza a través de la síntesis aditiva o de la síntesis sustractiva. Los modelos de color basados en síntesis o mezcla aditiva requieren una fuente de luz directa que forma los colores antes que un objeto la refleje. En este tipo de sistema se combinan ondas de luz para formar los colores. En la ausencia de todas las ondas del espectro de luz, el color percibido por el observador es negro, mientras que en la presencia de todas las ondas del espectro de luz el color percibido por el observador es blanco.

En la figura 2.4 a) se puede ver un esquema de la formación de colores con síntesis aditiva, donde los colores primarios son el rojo, el verde y el azul, mientras que los colores secundarios son el Magenta, Cyan y Amarillo.

Los modelos de color basados en síntesis o mezcla sustractiva requieren de algunos tipos de colorantes y de la luz reflejada. Un color sustractivo requiere de un objeto que refleje la luz como

un sustrato de papel o lienzo blanco, y utiliza colorantes como pigmentos o tintas, para sustraer parte de la luz que retorna al observador. Cuando el sustrato se colorea, parte del espectro de luz es absorbido por la tinta y el resto es reflejado. Estas combinaciones hacen que el observador perciba colores diferentes. Si por la ausencia de colorantes el sustrato refleja todas las ondas del espectro de luz, entonces el color percibido por el observador es blanco. Si la presencia de los colorantes cubre todo el espectro de la luz, ninguna onda de ésta será reflejada y en consecuencia, el color percibido por el observador es negro.

En la figura 2.4 b) se puede ver un esquema de la formación de colores con síntesis sustractiva, donde los colores primarios en los sistemas sustractivos son el Magenta, Cyan y Amarillo, mientras que los colores secundarios son el rojo, el verde y el azul.

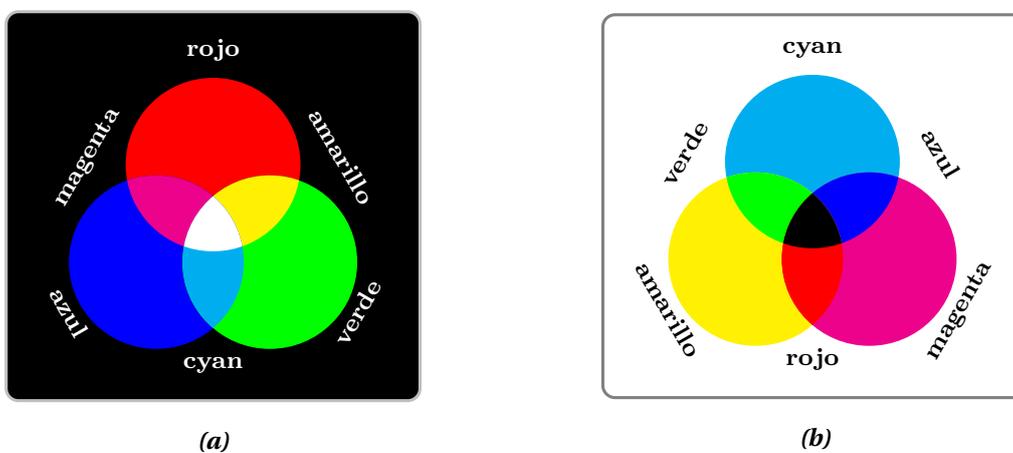


Figura 2.4: a) Esquema de formación de colores del sistema aditivo presentes en pantallas de computadoras. b) Esquema de formación de colores del sistema sustractivo presente en impresoras

Algo para resaltar es que el proceso de formación del color por síntesis sustractiva es el proceso natural que permite que todos podamos percibir los colores de los objetos que nos rodean.

2.1.4. Propiedades del color

Las propiedades del color son atributos que describen las características que hacen único a cada color. Estos atributos son tres y pueden recibir nombres diferentes con significados ligeramente distintos según el contexto, estos son el matiz, la saturación y la luminosidad.

Matiz

El matiz, también denominado tono, tinte, tonalidad o por su nombre en inglés *hue*, es la propiedad fundamental que nos permite diferenciar un color de otro y darle un nombre que permite identificarlo. La longitud de onda y el matiz están altamente correlacionados, tal como aparece en el arcoiris.

Saturación

La saturación representa la intensidad cromática o el grado de pureza de un color. La saturación de los colores cambia a medida que ese color tiene más o menos cantidad de gris. A mayor cantidad de gris, el color es más pálido y menos saturado. A menor cantidad de gris, es más vivo y más saturado.

Comúnmente suele confundirse el término *saturación* con el término *croma* ya que ambos se utilizan para indicar el grado de diferencia entre un estímulo cromático y otro acromático. Estos términos contrastan en que el *croma* requiere que los estímulos tengan el mismo brillo mientras que para la *saturación* el brillo no se toma en cuenta.

Luminosidad

La luminosidad, también denominada valor, luminancia o brillo, es la cantidad de luz reflejada por una superficie en comparación con la reflejada por una superficie blanca en iguales condiciones de iluminación. Este término se utiliza para describir el grado de claridad u oscuridad de un color, y se refiere a la cantidad de luz percibida. A mayor luminosidad de un color mayor luz reflejará y se percibirá más claro.

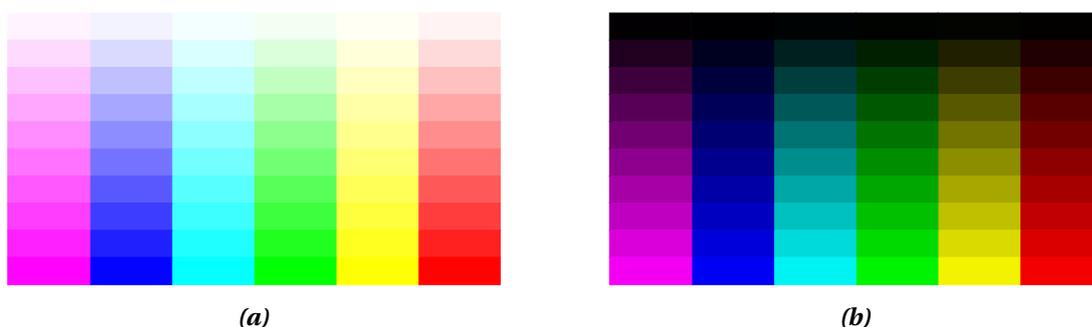


Figura 2.5: a) Variación de saturación a lo largo en los colores primarios y secundarios. b) Variación de luminosidad a lo largo en los colores primarios y secundarios.

2.1.5. Modelo de Color

Un modelo de color es un modelo matemático abstracto, utilizado para describir la forma en la que los colores son representados a través de una lista ordenada de valores. La mayoría de estos valores de color tienen tres o cuatro componentes diferentes, los cuales, al combinarse, ofrecen un espectro de colores. En consecuencia, cada modelo de color diferente compone una sección específica del espectro de color visible. En la práctica no existe un modelo de color digital que abarque todo el espectro electromagnético de luz visible.

Gamut de Color

El término gamut se refiere al subconjunto de colores o la porción del espacio de color del espectro visible que puede ser representado por un modelo o dispositivo de salida (monitor, televisor, impresora, etc.). Este concepto surge debido a la existencia de limitaciones físicas que imposibilitan que cualquier proceso de generación de color reproduzca por completo el espacio de color visible por el ser humano.

Espacio de Color

Se denomina espacio de color al conjunto resultante de los colores que pueden reproducirse utilizando todas las combinaciones de valores posibles de los componentes que forman el color para un modelo de color específico.

2.2. Modelos de Color Básicos

En la actualidad hay una importante cantidad de modelos que establecen la forma de describir o de reproducir un color. Estos modelos permiten definir colores utilizando valores numéricos y representaciones físicas. Las siguientes secciones resumen los espacios de color comúnmente utilizados en los artículos analizados para representar el color de la piel.

2.2.1. Modelo de Color RGB

Este modelo es el más extendido debido a que es el modelo de color que usa la tecnología para representar, almacenar y mostrar imágenes digitales. Hoy en día dispositivos como computadoras, tabletas, teléfonos móviles, cámaras de video, televisores y escáneres utilizan este sistema de manera parcial o total para representar los colores.

Su origen se basa en la teoría de visión de color tricromática de Young-Helmholtz, desarrollada por Thomas Young y Hermann Helmholtz durante la primera mitad del siglo XIX y el triángulo de color que posteriormente elaboró James Maxwell sobre dicha teoría. La elección de los colores primarios (rojo, verde y azul) de este sistema está relacionada a la fisiología del ojo humano debido a que las células del cono de la retina humana son mayormente sensibles a las longitudes de onda de luz correspondientes a los colores rojo, verde y azul.

Este modelo de color está basado en los tres colores primarios de la síntesis aditiva. Su nombre proviene de las iniciales en inglés de los componentes que se utilizan para formar el color, R es por red (rojo), G es por green (verde) y B es por blue (azul).

El espacio de color generado por este modelo representa un cubo donde cada coordenada está asociada a una proporción de rojo, verde y azul en el color asociado a cada punto dentro del mismo cubo (figura 2.6). En los vértices del cubo se encuentran los colores primarios y secundarios

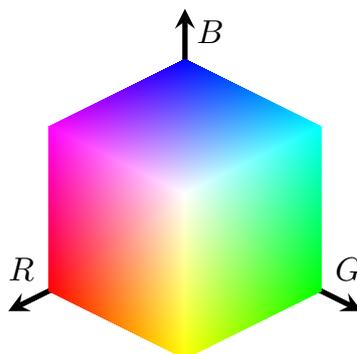


Figura 2.6: Geometría del espacio de color del modelo RGB

de la síntesis aditiva. Sobre la diagonal principal del cubo se encuentran los puntos con iguales intensidades en sus componentes que representan una escala de grises. Esta diagonal comienza con el color negro (cero intensidad en sus componentes RGB) y avanza aclarando los colores en su recorrido hasta llegar al color blanco (intensidad máxima en todos los componentes RGB).

Debido a que en este modelo la información de cromaticidad y luminosidad está mezclada en todos sus canales, no resulta ser el más apropiado para utilizar con algoritmos que reconocen colores. Por este motivo, suelen aplicarse transformaciones tanto lineales como no lineales para obtener otros modelos de representación de color a partir de este.

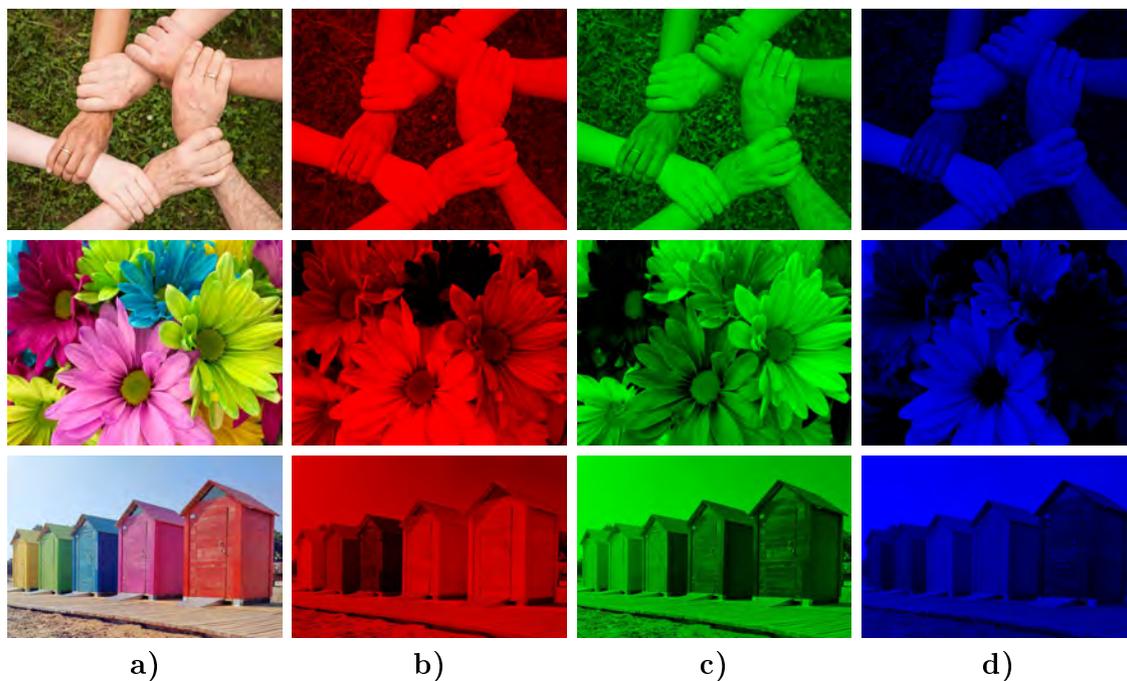


Figura 2.7: a) Imágenes originales. b) Imágenes del canal rojo. c) Imágenes del canal verde. d) Imágenes del canal azul

En la figura 2.7 se pueden observar tres fotografías con sus respectivas descomposiciones en los canales rojo, verde y azul del modelo RGB.

2.2.2. Modelo de Color RGB Normalizado

El modelo RGB normalizado es una variante del modelo RGB que pretende reducir la dependencia de la luminosidad. En este espacio de color, cada componente del modelo RGB es convertido en una versión normalizada de forma tal que la suma de sus componentes sea igual a uno ($R + G + B = 1$). De esta manera el modelo queda representado por sólo dos componentes cromáticas (comúnmente R y G) ya que la tercera componente (comúnmente B) es redundante debido a que puede inferirse con las dos anteriores ($B = 1 - R - G$). De ser necesario, en aplicaciones prácticas, este modelo puede complementarse con un canal adicional que contemple la componente de variación de la luminosidad (figura 2.8 c).

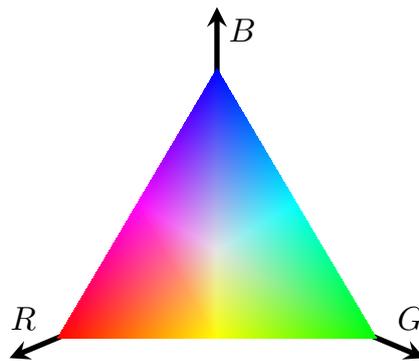


Figura 2.8: Geometría del espacio de color del modelo RGB normalizado.

Conversión RGB a RGB normalizado

La conversión al modelo RGB normalizado a partir del modelo RGB se realiza dividiendo cada componente por la suma de los 3 componentes. La siguiente fórmula muestra la conversión:

$$R' = \frac{R}{R + G + B} \quad G' = \frac{G}{R + G + B} \quad B' = \frac{B}{R + G + B} \quad (2.1)$$

En la figura 2.9 se pueden observar tres fotografías y sus correspondientes versiones en el sistema de color RGB normalizado junto con un canal de luminosidad.

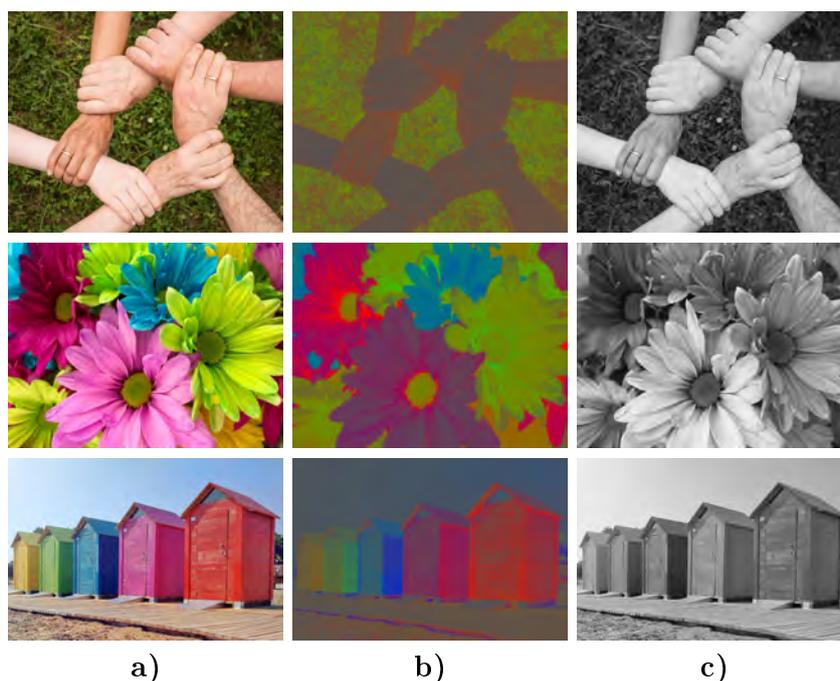


Figura 2.9: a) Imágenes originales. b) Imágenes del canal cromático RGB normalizado. c) Imágenes del canal adicional con luminosidad.

2.3. Espacios de Color Perceptivos

Los espacios de color perceptivos son aquellos en donde una variación en el valor de color produce una variación de igual relevancia en la percepción visual del observador. En la década de los '70 investigadores del área de la computación gráfica desarrollaron modelos alternativos al RGB buscando acercarse a la forma en que las personas perciben los atributos del color. De esta manera surgieron los modelos de color HSV (Hue, Saturation, Value), HSL (Hue, Saturation, Lightness) y HSI (Hue, Saturation, Intensity) para ofrecer al observador una experiencia más intuitiva y natural del color que el modelo de color RGB.

Las características perceptivas del color como matiz, saturación y luminosidad no pueden ser descritas de manera directa por el modelo de color RGB. Por este motivo, se han desarrollado muchas transformaciones no lineales con el objetivo de representar las características perceptivas del color a partir del modelo RGB. En las siguientes secciones se describen algunas de estas transformaciones.

2.3.1. Modelo de Color HSV

El modelo de color HSV (del inglés Hue, Saturation, Value) o HSB (del inglés Hue, Saturation, Brightness) caracteriza al color en términos de Matiz (Hue) que representa la esencia del color, Saturación (Saturation) que representa la pureza del color y la intensidad (Value, Brightness) que representa el brillo del color.

El matiz en el modelo HSV se representa con un valor numérico que corresponde a un desplazamiento angular alrededor del eje central en el espacio de color del sistema. Cada círculo que comprende todas las coordenadas de un mismo plano es similar al círculo cromático de la sección 2.1.2 y mantiene una percepción intuitiva del color para el observador. Normalmente, los valores típicos para el matiz pueden ir del rango de 0 a 1 o de 0° a 360°.

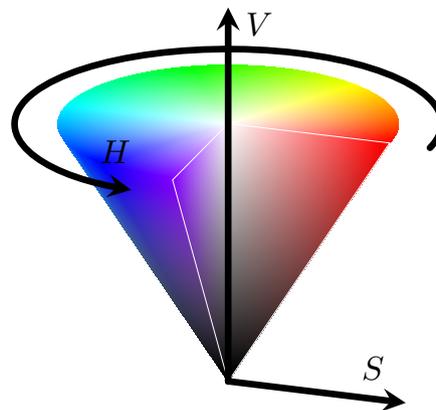


Figura 2.10: Geometría del espacio de color del modelo HSV.

La dimensión correspondiente a la saturación representa la pureza del color. Normalmente toma valores dentro del rango de 0 a 1 o de 0 a 100. A medida que el valor de saturación se acerca a su valor mínimo, el color se torna más puro o vivo (saturado). A medida que el valor de saturación se acerca a su valor máximo, el color se torna gris (no saturado).

El valor en el modelo HSV representa la luminosidad del color y describe qué tan oscuro es éste. Normalmente toma valores en el rango de 0 a 1 o de 0 a 100. A medida que el valor de luminosidad disminuye, el color se percibe más oscuro hasta tornarse negro. A medida que el valor de luminosidad incrementa el color se percibe más claro o brillante.

En la figura 2.11 se pueden observar tres fotografías con sus respectivas descomposiciones en los canales matiz, saturación y luminosidad del modelo HSV.

Conversión RGB a HSV

A continuación, se muestran las ecuaciones para transformar el espacio de color del modelo RGB al modelo HSV:

$$H = \cos^{-1} \left(\frac{(R - G) + (R - B)}{\sqrt{(R - G)^2 + (R - B)(G - B)}} \right) \quad (2.2)$$

$$S = \frac{MAX(R, G, B) - MIN(R, G, B)}{MAX(R, G, B)} \quad (2.3)$$

$$V = MAX(R + G + B) \quad (2.4)$$

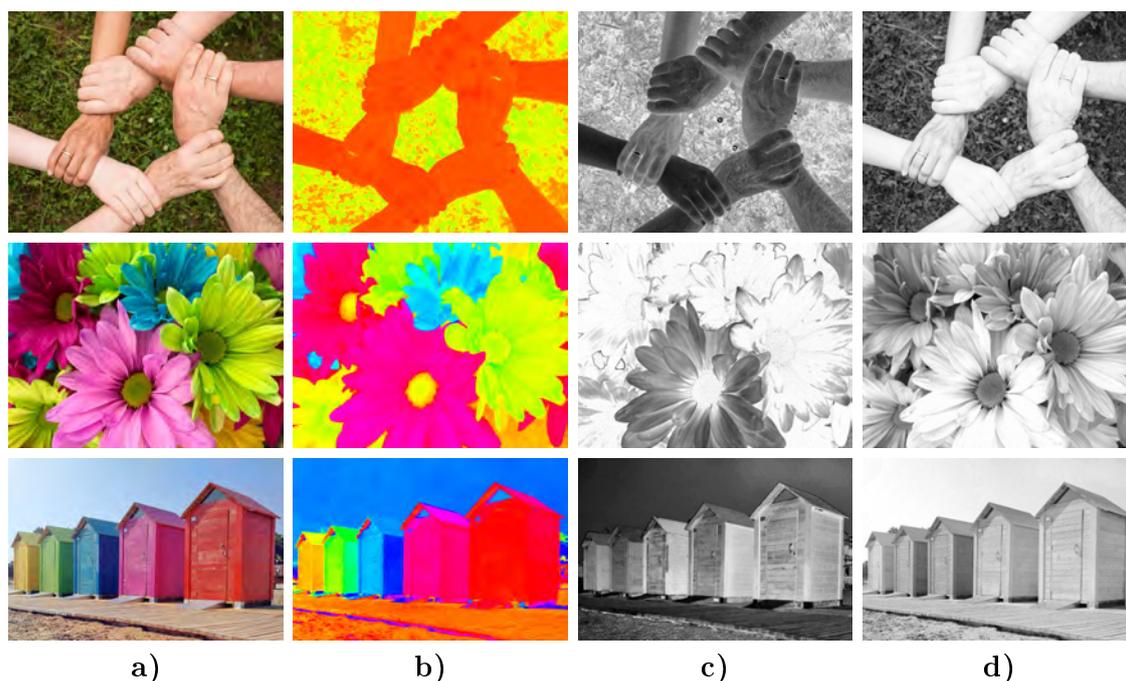


Figura 2.11: a) Imágenes originales. b) Imágenes del canal de matiz H . c) Imágenes del canal de saturación S . d) Imágenes del canal de luminosidad V .

2.3.2. Modelo de Color HSL

El modelo de color HSL (del inglés Hue, Saturation, Lightness) es similar al modelo HSV pero refleja mejor la noción intuitiva de la saturación y la luminosidad como dos parámetros independientes.

En el modelo de color HSL la componente del matiz guarda la misma relación que en el modelo de color HSV, pero varían las componentes de saturación y luminosidad. Respecto de la componente de saturación, en el modelo HSL ésta va desde completamente saturada hasta un gris equivalente, mientras que en el modelo HSV (con V al máximo), va desde completamente saturada hasta un blanco total, lo que no es muy intuitivo.

Respecto de la componente de luminosidad, el modelo de color HSL tiene simetría en el eje luz-oscuridad, lo que no sucede con el modelo HSV. Es por esto que el espacio geométrico de este modelo se muestra con un bi-cono (ver figura (2.12)). Normalmente, esta componente va desde el negro hasta el blanco pasando por todas las tonalidades de gris de una manera más intuitiva para el observador. En este sentido la componente de valor del modelo HSV se queda a mitad de camino ya que sólo refleja la variación entre el negro y la tonalidad escogida.

En la figura 2.13 se pueden observar tres fotografías con sus respectivas descomposiciones en los canales matiz, saturación y luminosidad del modelo HSL.

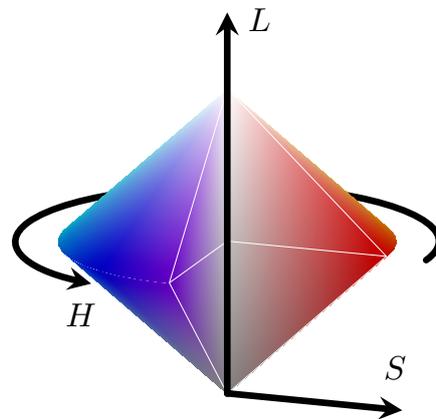


Figura 2.12: Geometría del espacio de color del modelo HSL.

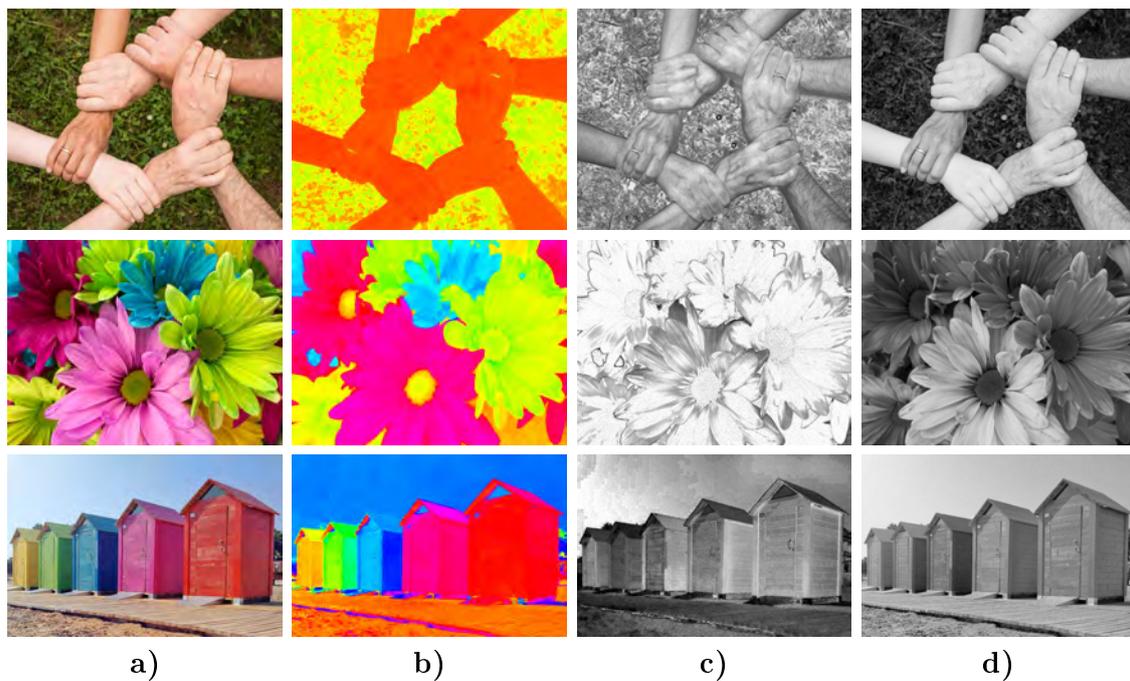


Figura 2.13: a) Imágenes originales. b) Imágenes del canal de matiz **H**. c) Imágenes del canal de saturación **S**. d) Imágenes del canal de luminosidad **L**.

Conversión RGB a HSL

A continuación, se muestran las ecuaciones para transformar el espacio de color del modelo RGB al modelo HSV:

$$H = \cos^{-1} \left(\frac{(R - G) + (R - B)}{\sqrt{(R - G)^2 + (R - B)(G - B)}} \right) \quad (2.5)$$

$$S = \frac{\text{MAX}(R, G, B) - \text{MIN}(R, G, B)}{\text{MAX}(R, G, B) + \text{MIN}(R, G, B)} \quad (2.6)$$

$$L = \frac{\text{MAX}(R, G, B) + \text{MIN}(R, G, B)}{2} \quad (2.7)$$

2.3.3. Modelo de Color HSI

El modelo de color HSI (del inglés Hue, Saturation, Intensity) es similar a los modelos HSV y HSL. Los tres modelos coinciden en la definición de la componente de matiz, pero se diferencian en la manera que calculan las componentes de saturación e intensidad.

La intensidad va desde el negro al blanco pasando por la escala de grises al igual que el modelo HSL, con la diferencia que este eje de luminosidad coincide con el eje transversal del modelo RGB. Por este motivo se define como el promedio de los componentes del modelo RGB.

Debido a la geometría del eje de la luminosidad, la saturación se define en función de la distancia radial de este eje. De esta manera valores iguales de saturación forman curvas concéntricas alrededor del eje de iluminación.

En la figura 2.14 se pueden observar tres fotografías con sus respectivas descomposiciones en los canales matiz, saturación e Iluminación del modelo HSI.

Conversión RGB a HSI

$$H = \cos^{-1} \left(\frac{(R - G) + (R - B)}{\sqrt{(R - G)^2 + (R - B)(G - B)}} \right) \quad (2.8)$$

$$S = 1 - \frac{3 \cdot \text{MIN}(R, G, B)}{R + G + B} \quad (2.9)$$

$$I = \frac{R + G + B}{3} \quad (2.10)$$

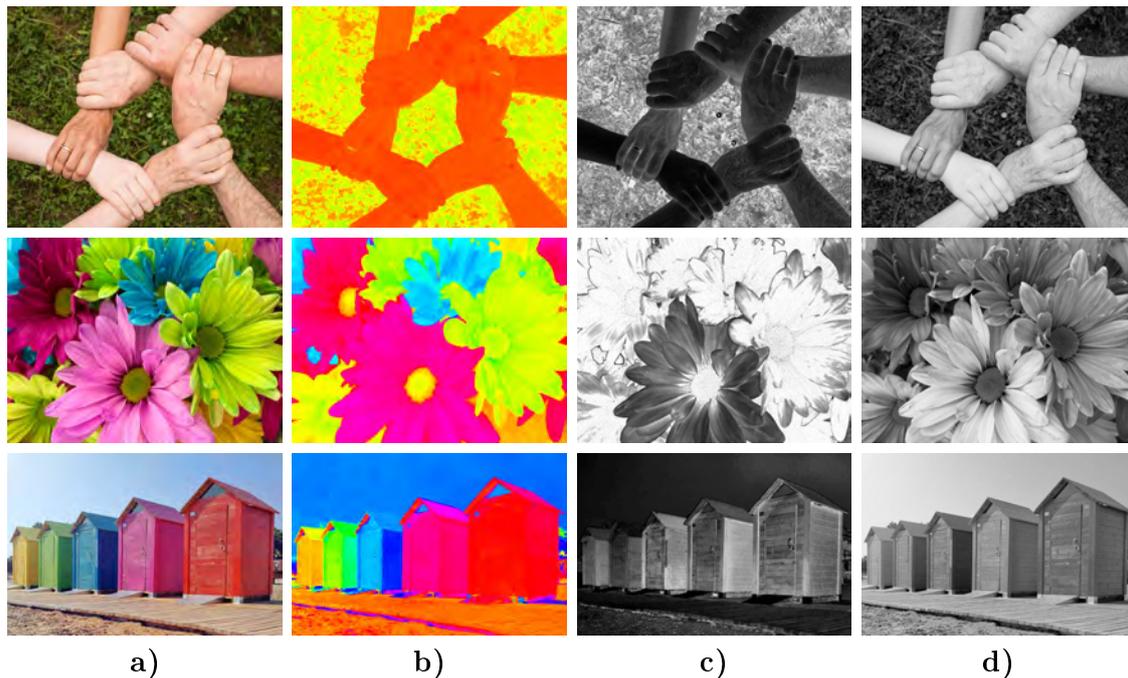


Figura 2.14: a) Imágenes originales. b) Imágenes del canal de matiz H . c) Imágenes del canal de saturación S . d) Imágenes del canal de luminosidad I .

2.4. Espacios de Color Ortogonales

Los espacios de colores ortogonales reducen la redundancia presente en el espacio de color RGB y representan al color con componentes estáticamente independientes, separando iluminación y cromaticidad. Una ventaja de estos espacios sobre los espacios de color perceptivos como el HSV es que la transformación desde RGB es más simple y en consecuencia tiene un costo computacional menor.

2.4.1. Modelo de Color YUV

La codificación YUV fue utilizada originalmente por el estándar de video analógico PAL (del inglés Phase Alternation Law). Esta codificación mantuvo la compatibilidad con los sistemas anteriores que sólo utilizaban un componente monocromático. La información de color fue añadida por separado mediante la modulación de una subportadora, de modo que un televisor blanco y negro pudiera recibir una transmisión de color en el formato monocromático propio del equipo. En este modelo de color el componente Y contiene información de luminosidad mientras que los componentes U y V contienen información sobre el color. La componente U codifica información de color en el rango del azul mientras que el canal V codifica información de color en el rango del rojo. Normalmente el rango de valores para Y está en el rango 0 a 1, mientras que el de U está en el rango -0,436 a 0,436 y el de V está en el rango -0,615 a 0,615.

Conversión RGB a YUV

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0,299 & 0,587 & 0,114 \\ -0,147 & -0,289 & 0,436 \\ 0,615 & -0,515 & -0,100 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2.11)$$

$$Y = 0,299 \cdot R + 0,587 \cdot G + 0,114 \cdot B \quad (2.12)$$

$$U = -0,147 \cdot R - 0,289 \cdot G + 0,436 \cdot B \quad (2.13)$$

$$V = 0,615 \cdot R - 0,515 \cdot G - 0,100 \cdot B \quad (2.14)$$

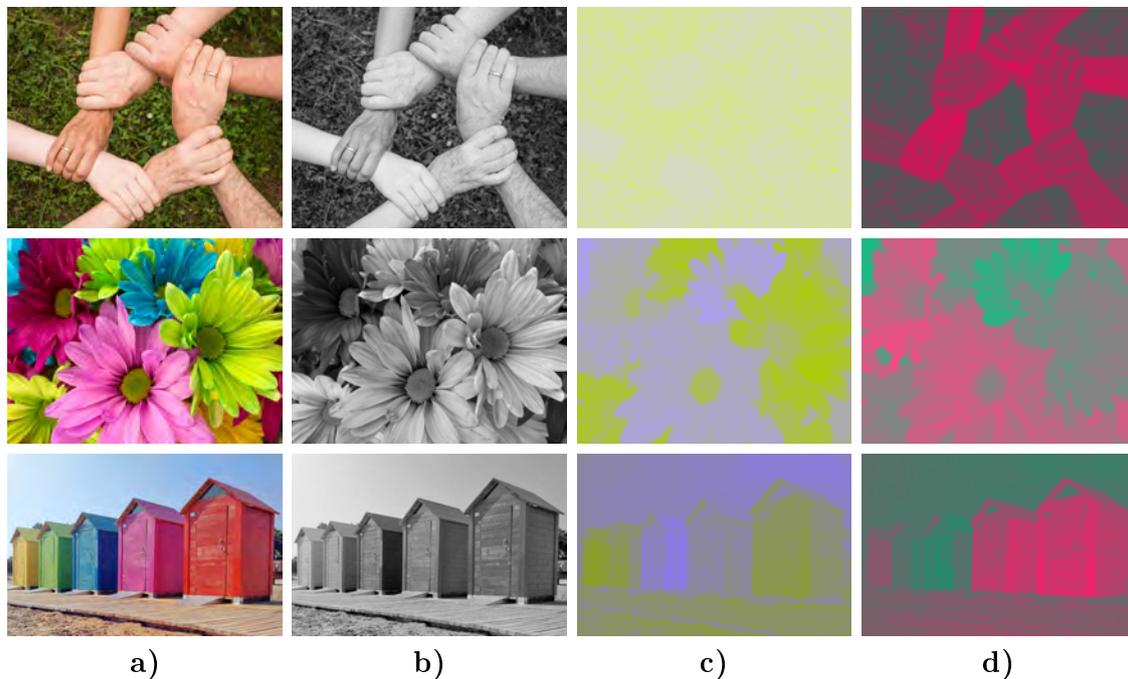


Figura 2.15: a) Imágenes originales. b) Imágenes del canal de luminosidad Y . c) Imágenes del canal de cromaticidad U . d) Imágenes del canal de cromaticidad V .

2.4.2. Modelo de Color YIQ

Este modelo de representación del color fue diseñado para separar la información de cromaticidad de la de iluminación. Utilizado por el estándar NTSC, este modelo surgió como un requerimiento en la época en que aparecieron los televisores a color para mantener la compatibilidad con los televisores blanco y negro. El componente Y contiene información de iluminación mientras que el componente I (del inglés In-phase) contiene información de color en el rango del naranja al azul y el componente Q (del inglés Quadrature) contiene información del color en el rango del púrpura al verde.

El plano IQ es una rotación de 33° respecto del plano UV del modelo YUV . Esta transformación permitió una mejor codificación de los tonos de piel ya que las personas percibían con mayor facilidad la falta de fidelidad de dichos tonos de color en otros sistemas.

En la figura 2.16 se pueden observar tres fotografías con sus respectivas descomposiciones en los canales Y , I y Q del modelo YIQ .

Conversión RGB a YIQ

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0,299 & 0,587 & 0,114 \\ 0,596 & -0,274 & -0,322 \\ 0,212 & -0,523 & 0,311 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2.15)$$

$$Y = 0,299 \cdot R + 0,587 \cdot G + 0,114 \cdot B \quad (2.16)$$

$$I = 0,596 \cdot R - 0,274 \cdot G - 0,322 \cdot B \quad (2.17)$$

$$Q = 0,212 \cdot R - 0,523 \cdot G - 0,311 \cdot B \quad (2.18)$$

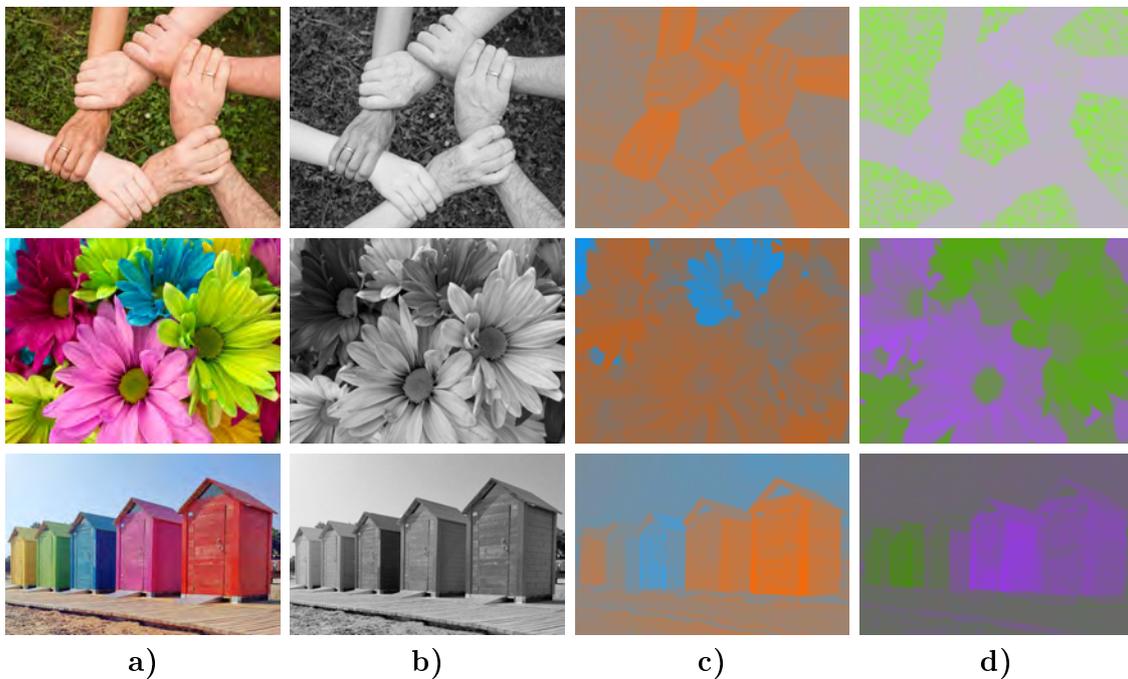


Figura 2.16: a) Imágenes originales. b) Imágenes del canal de luminosidad Y . c) Imágenes del canal de cromaticidad I . d) Imágenes del canal de cromaticidad Q .

2.4.3. Modelo de Color YCbCr

El modelo de color YC_bC_r debe su popularidad a su capacidad de compresión. Es usado comúnmente en el dominio del video digital. Debido a que esta representación facilita "deshacerse" de información del color redundante, es utilizado en estándares de compresión de imágenes y video como JPEG, MPEG1, MPEG2, MPEG4. El modelo de color YC_bC_r es una versión escalada y desplazada del plano de coordenadas UV del modelo de color YUV . Está compuesto por un único componente de luminosidad Y y dos componentes de cromaticidad C_b y C_r . El componente Y es la suma pesada de los valores de rojo, verde y azul, y coincide con el componente de luminosidad del estándar CIE. El componente C_b representa la diferencia entre el componente azul y el valor de referencia Y . El componente C_r representa la diferencia entre el componente rojo y el valor de referencia Y . Comúnmente, los valores del espacio de color YC_bC_r admiten valores en el rango 0 a 255 o de 0 a 1.

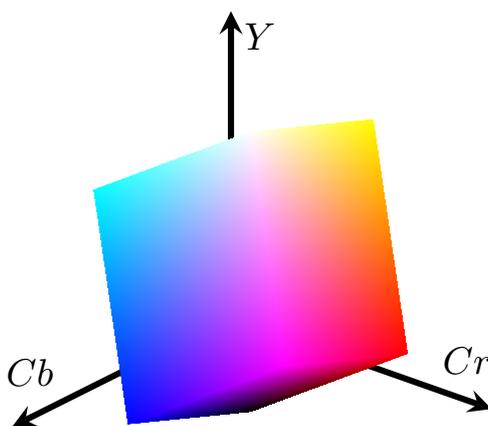


Figura 2.17: Geometría del espacio de color del modelo YCbCr.

Conversión RGB a YCbCr

$$\begin{bmatrix} Y \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0,299 & 0,587 & 0,114 \\ -0,169 & -0,331 & 0,500 \\ 0,500 & -0,419 & -0,081 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} \quad (2.19)$$

$$Y = 0,299 \cdot R + 0,587 \cdot G + 0,114 \cdot B + 0 \quad (2.20)$$

$$C_b = -0,169 \cdot R - 0,331 \cdot G + 0,500 \cdot B + 128 \quad (2.21)$$

$$C_r = 0,500 \cdot R - 0,419 \cdot G - 0,081 \cdot B + 128 \quad (2.22)$$

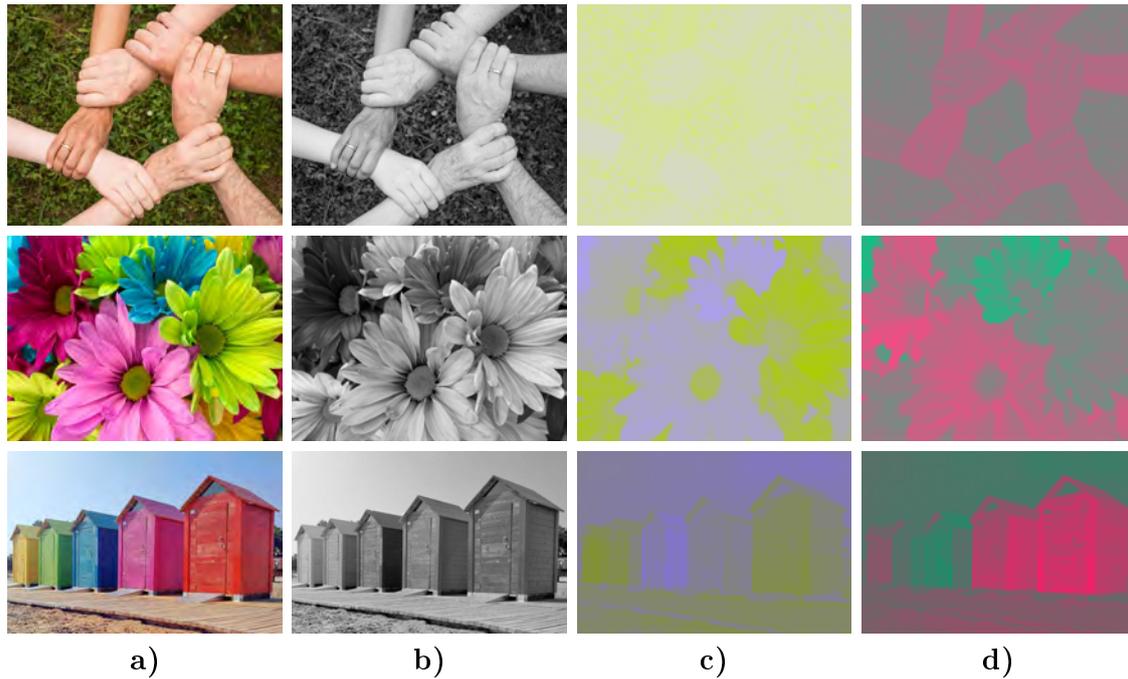


Figura 2.18: a) Imágenes originales. b) Imágenes del canal de luminosidad Y . c) Imágenes del canal de cromaticidad C_b . d) Imágenes del canal de cromaticidad C_r .

2.4.4. Modelo de Color $YCgCr$

Este espacio de color es una variación del espacio de color $YCbCr$ introducida en [dDG03]. La diferencia es que reemplaza el canal de cromaticidad C_b por el canal C_g , que consiste en la resta del canal Y al canal G del modelo RGB ($G - Y$). Este canal incluye información de color del canal G , que es más significativa para la representación del color de la piel que el canal C_b .

Conversión RGB a $YCgCr$

$$\begin{bmatrix} Y \\ C_g \\ C_r \end{bmatrix} = \begin{bmatrix} 65,481 & 128,553 & 24,966 \\ -81,085 & 112,000 & -30,915 \\ 112,000 & -93,786 & -18,214 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \quad (2.23)$$

$$Y = 65,481 \cdot R + 128,553 \cdot G + 24,966 \cdot B + 16 \quad (2.24)$$

$$C_g = -81,085 \cdot R + 112,000 \cdot G - 30,915 \cdot B + 128 \quad (2.25)$$

$$C_r = 112,000 \cdot R - 93,7689 \cdot G - 18,214 \cdot B + 128 \quad (2.26)$$

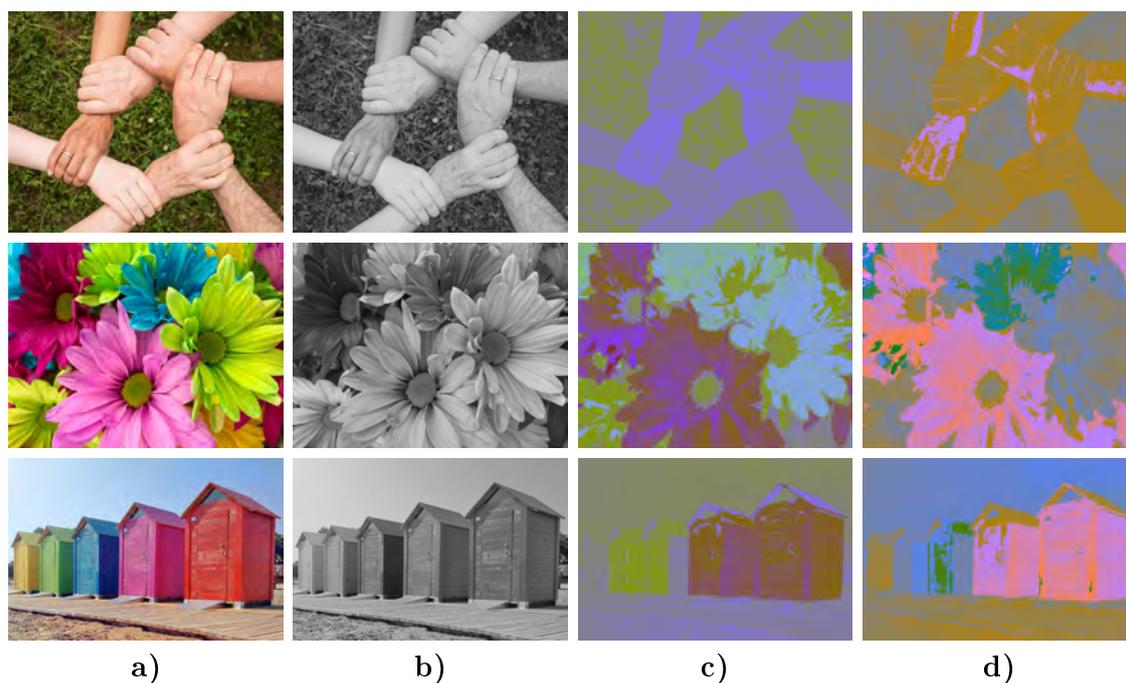


Figura 2.19: a) Imágenes originales. b) Imágenes del canal de luminosidad Y . c) Imágenes del canal de cromaticidad Cg . d) Imágenes del canal de cromaticidad Cr .

2.5. Espacios de Color Perceptiblemente Uniformes

Se dice que los espacios de color son perceptiblemente uniformes cuando una variación en el valor de color produce una variación de igual importancia visual para el observador humano. Para obtener esta percepción uniforme se requieren transformaciones de gran costo computacional. En los espacios de color perceptiblemente uniformes, el cálculo de la luminosidad y de la cromaticidad se obtienen a través del mapeo no lineal de coordenadas del espacio de color $CIE - XYZ$. Dos ejemplos de modelos de color perceptiblemente uniformes que frecuentemente se utilizan son los modelos $CIE - LAB$ y $CIE - LUV$.

2.5.1. Modelo de Color CIE-LAB

En estos modelos de representación del color la luminosidad y la cromaticidad se obtienen con una transformación no lineal del espacio de color $CIE - XYZ$. El componente L representa la luminosidad y los demás componentes (A, B y U, V) representan la cromaticidad.

En la versión $CIE - LAB$ de este modelo, L representa la luminosidad, A oscila entre verde y rojo y B oscila entre azul y amarillo.

Este modelo de color tiene como característica que define colores imaginarios que no pueden ser reproducidos en el mundo físico. La gama de colores definidas en el espacio de color Lab excede la gama de colores de los modelos RGB y $CMYK$ utilizados en monitores e impresoras,

respectivamente. La gama de colores que define es mayor a la que la visión humana puede percibir.

El objetivo principal del modelo *LAB* es definir un modelo independiente del dispositivo en el que se muestra. Es un espacio de color puramente teórico que suele ser utilizado como un estándar para comparar otros espacios de color.

En la figura 2.20 se pueden observar imágenes con los distintos canales que componen el modelo de color *LAB*.

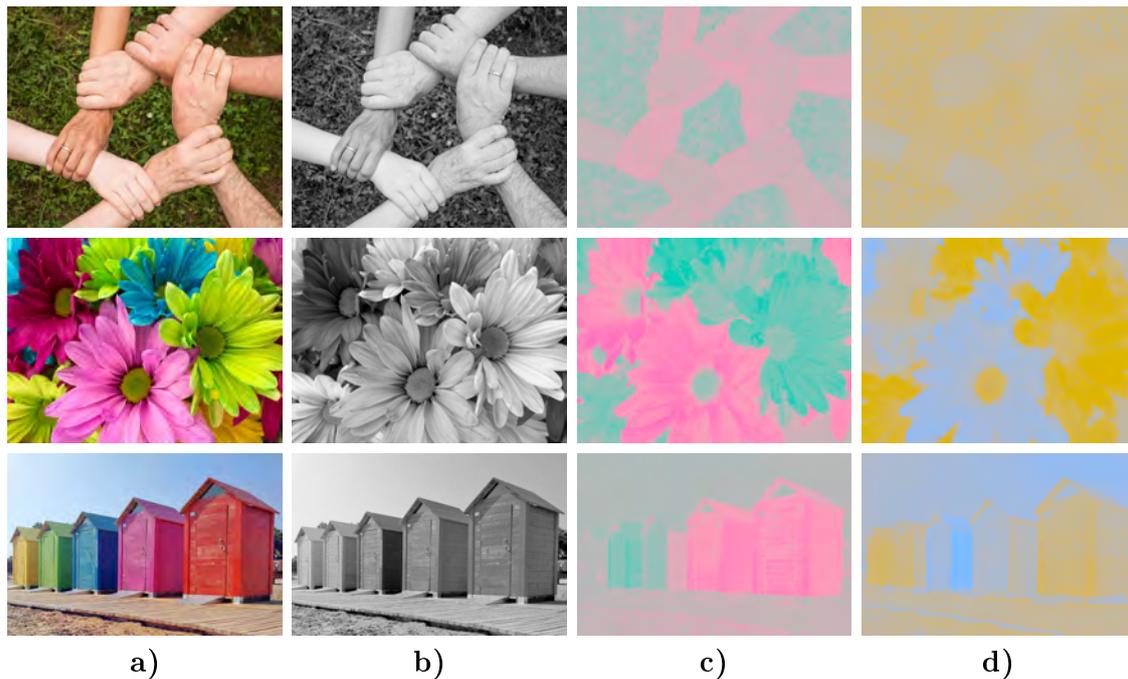


Figura 2.20: a) Imágenes originales. b) Imágenes del canal de luminosidad *L*. c) Imágenes del canal de cromaticidad *A*. d) Imágenes del canal de cromaticidad *B*.

Conversión XYZ a Lab

$$L = \begin{cases} 116 \cdot \sqrt[3]{\frac{Y}{Y_n}} - 16 & \text{si } \frac{Y}{Y_n} > 0,008856 \\ 903,3 \cdot \frac{Y}{Y_n} & \text{si } \frac{Y}{Y_n} \leq 0,008856 \end{cases} \quad (2.27)$$

$$A = 500 \cdot \left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right) \right] \quad (2.28)$$

$$B = 200 \cdot \left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right) \right] \quad (2.29)$$

donde:

$$f(t) = \begin{cases} \sqrt[3]{t} & \text{si } t > 0,008856 \\ 7,787 \cdot t + \frac{16}{116} & \text{si } t \leq 0,008856 \end{cases}$$

(X_n, Y_n, Z_n) = coordenada del punto de referencia del color blanco

Conversión RGB a XYZ

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0,4125 & 0,3576 & 0,1804 \\ 0,2127 & 0,7152 & 0,0722 \\ 0,0193 & 0,1192 & 0,9502 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2.30)$$

ALGORITMOS PARA SEGMENTACIÓN

Resumen

En éste Capítulo se describen en detalle los algoritmos utilizados para hacer una segmentación basada en píxeles sobre la piel y que son incluidos por los artículos de estudio. Estos están agrupados por categorías y subcategorías según sus características.

3.1. Modelos No Paramétricos

Los métodos o modelos no paramétricos se caracterizan por no asumir ninguna hipótesis sobre las características subyacentes de la distribución de los datos para construir un modelo. A diferencia de los modelos paramétricos, tienen la capacidad de representar con fidelidad cualquier función de distribución de probabilidad. Ejemplos típicos de estos modelos son los Basados en Umbrales, la Cuantificación de Vectores (VQ) y la Máquina de Soporte Vectorial (SVM).

3.1.1. Basados en Umbrales

En el contexto de procesamiento de imágenes, los métodos de umbralización tienen como objetivo separar dos objetos que comúnmente reciben la denominación de *fondo* y *objeto*. A través de un valor de umbral que se aplica a una imagen, se obtiene una máscara (imagen binaria) de un objeto de interés para su posterior procesamiento.

Frecuentemente el término de umbralización se encuentra asociado al área de procesamiento de imágenes, los conceptos asociados a dicho término se pueden generalizar y aplicar a otras áreas. De

manera general las técnicas de umbralización pueden utilizarse básicamente para dividir en dos un conjunto de datos.

La característica que hace popular a los métodos tradicionales de umbralización es que ofrecen un bajo costo computacional y son de implementación rápida y sencilla. En condiciones de aplicación favorables permiten obtener buenos resultados con relativamente poco esfuerzo. No obstante, cuando estas condiciones no se dan o se dan parcialmente, la eficacia de estos métodos se degrada y deben ser reemplazados por métodos más robustos y de mayor costo computacional.

Un problema que tienen los métodos de umbralización es que cuando el umbral se aleja del valor “óptimo” incluye muchos datos que no corresponden al objeto de interés (por debajo) y quedan pocos datos del objeto de interés cuando (por arriba).

Métodos de Umbralización Global

Los métodos basados en umbrales pueden clasificarse en globales y locales. Los métodos de umbralización global utilizan un único valor que puede ser de naturaleza estática o dinámica para separar el espacio de datos. Esta estrategia puede resultar muy útil cuando el objeto a separar presenta una distribución relativamente uniforme en el espacio de datos. Como la umbralización global tiene en cuenta el valor de un dato sin considerar la relación con otros datos, no hay garantía de que dos datos identificados como el objeto de interés correspondan realmente a esta situación. Fácilmente pueden incluirse datos que no corresponden y excluirse datos que corresponden. Además, este efecto empeora cuando los datos a segmentar contienen algún tipo de ruido.

Métodos de Umbralización Local

Los métodos de umbralización local utilizan un valor diferente según el área del espacio donde se aplica. Normalmente establece el valor de umbral por medio de un algoritmo adaptativo o dinámico que explora un área particular o analiza una característica local de los datos. La idea en que se basan estos métodos es que diferentes subconjuntos del total de los datos están localmente relacionados, por lo que en vez de aplicar un umbral a todos los datos se aplica un umbral individual para cada subconjunto. Una característica interesante es que los métodos de umbralización global pueden adaptarse fácilmente a métodos de umbralización local, dividiendo el espacio de datos en subconjuntos donde se aplica el mismo algoritmo para calcular el umbral sobre estos datos.

Método de Umbralización de Otsu

Nobuyuki Otsu desarrolló en 1979 un método automático de umbralización [Ots79] basado en técnicas estadísticas con el objetivo de segmentar un objeto dentro de una imagen en escala de grises. En concreto, utiliza la distribución de probabilidad, la media y la varianza de las intensidades del histograma normalizado de la imagen para calcular un umbral óptimo mediante un algoritmo iterativo.

El método de Otsu establece segmentos dentro del histograma de la imagen a binarizar y obtiene un valor del umbral que por un lado minimiza la dispersión dentro de cada segmento, pero por otro lado maximiza la dispersión entre segmentos diferentes. A partir de la premisa de que existen dos clases separables (fondo y objeto) cada una con su función de distribución de probabilidad, se calcula el cociente de ambas varianzas y se exploran distintos valores de umbrales en busca del óptimo que maximiza dicho cociente.

Formalmente, la varianza pesada de las 2 clases a ser separadas (objeto y fondo) esta dada por la siguiente ecuación:

$$\sigma_w^2(t) = q_1(t) \cdot \sigma_1^2(t) + q_2(t) \cdot \sigma_2^2(t) \quad (3.1)$$

donde:

- La probabilidad de cada clase es estimada con:

$$q_1(t) = \sum_{i=1}^t P(i) \quad q_2(t) = \sum_{i=t+1}^N P(i) \quad (3.2)$$

- La media de cada clase es estimada con:

$$\mu_1(t) = \sum_{i=1}^t \frac{i \cdot P(i)}{q_1(t)} \quad \mu_2(t) = \sum_{i=t+1}^G \frac{i \cdot P(i)}{q_2(t)} \quad (3.3)$$

- La varianza de cada clase es determinada con:

$$\sigma_1^2(t) = \sum_{i=1}^t [i - \mu_1(t)]^2 \cdot \frac{P(i)}{q_1(t)} \quad \sigma_2^2(t) = \sum_{i=t+1}^G [i - \mu_2(t)]^2 \cdot \frac{P(i)}{q_2(t)} \quad (3.4)$$

A partir de la ecuación 3.1, el algoritmo desarrollado por Otsu obtiene un umbral óptimo que maximiza la dispersión total y permite segmentar un objeto dentro de una imagen en escala de grises. Formalmente, la ecuación que obtiene dicho umbral es la siguiente:

$$T = \operatorname{argmax}_{0 < t \leq N} (\sigma_w^2(t)) \quad (3.5)$$

Las ventajas de usar este algoritmo residen en que es relativamente rápido y eficaz para imágenes con iluminación relativamente uniforme y con contraste aceptable entre los píxeles que representan el fondo y los que representan al objeto a separar. En condiciones menos favorables suelen aplicarse métodos más sofisticados, pero más costosos para mejorar la calidad de la segmentación.

3.1.2. Cuantificación Vectorial (VQ)

La Cuantificación Vectorial (VQ, Vector Quantization) es un proceso que consiste en mapear un espacio con un gran número de vectores en un número reducido de regiones dentro del mismo espacio. Cada una de estas regiones recibe el nombre de *cluster* (agrupamiento) y es representada por un vector principal que recibe el nombre de centroide. A la colección de centroides se la denomina

libro de códigos o *codebook*. El resultado de la aplicación de VQ produce una cantidad de información significativamente menor respecto de la cantidad de vectores de entrada originales. La principal ventaja de esta transformación es la reducción en la cantidad de cómputo que se realiza en etapas posteriores de procesamiento. No obstante, a pesar de que el número de centroides del *codebook* es muy reducido, aún representa de manera significativa las características de los datos originales.

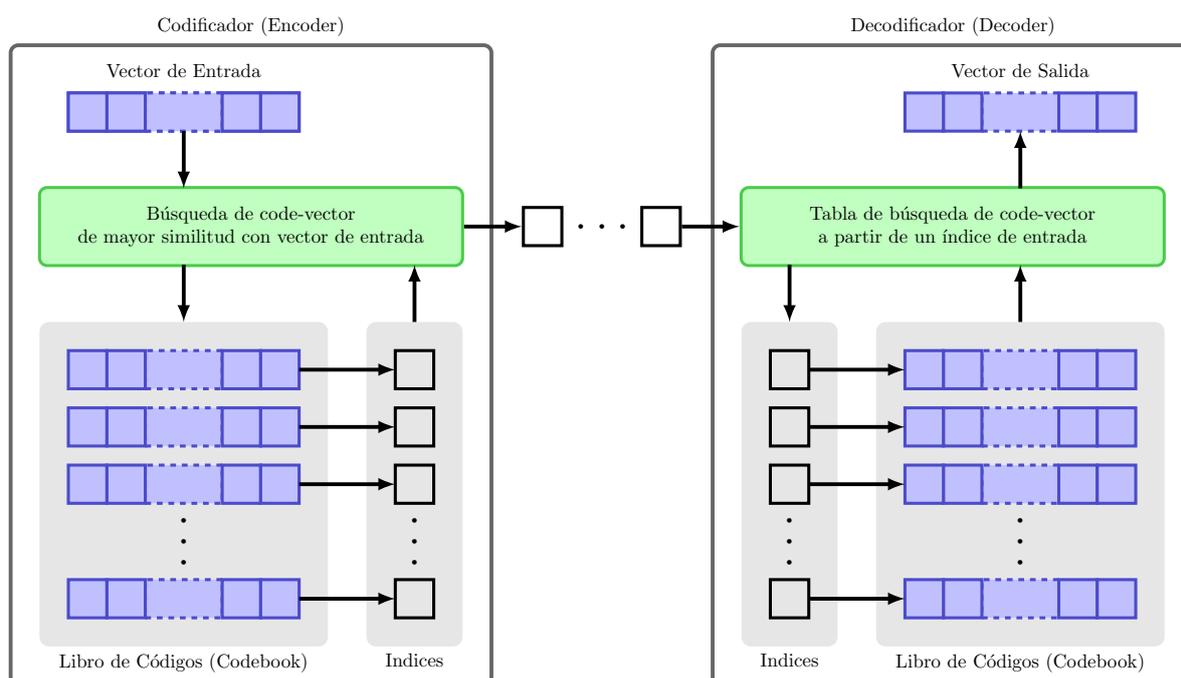


Figura 3.1: Esquema codificador/decodificador utilizado en cuantificación vectorial.

VQ requiere de dos fases para su funcionamiento. La primera fase es la de entrenamiento y que consiste en la aplicación de un algoritmo de agrupamiento (*clustering*) como k-medias o LBG-VQ para generar un *codebook* específico para cada grupo a identificar. La segunda fase es la de reconocimiento donde se compara un vector o conjunto de vectores de un grupo a reconocer con todos los grupos de referencia existentes en *codebooks*. Para realizar la comparación entre dos grupos se utiliza una estimación de distorsión basada en alguna medida de distancia como la Euclídea o la de Chebichev.

En la actualidad, el modelo basado en la Cuantificación de Vectores ha sido desplazado por el Modelo de Mezcla de Gaussianas y el Modelo Oculto de Markov. No obstante, es importante destacar que es un método que reduce el volumen de información a procesar y que presenta características de robustez cuando el conjunto de datos disponible sobre un grupo es escaso para construir otro tipo de modelo.

Aplicaciones

VQ es una técnica utilizada en contexto muy disímiles. En general se aplica a problemas de compresión de datos con pérdida, corrección de pérdida de datos, reconocimiento de patrones, estimación de densidad y agrupamiento de datos. En particular se ha utilizado para almacenamiento y transmisión de imágenes (ej.: médicas y satelitales) y de señales de audio. También han sido utilizadas para el reconocimiento del habla y de locutores independiente del texto [ZZGL08, ZME09, TtSx13] y en el reconocimiento de rostros[KO02, AM07, CKLO08]. También ha sido utilizada en aplicaciones de esteganografía (técnica para ocultar mensajes u objetos dentro de otros).

3.1.3. K Vecinos Cercanos (KNN)

K Vecinos más Cercanos (en inglés K-Nearest Neighbor o KNN) es uno de los algoritmos más simples del aprendizaje automático. Es una técnica ampliamente utilizada tanto en clasificación supervisada como estimación y predicción. Este algoritmo se basa en un método de aprendizaje basado en instancias (lazy learning) donde la función objetivo es aproximada de manera local y donde todo el cómputo de la misma es diferido hasta el momento de procesar un nuevo ejemplo. Utiliza una medida de distancia o similitud para determinar los k ejemplos o vecinos más próximos (de ahí su nombre). Estos k vecinos son tomados sobre un conjunto de datos conocidos que suele ser denominado conjunto de entrenamiento, aunque para KNN no hay una etapa de entrenamiento explícito como lo hay en otros algoritmos.

Entre las ventajas que ofrece KNN se pueden mencionar:

- robustez frente a datos dinámicos porque se adapta a los cambios rápidamente.
- permite aprender funciones objetivo complejas.
- no hay pérdida de información ya que su cálculo se basa en los ejemplos de entrenamiento originales.

Entre las desventajas de KNN cabe mencionar:

- su sensibilidad a la estructura local de los datos.
- requiere tiempo de cómputo al consultar por un nuevo ejemplo, ya que difiere su cálculo.
- sensibilidad a atributos con información irrelevante.
- requiere de una buena cantidad de ejemplos de entrenamiento.

Como se mencionó anteriormente, KNN es un método que puede utilizarse tanto para clasificación como regresión. En ambos casos, a partir de una instancia de entrada, se toman los k vecinos más próximos de los ejemplos de entrenamiento. No obstante, la respuesta del algoritmo se calcula de manera diferente dependiendo si se está utilizando para clasificación o para regresión.

En la clasificación la respuesta es la pertenencia a una clase. Un nuevo ejemplo será clasificado por voto mayoritario de los k vecinos más cercanos a éste. Al nuevo patrón se le asigna la clase más común de los k vecinos. Por ejemplo, si $k = 1$ la clase que se asignará será la del único vecino. La clase

de pertenencia de una nueva instancia x a clasificar se calcula como:

$$\operatorname{argmax}_{1 \leq j \leq \#C} \left(C_j = \sum_{i=1}^k f(n_i, x, j) \right) \quad (3.6)$$

donde:

- x es el patrón cuyo valor se quiere determinar.
- k es el número de vecinos de x a tomar en cuenta.
- $\#C$ es la cantidad de clases o grupos.
- n_i es la función que identifica al i -ésimo vecino más cercano a x .
- $f(n_i, x, j)$ vale 1 si $n_i(x)$ pertenece a la clase j y 0 en caso contrario.

En la regresión, la respuesta del algoritmo es un valor para el ejemplo de entrada. Este valor se calcula como el promedio de los valores de los k vecinos más cercanos a él. El valor de regresión de una nueva instancia x se calcula como:

$$\frac{1}{k} \cdot \sum_{i=1}^k \|x - n_i(x)\| \quad (3.7)$$

donde:

- x es el patrón para determinar su valor.
- k es el número de vecinos a tomar en cuenta.
- n_i es el i -ésimo vecino más cercano.

Tanto para clasificación como regresión, una técnica útil puede ser usada para asignar pesos en la contribución de cada vecino. De esta manera los vecinos más próximos contribuyen más al promedio que aquellos que están más distantes. Una estrategia de uso común consiste en dar a cada vecino un peso en función de su distancia (comúnmente $1/d$ o $1/d^2$), de manera que a mayor distancia menor influencia tendrá al momento de calcular el valor de salida o determinar la clase de pertenencia.

3.1.4. KMedias

K Medias (K-Means en inglés) es uno de los algoritmos de aprendizaje no supervisado que más se emplea para hacer agrupamiento (clustering) en una amplia gama de problemas. Es un procedimiento fácil y simple para clasificar un conjunto de datos en un número de grupos conocidos a priori.

La idea principal es definir K centros en el espacio de datos, uno por cada grupo. Inicialmente, estos centros deben elegirse de manera astuta debido a que diferentes ubicaciones iniciales provocan diferentes resultados. Por este motivo debe utilizarse alguna estrategia para elegirlos lo más alejados unos de otros. Una vez inicializados los centros se realiza algoritmo iterativo de 2 pasos. El primer paso es el de asignación y consiste en procesar cada elemento del conjunto de datos para asociarlo al centro del grupo más cercano. El segundo paso es el de actualización y consiste en calcular el baricentro o centroide para cada grupo. Finalmente, el centro de cada grupo es actualizado con el nuevo valor. A partir de este punto comienza a repetirse el proceso, procesando nuevamente los datos

y reasignándolos a los grupos y realizando un nuevo ajuste de los centros de los grupos. El proceso continúa hasta que no se producen cambios en los centros.

$$ECS = \operatorname{argmin} \left(\sum_{i=1}^k \sum_{j=1}^{\#G_i} \|x_j - \mu_i\|^2 \right) \quad (3.8)$$

donde:

- k es el número de grupos o clases.
- $\#G_i$ es el número de datos asociados al grupo G_i .
- x_j es el j -ésimo patrón asociado al grupo G_i .
- μ_i es la media de los datos asociados al grupo G_i .

3.1.5. Máquinas de Vectores de Soporte (SVM)

Un Máquina de Vectores de Soporte (SVM)[CV95] es un clasificador binario que construye un hiperplano óptimo para separar dos conjuntos de datos que pertenecen a dos clases diferentes. SVM es una implementación basada en el principio de minimización del riesgo estructural (SRM, Structural Risk Minimization), originado de la teoría de aprendizaje estadístico desarrollado por Vapnik y Chervonenkis. El algoritmo de entrenamiento de SVM mapea datos de entrada en un espacio de características de una dimensión mayor (ej.: de dos a tres dimensiones) y localiza un hiperplano que separe dichos datos, maximizando el margen entre las dos clases en este nuevo espacio (ver figura 3.2). A partir de este hiperplano la SVM puede discriminar datos con los que nunca fue entrenada y determinar la clase a la que pertenecen.

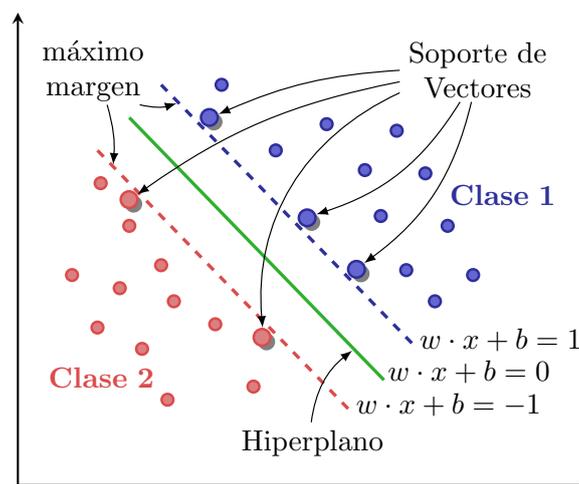


Figura 3.2: Esquema de un hiperplano que maximiza el margen de separación entre conjuntos de patrones pertenecientes a dos clases para una Máquina de Soporte Vectorial (SVM).

Las SVM intentan predecir a partir de lo ya conocido. Fueron presentadas en 1992 y adquirieron fama cuando dieron resultados muy superiores a las redes neuronales en el reconocimiento de letra manuscrita, usando píxeles como entrada.

Las SVM son un modelo exitoso, que ofrecen las siguientes ventajas:

- La minimización del riesgo estructurado le da una excelente capacidad de generalización.
- No requiere de una cantidad importante de datos y solo utiliza aquellos que proveen mayor información. Tiene pocos parámetros a ajustar.
- Evita la existencia de mínimos locales al estimar los parámetros con la optimización de una función de costo convexa.
- La solución resultante de SVM es característica de vectores raros (sparsed). El modelo resultante se expresa como una combinación de un número pequeño de vectores de entrada (denominados de soporte) ya que la mayoría de las variables de la solución tienen valor cero.

La función de decisión de una SVM para encontrar el hiperplano que maximice el margen entre las clases tiene la siguiente expresión:

$$L_{SVM}(U) = \sum_{j=1}^J y_j \alpha_j K(U_j, U) + b \quad (3.9)$$

donde:

- U_j es una muestra de entrenamiento con $j \in [1, J]$.
- y_j es 1 para la hipótesis H_0 o -1 para la hipótesis H_1 .
- α_j y b son coeficientes, resueltos con técnicas de programación cuadráticas.
- $K(U_j, U)$ es la función kernel del estilo de la figura 3.3.

La función kernel o núcleo K es un parámetro que permite convertir el espacio de entrada en uno de mayor dimensión cuando los datos no son linealmente separables. Ejemplos de funciones típicas se pueden observar en la figura 3.3.

$$\begin{array}{ll} K(x, y) = x \cdot y & \text{(a)} \\ K(x, y) = \exp(-\gamma |x - y|^2) & \text{(c)} \end{array} \quad \begin{array}{ll} K(x, y) = ((x \cdot y) + 1)^d & \text{(b)} \\ K(x, y) = \tanh(s(x \cdot y) + r) & \text{(d)} \end{array}$$

Figura 3.3: Funciones de Kernel comunes para SVM. a) Lineal. b) Polinomial. c) Base Radial (RBF). d) Sigmoidal

Como es un clasificador binario existen dos aproximaciones de solución para el reconocimiento:

- uno contra uno: para esta aproximación la SVM debe tener combinaciones de pares de locutores de manera de ir discriminando. El problema de esta aproximación es que para un total de n locutores se producen $n(n-1)/2$ combinaciones. Por ejemplo para $n = 100$ se tienen $100 * 99/2 = 4950$ combinaciones.
- uno contra todos: para cada locutor la SVM debe discriminar dicho locutor del resto. En este tipo de aproximación el costo computacional en la etapa de entrenamiento de la SVM es aproximadamente cuadrático [Pla98] en base al número de muestras utilizadas. Esto dificulta este tipo de aproximación si el número de locutores es alto.

Aplicaciones

SVM ha sido utilizada en las más diversas áreas de aplicación. En el ámbito de la seguridad se ha utilizado para el reconocimiento de rostros, del locutor [LLX08, DZSD08, RKRSRVK12] y de firmas, detección de intrusos y predicción del tráfico [PFS⁺07, GM12]. En bio-informática proteínas. En Medicina, detección de enfermedades relacionadas a la diabetes [RKP14], diagnóstico y pronóstico de cáncer de mamas. En geología potencial de licuefacción sísmica [GG07] y clasificación de terrenos. En clasificación general de datos, como recuperación de contenido a través de imágenes, clasificación imágenes [MB04], texturas y textos.

3.2. Modelos Paramétricos

Los modelos paramétricos presumen una clase particular de modelo estadístico que puede ser caracterizado por un conjunto de parámetros. Para cumplir con un requerimiento determinado, se define una estructura paramétrica que limita y establece la forma del modelo. Esto asegura un uso eficiente de los datos al momento de estimar los parámetros del modelo.

Una de las principales ventajas de este tipo de modelos es que los parámetros se establecen con facilidad a partir de los cambios en los datos. La principal desventaja es que a veces no pueden representar fielmente la distribución de probabilidad de los datos.

Representantes típicos de modelos paramétricos, que se desarrollarán en las secciones siguientes, son los Modelos de Mezclas de Gaussianas (GMM), los Modelos Ocultos de Markov (HMM) y las Redes Neuronales Artificiales (ANN).

3.2.1. Modelo de Mezclas Gaussianas (GMM)

El Modelo de Mezclas de Gaussianas (en inglés Gaussian Mixture Model o GMM) es un modelo probabilístico que asume que los datos de un universo pueden generarse a través de la combinación de un número finito de distribuciones Gaussianas de parámetros desconocidos. De esta manera es posible representar sub-poblaciones normalmente distribuidas dentro del total de una población. Una característica interesante de este tipo de modelo es que al no requerir de conocimiento previo sobre la pertenencia de los datos respecto de una sub-población permite realizar un aprendizaje automático y no supervisado.

El Modelo de Mezclas de Gaussianas ha sido utilizado para el modelado de características de datos de la voz en sistemas de reconocimiento del habla (ver ejemplo en 3.4). También ha sido muy usado en el seguimiento de múltiples objetos, donde el número de los componentes de mezcla y sus medias predicen la ubicación de los objetos en cada cuadro de una secuencia de video. Otra área en el que ha sido utilizado en la segmentación de imágenes

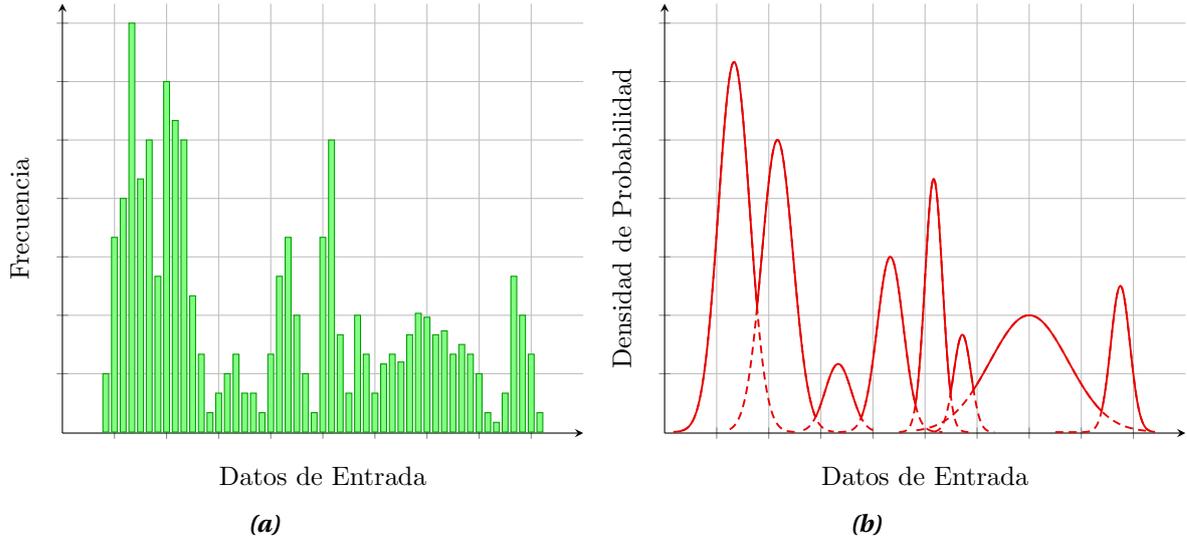


Figura 3.4: Modelo de Mezcla de Gaussianas (GMM). a) Gráfico de un histograma. b) Gráfico con la aproximación del histograma utilizando 8 componentes gaussianas para un espectro de audio.

Arquitectura

Un modelo de mezcla de gaussianas tiene como parámetros dos tipos de valores, los componentes de mezcla de los pesos y el componente de media y varianza/covarianza.

Formalmente, la distribución gaussiana de un vector \vec{x} de D dimensiones es definida por la siguiente expresión:

$$b_i(\vec{x}) = \frac{\exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu}_i)' \Sigma_i^{-1} (\vec{x} - \vec{\mu}_i)\right)}{\sqrt{(2\pi)^D |\Sigma_i|}} \quad (3.10)$$

donde:

- \vec{x} es un vector de evaluación espectral de dimensión D .
- $\vec{\mu}_i$ es el vector de medias.
- Σ_i es la matriz de covarianza.

En problemas de clasificación, la probabilidad de que un patrón pertenezca a un modelo estadístico de un clase particular está representado con una combinación lineal de distribuciones gaussianas. La expresión matemática de la probabilidad de una mezcla de M gaussianas se define como:

$$p(\vec{x}|\lambda) = \sum_{i=1}^M w_i b_i(\vec{x}) \quad (3.11)$$

donde:

- λ representa la clase a evaluar.
- \vec{x} es un vector de evaluación de dimensión D .
- M es el número de componentes gaussianas.
- w_i es el peso de la componente gaussiana i , con restricción $\sum_{i=1}^M w_i = 1$ y $0 \leq w_i \leq 1$.
- $b_i(\vec{x})$ es la función definida en la ecuación 3.10.

La función de densidad de probabilidad que modela a un clase puede aproximarse como una mezcla de M gaussianas tal que $\lambda = (\mu_i, \Sigma_i, p_i)_{i=1}^M$.

Aprendizaje de los Parámetros

Si el número de componentes gaussianos es conocido, la maximización de la expectativa (expectation maximization) es la técnica mayormente utilizada para estimar los parámetros del modelo de mezcla. En la teoría de probabilidad frecuentista, los modelos son típicamente entrenados con técnicas de estimación de máxima verosimilitud, que buscan maximizar la probabilidad, o verosimilitud, de los datos observados dados los parámetros del modelo. Desafortunadamente, encontrar la solución que maximiza la verosimilitud para los modelos de mezcla a través de la diferenciación de la verosimilitud logarítmica y resolverlo para w es usualmente imposible.

La maximización de la expectativa (EM, del inglés Expectation Maximization) es una técnica numérica para la estimación del máximo de verosimilitud, y es usualmente utilizada cuando las expresiones de forma cerrada se pueden calcular para actualizar los parámetros del modelo. Esta técnica se basa en un algoritmo iterativo y tiene la conveniente propiedad de que el máximo de verosimilitud de los datos es estrictamente creciente con cada iteración subsecuente, lo que garantiza que se aproximará al máximo local.

Maximización de Expectativa

El algoritmo de maximización de expectativa para el modelo de mezcla de gaussianas consiste de dos pasos: El primer paso, conocido como paso de expectativa o paso E, consiste en el cálculo de la expectativa de las asignaciones de los componentes de cada dato, dados los parámetros actuales del modelo. El segundo paso es conocido como paso de maximización o paso M, y consiste en la maximizar las expectativas calculadas en el paso E respecto de los parámetros del modelo. Estos parámetros obtenidos del paso M se utilizan para comenzar un nuevo paso E. Este proceso iterativo se repite hasta que el algoritmo converge y se obtiene una estimación del máximo de verosimilitud.

Aplicaciones

GMM es un modelo muy utilizado en una amplia variedad de aplicaciones. Una de las principales se encuentra en el área de procesamiento de señales, específicamente en sonido con aplicaciones de reconocimiento del locutor [CSR06], recuperación de información y clasificación de audio [TBTL08] y señales como las de radar, por ejemplo. También se ha utilizado ampliamente en el área de procesamiento de imágenes en la predicción de tráfico [OKH15], detección automática de

vehículos[CEV12], recuperación y segmentación de imágenes[GRG06], sustracción de fondo[TM07], clasificación de texturas.

3.2.2. Modelos Ocultos de Markov (HMM)

El Modelo Oculto de Markov (HMM, Hidden Markov Model) es una herramienta ampliamente utilizada en el modelado de procesos estocásticos. La base teórica de este modelo fue desarrollada en una serie de artículos publicados entre 1966[BP66] y 1972[Bau72] por Leonard Baum y su equipo de investigadores.

Un modelo convencional de Markov es un conjunto de estados conectados a través de arcos de transición. Cada estado se encuentra asociado a un evento observable y cada transición establece la probabilidad de cambio de un estado a otro. En un Modelo Oculto de Markov, en cambio, los estados no son directamente observables. Por un lado el modelo describe un proceso de probabilidad que produce una secuencia de símbolos observables. Por otro lado describe un proceso de probabilidad subyacente u oculto que no puede ser observado pero que influye sobre la secuencia de los eventos observados.

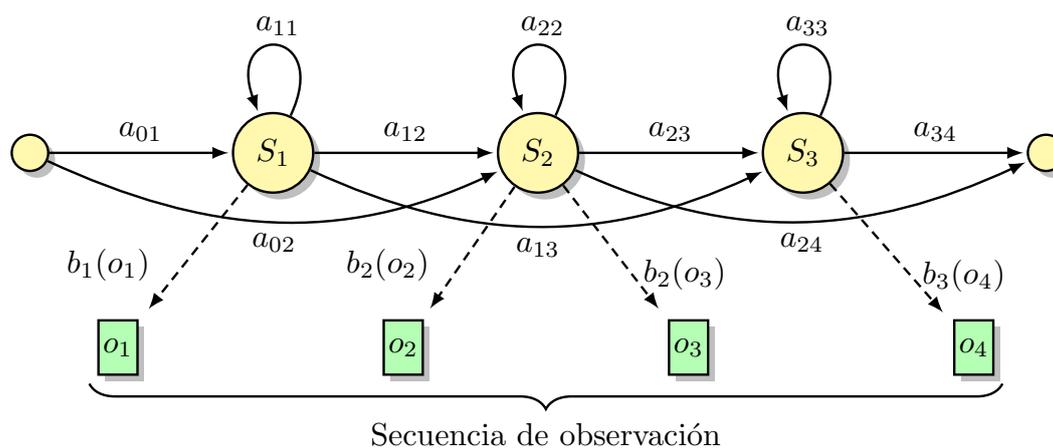


Figura 3.5: Elementos de la arquitectura general de un Modelo de Oculto de Markov (HMM).

En la figura 3.5 se pueden observar los distintos elementos que componen la arquitectura general de un Modelo Oculto de Markov. A continuación, se enumeran y detallan los elementos que permiten componer un modelo:

- Un conjunto de estados $S = \{S_1, S_2, \dots, S_N\}$, siendo N la cantidad de estados. Estos estados corresponden a una cadena de Markov de primer orden y aunque están ocultos, en las aplicaciones prácticas generalmente tienen un significado físico.
- Un conjunto de probabilidades de transición entre estados $A = \{a_{ij}\}$ con $i, j = 1, \dots, N$. Definiendo $a_{ij} = P(S_j|S_i)$ y verificando que $a_{ij} \geq 0$ y $\sum_{j=1}^N a_{ij} = 1$, con $i = 1, \dots, N$. Como estas probabilidades determinan si dos estados están conectados, en consecuencia, determinan la topología.

- La distribución de probabilidad de estados iniciales $\Pi = \{\pi_i\}$ con $i = 1, \dots, N$, verificando que $\pi_i \geq 0$ y $\sum_{i=1}^N \pi_i = 1$.
- La distribución de probabilidad de los símbolos o eventos de observación que caracterizan el proceso asociado a cada estado del modelo $B = \{b_j(O_k)\}$ con $j = 1, \dots, N$ y $k = 1, \dots, M$. Definiendo $b_j(O_k) = P(O_k|S_j)$ y O_k una observación de una secuencia de M observaciones.

Los HMM son especialmente aplicados a reconocimiento de formas temporales y por este motivo se han convertido en una herramienta muy utilizada en diversas áreas de aplicación. Como ejemplos de sus aplicaciones se pueden mencionar el reconocimiento del habla, de escritura manual y de gestos, traducción automática, etiquetado gramatical, alineación de cadenas de aminoácidos de proteínas y de ADN y pronóstico del tiempo, entre muchas otras.

3.3. Clasificador de Bayes Ingenuo

Bayes es una técnica de clasificación que se ha vuelto popular en el área de aprendizaje automático. Aprende con datos de entrenamiento y luego predice la clase de mayor probabilidad posterior en con los datos de prueba. Un *Clasificador de Bayes "Ingenuo"* (del inglés o NBC) es un algoritmo probabilista inductivo fundamentado en el teorema de Bayes que asume una fuerte hipótesis de independencia de los atributos o variables predictoras. Es por este motivo que recibe el apelativo de *ingenuo* (naive).

Si bien este clasificador se aplica en diversas áreas de aplicación, las que mas sobresalen son: dentro de la clasificación de texto la clasificación de documentos y detección de correo no deseado. En aplicaciones de diagnóstico médico se encuentran detección de Alzheimer, tumores cerebrales y enfermedades linfáticas y predicción de enfermedades cardíacas. También predicción del clima, de peligros sísmicos En el área social, se pueden mencionar la clasificación de sentimientos, análisis y clasificación de actividades humanas, predicción de rendimiento académico. También ha sido utilizado en el área de seguridad, realizando la detección de accesos anormales.

Para el algoritmo de Bayes Ingenuo una instancia puede verse como un vector x de n atributos, tal que $x = \langle x_1, x_2, \dots, x_n \rangle$. Aplicando la regla de la probabilidad condicional la probabilidad de $p(x, y)$ de una instancia x etiquetada como y puede expresarse como:

$$p(x, y) = p(y) \cdot p(x_1|y) \cdot \dots \cdot p(x_n|y, x_1, x_2, \dots, x_{n-1}) \quad (3.12)$$

Aplicando la fuerte asunción que la elección de cualquier atributo es condicionalmente independiente de la elección del valor de cualquier otro atributo dada una etiqueta, la probabilidad $p(x, y)$ se puede expresar como:

$$p(x, y) = p(y) \cdot \prod_{i=1}^n p(x_i|y) \quad (3.13)$$

A pesar de realizar la simplificación anterior, el algoritmo de Naive Bayes ha demostrado ser muy efectivo en un amplio rango de problemas de aprendizaje. Como ventaja adicional reduce el costo computacional de manera significativa.

El algoritmo estima $p(y)$ y $p(x_i|y)$ con la frecuencia relativa de los datos de entrenamiento. Dada una nueva instancia x , la predicción del clasificador será la etiqueta que maximice la probabilidad condicional:

$$p(y|x) = \frac{p(x, y)}{p(x)} \quad (3.14)$$

Ya que el denominador es constante a lo largo de la elección de la etiqueta, esta puede ser ignorada. Entonces la predicción del clasificador puede expresarse como:

$$C(x) = \arg \max_y \left(p(y) \cdot \prod_{i=1}^n p(x_i|y) \right) \quad (3.15)$$

3.4. Modelos de Redes Neuronales Artificiales (ANN)

3.4.1. Redes Neuronales Artificiales (ANN)

Las Redes Neuronales Artificiales han sido ampliamente usadas para resolver todo tipo de problemas de reconocimiento de patrones. Su capacidad para discriminar diferentes clases a partir de patrones ha sido explotada en el área de la segmentación de imágenes.

Entre los tipos de redes más utilizados se pueden mencionar el Perceptrón Multicapa (MLP), Redes de Función de Base Radial (RBF), Redes Neuronales time-Delay (TDNN) y Mapas auto-organizados (SOM).

3.4.2. Red de Perceptrones Multicapa

La Neurona Artificial

Una neurona artificial no es más que un modelo matemático desarrollado para simular la función de una neurona biológica. En 1947 Warren McCulloch and Walter Pitts presentaron un modelo neuronal que denominaron perceptrón [PM47] basado en el modelo neurona de un trabajo previo [MP43] realizado en 1943.

La estructura de una neurona artificial es similar a la de la neurona biológica. Esta recibe información a través una o más entradas de forma similar a las dendritas. Luego procesa los valores de entrada como sucede en el soma. Finalmente tiene una terminal de salida similar al axón que transmite el estímulo como resultado del procesamiento de las entradas.

En el modelo neuronal denominado perceptrón, cada entrada de una neurona artificial es ponderada de forma independiente y sumada para pasarla a una función de activación. Esta función suele cumplir con las propiedades de ser monótonamente creciente, continua, diferenciable y

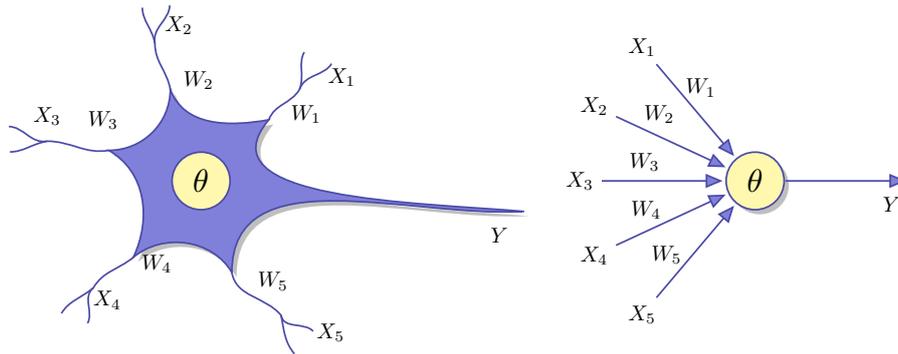


Figura 3.6: Esquema de una neurona biológica (izquierda) y una neurona artificial (derecha)

acotada. El modelo matemático de una neurona de este tipo puede representarse como:

$$a = f\left(b + \sum_{i=1}^q w_i \cdot p_i\right) \tag{3.16}$$

donde:

- q es la cantidad de entradas de la neurona.
- p_i es la entrada i -ésima de la neurona.
- w_i es el peso i -ésimo que pondera la entrada p_i .
- b es el bias.
- f es una función de activación.
- a la salida de la neurona.

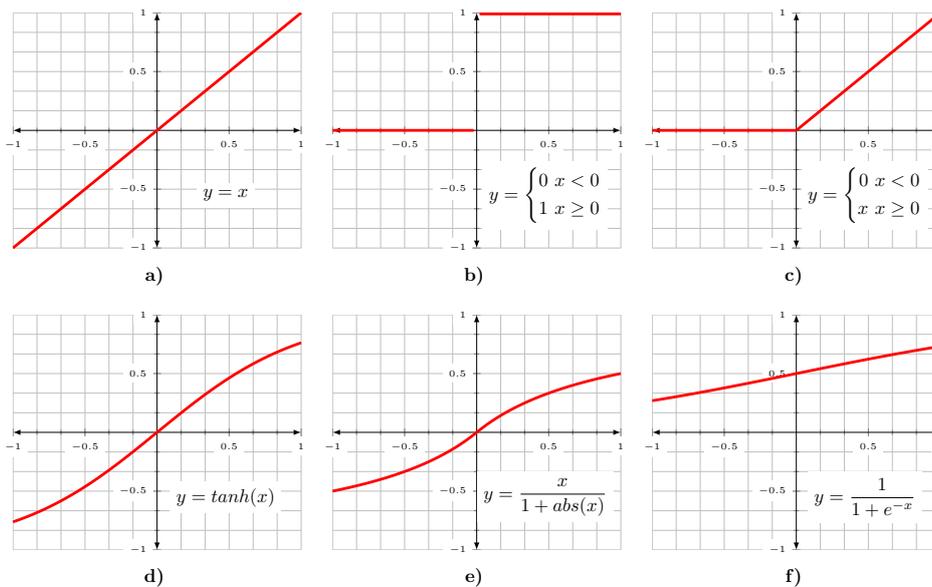


Figura 3.7: Funciones de activación de neuronas artificiales. a) identidad. b) paso binario. c) unidad lineal rectificada (ReLU). d) tangente hiperbólica. e) softsig. f) logística.

Arquitectura de la Red

Un perceptrón multicapa (multilayer perceptron) es un modelo donde las neuronas artificiales se encuentran interconectadas para formar una red. Debido a la manera en que se conectan, a este tipo de red se la denomina feed-forward, lo que significa que la información se propaga en un único sentido, desde una neurona origen o presináptica hacia una neurona destino o postsináptica sin retroalimentación alguna. La arquitectura de la red está formada por neuronas que se agrupan en tres tipos de capas según la función que desempeñan:

- **capa de entrada:** esta capa es la primera de la red y no realiza cómputo alguno. Su función es recibir los datos de entrada externos para transmitirlos a la siguiente capa de la red.
- **capa oculta:** este tipo de capa es interna a la red y no tiene contacto directo con el exterior. Es la capa que realiza la mayor parte del cómputo de la red y puede haber una o varias instancias de esta. Cuanto mayor sea la cantidad de capas mayor serán los grados de libertad y en consecuencia la capacidad de aprender. La primera capa oculta recibe la información de la capa de entrada, la procesa y genera nueva información de salida para que sea procesada por la siguiente capa.
- **capa de salida:** este tipo de capa toma la información de la última capa oculta y la procesa para generar la información de salida que se toma como respuesta de la red para el patrón provisto como entrada.

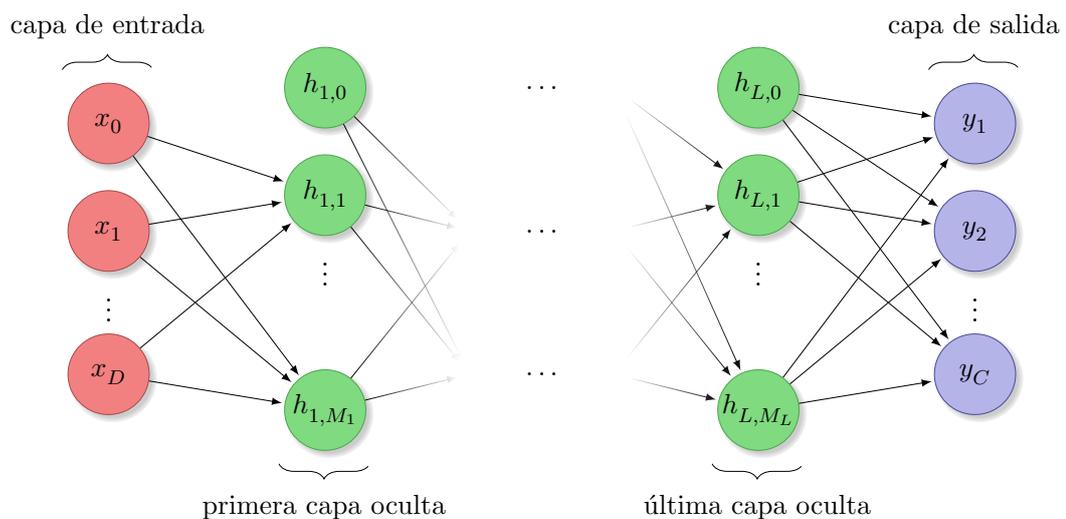


Figura 3.8: Arquitectura de una red de perceptrones multicapa.

Aprendizaje de la Red

Este tipo de red neuronal comenzó a comprobar su potencial en el año 1974 cuando Paul Werbos [Wer74] desarrolló la idea elemental del algoritmo de retro-propagación (back-propagation). A partir del año 1986, con la contribución de David Rumelhart, Geoffrey Hinton y Ronald Williams [RHW86], este algoritmo se consolida y comienza a utilizarse con éxito tanto en investigación como en desarrollo.

El algoritmo de retro-propagación realiza un cálculo para ajustar los pesos de las conexiones de todas las neuronas luego que un patrón de los datos de entrenamiento es procesado por la red. Se utiliza una generalización del algoritmo de mínimos cuadrados del perceptrón lineal para obtener un valor de ajuste basado en el error cometido por cada neurona al comparar la respuesta de la red con el resultado esperado.

El error en la salida de una neurona de la red sobre un patrón n se expresa mediante la siguiente ecuación:

$$\mathcal{E}(n) = \frac{1}{2} \cdot \sum_{j=1}^q \left(d_j(n) - y_j(n) \right)^2 \quad (3.17)$$

donde:

- q es la cantidad de entradas de la neurona.
- $d_j(n)$ es el valor esperado de la j -ésima neurona para el patrón n -ésimo.
- $y_j(n)$ es el valor producido por la j -ésima neurona para el patrón n -ésimo.

El valor de ajuste de un peso asociado a una entrada de una neurona se obtiene de la siguiente expresión:

$$\Delta \omega_{ji}(n) = -\alpha \cdot y_i(n) \cdot \frac{\delta \mathcal{E}(n)}{\delta f(v_j(n))} \quad (3.18)$$

donde:

- α es un coeficiente constante de aprendizaje.
- $\mathcal{E}(n)$ es el error de la ecuación 3.17 para el patrón n -ésimo.
- $y_i(n)$ es el valor producido por la i -ésima (previa) neurona para el patrón n -ésimo.
- f es la función de activación utilizada para la salida de las neuronas.
- $v_j(n)$ es la suma ponderada de la j -ésima neurona para el patrón n -ésimo.

3.4.3. Mapas Auto-organizativos

Motivado por las funciones de ciertas partes de la corteza del cerebro humano que se encuentran representadas por mapas topológicamente ordenados, Teuvo Kohonen desarrolló los mapas auto-organizativos [Koh82] o SOM¹ (del inglés Self-Organizing Map).

Básicamente, estos mapas proveen de un método automático y no supervisado de escala multidimensional que proyecta un espacio de entrada amplio a un espacio de salida reducido,

¹Otros nombres que suelen utilizarse: SOFM (Self-Organizing Feature Map)

comprimiendo dicho espacio de entrada para generar un conjunto de vectores de salida codificados y representativos. Además, los mapas auto-organizados mantienen las características topológicas de los datos, de modo que dos patrones cercanos en el espacio de entrada estarán representados por un mismo patrón en el espacio de salida.

Los mapas de Kohonen se caracterizan por tener un algoritmo simple, de fácil implementación y con la capacidad de procesar y organizar automáticamente importantes volúmenes de datos sobre los cuales no se dispone de información previa. Entre algunos ejemplos de aplicaciones que pueden mencionarse están la minería y clasificación de documentos, la autenticación e identificación a través de voz, rostros y comportamiento, el análisis de imágenes meteorológicas y oceanográficas, el análisis de datos económicos, el ruteo dinámico de vehículos y la clasificación de genes, tejidos y células.

Arquitectura de la Red

Los mapas auto-organizativos de Kohonen tienen una arquitectura de dos capas: una capa de entrada y una capa competitiva.

La capa de entrada es una capa simple que tiene por función transmitir los datos de entrada hacia la capa competitiva. La cantidad de atributos que tengan los patrones de entrada determinará la dimensión de la capa de entrada y la cantidad componentes que tendrá el vector de pesos de cada neurona de la capa competitiva.

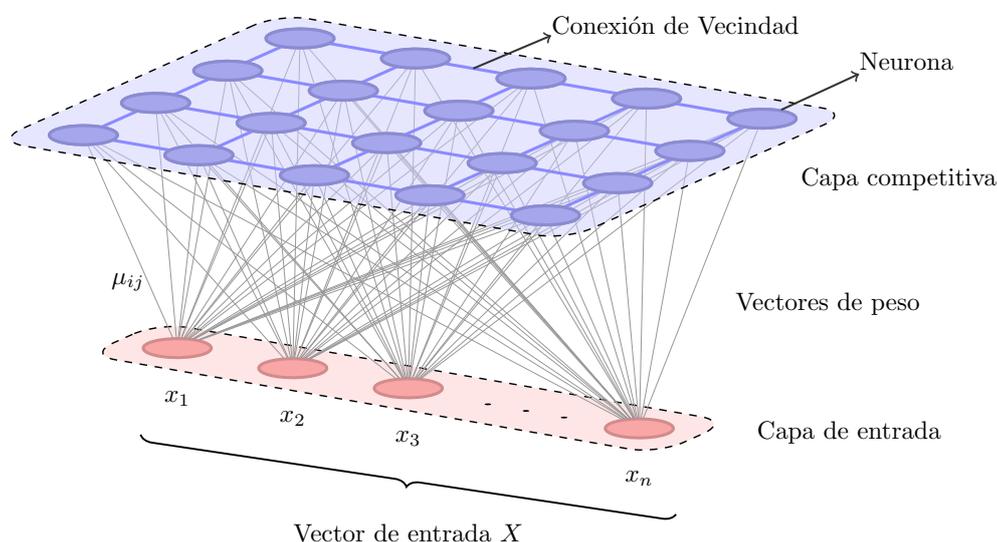


Figura 3.9: Esquema de arquitectura de Mapa auto-organizado (SOM).

La capa competitiva está formada por un conjunto de neuronas interconectadas que se organizan en una estructura n-dimensional. En general suele tener una distribución regular bidimensional en forma de grilla aunque pueden utilizarse distribuciones tanto unidimensionales como tridimensionales y además con forma irregular (ver figura 3.10). Cada neurona de la red tiene

un lazo fijo que la conecta con sus vecinas y que tiene fundamental importancia durante el proceso de entrenamiento.

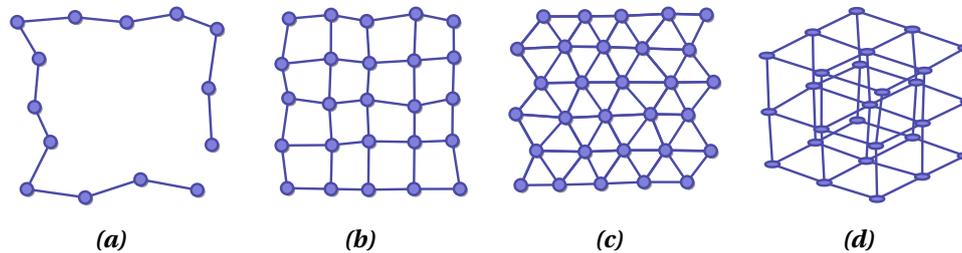


Figura 3.10: Capa competitiva de redes SOM con diferentes dimensiones y vecindades. a) Red uni-dimensional con vecindad 2. b) Red bi-dimensional de vecindad 4. c) Red bi-dimensional de vecindad 6. d) Red tri-dimensional de vecindad 6.

Funcionamiento de la Red

Cada vez que un patrón se presenta a la capa de entrada esta lo propaga a la capa competitiva donde se localiza a la neurona que mejor lo aproxima o representa. A esta neurona se la denomina *ganadora* o *BMU* (del inglés Best Matching Unit).

Para encontrar la neurona ganadora se utiliza una función de distancia o similitud que estima la relación entre cada neurona de la red y un patrón de entrada. La neurona con menor distancia al patrón de entrada es la que mejor lo aproxima y es etiquetada como ganadora.

Formalmente, la siguiente expresión permite modelar la búsqueda de la neurona ganadora para un patrón de entrada determinado:

$$\mu_g = \gamma(x, \boldsymbol{\mu}) \quad (3.19)$$

donde:

- γ es la función que retorna los pesos de la neurona que mejor representa al patrón x .
- x es un vector o patrón de entrada.
- $\boldsymbol{\mu}$ son los vectores de peso de todas las neuronas de la red.
- μ_g es el vector de pesos de la neurona ganadora.
- g es la neurona ganadora tal que $g = \operatorname{argmin}(\delta(x, \mu_i)) \forall i \in [1, \#\mu]$.
- δ es una función de distancia o similitud.

Aprendizaje de la Red

Durante la etapa de entrenamiento o aprendizaje se procesan los patrones de entrada y se modifican los pesos de las neuronas para que la red aproxime la distribución de los datos de entrada, preservando las relaciones que existen entre ellos.

El algoritmo de entrenamiento requiere de una estructura de neuronas asociadas en vecindad, de un radio de vecindad, de una tasa de aprendizaje y de un coeficiente de vecindad. El primer paso en el proceso de entrenamiento consiste en la definición de la estructura de vecindad, la inicialización de los pesos de las neuronas y del radio de vecindad.

Luego de establecer los parámetros iniciales comienza una serie de épocas. Una época consiste en el procesamiento de todos los patrones de entrenamiento por parte de la red. Cada vez que un patrón es presentado a la red, se localiza la neurona ganadora y se procede al ajuste sus pesos y de los pesos de todas las neuronas dentro del radio de vecindad. Este ajuste consiste en desplazar los pesos en dirección al patrón de entrada, ponderando su magnitud por el factor de aprendizaje y el factor de vecindad. La ecuación 3.20 desarrolla la fórmula de actualización de los pesos de una neurona:

$$\mu_i(t+1) = \mu_i(t) + [x(t) - \mu_i(t)] \cdot \tau(t, g, i) \cdot \alpha(t) \quad (3.20)$$

donde:

- t es el tiempo discreto o paso del entrenamiento.
- i es el índice de la neurona a actualizar.
- g es el índice de la neurona ganadora.
- $\mu_i(t)$ es el vector de pesos de la neurona i para el tiempo t .
- $x(t)$ es el vector del conjunto de vectores entrada para el tiempo t .
- $\tau(t, g, i)$ es coeficiente de vecindad para el tiempo t entre las neuronas g e i .
- $\alpha(t)$ es coeficiente de aprendizaje para el tiempo t .

Cuando todos los patrones fueron procesados se finaliza la época actual, se actualiza la tasa de aprendizaje, el radio de vecindad y el coeficiente de vecindad. Luego comienza una nueva época que itera nuevamente sobre todos los patrones de entrada. El entrenamiento termina cuando la secuencia iterativa de épocas alcanza la condición preestablecida de una cantidad máxima de iteraciones o de una cota de error mínimo de representación de la red.

3.4.4. Redes Neuronales de Energía de Coulomb Restringida (RCE)

En la década de 1980 surge una propuesta de un nuevo tipo de red neuronal competitiva [RCE82, SREC87] con la capacidad de separar patrones no lineales dentro de un espacio de entrada compuesto por regiones arbitrarias. Este modelo fue evolucionando hasta convertirse en lo que se denominó Red Neuronal de Energía de Coulomb. Con el transcurso de los años fueron surgiendo estudios que analizaron las deficiencias del modelo y propusieron distintas soluciones que dieron origen a nuevas variantes de este tipo de red neuronal [BD94, YGX01, GM05, SKR11].

Entre las características que hacen interesantes el uso de las redes neuronales RCE se pueden mencionar la capacidad modificar su estructura para generar neuronas solo cuando son realmente necesarias. Esto se logra adaptando las neuronas existentes y generando dinámicamente nuevas neuronas para cubrir regiones de decisión complejas mediante el uso de reglas simples. Otra característica que resaltan varios autores es que tiene una mejor velocidad de adaptación y un menor requerimiento de memoria que otros algoritmos de clasificación como KNN y redes de retro-propagación.

Como cualquier otra red neuronal tiene dos etapas con modos de funcionamiento diferentes. La primera etapa consiste en un proceso de aprendizaje o entrenamiento por el cual la red adapta los pesos de sus neuronas para relacionar el espacio de entrada con el espacio de salida. La segunda etapa consiste en un proceso de clasificación que determina la clase de pertenencia para un patrón o ejemplo de entrada presentado a la red.

Arquitectura de la Red

Una red neuronal RCE está formada por una arquitectura de tres capas de neuronas interconectadas: una de entrada, una oculta y una de salida. Un esquema genérico de esta arquitectura puede observarse en la figura 3.11. A continuación se describen las neuronas de las distintas capas:

- **Capa de entrada:** está compuesta por neuronas que reciben y transmiten un patrón de entrada a cada neurona de la capa oculta. La cantidad de neuronas que la forman está determinada por la cantidad de características de los patrones de entrada.
- **Capa oculta:** también denominada capa intermedia o de prototipo. Está formada por una cantidad variable de neuronas que representan un prototipo asociado a un subconjunto de los patrones utilizados durante la fase de aprendizaje. Normalmente, el campo receptivo de cada neurona abarca una parte del espacio de entrada con una hiper-esfera y puede estar superpuesto de manera de cubrir parcialmente un mismo espacio de entrada.
- **Capa de salida:** está compuesta por una neurona por cada clase a clasificar. A diferencia de otras arquitecturas neuronales donde cada neurona de la capa oculta se conecta con todas las de salida, cada neurona de la capa oculta se conecta exclusivamente con solo una neurona de la capa de salida.

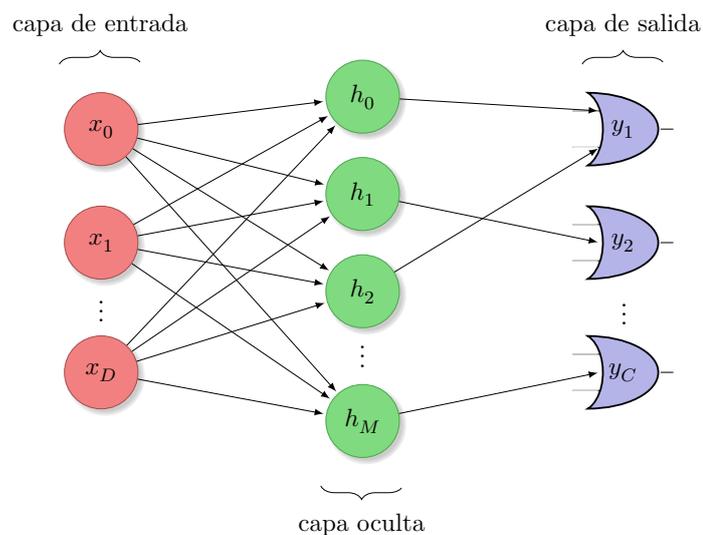


Figura 3.11: Arquitectura de una red neuronal de Energía de Coulomb Restringida (RCE).

Funcionamiento de la Red

El modo de funcionamiento normal de la red tiene lugar una vez que ha finalizado la etapa de entrenamiento o aprendizaje. Durante esta fase, las neuronas de la capa oculta han adaptado sus pesos para obtener un modelo neuronal que relaciona el espacio de entrada con el espacio de salida. A partir de este momento, la red tiene la capacidad de tomar decisiones basadas en las activaciones de las neuronas y así determinar la clase de pertenencia de un patrón o ejemplo de entrada dado.

Como se dijo anteriormente, el campo receptivo de una neurona de la capa oculta está determinado por el espacio hiperesférico abarcado por un punto del hiperespacio donde está ubicada la neurona (normalmente denominado centro o centroide) y su radio de cobertura.

Formalmente, cada neurona de la capa oculta o capa de prototipos P^j está formada por los tres componentes que se describen en la ecuación 3.21:

$$P^j = [C^j, R^j, N^j] \quad (3.21)$$

donde:

- C^j es el centro de la región hiperesférica del prototipo o neurona j .
- R^j es el valor del radio de la hiperesfera de la neurona j que establece su campo receptivo.
- N^j indica la cantidad de patrones o ejemplos de entrenamiento que fueron asociados a la neurona j durante el aprendizaje.

El esquema de aprendizaje utiliza una métrica de distancia euclídea entre un patrón de entrada y el centro de una neurona para determinar su activación. Si esta distancia es menor al radio de cobertura de la neurona, entonces el patrón recae en su campo receptivo, por lo que la neurona se activa y propaga una señal de excitación a través de sus conexiones hacia las neuronas de la capa de salida que finalmente determinan el prototipo o clase de pertenencia de cada ejemplo de entrada.

Aprendizaje de la Red

El proceso de aprendizaje consiste en ajustar los pesos de las neuronas de la capa oculta para estas representen una partición del espacio de entrada o de características. Como se mencionó en 3.4.4 existen diferentes variantes que implementan de distinta manera el algoritmo de aprendizaje.

Como en cualquier etapa de aprendizaje de una red neuronal, el tiempo de procesamiento donde todos los ejemplos destinados al entrenamiento son procesados por la red se denomina época. Durante cada época surgen nuevas neuronas y se eliminan neuronas existentes. Este proceso iterativo de adaptación finaliza cuando no se han producido cambios en la estructura de la red, es decir que no se han creado o eliminado neuronas en la última época.

Si durante la fase de aprendizaje de la red se cuenta con m ejemplos para su entrenamiento, entonces se denotará X a dicho conjunto y se lo define formalmente como:

$$X = \{x_1, x_2, \dots, x_m\} \quad (3.22)$$

Cuando un ejemplo de entrenamiento x_i se presenta a la capa de entrada de la red, esta lo transmite a la capa oculta para calcular la distancia euclídea entre éste y las neuronas C^j utilizando la siguiente ecuación:

$$\delta(x_i, C^j) = \sqrt{\sum_{k=1}^K (x_{ik} - C_k^j)^2} \quad (3.23)$$

donde:

- C^j es el centro de la región hipersférica del prototipo o neurona oculta j .
- x_i es el i -ésimo ejemplo de entrenamiento.
- K es el número de características de cada ejemplo de entrenamiento y de pesos de cada neurona.

Si la distancia $d(x_i, C^j)$ es menor que el radio R^j de la neurona j -ésima, el ejemplo x_i se encuentra dentro de la hipersfera del campo receptivo y en consecuencia se considera que dicho patrón es similar al prototipo P^j . Por lo que se registra su asignación al grupo representado por dicho prototipo y se incrementa en 1 el valor del contador N^j . Por el contrario, si la distancia calculada supera el radio indicado por R^j se creará un nuevo prototipo $P^{n+1} = [C^{n+1}, R^{n+1}, N^{n+1}]$ con centro en x_i , radio R^{n+1} y contador de ejemplos $N^{n+1} = 1$.

Finalizada una época de entrenamiento se calcula la densidad para cada una de las neuronas de la capa oculta. El objetivo es determinar si el agrupamiento abarcado por cada neurona es suficiente para que esta permanezca en la estructura de la red. Sólo las neuronas que superen un umbral mínimo de densidad preestablecido serán conservadas, el resto serán descartadas y los ejemplos que habían sido asignados a ellas volverán a redistribuirse. Este proceso se repite hasta que no se generen nuevos prototipos, es decir que no haya cambios en la estructura de la red.

La ecuación (3.24) indica la forma de calcular la densidad de la j -ésima neurona o prototipo.

$$D_{P^j} = \frac{3N^j}{4\pi R^j K} \quad (3.24)$$

donde:

- N^j es el contador de ejemplos de la neurona oculta j .
- R^j es el radio actual de la neurona oculta j
- K es el número de pesos de cada neurona de la red.

Uno de los factores más importantes para un buen funcionamiento de la red es establecer los valores de los parámetros de densidad y radio que deben ser utilizados. Respecto de la densidad, ésta define la cantidad mínima de ejemplos de entrenamiento que debe abarcar una neurona para formar parte de la red. Un valor de densidad alto requiere de una mayor cantidad de ejemplos para que la neurona se mantenga en la red, ya que de lo contrario será eliminada. En el caso de un valor de densidad bajo, requiere de una pequeña cantidad de ejemplos la neurona sea aceptada. Respecto del radio, éste establece la región del hiperespacio que abarca una neurona y determina que ejemplos pertenecerán a la misma. Por un lado, un valor de radio alto requiere de menos

neuronas para cubrir un determinado espacio de entrada y tiene un costo computacional menor ya que se requieren menos cálculos de distancia. Por otro lado, incorpora falsos positivos ya que por su amplitud tiende a abarcar ejemplos que pertenecen a diferentes clases. Sin embargo, un valor de radio bajo reduce la posibilidad de tener falsos positivos pero produce una cantidad de neuronas mayor y en consecuencia incrementa el costo computacional.

DISTRIBUCIÓN DEL COLOR DE LA PIEL EN DIFERENTES SISTEMAS DE COLOR

Resumen

En este capítulo se muestra el comportamiento de la distribución de la piel humana en los diferentes modelos de representación del color que se incluyen en los artículos de estudio. Se detalla el juego de datos utilizados para generar las distribuciones y se discute la normalización de escalas entre los diferentes modelos de color. Luego se muestra y comenta una serie de gráficos por cada modelo de color. Finalmente se realiza un análisis sobre las ventajas y desventajas de la distribución de cada modelo de color.

4.1. Introducción

La conversión entre espacios de color es una transformación que afecta la manera en que se forma el color y en consecuencia como éste se distribuye en el espacio de representación. Por este motivo, en éste capítulo se realiza un breve análisis de la distribución del color de la piel en los diferentes sistemas de representación de color utilizados en los artículos de estudio.

Para tener una referencia más intuitiva y facilitar el análisis de distribución del color de la piel en los diferentes sistemas de representación del color se construyó una serie de gráficos tridimensionales a partir componentes de cada color correspondiente a la piel.

Varios de los sistemas de representación del color estudiados tienen diferentes rangos de valores y unidades en los componentes que forman el color. Para realizar una comparación justa entre los diferentes sistemas es necesario la unificación de dichas unidades de manera que el volumen del

espacio sea representativo para compararlo.

4.1.1. Base de Datos de Piel

Para obtener los gráficos con la distribución de color de la piel se utilizó el repositorio UCI [DKT17] que se encuentra disponible de manera libre y gratuita. Este repositorio reúne una importante colección de bases de datos destinadas al uso de algoritmos y aplicaciones de aprendizaje automático. En particular se utilizó una base de datos de piel [BD12] que se compone de unas 245.057 muestras de píxeles codificados en el sistema RGB de los cuales 50.859 corresponden a píxeles extraídos de muestras de piel.

4.1.2. Gráficos para Analizar

Como todos los espacios de color se pueden representar en un espacio tridimensional, es posible hacer un análisis visual de la distribución de la piel en los distintos sistemas. Para obtener estos gráficos, los componentes de cada muestra de la base de datos fueron convertidos a la respectiva representación en cada uno de los sistemas de estudio. Esta conversión genera un modelo tridimensional de la distribución de la piel que se muestra en cuatro vistas diferentes. Por un lado, una vista del espacio tridimensional que captura la distribución mostrando los tres ejes. Por otro lado, tres vistas del espacio bidimensional formado por los planos que surgen de las combinaciones de los tres ejes que componen el espacio.

4.1.3. Normalización de Escalas

Cuando se aplica una transformación entre espacios de color, puede suceder que el rango de los valores de los componentes quedan en una escala diferente, lo que dificulta la comparación. Es por este motivo que una normalización de los valores de cada canal, componente o eje fue aplicada para obtener una escala de espacios con volúmenes comparables que ofrecieran una mejor percepción del espacio. Todos los valores de conversiones se acotan al rango que va de 0 a 100, indicando un porcentaje del espacio que ocupa ese componente, donde 0% es la mínima cantidad que puede tener y un 100% es la máxima cantidad que puede tener. Así por ejemplo, para el canal o componente H del sistema HSV que va de 0 a 359 grados, se transformó a la escala que va desde 0% a 100%. La unidad utilizada en los ejes de todos los gráficos es el porcentaje (%) del espacio que ocupa en el sistema en que se encuentra.

4.2. Distribución de la Piel en Sistemas de Color

En la siguiente sección se presenta una serie de gráficos de la distribución de la piel junto con un breve análisis de las mismas por cada sistema de representación de color relevado en los distintos artículos de estudio.

4.2.1. Distribución de la Piel en RGB

En la figura 4.1 se muestra la distribución de la piel en el espacio de color RGB. En esta se puede observar como la distribución de la piel en cada plano se expande diagonalmente y ocupa una región de 59% del componente (R), 69% del componente (G) y 70% del componente (B), siendo la más amplia de todos los sistemas. Además como la información de cromaticidad y luminosidad está mezclada no permite discriminar cuando varía el color o la iluminación de la piel. No parece ser el mejor sistema para representar el color de la piel. Por un lado para algoritmos de segmentación elementales como los basados en umbrales no permite una cobertura eficiente o requiere alguna transformación adicional. Por otro lado para algoritmos más complejos como los basados en agrupamiento el espacio a cubrir es significativo e irregular.

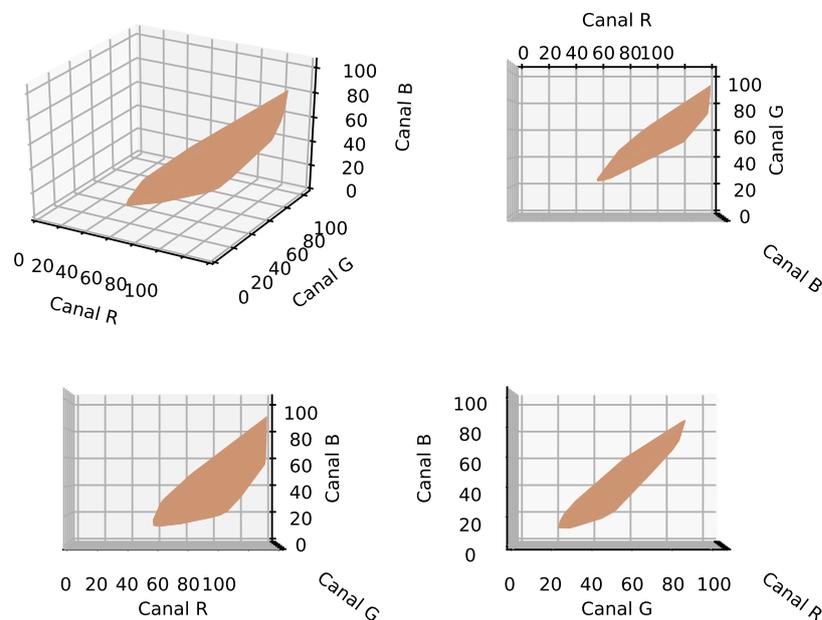


Figura 4.1: Distribución de la piel en el espacio de color RGB según base de datos [BD12].

4.2.2. Distribución de la Piel en RGB Normalizado

Es importante recordar que este espacio de representación se reduce a un triángulo dentro del plano $R + B + G = 1$ por lo que la distribución del color no es tridimensional sino bidimensional. Los valores de los componentes de color son muy diferentes a los demás espacios. En la figura 4.2 se muestra la distribución de la piel en este espacio. En esta figura se puede ver que la distribución de la piel es más compacta, aunque a diferencia de otros sistemas de representación, el plano en que se encuentra también es más compacto. Ocupa una región de 21% del componente (R), 9% del componente (G) y 19% del componente (B). En los planos RG y GB se puede observar como la distribución ocupa una región relativamente regular y compacta. No obstante en el plano RB la distribución se expande de manera diagonal lo que perjudica a algoritmos elementales de

segmentación. Respecto del sistema RGB, la distribución se muestra más compacta y uniforme en sentido paralelo a los ejes lo que favorece a los algoritmos de segmentación simple.

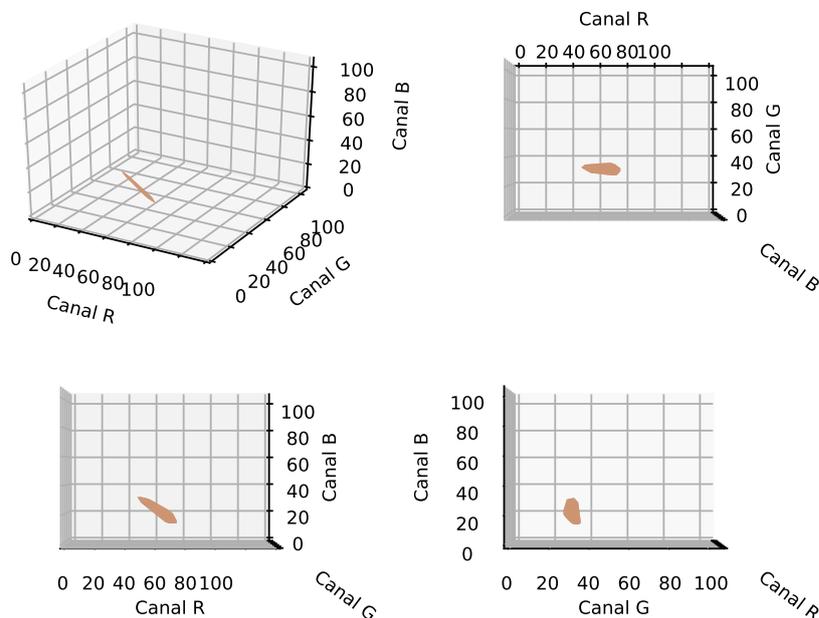


Figura 4.2: Distribución de la piel en el espacio de color RGB normalizado según base de datos [BD12].

4.2.3. Distribución de la Piel en HSV, HSL y HSI

Como se describió en el Capítulo 2, los sistemas de representación HSV, HSL y HSI se diferencian por la forma en que se calculan los componentes de saturación e intensidad mientras que el componente de matiz se calcula de la misma forma para todos los sistemas.

En el espacio de color HSV de la figura 4.3 se puede observar que la distribución del color de la piel está comprendida dentro de una región que abarca un 13% en el componente de Matiz (H), un 69% del componente de Saturación (S) y un 59% del componente de Luminosidad (V).

En la figura 4.4 se muestra la distribución del color de la piel en el espacio de color HSL dentro de una región que abarca un 13% en el componente de Matiz (H), un 58% del componente de Saturación (S) y un 68% del componente de Luminosidad (L).

En el espacio de color HSI de la figura 4.5 se muestra la distribución del color de la piel dentro de una región que abarca un 13% en el componente de Matiz (H), un 59% del componente de Saturación (S) y un 68% del componente de Luminosidad (I).

A simple vista se puede observar en los gráficos de los tres sistemas que la distribución de la piel forma una figura muy similar. En particular se puede concluir que la distribución en los sistemas HSL y HSI son prácticamente iguales, mientras que en el HSV es ligeramente más amplia respecto del

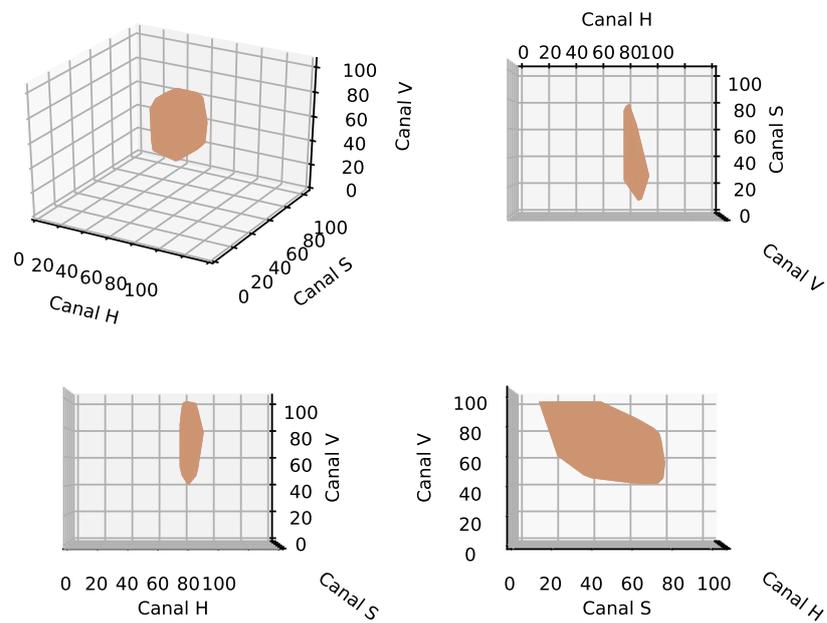


Figura 4.3: Distribución de la piel en el espacio de color HSV según base de datos [BD12].

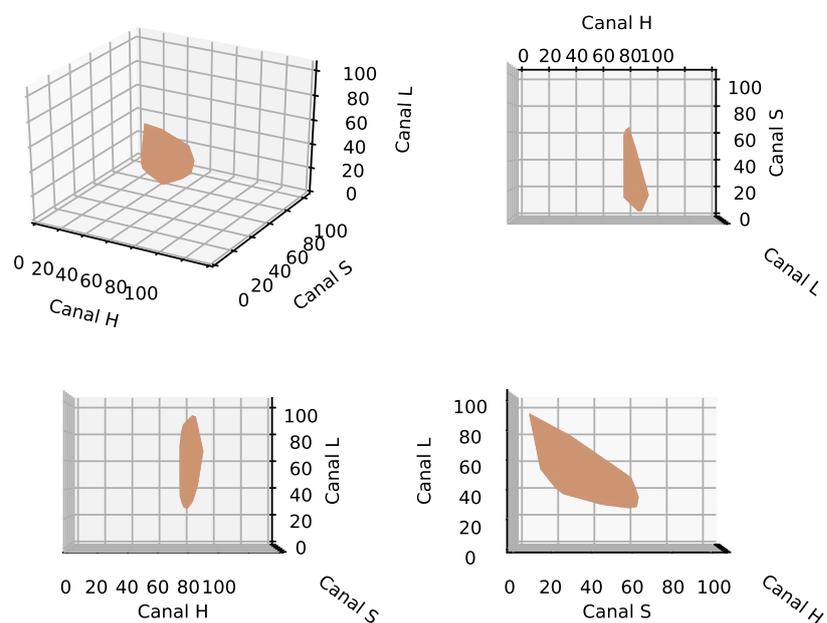


Figura 4.4: Distribución de la piel en el espacio de color HSL según base de datos [BD12].

plano Saturación/Luminosidad (SV).

Por lo acotado del rango de valores del componente de Matiz se ve que la distribución de los diferentes tonos de piel tienen un valor similar. Esto es algo favorable para la tarea de segmentación ya que este componente es el que más información aporta de los tres componentes de cada sistema.

Respecto de los componentes de Saturación y Luminosidad el rango de valores que abarcan son muy amplios lo que indica que los colores de piel de distintas etnias difieren mas en estos componentes que en el Matiz. Esto coincide con el análisis realizado por Yang y Waibel en [YW96] sobre la distribución de la piel de rostros de razas que incluyeron la raza asiática, afroamericana y caucásica.

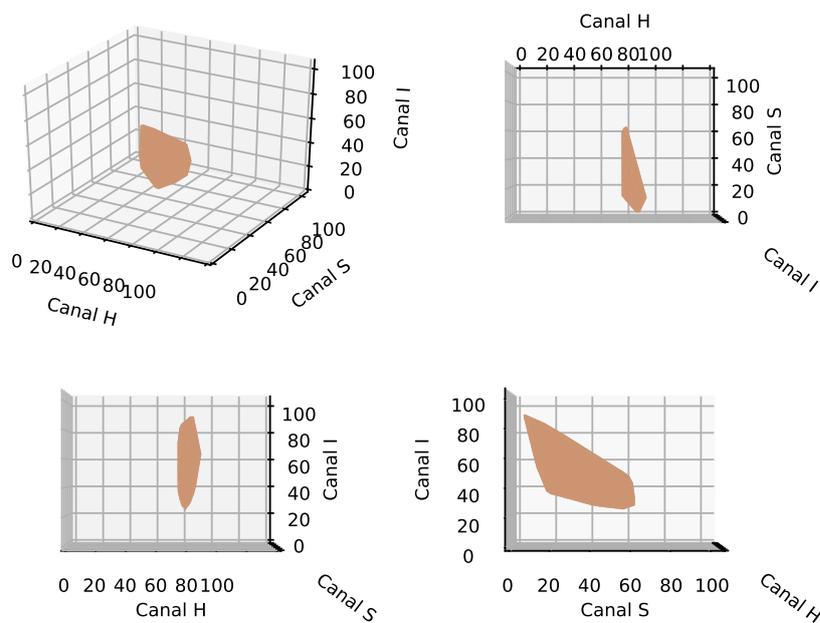


Figura 4.5: Distribución de la piel en el espacio de color HSI según base de datos [BD12].

4.2.4. Distribución de la Piel en YUV e YIQ

En los sistemas de representación de color YUV e YIQ, el componente (Y) que corresponde a la información de luminosidad es el mismo, por lo que mantiene en el rango de valores para ambos sistemas. Los demás componentes que varían de acuerdo a sus componentes de cromaticidad, tienen una leve diferencia.

Por un lado, como se observa en la figura 4.6, la distribución del color de la piel en el espacio de color YUV se encuentra dentro de una región que abarca un 70% en el componente de Luminosidad (Y), un 3% del componente de Cromaticidad (U) y un 6% del componente de Cromaticidad (V).

Por otro lado en la figura 4.7 se puede observar la distribución del color de la piel en el espacio de color YIQ dentro de una región que abarca un 70% en el componente de Luminosidad (Y), un 6% del componente de Cromaticidad (I) y un 3% del componente de Cromaticidad (Q).

Comparando los gráficos de los dos sistemas no parece haber diferencias substanciales. Tanto los componentes (Y) de ambos, como el componente (U) con el componente (Q) y el componente (V) con el componente (I) coinciden en el mismo rango de valores. La única diferencia apenas significativa se puede observar en el plano IQ donde la distribución se asemeja a un cuadrado

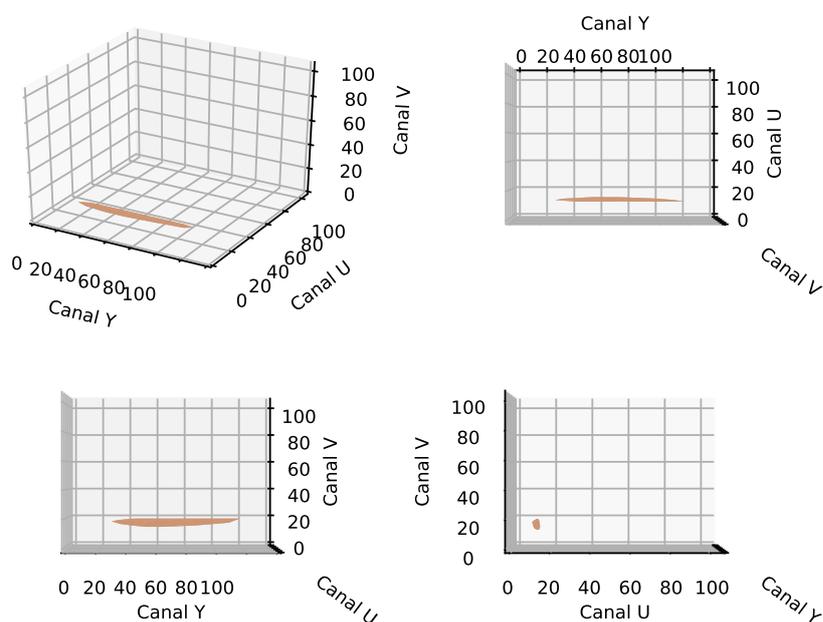


Figura 4.6: Distribución de la piel en el espacio de color YUV según base de datos [BD12].

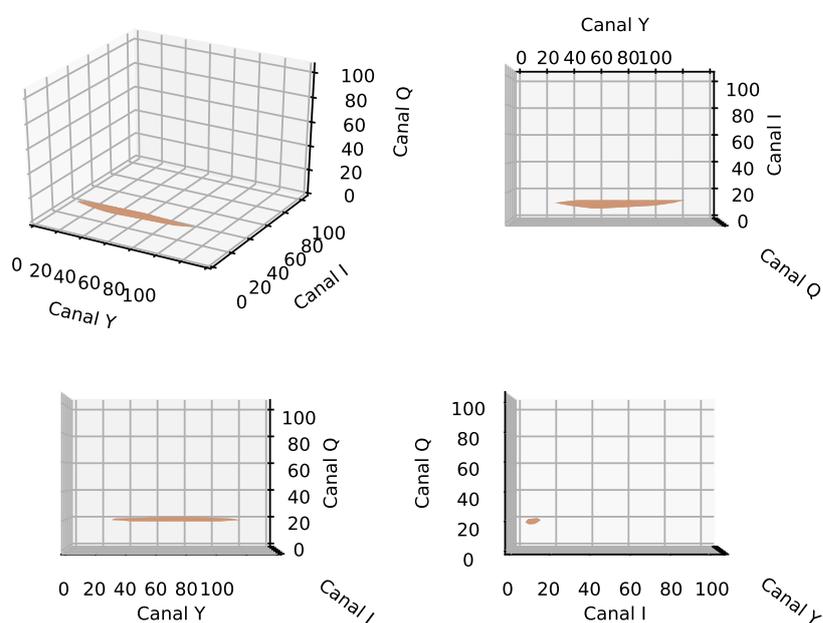


Figura 4.7: Distribución de la piel en el espacio de color YIQ según base de datos [BD12].

alineado a los ejes mientras que en el plano UV la distribución se asemeja más a un círculo lo que podría favorecer a los algoritmos de segmentación elementales.

Comparando con los demás sistemas de representación el rango de valores del componente de luminosidad es similar. Respecto de los componentes de color es importante destacar que los

sistemas YUV e YIQ los componentes de cromaticidad se encuentran compactados como se puede observar en el plano UV como IQ respecto de los demás sistemas. Esta característica hace que estos sistemas favorezcan en general a todos los algoritmos de segmentación y parecerían ser los más prometedores. (REVISAR REDACCION)

4.2.5. Distribución de la Piel en YCbCr e YCgCr

En los sistemas de representación de color YCbCr e YCgCr sucede algo similar a los sistemas YUV e YIQ. El rango de valores es similar en sus tres componentes. Tanto los valores de los componentes de luminosidad como el componente de cromaticidad (Cr) son iguales para ambos sistemas. Respecto del componente de cromaticidad (Cg) parece estar un poco más comprimido que el componente (Cb) lo cual sería una ventaja aparente.

En la figura 4.8 se puede observar el espacio YCbCr con una distribución de la piel que abarca una zona más compacta que el espacio de color HSV. Esta ocupa una región de 60% en el componente (Y), un 15% en el componente de cromaticidad asociado al azul (Cb) y 16% en el componente de cromaticidad rojo (Cr).

Respecto de la forma de la distribución en el plano parece algo irregular desde los planos YCr e YCb aunque está relativamente alineado a los ejes, lo que favorece a algoritmos elementales de segmentación. En el plano CrCb tiene una forma un poco más regular pero está distribuida en diagonal a los ejes lo que desfavorece a algoritmos elementales.

VER: Por este motivo y porque el costo de la transformación desde RGB es menor que el de HSV, éste modelo se perfila como una buena alternativa.

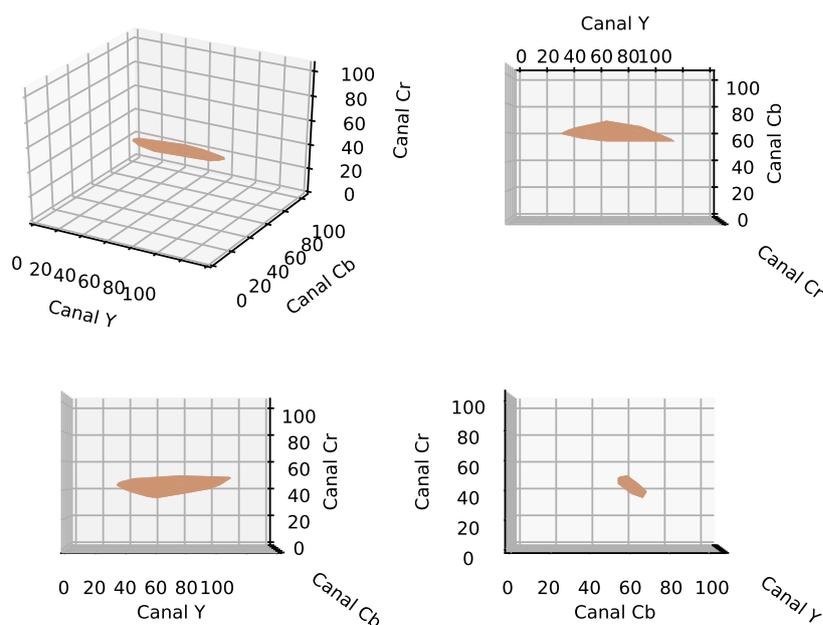


Figura 4.8: Distribución de la piel en el espacio de color YCbCr según base de datos [BD12].

Como se observa en la figura 4.9 se muestra la distribución del color de la piel en el espacio de color YCbCr dentro de una región que abarca un 60% en el componente de Luminosidad (Y), un 10% del componente de Cromaticidad (Cg) y un 15% del componente de Cromaticidad (Cr).

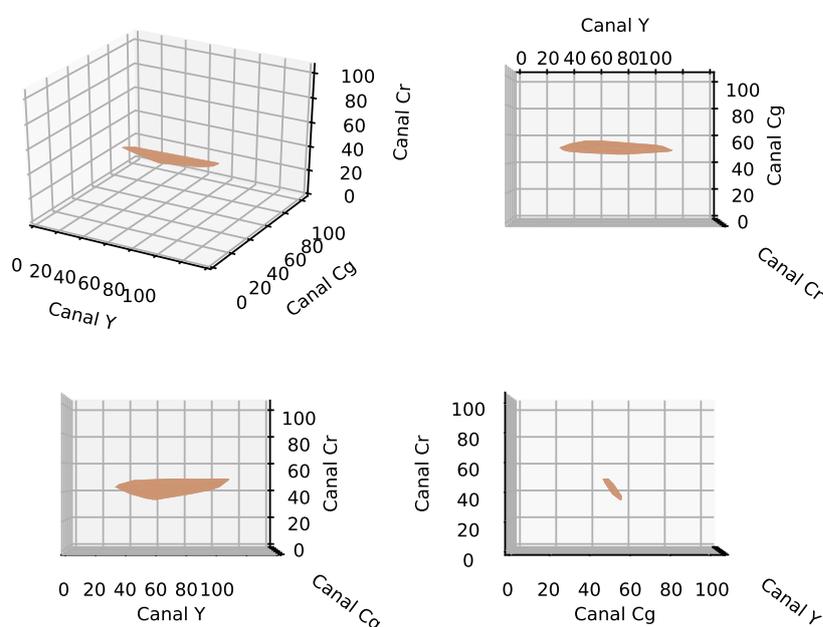


Figura 4.9: Distribución de la piel en el espacio de color YCgCr según base de datos [BD12].

Respecto de la forma de la distribución se ve similar al sistema YCbCr. En el plano YCr la distribución se ve exactamente igual. En el plano YCg la distribución se ve un poco más compacta y su forma es un poco más regular que el plano YCb. En el plano CgCr la distribución se ve con una forma similar pero parece algo más compacta. La distribución diagonal en este plano tiene las mismas desventajas que el plano CbCr.

4.2.6. Distribución de la Piel en LAB

En la figura 4.10 se puede observar que el espacio CIE-Lab ocupa una región que abarca un 67% en el componente de Luminosidad (Y), un 11% del componente de Cromaticidad (A) y un 21% del componente de Cromaticidad (B). Tiene un volumen de distribución del color de la piel similar a los sistemas YCbCr e YCbCr, y si bien este espacio es el que más se aproxima a la percepción del observador es el que tiene el costo computacional más alto. Este balance es un aspecto importante a tener en cuenta al momento de elegir este sistema para detectar el color de la piel.

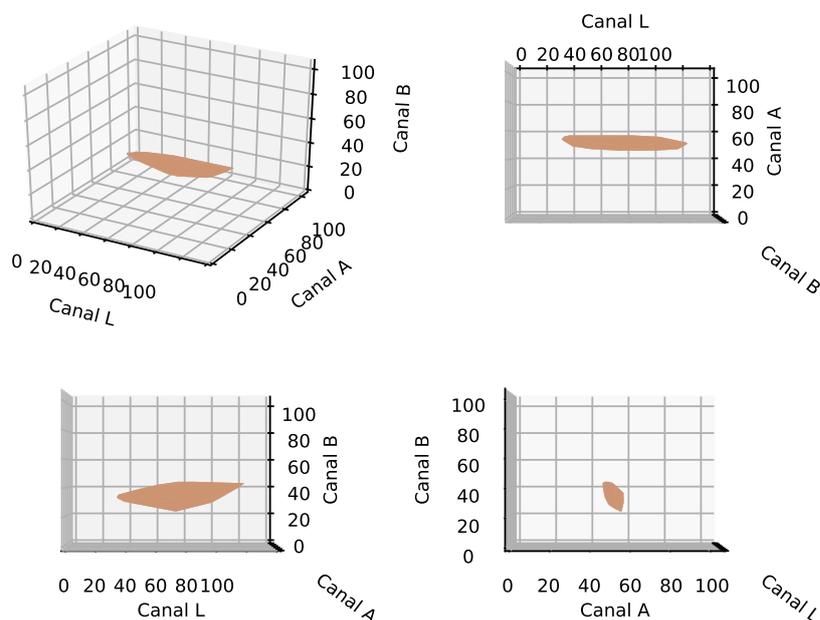


Figura 4.10: Distribución de la piel en el espacio de color CIE-LAB según base de datos [BD12].

4.3. Análisis de la Distribución de la Piel

Teniendo en cuenta los sistemas de representación del color estudiados, se podría realizar un análisis haciendo una división de éstos teniendo en cuenta por un lado los componentes asociados al color y por otro el componente asociado a la luminosidad.

Respecto del componente de luminosidad hay tres casos para tener en cuenta. El primer caso es que en el sistema RGB la información relativa a la luminosidad está mezclada en los tres componentes (R), (G) y (B) lo que no permite una comparación directa con los demás sistemas desde este punto de vista. El segundo caso es que en el sistema RGB normalizado la información asociada a la luminosidad ha sido eliminada. De ser necesario esta información podría obtenerse de manera similar a la que se obtiene en los demás sistemas a partir de los componentes (R), (G) y (B). El tercer caso comprende a todos los restantes sistemas analizados que tienen en alguno de sus componentes la información correspondiente a la luminosidad. Teniendo en cuenta esto, en líneas generales se puede afirmar que en todos los sistemas, el componente correspondiente a la luminosidad mantiene un rango similar que ronda entre el 59% y el 70%. Como no hay una diferencia significativa entre estos valores, tampoco hay una ventaja aparente respecto de este componente entre todos los sistemas.

Para el análisis de los componentes asociados al color hay dos cosas a tener en cuenta. Por un lado se encuentra la complejidad del algoritmo de transformación de un sistema de color a otro que impacta en el tiempo de cómputo. Por otro lado está la forma de la distribución de la piel en cada sistema, luego de su transformación, que impacta en la eficiencia del algoritmo de segmentación.

La principal ventaja del sistema de representación RGB es que es el sistema nativo de adquisición

y representación del color por lo que no tiene costo computacional de transformación. Por otro lado al tener una distribución del color de piel que se extiende de forma diagonal a los ejes impide que algoritmos elementales de segmentación como los basados en umbrales pueden aplicarse de forma eficiente y pone a prueba la robustez de algoritmos más complejos.

El sistema de representación RGB normalizado es el más particular de todos los sistemas ya que reduce los colores a un triángulo comprendido dentro del plano $R + G + B = 1$. El costo computacional de conversión es bajo, aunque no cuenta con información de luminosidad, componente que habría que agregar de ser necesario con un incremento bajo respecto del costo computacional. En general es una opción mucho más conveniente para algoritmos elementales de segmentación comparada con el sistema RGB. La única desventaja en este sentido es que la distribución se expande diagonalmente en el plano RB lo que perjudica a algoritmos simples como los basados en umbrales.

Respectos de los sistemas HSV, HSL y HSI, son modelos que tienen un costo computacional de conversión mas elevado que los anteriores. Respecto de los algoritmos de segmentación, el componente (H) contiene la representación mas significativa del color de la piel respecto de los demás componentes. Por un lado, la distribución de la piel toma valores muy acotados lo que favorece a algoritmos elementales de segmentación. No obstante, como en estos sistemas la distribución de la piel se extiende de forma diagonal a los ejes en el plano SL, perjudica la efectividad de estos tipos de algoritmos.

Los modelo de representación YUV e YIQ tienen un costo de conversión intermedio y ofrecen una distribución relativamente compacta en los planos de cromaticidad (UV e IQ). En principio parece conveniente para todo tipo de algoritmo.

Los sistemas YCbCr e YCgCr tienen características similares al YUV e YIQ, salvo que la distribución de la piel en el plano de cromaticidad es ligeramente más amplia que en dichos sistemas y por este motivo pierden un poco de competitividad.

El modelo de representación LAB es el que requiere mayor costo computacional y tiene una distribución similar a los sistemas YCbCr e YCgCr. Por este motivo es preferible utilizar modelos como YUV e YIQ ya que presentan un mejor equilibrio entre costo y efectividad.

JUEGOS DE DATOS PARA PRUEBAS

Resumen

Este capítulo tiene como objetivo documentar las características de los juegos o bases de datos utilizados para realizar pruebas en los distintos artículos de estudio. Se describen las características que debería tener un buen juego de datos utilizado para segmentación de piel. También se detallan los requerimientos necesarios para sistematizar las pruebas de segmentación y así poder hacer una comparación justa entre los modelos de representación de color y los algoritmos utilizados para la segmentación.

Después se hace una descripción de cada uno de los juegos de datos de cada artículo de estudio y se realiza un breve análisis sobre las capacidades de los mismas para utilizarse de manera justa en la comparación de los resultados de los distintos artículos.

5.1. Características de un Buen Juego de Datos

En este capítulo se recopilan, describen y analizan brevemente los distintos juegos o bases de datos utilizadas por los artículos estudiados. También se incluyen algunas bases de datos adicionales, no incluidas en los artículos, con potencial para su uso en pruebas que faciliten la comparación tanto de los sistemas de representación del color como de los algoritmos de segmentación mencionados en los capítulos 2 y 3, respectivamente.

Para que los sistemas de representación del color y los algoritmos de segmentación aplicados a la extracción de piel puedan ser evaluados y comparados son necesarias bases de datos accesibles, representativas y precisas. En este sentido, dichas bases de datos deben cumplir con ciertos requisitos o reunir ciertas características que posibiliten la obtención de resultados confiables, comparables,

pero sobre todo reproducibles. Entre los aspectos que se deben tener en cuenta para que una base de datos asegure una mínima calidad de los resultados se deben considerar:

- **Imágenes en color.** Como este trabajo evalúa los diferentes sistemas de representación de color, indefectiblemente el conjunto de las imágenes de los datos de prueba debe incluir obligatoriamente imágenes a color.
- **Cantidad de Imágenes.** Para que las mediciones de los resultados sean significativos es necesario contar con un número importante de imágenes. La cantidad también es relevante para dividir en un grupo de entrenamiento y un grupo de prueba para los algoritmos que lo requieren.
- **Tipos de piel.** Si bien para la comparación de resultados la diversidad de la piel proveniente de diferentes partes del cuerpo, sería deseable incluir una amplia variedad de tipos de piel para cubrir el espectro posible, incluyendo por ejemplo, piel de las manos, brazos, piernas, rostros, etc.
- **Tipos de razas.** Es recomendable la inclusión de la piel de distintas razas para la generalización de los algoritmos. Es importante destacar que el un número de muestras incluidas debe ser similar para cada raza para mantener una proporción justa.
- **Condiciones de Iluminación.** Imágenes con intensidades y tipos de luz variables deben ser incluidas. Esto permite evaluar la versatilidad o capacidad de adaptación de los algoritmos y de los sistemas de color en diferentes condiciones de iluminación.
- **Tipos de Fondos.** Para medir la capacidad de los algoritmos de segmentación es interesante que las imágenes tengan distintas complejidades respecto al fondo. Deberían incluirse ejemplos de datos con imágenes de piel con fondos de diferentes grados complejidad.
- **Sistematización de pruebas.** Un aspecto de fundamental relevancia al momento de comparar sistemas de representación de color y algoritmos de segmentación es la sistematización de las pruebas. Es fundamental que cada imagen tenga asociada información de la porción de ésta que corresponde al recorte de los píxeles de la piel (máscara de segmentación).
- **Disponibilidad.** Otro factor determinante es la disponibilidad de los datos para que los autores puedan realizar pruebas que permitan comparar resultados entre diferentes sistemas de color y/o algoritmos de segmentación. Por este motivo las bases de datos deben estar disponibles de manera gratuita y libre de uso para fines no comerciales, ya que aquellas que son propietarias hace imposible o limita la comparación de los resultados.

En general todas las bases de datos de los artículos de estudio carecen de muchos de los aspectos y características mencionados anteriormente. Tal vez la dificultad mas importante es la falta de disponibilidad de los juegos de datos que los autores de los artículos utilizan para evaluar los sistemas de representación del color y los algoritmos de segmentación. Incluso una buena parte de los artículos que utilizan bases de datos que están disponibles incluyen imágenes propias en los resultados, haciendo imposible la comparación con otros autores.

También suele suceder que los autores de los artículos usan bases de datos que no tienen información de ground-truth o máscaras con los píxeles correspondientes a la piel segmentada. Como para obtener dichas máscaras, usualmente aplican un procedimiento manual o automático bajo supervisión que no comparten, también imposibilita la comparación justa entre resultados de diferentes artículos.

Otro problema que puede observarse con relativa frecuencia es que para aquellos algoritmos que requieren grupos de datos de entrenamiento y de prueba, los autores no especifican como es la división de los datos de la base de datos.

A continuación se muestra una tabla con el detalle de cada artículo estudiado con una breve descripción de la información disponible que ofreció el autor o que se pudo recolectar de las referencias para hacer la evaluación tanto de los sistemas de representación de color como de los algoritmos de segmentación.

<i>Bases de Datos Utilizadas en la Segmentación de Piel</i>	
<i>Artículo</i>	<i>Base de Datos</i>
[ZMXNJS06]	120 imágenes descargadas de Internet
[NKN16]	500 imágenes propias de estudiantes
[XP17]	No hay datos
[HSW08]	No hay datos
[GCRÁCP ⁺ 09]	Imágenes propias de webcams, 224 en condiciones controladas y 129 condiciones no controladas
[Lei13]	Recolección de 80 imágenes descargadas de Internet y propias capturadas con cámara
[TKSY17]	Videos propios
[BAO17]	Pratheepan dataset
[YMS17]	369 imágenes de Aberdeen Pilot Database - PICS
[HBGR09]	800 imágenes provistas por Computer Vision Laboratory (Ljubljana University, Slovenia) y 1.600 imágenes de varias fuentes para ejemplos negativos
[IK13]	100 imágenes. No hay detalles
[BV15]	22 imágenes propias
[HATWSS14]	300 imágenes de Faces in the Wild Home Dataset
[TPM ⁺ 11]	16 imágenes descargadas de internet
[SGK ⁺ 15]	30 imágenes. No especifica origen
[KKS ⁺ 17]	Pratheepan dataset
[RCGS16]	ECU Dataset y HGR Dataset
[BK10]	500 imágenes descargadas de Internet
[XS13]	100 imágenes descargadas de Internet
[GMX11]	200 imágenes descargadas de Internet
[BCL01]	396 imágenes de Aberdeen Pilot Database y 113 imágenes descargadas de Internet
[PCB01]	100 imágenes propias
[CPB03]	ECU Dataset e imágenes descargadas de Internet
[CZL ⁺ 02]	BIOID dataset, imágenes propias, imágenes descargadas de Internet
[HTMK02]	Imágenes descargadas de Internet
[RTBA12]	Recolección 41 imágenes propias y descargadas de Internet
[LzP10]	800 imágenes propias
[KK12]	No hay datos
[SCVS03]	Recolección de imágenes propias, descargadas de Internet e IITK Face Database

Cuadro 5.1: Bases de datos utilizadas por los artículos de estudio para segmentación de piel.

5.2. Juegos de Datos Utilizados por Artículos Estudiados

5.2.1. IITK Face Database

De esta base de datos parece no haber mucha información sobre el estado actual y las condiciones de uso ni sitios de descargas. No hay información que la describa unívocamente en los diferentes publicaciones encontradas. Si bien hay varios documentos que hacen referencia a esta base de datos para reconocimiento de rostros, por la descripción que se puede encontrar en éstos parece que ha ido evolucionando en el tiempo. Uno de los documentos [KRTG08] que hace referencia a la base de datos la describe con unas 1.000 imágenes, con diferentes fondos y condiciones de iluminación, de 200 sujetos de raza india con 5 poses cada uno. Otro de los documentos [KGMS09] que la cita, la describe con 400 sujetos con 2 imágenes por cada uno.

Obviamente, esta base de datos no es la indicada para utilizar en aplicaciones de segmentación de piel. Mas allá de no haber información sobre la disponibilidad para conseguirla, solo contiene imágenes de raza India lo que perjudica la efectividad de los algoritmos de segmentación de piel al aplicarlo en otras razas. Además no es posible la comparación de algoritmos por la falta de imágenes de máscaras (ground-truth) de segmentación de los píxeles de la piel.

5.2.2. Pratheepan Dataset

Las imágenes de la base de datos “Pratheepan Dataset” fueron descargadas de manera aleatoria a través del buscador de Google. Fueron elegidas para investigación de algoritmos de detección de piel humana y usada por primera vez en la publicación “A Fusion Approach for Efficient Human Skin Detection” [TCYC12]. Dichas imágenes han sido capturadas con un rango de diferentes cámaras con varias mejoras de color y bajo diferentes condiciones de iluminación. La información está estructurada en carpetas y contiene 32 imágenes a color de 1 sujeto bajo fondo simple y 46 imágenes de múltiples sujetos con fondos complejos. Todas las imágenes tienen asociadas su correspondiente máscara de píxeles (ground-truth) que corresponden a la piel segmentada. Esta base de datos ha ido recibiendo varios aportes de terceros con lo que su última actualización fué en marzo de 2017. Tiene imágenes de distintas razas pero no hay una proporción aparentemente balanceada y tampoco están documentadas. La base de datos está disponible de manera pública en Internet su uso es gratuito si es con fines no comerciales.

La base de datos “Pratheepan Dataset” cumple, en general, con los puntos descritos en 5.1 por lo que es una opción aceptable para el uso en aplicaciones de segmentación de piel y en la comparación de los resultados. Como puntos relativamente débiles se pueden mencionar dos. El primer punto es que, si bien hay imágenes, no hay información sobre distintas razas que se pueden encontrar en los datos. El segundo punto y tal vez mas débil, es la baja cantidad de imágenes disponibles. En los ambos casos sería interesante ampliar la cantidad de imágenes con una proporción adecuada de razas y con sus respectivas máscaras de segmentación.

5.2.3. Aberdeen Pilot Dataset

“Aberdeen Pilot Dataset” [Psy17] es una base de datos formada por un grupo de bases de datos denominada “Psychological Image Collection Stirling(PICS) Database”. Esta colección tiene como objetivo principal el uso en experimentos de conducta en psicología, primariamente en rostros. Su uso y descarga es libre y se encuentra disponible para propósitos de investigación.

Esta base de datos se compone de 687 imágenes a color de rostros de varios sujetos. No está estandarizada ya que no hay la misma cantidad de fotos, ni las mismas condiciones de iluminación, ni los mismos ángulos de captura de las imágenes de cada sujeto. Cuenta con 90 sujetos con una cantidad de fotografías de cada uno que varía entre 1 y 18. Algunas de ellas tienen variaciones de luminosidad y diferentes ángulos del rostro. No hay información de raza ni máscaras que permitan la comparación de la segmentación de la piel entre diferentes algoritmos.

Esta base de datos es antigua y poco estructurada. La falta de imágenes de ground-truth o máscaras de correspondencia de la segmentación de piel imposibilita la sistematización en la comparación de distintos algoritmos de segmentación. La cantidad de imágenes de prueba y las condiciones de iluminación es aceptable, pero la variabilidad en la cantidad de imágenes de cada sujeto, sumado a lo anteriormente expuesto indica que no es una buena candidata para comparar algoritmos.

5.2.4. Computer Vision Laboratory Face Database

“Computer Vision Laboratory (CVL) Face Database” [Com99] es una base de datos de dominio público creada por el Laboratorio de Computación Gráfica de la Universidad de Liubliana, Eslovenia. Cuenta con un conjunto de imágenes a color de rostros de unos 114 sujetos donde cada uno tiene 7 poses diferentes, 4 con variaciones del ángulo del rostro y 3 poses frontales con variaciones de sus expresiones. Las imágenes corresponden en un 90% a hombres mientras que el resto corresponden a mujeres, en su mayoría son estudiantes de alrededor de unos 18 años, aunque hay algunas fotos de profesores. No hay información referente a las razas incluidas y aparentemente no hay imágenes con variantes de éstas. Los fondos de las imágenes son simples y tampoco hay máscaras de la segmentación de piel.

Esta base de datos no es una buena candidata para utilizar en la comparación de algoritmos de segmentación. No tiene máscaras de píxeles de la piel (ground-truth) por lo que no se pueden sistematizar las pruebas para comparar resultados de segmentación entre algoritmos. Además, al contener imágenes con fondo simple, produce un impacto negativo en la capacidad de adaptación de los algoritmos de segmentación.

5.2.5. Labeled Faces in the Wild Dataset

“Labeled Faces in the Wild Dataset (LFW)” [HRBLM07] contiene imágenes de los rostros de personas conocidas a nivel mundial y está diseñadas para el estudio de problemas de reconocimiento

de rostros sin restricciones. Esta está formada por más de 13.233 imágenes de rostros recolectadas de Internet. Las imágenes se organizan en 5.749 carpetas con los nombres de los sujetos que se encuentran en ellas. La cantidad de imágenes de cada sujeto varía y dentro de cada carpeta se encuentran las fotos del mismo. Existen unas 1.680 personas fotografiadas con 2 o más imágenes diferentes, mientras que el resto tiene solo 1 imagen. Solo hay información asociada al nombre del sujeto y no se dispone de información de donde se encuentra el rostro. Tampoco se encuentra disponible las máscaras de segmentación o ground-truth.

Si bien al explorar las carpetas se pueden encontrar imágenes que contienen rostros de varias razas, no hay una información sobre la proporción de cada una de ellas. Respecto de los fondos de las imágenes, aparentemente no hay una restricción y en general se puede observar que son complejos.

Características como la falta de normalización de la cantidad de imágenes de cada sujeto, la incierta variabilidad de razas, junto con la falta de máscaras de segmentación, determinan claramente que esta base de datos no es una candidata aceptable para la comparación justa de los resultados de los artículos estudiados.

5.2.6. Face Detection Data Set and Benchmark

“Face Detection Data Set and Benchmark (FDDB)” es una base de datos que contiene unos 5.171 rostros en un total de 2.845 imágenes y fue diseñada para uso de algoritmos de detección de rostros. Las imágenes fueron extraídas de la base de datos “Labeled Faces in the Wild Dataset (LFW)”, por lo que este subconjunto tiene las mismas características pero incluye información asociada de la posición del rostro en la imagen. Cuenta con una descripción geométrica de un rectángulo y de una elipse que se ajusta al área del rostro de forma que es posible conocer el espacio donde está ubicado.

Esta base de datos no contiene información de las máscaras de píxeles que se corresponden con la piel pero, según lo observado al explorar las imágenes, las elipses con los recortes se ajustan relativamente bien al área y ángulo del rostro. Por este motivo, si bien no es la situación ideal, podría utilizarse para comparar los resultados de distintos algoritmos de segmentación con algún margen de error.

5.2.7. Edith Cowan University Dataset

El juego de datos “Edith Cowan University (ECU) Face and Skin Detection Dataset” [PBC05] fue creado en dicha universidad y está compuesto mayoritariamente por un conjunto de 4.000 imágenes a color recolectadas de Internet. Dichas imágenes fueron recopiladas de forma manual para asegurar la diversidad en términos de fondos de escena, condiciones de iluminación, rostros y tipos de piel. Cada imagen posee información de las máscaras de píxeles correspondientes a la piel (ground-truth) y fueron preparadas meticulosamente de manera manual, segmentando los rostros y demás regiones de la piel como cuello, brazos y manos. La base de datos se encuentra disponible para uso en investigación y se requiere el llenado de un formulario de consentimiento ¹.

¹<https://www.uow.edu.au/%7Ephung/download.html>

Debido a que esta base de datos cumple con todas las características recomendadas en la sección 5.1 se puede decir que es una buena candidata para la comparación de los resultados tanto de distintos sistemas de representación del color como de los algoritmos de segmentación de piel de los artículos estudiados.

5.2.8. Hand Gesture Recognition Dataset

“Hand Gesture Recognition Dataset” [KKNS14] es un juego de datos que contiene mayoritariamente gestos de la Lengua de Señas Polaca y Americana. Éste fue desarrollado como parte del proyecto de estimación de la pose y detección de manos soportado por el Ministerio de Ciencia y Educación Polaco. Respecto de las razas que componen la base de datos no hay información disponible y tampoco parece haber la variabilidad que se requiere para el tipo de problemas que se trata en este documento.

El juego de datos está dividido en 3 grupos (HGR1, HGR2A and HGR2B) que incluyen: la imagen original en formato jpeg en RGB, la máscara de los píxeles correspondientes a la piel en formato bmp y la ubicación de los puntos de interés en archivo xml.

El primer grupo de la serie se denomina HGR1 y está compuesto por 899 imágenes correspondientes a 25 gestos de 12 sujetos con fondo de escena e iluminación no controlados. El segundo grupo se denomina HGR2A y está compuesto por 85 imágenes correspondientes a 13 gestos de 3 sujetos con fondos de escena gris e iluminación uniforme. El último grupo se denomina HGR2B y está compuesto por 574 imágenes correspondientes a 32 gestos de 18 sujetos con fondos de escena verde e iluminación uniforme. Los 3 grupos del juego de datos se encuentran disponibles para descarga en el sitio ² de difusión y es de uso libre para propósitos no comerciales.

Esta base de datos reúne los requerimientos establecidos en la sección 5.1. Por este motivo es una buena candidata para comparar tanto los sistemas de representación del color como los algoritmos de segmentación relevados de los artículos estudiados.

5.2.9. BioID Face Database

“BioID Face Database” fue diseñada y publicada para posibilitar que los investigadores en el área de detección de rostros tengan la posibilidad de comparar la calidad de sus algoritmos de detección de rostros. La base de datos consta de 1.521 imágenes en escala de grises que contienen los rostros de unos 23 sujetos diferentes. Las imágenes fueron grabadas haciendo especial énfasis en las condiciones del “mundo real”, con buena variedad de iluminación, fondos y tamaños de rostros. Puede ser usada con el propósito que fue diseñada sin mayores permisos.

En la publicación [CZL⁺02] se indica que éste juego de datos se encuentra en colores pero esto, aparentemente no es así. Como no se indica ninguna referencia de la localización no se pudo corroborar que el juego de datos accesible en Internet se corresponda con éste. Por lo tanto, o bien el juego de datos en cuestión no esta disponible o si está disponible es en escala de grises. Por ende, ésta

²<http://sun.aei.polsl.pl/~mkawulok/gestures/>

base de datos no es una candidata para comparar los resultados de los diferentes artículos estudiados porque no permite la comparación de los sistemas de representación del color.

5.2.10. SFA Database

“SFA Database”[CMR13] es una base de datos de piel que tiene como objetivo asistir en las investigaciones que usan el color o la textura de la piel como característica. Esta compuesta por imágenes de rostros provenientes de 876 imágenes de la base de datos FERET³ y 242 imágenes de la base de datos AR⁴. De éstas imágenes se han extraído 3.354 muestras de píxeles correspondientes a la piel y unos 5.590 píxeles correspondientes a muestras que no corresponden a la piel. Los fondos incorporados en las imágenes son simples y homogéneos. Todas las imágenes cuentan con las máscaras de píxeles (ground-truth) correspondientes a la piel.

No se cuenta con información disponible acerca de las razas que están incluidas en la base de datos, y si bien es un base de datos de solo rostros, debido a que las máscaras están asociadas a la piel y no solo al rostro, ésta base de datos es una candidata aceptable para ser utilizada en comparaciones entre diferentes sistemas de representación del color y de los algoritmos de segmentación de piel.

³<https://www.nist.gov/programs-projects/face-recognition-technology-feret>

⁴<http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html>

DISCUSIÓN Y ANÁLISIS DE LOS ARTÍCULOS DE ESTUDIO

Resumen

En este capítulo, a través de sus distintas secciones, se detalla, analiza y discute el funcionamiento y la efectividad expresada en los más de 30 trabajos investigados.

Primeramente, se hace una apreciación sobre la calidad de los resultados obtenidos por los autores de los artículos estudiados y el impacto que esto tiene en el análisis de los mismos.

Seguidamente, se muestra el relevamiento y se realiza un análisis de los distintos modelos de representación de color y el comportamiento que la distribución de la piel tiene en éstos espacios. También se presenta el relevamiento de los algoritmos de segmentación implementados en los artículos y se realiza análisis de la efectividad reportada en los mismos.

Se muestran los resultados de una experiencia realizada durante el desarrollo de este trabajo para segmentación de piel sobre distintos modelos de color utilizando redes neuronales RCE para el reconocimiento de gestos.

Finalmente, se realiza una discusión sobre cual puede ser el mejor sistema de representación del color en base a la información publicada por los autores de los trabajos estudiados y a la experiencia obtenida con redes neuronales RCE.

6.1. Sobre los Artículos Estudiados

En esta sección se realizan unos breves comentarios sobre la calidad los resultados y el camino realizado por los autores de los artículos estudiados para obtenerlos. Durante el análisis de los documentos surgieron algunos inconvenientes al momento de comparar de manera objetiva los resultados obtenidos de aplicar los distintos algoritmos de segmentación sobre los distintos modelos de representación del color.

El único punto que no presenta mayores inconvenientes es que todos los artículos estudiados utilizan las mismas ecuaciones en las transformaciones entre distintos espacios de representación del color. Generalmente las describen con suficiente detalle y no hay variaciones significativas entre ellas.

Un punto que presenta alguna dificultad es la falta de detalles sobre la implementación o parametrización de los algoritmos utilizados para la segmentación. En ocasiones los autores nombran los algoritmos sin dar mayor detalle, omitiendo configuraciones o información de parámetros importantes. También suele suceder que para aquellos algoritmos que generan algún tipo de modelo que requiere de aprendizaje no hay información de como se obtiene. Por estos motivos resulta difícil realizar una comparación justa debido a que algunos los algoritmos de segmentación podrían no ser utilizados en las mismas condiciones o en contextos similares.

El punto realmente crítico son las pruebas realizadas por los autores de los artículos para obtener los resultados. Muchas veces no hay información suficiente para determinar cuando una comparación puede ser justa y otras veces cuando la misma está presente, se puede advertir que la comparación no es completamente justa. En general se pueden observar varios problemas, entre ellos la falta de información sobre los juegos de datos de prueba, el uso de imágenes propias o descargadas de Internet si especificar cuales, la cantidad insuficiente de imágenes para asegurar la calidad de los resultados, entre otros.

Por último, algunos autores incorporan una etapa adicional de verificación para mejorar el grado de detección. Por ejemplo, en el caso de la detección de rostros algunos autores incorporan la localización de los ojos o de la nariz para incrementar la robustez de la detección. Esta etapa adicional incorpora una variante que dificulta aún mas la comparación justa de los resultados de las pruebas incluidas en los artículos.

Algunos de los inconvenientes mencionados anteriormente podrían ser resueltos replicando total o parcialmente los experimentos realizados por los autores. No obstante ante la imposibilidad de contar con el juego de datos original en una buena cantidad de casos, hace que no sea viable la replica de estos experimentos. Para resolver los problemas anteriores sería factible realizar pruebas que combinen de manera exhaustiva todos modelos de representación del color con todos los algoritmos de segmentación documentados en los artículos, utilizando además juegos de datos con información de máscaras de segmentación para sistematizar la obtención y comparación objetiva de los resultados. Lamentablemente, por el esfuerzo y el requerimiento de tiempo que supone esta solución deja su implementación fuera del alcance de este trabajo, aunque queda planteada para uno próximo.

6.2. Algoritmos de Segmentación de los Artículos de Estudio

Del estudio realizado por los trabajos se relevó el uso de los algoritmos Vector Quantization (VQ), Gaussian Model (GM), Mixture Gaussian Model (GMM), Fixed Threshold (FT), Automatic Threshold (AT), Rules of Thresholds (RT), Naives Bayes Classifier (NBC), Support Vector Machine (SVM), K Nearest Neighbour (KNN), K-Means, Multi-layer Perceptron (MLP), Self Organizing Maps (SOM) y Restricted Coulomb Energy (RCE).

En el cuadro 6.1 se presentan las referencias a los artículos estudiados y se marcan los algoritmos que se utilizan para realizar la segmentación de la piel en el artículo correspondiente. En las columnas relacionadas a los algoritmos, estos se encuentran agrupados por categoría similar de implementación, aunque esta última no sea exactamente la misma. Cada uno tiene variantes propias, ajustes particulares y/o parametrizaciones diferentes. Por ejemplo en el caso de los algoritmos basados en umbrales están agrupados los de umbrales fijos, los de umbrales dinámicos y los basados en reglas. Respecto de los algoritmos de Mezclas de Gaussianas, éstos se agrupan sin tener en cuenta las distintas cantidades de Gaussianas que puedan utilizar.

En la figura 6.1 se muestran dos gráficos de torta con el porcentaje de algoritmos utilizados en los artículos de estudio. En la parte *a*) de la figura 6.1 se encuentra un gráfico con la distribución de los algoritmos utilizados en los artículos de estudio agrupados en categorías de mas relevancia según la cantidad artículos. En éste se puede observar que los algoritmos de segmentación basados en umbrales son los mas ampliamente utilizados con casi un 49%. Luego siguen los algoritmos basados en Redes Neuronales con casi un 18% y en tercer lugar se encuentran los algoritmos que utilizan Gaussianas y Mezclas de Gaussianas con un poco mas de un 15%. Si bien el porcentaje de los algoritmos restantes alcanza al de las Redes Neuronales, se los pone en último lugar por la diversidad de sus diferentes implementaciones.

Las razones por las cuales los métodos de segmentación Basados en Umbrales dominan las implementaciones, posiblemente radiquen en que son algoritmos simples, de bajo costo computacional, de implementación directa y rápida. Además si las condiciones de captura de las imágenes a segmentar son relativamente controladas respecto de la iluminación, las sombras y/o el fondo, suelen garantizar buenos resultados. Otra cuestión que influye de manera significativa es el uso de la transformación a un modelo de color donde la distribución de la piel se encuentra alineada en el sentido de los ejes. Si bien esta aproximación tiene la desventaja de requerir un mayor costo computacional por la transformación, tiene la ventaja de facilitar un mejor ajuste a la distribución de la piel para métodos de segmentación poco robustos como es el caso de este tipo de algoritmos.

En la parte *b*) de la figura 6.1 se muestra un gráfico de torta con los algoritmos agrupados por su tipo según la cantidad de veces que fueron utilizados en los artículos de estudio. Aquí se puede observar que los algoritmos de segmentación basados en umbrales fijos son los mayoritariamente usados con mas de un 28%. Si bien no tienen los mismos porcentajes, en un segundo nivel se encuentran los algoritmos basados en umbrales de reglas con casi un 13% y es escalón apenas mas abajo las mezclas de gaussianas y las redes neuronales multi-perceptrón con poco mas del 10%. Finalmente en un tercer nivel los algoritmos basados en umbrales automáticos con casi un 8% seguido por el resto de los algoritmos con un valor que va desde el 2.6% al 5.1%. Al igual que sucede en

Artículo	Algoritmos Utilizados en la Segmentación de Piel												
	VQ	GM	GMM	FT	AT	RT	NBC	SVM	KNN	K-Means	MLP	SOM	RCE
[ZMXNJS06]	✓												
[NKN16]				✓									
[XP17]				✓									
[HSW08]			✓										
[GCRÁCP ⁺ 09]								✓					
[Lei13]		✓			✓								
[TKSY17]				✓									
[BAO17]				✓						✓			
[YMS17]				✓									
[HBGR09]						✓			✓				
[IK13]			✓										
[BV15]				✓									
[HATWSS14]						✓							
[TPM ⁺ 11]						✓							
[SGK ⁺ 15]				✓									
[KKS ⁺ 17]						✓							
[RCGS16]						✓							
[BK10]											✓		
[XS13]				✓									
[GMX11]		✓		✓									
[BCL01]												✓	
[PCB01]											✓		
[CPB03]							✓						
[CZL ⁺ 02]											✓		
[HTMK02]			✓										
[RTBA12]				✓									
[LzP10]					✓								
[KK12]				✓									
[SCVS03]				✓									
[YGX01]													✓
[SKR11]													✓

Cuadro 6.1: Algoritmos de segmentación de la piel implementados por los artículos de estudio. VQ: Cuantificación Vectorial. GM: Modelo de Gaussiana. GMM: Modelo de Mezcla de Gaussianas. FT: Umbral fijo. AT: Umbral automático. RT: Umbral por reglas. NBC: Clasificador Bayesiano. SVM: Máquinas de Vectores de Soporte. KNN: K Vecinos Cercanos. K-Means: K Medias. MPL: Perceptrón Multicapa. SOM: Mapas Auto-organizativos. RCE: Redes de Energía de Coulomb.

el gráfico 6.1 a) los algoritmos más utilizados son los basados en umbrales fijos y umbrales con reglas que son los de menor complejidad y los de menor costo computacional. Luego siguen algoritmos mas robustos como las Redes Multi-perceptrón conocidas por su capacidad de adaptación y las Mezclas de Gaussianas que particularmente pueden adaptarse bien a la distribución de la piel de varios de los modelos de representación de color. También cabe aclarar que otros algoritmos, tal vez menos populares, si bien no son muy utilizados han demostrado tener buenos resultados para la segmetación de piel con diferentes modelos de color como por ejemplo es el caso de la Redes Neuronales RCE comentado en la sección 6.3.3.

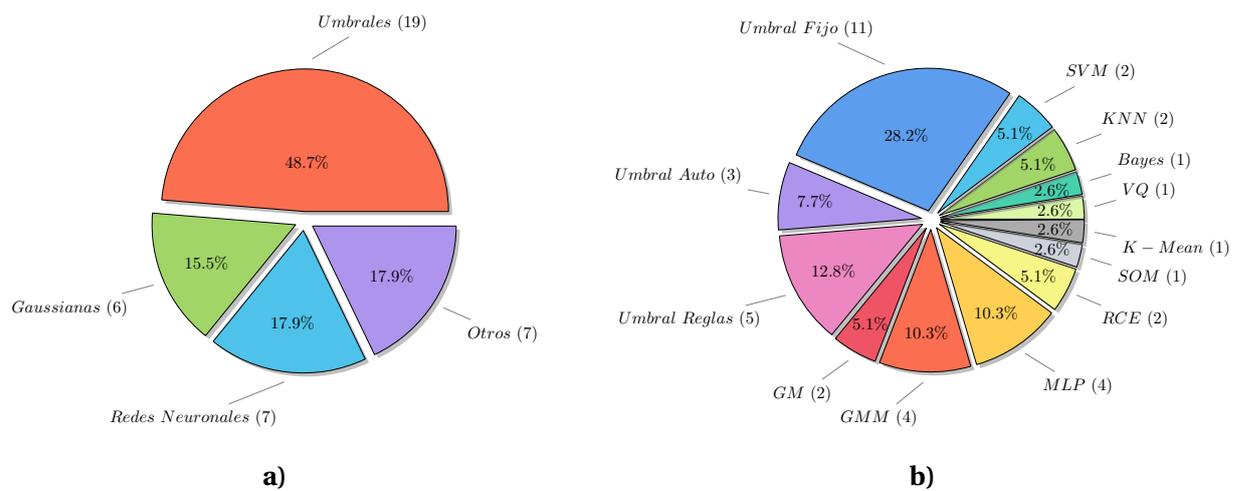


Figura 6.1: a) Categorización de algoritmos en grupos de relevancia. b) Desglose de los tipos de algoritmos utilizados en los trabajos estudiados.

6.3. Modelos de Color de los Artículos de Estudio

6.3.1. Uso de los Modelos de Color

En la sección 2.2 de este documento se describió en detalle los modelos de color RGB, RGB normalizado, HSV, HSL, HSI, YUV, YIQ, YCbCr, YCgCr y Cie-LAB, utilizados para la segmentación de la piel en los artículos de estudio. En esta sección y en las subsiguientes se desarrolla un análisis sobre su uso y efectividad.

En el cuadro 6.2 se presenta de manera compacta un resumen con la información relacionada a los sistemas de representación de color de los artículos. Respecto de las filas, hay una por cada artículo relevado. Respecto de las columnas, una muestra la referencia al artículo, un grupo de columnas para todos los modelos de color que contiene una marca cuando éste es utilizado por el artículo correspondiente y finalmente una columna para indicar cuando los modelos de color marcados en las columnas anteriores se utilizan de manera *combinada* o *independiente*. En particular el término *independiente* indica que los autores del artículo han hecho experimentos individuales con el objetivo de comparar los resultados, mientras que el término *combinada* se refiere a que en

Artículo	Modelo de Representación del Color									
	Modelo Combinado	RGB	RGB _{Norm}	HSV	HSI	Y _{C_bC_r}	Y _{C_gC_r}	YUV	YIQ	Cie-LAB
[ZMXNJS06]		✓								
[NKN16]	✓			✓		✓				
[XP17]							✓			
[HSW08]						✓				
[GCRÁCP ⁺ 09]			✓			✓				✓
[Lei13]						✓				
[TKSY17]						✓				
[BAO17]	✓			✓		✓				
[YMS17]					✓	✓		✓	✓	
[HBGR09]		✓								
[IK13]	✓		✓	✓		✓				
[BV15]	✓			✓		✓				
[HATWSS14]	✓	✓						✓		
[TPM ⁺ 11]	✓	✓		✓		✓				
[SGK ⁺ 15]				✓		✓				
[KKS ⁺ 17]	✓	✓		✓		✓				
[RCGS16]	✓		✓	✓		✓				
[BK10]		✓								
[XS13]	✓	✓						✓		
[GMX11]	✓			✓			✓			
[BCL01]			✓	✓						
[PCB01]						✓				
[CPB03]		✓								
[CZL ⁺ 02]			✓							
[HTMK02]				✓						
[RTBA12]	✓	✓		✓		✓				✓
[LzP10]				✓						
[KK12]						✓				✓
[SCVS03]	✓	✓			✓	✓				
[YGX01]	✓				✓					✓
[SKR11]						✓				

Cuadro 6.2: Sistemas de representación del color utilizados por los artículos de estudio.

sus experimentos los autores combinaron en un modelo al menos dos componente de diferentes sistemas de representación del color marcado en las columnas correspondientes.

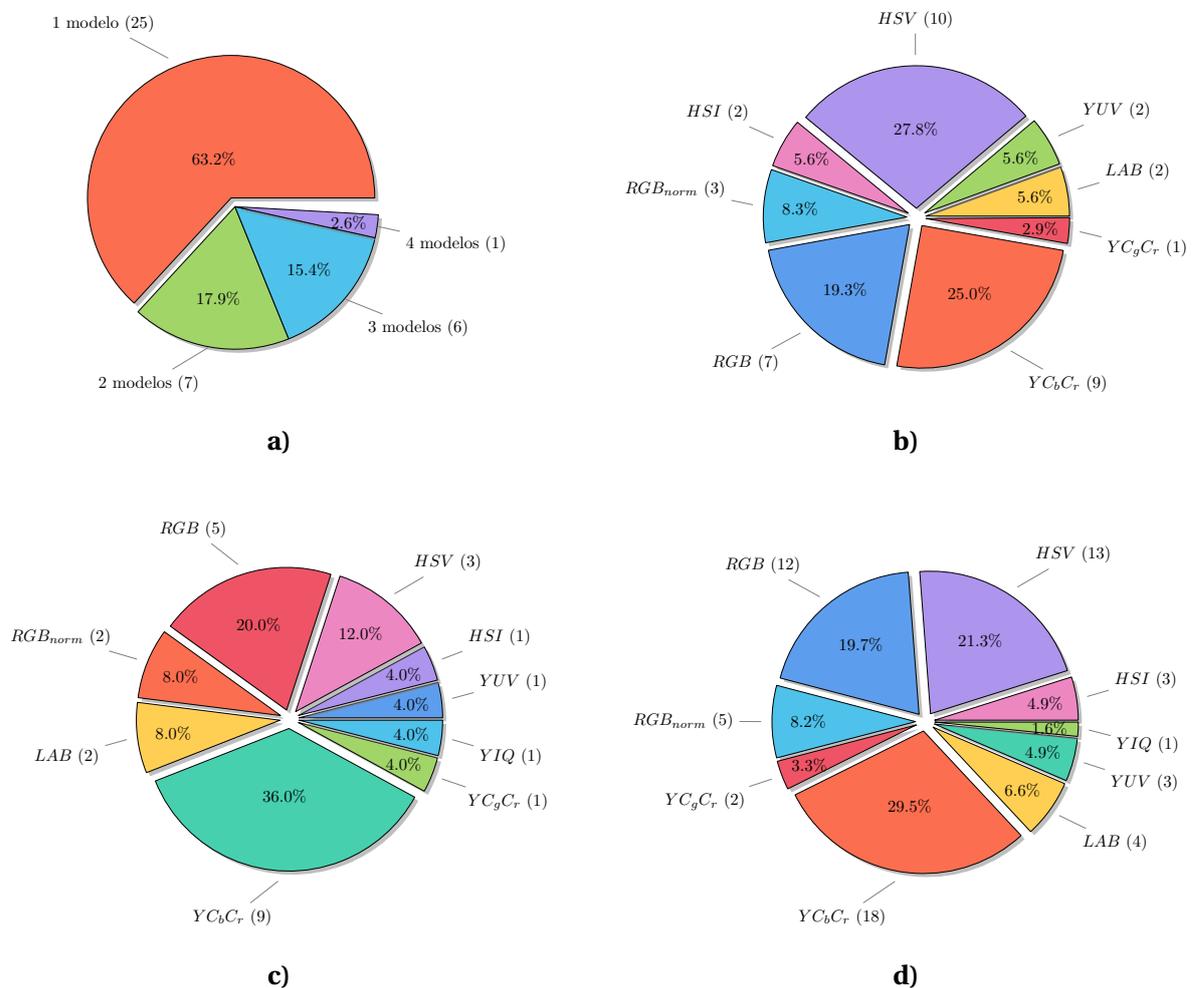


Figura 6.2: a) Experimentos agrupados por cantidad de modelos de color que se utilizan para segmentar la piel. b) Experimentos que usan 2 o más modelos de color agrupados por modelo. c) Experimentos que usan 1 modelo de color agrupados por modelo. d) Experimentos agrupados por modelos de color sin distinguir la cantidad que utilizan.

En la figura 6.2 se presenta una serie de gráficos de torta que muestran la frecuencia de uso de los diferentes sistemas de representación del color en los distintos experimentos realizados en los artículos. En el gráfico 6.2a) se contabilizan la cantidad de sistemas de representación utilizados en los artículos. En este se puede apreciar que poco más del 63% de los artículos utiliza un único sistema de color. Del 37% restante que utilizan al menos una combinación de componentes de 2 o mas sistemas, prácticamente un 18% utilizan 2 sistemas, un 15% utilizan 3 sistemas y solo un 2.6% utilizan componentes de 4 sistemas.

En el gráfico 6.2b) se contabilizan los sistemas de color que han sido combinados en los experimentos con componentes de color de 2 o mas sistemas de representación. Se puede observar que el mas utilizado es el sistemas HSV con casi un 28%, lo que es razonable debido a las

características que posee su componente de matiz H. Cabe destacar que si se agrupa con el sistema HSI que tiene similares características llega a superar el 33%. El segundo más utilizado es el sistema YCbCr con un 25% pero si se lo agrupa por similitud con otros sistemas como el YCgCr y el YUV se obtiene un 33%. El tercer sistema más utilizado es el RGB con un 19% pero si se lo agrupa también por similitud con su versión normalizada se obtiene un poco más del 27%. En cualquiera de los casos se observa que la proporción de uso es similar en los tres sistemas.

En el gráfico 6.2c) se contabilizan los experimentos donde se utiliza un solo sistema de color en la segmentación. Se puede observar que el 36% corresponde al modelo YCbCr seguido por el sistema RGB con un 20% y en tercer lugar se encuentra algo más lejos el sistema HSV con un 12%. Claramente el más utilizado es el YCbCr por un buen margen.

Finalmente, en el gráfico 6.2d) se contabilizan los experimentos que involucran algún componente de cada sistema de representación del color sin tener en cuenta si se usa en combinación o no con otros sistemas. Nuevamente el sistema YCbCr es el que cuenta con un mayor uso con más de un 29%, luego es seguido por el HSV con poco más de 21% y finalmente en tercer lugar el sistema RGB con casi un 20%.

6.3.2. Distribución de la Piel en los Modelos de Color

En esta sección se muestra en el cuadro 6.3 una tabla con un resumen donde se puede apreciar para cada artículo la referencia del autor, el tipo de segmentación (rostro, manos, piel), el sistema de color o la combinación que utiliza de estos, el algoritmo de segmentación y el porcentaje de acierto.

Se realiza un análisis de la efectividad de los distintos sistemas de representación del color. Es importante remarcar dos cuestiones a la hora de analizar los resultados. La primera es que los valores utilizados en el análisis son los reportados por los autores de los artículos siempre y cuando estos estuvieran disponibles, sin importar los condicionamientos mencionados en la sección 6.1. La segunda cuestión es que la mayoría de los autores no especifican el juego de datos o las imágenes que utilizan ni las condiciones de éstas al momento de realizar las mediciones. En este sentido algunos autores podrían asumir condiciones muy controladas como buena condición de luz, fondos simples, poco ruido, buena resolución para realizar las pruebas. Mientras que otros autores podrían utilizar imágenes con variabilidad de luz o condiciones variables de fondo. Está claro que la comparación de estos no sería ciento por ciento justa porque la efectividad de los artículos incluidos en el primer grupo será mejor que la de los incluidos en el segundo y no reflejarán de manera justa la efectividad de la representación.

A continuación, en las siguientes subsecciones se realiza un análisis basado en el cuadro 6.3. Algunos sistemas de representación del color que comparten sus características más importantes se encuentran agrupados por su similitud con el objetivo de simplificar el análisis.

Efectividad de la Segmentación de Piel en los Sistemas de Color				
Artículo	Aplicación	Sistema de Color	Algoritmo	Acierto (%)
[NKN16]	Rostro	HSV + YCbCr	SVM	86,00
[ZMXNJS06]	Rostro	RGB	VQ	98,33
[XP17]	Mano	YCgCr	Threshold	—
[HSW08]	Piel	YCbCr	GMM	92,00
[GCRÁCP+09]	Piel	YCbCr	GMM	77,40
			MPL	82,80
			SVM	83,80
		LAB	GMM	85,80
			MPL	89,00
			SVM	89,40
		RGB _{norm}	GMM	85,20
			MPL	87,40
			SVM	88,40
[Lei13]	Rostro	YCbCr	Automatic Threshold	90,00
[TKSY17]	Rostro	YCbCr	Threshold	—
[BAO17]	Piel	HSV + YCbCr	Automatic Threshold + K-Mean	81,00
[YMS17]	Piel	RGB	Threshold	88,11
		YCbCr	Threshold	97,22
		YUV	Threshold	96,11
		YIQ	Threshold	93,06
		HSI	Threshold	92,50
[HBGR09]	Rostro	RGB	Threshold	—
[IK13]	Piel	RG _{norm} +HS+CbCr	GMM	98,08
[BV15]	Rostro	HSV+YCbCr	Threhold	95,00
[HATWSS14]	Piel	RGB+YUV	Threshold	95,11
[TPM+11]	Rostro	RGB+HSV+YCbCr	Threshold Rules	95,81
[SGK+15]	Piel	HSV + YCbCr	Threshold	—
[KKS+17]	Piel	RGB+HSV+YCbCr	Threshold Rules	94,43
[RCGS16]	Piel	RGB _{norm} +HSV+YCbCr	Threshold Rules	91,05
[BK10]	Piel	RGB	MLP	94,32
[XS13]	Piel	RGB+YUV	Threshold	—
[GMX11]	Piel	HSV+YCgCr	GM	95,45
[BCL01]	Piel	RGB+HSV	SOM	94,00
[PCB01]	Piel	YCbCr	MLP	91,60
[CPB03]	Piel	RGB	Bayesian Classifier	95,36
[CZL+02]	Rostro	RGB Norm	MLP	—
[HTMK02]	Rostro	HSV	GMM	—
[RTBA12]	Rostro	RGB+HSV+YCbCr+LAB	Threshold	92,30
[LzP10]	Rostro	HSV	Automatic Threshold	88,90
[KK12]	Rostro	YCbCr	Threshold	—
[SCVS03]	Rostro	RGB+HSI+YCbCr	Threshold	95,18
[YGX01]	Mano	RGB+HSI+LAB	RCE	95,00
[SKR11]	Mano	YCbCr	RCE	—

Cuadro 6.3: Efectividad de la segmentación reportada en los artículos de estudio.

Sistema de Representación RGB

Sobre éste sistema de representación se puede verificar que en general se obtienen muy buenos resultados, superando el 94 % de efectividad en algoritmos de segmentación como VQ[ZMXNJS06], MLP[BK10] y Bayes Classifier[CPB03]. También se puede verificar que tiene valores de efectividad mucho menor cuando se aplica un algoritmo basado en umbrales[YMS17]. No obstante la efectividad mejora cuando este sistema se combina con componentes de otros sistemas de representación con algoritmos de segmentación basados en umbrales [HATWSS14, BV15, SCVS03, YGX01] y en redes neuronales[BCL01, YMS17].

Sistema de Representación RGB Normalizado

No hay muchos artículos que incluyan este sistema de representación por lo que no hay mucha información para determinar su efectividad. Este tipo de representación muestra los valores mas bajos de todos los artículos. Se utiliza con algoritmos GMM, MPL y SVM con una efectividad promedio de 87 % según[GCRÁCP⁺09], que es una de las mas bajas relevadas. En general este sistema no se combina con otros y en particular cuando este se combinó con componentes de HS y CbCr los autores [IK13] reportaron una efectividad del 98 % para GMM.

Sistema de Representación YCbCr e YUV

Estos dos sistemas de representación pueden agruparse ya que simplemente tienen una diferencia mínima en la manera en que son codificados. YCbCr es una versión escalada y desplazada del espacio de color YUV, por este motivo se incluyen juntos en esta sección. De todos los sistemas de representación relevados, éste es uno de los más utilizados tanto de forma individual como combinada. Respecto de la segmentación sobre éste sistema alcanza el 90 % de efectividad con GMM[HSW08] y supera el 91 % con MPL[PCB01]. Respecto de la segmentación combinada generalmente se utiliza con componentes de los sistemas RGB y HSV. Con estos sistemas la segmentación alcanza una efectividad del 95 % con algoritmos basados en umbrales[SCVS03, TPM⁺11, BV15] y supera el 98 % con GMM[IK13].

Sistema de Representación YCgCr

El sistema YCgCr fue creado en 2003 para aplicarlo en la detección de rostros[dDG03]. Como no ha se han encontrado otros artículos que incluyan este tipo de modelo no hay mucha información sobre su uso en segmentación de piel con distintos algoritmos. En este trabajo los autores reportaron una efectividad superior al 95 % para la segmentación basada en GM[GMX11] en combinación con componentes del espacio HSV.

Sistema de Representación HSL, HSI y HSV

Los sistemas basados en un componente de matiz como HSV, HSI y HSL son muy similares (el componente H es igual en todos) y su diferencia principal se encuentra en la forma de representación del componente de luminosidad (V, L e I) que frecuentemente suele no utilizarse. Según varios de los trabajos relevados, estos modelos presentan buenos resultados con algoritmos basados en umbrales alcanzando una efectividad promedio de más del 90% [LzP10, YMS17]. Respecto de su uso en combinación con componentes de otros sistemas es uno de los más utilizados (en general se usan los componentes H y S) y los autores reportan una efectividad de más del 95% en la segmentación con algoritmos RCE [YGX01], GM [GMX11] y los basados en umbrales [SCVS03, GMX11, TPM⁺11, BV15]

Sistema de Representación Cie-LAB

No hay muchos artículos que utilicen el sistema de representación Cie-LAB, posiblemente por ser uno de los que requiere mayor costo computacional para su transformación a partir del modelo RGB. En el artículo [GCRÁCP⁺09] se realizan una serie de pruebas entre varios sistemas de color y varios algoritmos de segmentación y si bien en general los resultados son de los peores de todos los artículos (ninguno llega al 90% de acierto), el sistema LAB presenta mejores resultados que YCbCr y RGB normalizado para MPL y SVM. Al igual que otros sistemas ha presentado mejores resultados cuando se combinan con otros componentes de representación del color. Para algoritmos como los basados en umbrales ha obtenido una efectividad de más del 92% según [RTBA12] mientras que para RCE ha alcanzado un 95% según [YMS17].

Sistemas Combinados

Hay una buena parte de los artículos estudiados que utilizan componentes de dos, tres y hasta cuatro modelos de representación de color para construir un modelo híbrido. Del análisis de estos artículos parecería haber una tendencia general donde la efectividad en la segmentación mejora. Como evidencia de esto se pueden mencionar dos casos particulares. El primero consiste en modelos que utilizan distintos algoritmos de segmentación con una efectividad que va del 94% al 98% como es el caso de redes RCE [YGX01], redes SOM [BCL01], GM [GMX11], GMM [IK13]. El segundo caso son los algoritmos basados en umbrales que muestran una efectividad que va del 94% al 96%, entre ellos [SCVS03, TPM⁺11, HATWSS14, BV15, KKS⁺17]. Teniendo en cuenta ésto último y lo expuesto en los apartados anteriores, en general se puede afirmar que los algoritmos de segmentación basados en umbrales solo producen resultados aceptables cuando se utilizan transformaciones del espacio de color de por lo menos un sistema, teniendo mejores resultados cuando se utilizan dos o más.

6.3.3. Un Experimento sobre Redes Neuronales de Energía de Coloumb

Durante el desarrollo de este trabajo se realizó de forma paralela una tesina de grado [Gen17] y se publicó un artículo [GEL17] donde se realizaba la segmentación de piel de manos utilizando redes neuronales de Energía de Coulomb (RCE) para el reconocimiento de gestos. En estos trabajos

se desarrollo un proceso de varias etapas donde primeramente se captura una imagen a través de una webcam HD, posteriormente se segmenta la piel de las manos, luego se determina la configuración de la mano y finalmente se realiza el seguimiento de la mano a lo largo del tiempo para determinar el gesto.

Como parte del desarrollo de ambos trabajos se tomaron los modelos de representación del color RGB, HSL, HSV, Cie-LAB y YCbCr para realizar una serie de pruebas que los combinan con una red neuronal de Energía de Coulomb para segmentar la piel de las manos. El objetivo de las mismas fue determinar cuál de los modelos ofrecía una mayor efectividad en la segmentación de piel al utilizar una red neuronal de tipo RCE. Para realizar las pruebas se utilizó la base de datos MOHI [HAAB⁺ 15] que contiene varias muestras de manos pertenecientes a 200 personas diferentes con variantes de rotación, escala y condiciones de iluminación.

Para determinar la efectividad de cada uno de los modelos de color se llevó adelante un procedimiento que constó de cuatro fases. Primeramente se construyeron cinco versiones de la base de datos MOHI con las imágenes transformadas en los sistemas de representación de color RGB, HSV, HSL, Cie-LAB y YCbCr. Seguidamente se realizaron una serie de entrenamientos variando los valores de los parámetros para obtener distintas configuraciones de redes RCE para cada uno de los cinco modelos de color. En la tercera fase se realizó la segmentación de las imágenes con cada variante de la red entrenada para cada uno de los cinco modelos de color. Finalmente se evaluó el rendimiento de cada variante de las red RCE entrenadas ponderando sus características y midiendo su efectividad con el objetivo de seleccionar las más exitosas para cada modelo de color.

En la figura 6.3 se expone parte de los resultados obtenidos en las diferentes pruebas realizadas, en particular se muestran dos de los gráficos de barras más relevantes. Por un lado, el gráfico de la izquierda muestra el promedio de píxeles segmentados por las cinco redes neuronales RCE (la red de mayor efectividad por cada modelo de color). En este se puede observar que no hay diferencias significativas en los resultados de la segmentación con las diferentes redes entrenadas para cada sistema de representación. De esto se desprende que en cada uno de los sistemas de representación del color, las redes se adaptaron a cada espacio de distribución de la piel sin importar el grado de complejidad de éste. Por otro lado, el gráfico de la derecha muestra la cantidad de neuronas generadas para cada red (la más exitosa por cada modelo de color) durante la fase de aprendizaje. En éste se puede observar como influye la complejidad del espacio de distribución de la piel (ver sección 4.2) sobre la cantidad de neuronas requeridas para cubrirlo. En otras palabras, una menor o mayor complejidad del espacio implica una menor o mayor cantidad de neuronas para abarcarlo, respectivamente.

De los gráficos mencionados anteriormente se pueden concluir dos cosas. La primera esta relacionada a la efectividad y es que este tipo de red tiene la capacidad de adaptarse al espacio a cubrir independientemente de la complejidad de este. En el marco de este trabajo esto es un indicio de que si el algoritmo utilizado para la segmentación es suficientemente robusto para adaptarse a la distribución de la piel, no resultaría de relevancia el sistema de representación del color usado. La segunda esta relacionada a la conveniencia de la conversión o no a otro modelo de color. Según lo observado, parecería que los modelos mas prometedores o convenientes serían el RGB y el YCbCr teniendo en cuenta tanto la efectividad como la cantidad de neuronas requeridas, aunque cada

uno tiene sus ventajas y desventajas sobre el otro. El modelo RGB tiene la ventaja de ser el modelo utilizado naturalmente para representar imágenes, lo que ahorra el tiempo y costo de conversión a otro modelo. Sin embargo al tener un espacio de distribución de la piel mas complejo, requiere más tiempo de entrenamiento para generar la red y más tiempo al momento de segmentar la piel. En el caso del sistema YCbCr si bien es uno de los que requiere menor tiempo y costo computacional de conversión, esto es una desventaja respecto del RGB ya que no lo tiene. No obstante al tener un espacio de distribución de piel mas simple requiere de menos neuronas, lo que implica menos tiempo de entrenamiento y menos tiempo para segmentar. Respecto de los demás sistemas parecería que no tienen ventajas sobre los dos anteriores. Por un lado tienen un costo computacional igual o mayor que el YCbCr para convertir las imágenes y por otro al tener espacios de distribución al menos moderadamente complejos implica un costo mayor para generar y usar las redes en la segmentación. En consecuencia como no tienen ventaja en la efectividad y el costo computacional total es mayor, en principio, parecería que no es conveniente su uso, aunque esto es algo que debería ser comprobado a través de experimentos para despejar esta duda de manera definitiva.

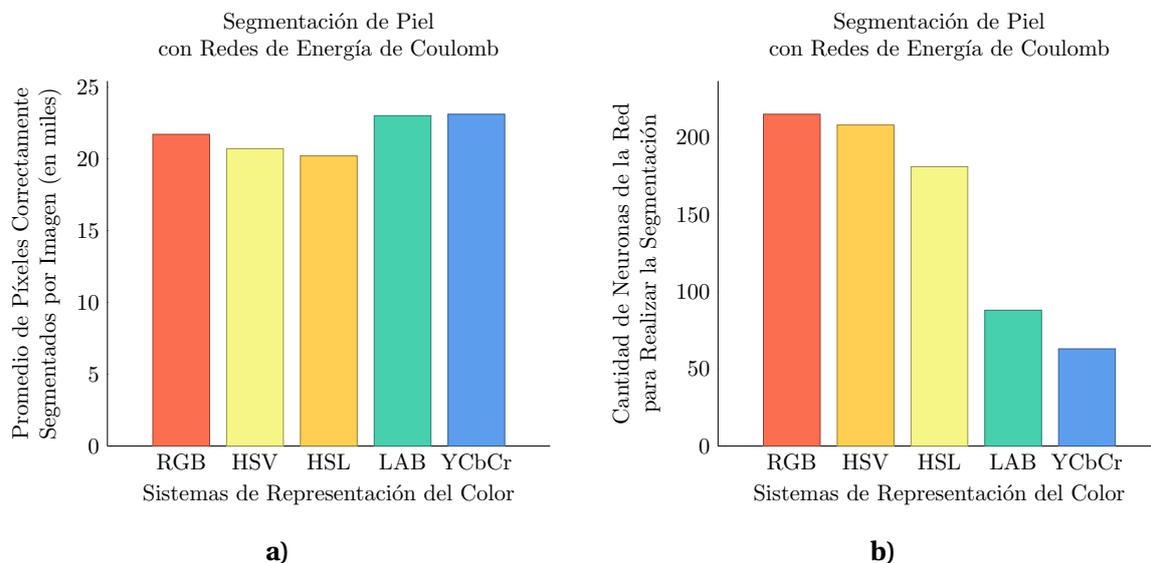


Figura 6.3: a) Promedio de píxeles segmentados para cada sistema. b) Cantidad de Neuronas requeridas para la segmentación de las imágenes para cada sistema.

6.4. El Mejor Modelo de Color para Segmentar Piel

En las dos últimas décadas se han escrito una gran cantidad de trabajos de investigación que abordan la segmentación de la piel humana utilizando diferentes modelos o espacios de color. Resulta claro que si existiera un espacio de color óptimo para representar la distribución del color de la piel, en un punto ya debería haber sido propuesto y a partir de ese punto todos los demás trabajos realizados deberían usarlo.

Es obvio que todos los espacios o modelos de color tienen sus ventajas y desventajas. Asumiendo que exista un modelo de color que tenga una representación óptima para la distribución del color de

la piel, no hay una implicación justificada de que ese modelo sirva para segmentar mejor la piel en si mismo. Más allá de todas las transformaciones entre todos los modelos de color, la información necesaria para determinar si un pixel es o no piel, está presente en todos ellos, sin importar la distribución.

La elección de un modelo de color determinado con una distribución más compacta, uniforme y/o alineada a los ejes del espacio de color influye en el algoritmo de clasificación. Por un lado, una distribución compacta posiblemente impacte en la velocidad de ejecución, en la cantidad de memoria utilizada, en el costo computacional o en su rendimiento general. Por otro lado, puede mitigar la falta de robustez de algoritmos de segmentación relativamente simples. De cualquier manera, ninguna de estas cuestiones implica que un modelo de color sea mejor que otro en si mismo.

En el cuadro 6.3 se encuentra el resumen de todos los trabajos estudiados y se muestra la efectividad (cuando estuvo disponible) de las combinaciones del espacio de color junto con el algoritmo de clasificación utilizado. Entre los algoritmos utilizados en los trabajos se encuentran Bayes Classifier (Bayes), Multilayer Perceptron (MLP), Self Organizing Maps (SOM), Restricted Energy Coulomb Network (RCE), Single Gaussian Model (GM), Gaussian Mixture Model (GMM), Simple Thresholding, Automatic Thresholding, Thresholding Rules, Vector Quantization (VQ), Support Vector Machine (SVM) y K-Means. En general puede observarse que se obtienen buenos resultados con diferentes combinaciones de modelos de color y algoritmos de segmentación lo cual es un indicio de que el sistema de representación de color no es determinante, por lo menos para la mayor parte de los algoritmos. Por ejemplo, para el modelo RGB se obtuvieron resultados que van del 94 % a más del 98 % con algoritmos como VQ, MLP y Bayes. Algo que aparentemente si se puede observar es que los resultados de la segmentación mejoran cuando se utilizan componentes de varios modelos de color.

A partir de todo lo estudiado, no puede concluirse que hay un modelo de color ideal para representar la distribución de la piel de manera que facilite la segmentación de la misma. En principio, éste se complementa con la capacidad del algoritmo de clasificación para determinar si un pixel de una imagen es piel o no es piel. Por lo tanto el éxito o fracaso queda determinado por la combinación de la dupla espacio de color-clasificador. En este sentido sería de esperar que en general diferentes combinaciones espacio-clasificador tuvieran un éxito similar [HSW08, YMS17, GCRÁCP⁺09], mas allá de que haya alguna combinación puntualmente superior a las demás [ZMXNJS06, YMS17]. Tampoco sería de sorprender que tengan tanto valores similares de coincidencias como de errores al clasificar un pixel como piel. Para expresar la idea con un ejemplo concreto sobre lo anterior, en el espacio HSV la distribución del color de la piel tiene una representación más compacta (alineada a los ejes del espacio) que en el espacio RGB. Si se utiliza como algoritmo de clasificación un par de umbrales por cada eje es simple de comprobar visualmente revisando las figuras 4.1 y 4.3 que el volumen acotado por los umbrales se ajustará mejor al modelo HSV y sería simple concluir que este modelo de color es mejor. En cambio, si se utilizan un par de rectas por cada plano en vez de umbrales, estas ajustarían de mejor manera al volumen en el espacio RGB y la conclusión anterior no sería, por lo menos, tan directa. No obstante, con cualquiera de los dos algoritmos, concluir que uno de los dos modelos de color es mejor que el otro no tiene en principio una justificación fundada, y es algo que la gran parte de los autores de los trabajos analizados no hacen.

Para algoritmos de segmentación poco robustos como los basados en umbrales, pareciera que se requiere la conversión a un sistema de color donde la distribución de la piel sea “uniforme” y en lo posible, paralela a los ejes. Pareciera que cuando más compleja y con mayor costo computacional es la conversión mejores resultados se obtienen. En cambio, cuando los algoritmos de segmentación son mas robustos, parecería que requieren (en algunos casos o no) conversiones de menor costo computacional para obtener un éxito similar. Dicho en otras palabras, cuanto mayor es la capacidad de adaptación de un algoritmo para cubrir el espacio de representación de la piel menor es la dependencia de conversión de color a otro sistema.

Nada de lo comentado anteriormente es totalmente definitivo, al menos teniendo en cuenta lo que se desprende de los resultados mostrados por los autores de los artículos estudiados. Solo son indicios pero para llegar a una conclusión definitiva, la única solución sería realizar la implementación de los distintos algoritmos, aplicarlos sobre todos los sistemas de color y contrastar los resultados con un juego de datos que garantice una comparación justa. Para esto es necesario incorporar pruebas que incluyan juegos de datos con información de máscaras de segmentación (ground-truth), variabilidad de razas, equidad en las variantes de imágenes, condiciones de iluminación variables, cantidad de ejemplos suficientes para garantizar los resultados y fondos de diferentes complejidades.

CONCLUSIONES Y TRABAJOS FUTUROS

7.1. Conclusiones

A lo largo del desarrollo de este trabajo se recopilaron y analizaron mas de 30 artículos sobre segmentación de piel basada en píxeles utilizando diferentes sistemas de representación del color y distintos algoritmos de segmentación.

Como parte del aporte realizado por este trabajo caben destacar tres puntos interesantes sobre el relevamiento realizado, que sirven como puntapie para futuros trabajos o investigaciones. El primer punto es la recopilación de los sistemas o modelos de representación del color frecuentemente utilizados junto con las fórmulas de conversión para transformar los distintos espacios de color a partir del modelo de color RGB que es el formato estándar para representación de imágenes. El segundo punto es el relevamiento de los algoritmos de segmentación aplicados sobre los diferentes modelos del color junto con los resultados de efectividad reportados por los autores de los artículos estudiados. Finalmente, el tercer punto a destacar es el relevamiento de los juegos o bases de datos reportado tanto en los artículos de estudio como los recopilados por el autor de este documento. De esto surgieron tres juegos de datos que cumplen con todas las características recomendadas en la sección 5.1, “Edith Cowan University (ECU) Face and Skin Detection Dataset” [PBC05], “Hand Gesture Recognition Dataset” [KKNS14] y ‘SFA Database’[CMR13]. Dichas características permiten realizar una comparación justa entre los resultados de distintos algoritmos de segmentación de piel y en consecuencia, una justa evaluación de la efectividad de modelos del color.

Otro aporte de este trabajo es análisis de la distribución de piel en los sistemas de representación de color realizado en 4.3. De aquí se puede concluir que en algunos sistemas la distribución de los píxeles correspondientes a la piel se ven más compactas y acotadas. Un ejemplo de esto son los modelos HSV y el YCbCr que presentan unas de las distribuciones mas compactas y alineadas de forma paralela a los ejes del espacio de representación.

Respecto de cuál es el modelo de color mas apropiado para utilizar en la segmentación de piel, no se ha obtenido una respuesta contundente. Después de haber recopilado una buena cantidad de publicaciones y realizado un análisis de los resultados expuestos por los respectivos autores, parecería que no hay un modelo de color que supere claramente a los demás. En los artículos el modelo de color YCbCr es el más utilizado de manera individual y los modelos HSV, YCbCr y RGB son los mas utilizados de manera combinada pero en general los autores no argumentan con sustento suficiente las razones que justifiquen el uso de un modelo particular. Si bien es cierto que los distintos espacios de color presentan ventajas como ser más compactos, ser más uniformes, separar mejor la luminosidad y la cromaticidad del color, no son propiedades fundamentales para determinar si un pixel de una imagen es piel humana o no. El éxito de la clasificación, aparentemente radica más en la capacidad de los algoritmos de clasificación utilizados que en el modelo de color elegido. En general distintos autores han tenido éxitos similares con diferentes combinaciones de modelos de representación del color y de algoritmos de segmentación. No obstante parecería que hay una tendencia que indica que los resultados de los artículos mejoran cuando se utilizan combinaciones de componentes de dos o más modelos de representación del color.

Respecto de cómo influye la elección del modelo de color en el algoritmo de clasificación de piel, parecería que un sistema de color que ajusta mejor la distribución del color de la piel mitiga la limitación de aquellos algoritmos que no tienen la robustez para adaptarse o cubrir el espacio de color RGB, que es el más amplio de todos. Como ejemplo de esto, por un lado se pueden mencionar los algoritmos basados en umbrales, menos robustos que otros, cuya aplicación resulta menos efectiva en la segmentación en el modelo de color RGB que en los demás modelos. Por otro lado, a partir del experimento realizado con redes neuronales RCE en la sección 6.3.3, más robustas que los modelos basados en umbrales, muestra que no hay diferencias significativas al realizar la segmentación cuando se aplica a distintos modelos de color.

7.2. Trabajos Futuros

Respecto a los trabajos futuros, hay mucho por hacer para conseguir una respuesta definitiva a cuál es el modelo de color óptimo para representar la piel humana. La primera tarea a realizar consiste en implementar los distintos algoritmos utilizados con cada una de las variantes introducidas por los autores de los artículos de estudio. La segunda tarea consiste en la aplicación de cada algoritmo a cada modelo de representación del color para medir la efectividad de la segmentación. Es importante destacar que en este punto es de fundamental importancia el uso de las bases de datos que aseguran la disponibilidad de máscaras de la piel segmentada (ground-truth) para realizar pruebas de comparación justas y con resultados reproducibles.

Otro aspecto interesante para cubrir es el cálculo del costo computacional tanto para las transformaciones entre modelos de representación del color como del algoritmo de segmentación utilizado, como el costo final de las combinaciones de ambos. Esto permitiría asociar a los resultados de la efectividad de segmentación un aspecto no menor que normalmente los autores no tienen en cuenta, o al menos no mencionan en los artículos.

Desarrolladas las tareas anteriores, es posible realizar un análisis mas amplio, exhaustivo y

justo para determinar por un lado, el modelo de representación de color más conveniente para aplicaciones de segmentación de piel y por otro cómo éstos algoritmos influyen en los algoritmos de segmentación.

ÍNDICE DE FIGURAS

2.1. Sensibilidad normalizada de las células retinales (bastones y conos) del ojo humano . . .	10
2.2. Espectro de luz visible.	11
2.3. a) Círculo cromático escalonado de 12 colores. b) Círculo cromático escalonado de 24 colores. c) Círculo cromático escalonado de 48 colores. d) Círculo cromático continuo de colores	12
2.4. a) Esquema de formación de colores del sistema aditivo presentes en pantallas de computadoras. b) Esquema de formación de colores del sistema sustractivo presente en impresoras	13
2.5. a) Variación del saturación a lo largo en los colores primarios y secundarios. b) Variación del luminosidad a lo largo en los colores primarios y secundarios.	14
2.6. Geometría del espacio de color del modelo RGB	16
2.7. a) Imágenes originales. b) Imágenes del canal rojo. c) Imágenes del canal verde. d) Imágenes del canal azul	16
2.8. Geometría del espacio de color del modelo RGB normalizado.	17
2.9. a) Imágenes originales. b) Imágenes del canal cromático RGB normalizado. c) Imágenes del canal adicional con luminosidad.	18
2.10. Geometría del espacio de color del modelo HSV.	19
2.11. a) Imágenes originales. b) Imágenes del canal de matiz H . c) Imágenes del canal de saturación S . d) Imágenes del canal de luminosidad V	20
2.12. Geometría del espacio de color del modelo HSL.	21
2.13. a) Imágenes originales. b) Imágenes del canal de matiz H . c) Imágenes del canal de saturación S . d) Imágenes del canal de luminosidad L	21
2.14. a) Imágenes originales. b) Imágenes del canal de matiz H . c) Imágenes del canal de saturación S . d) Imágenes del canal de luminosidad I	23
2.15. a) Imágenes originales. b) Imágenes del canal de luminosidad Y . c) Imágenes del canal de cromaticidad U . d) Imágenes del canal de cromaticidad V	24
2.16. a) Imágenes originales. b) Imágenes del canal de luminosidad Y . c) Imágenes del canal de cromaticidad I . d) Imágenes del canal de cromaticidad Q	25
2.17. Geometría del espacio de color del modelo YCbCr.	26
2.18. a) Imágenes originales. b) Imágenes del canal de luminosidad Y . c) Imágenes del canal de cromaticidad Cb . d) Imágenes del canal de cromaticidad Cr	27
2.19. a) Imágenes originales. b) Imágenes del canal de luminosidad Y . c) Imágenes del canal de cromaticidad Cg . d) Imágenes del canal de cromaticidad Cr	28

2.20. a) Imágenes originales. b) Imágenes del canal de luminosidad L. c) Imágenes del canal de cromaticidad A . d) Imágenes del canal de cromaticidad B	29
3.1. Esquema codificador/decodificador utilizado en cuantificación vectorial.	34
3.2. Esquema de un hiperplano que maximiza el margen de separación entre conjuntos de patrones pertenecientes a dos clases para una Máquina de Soporte Vectorial (SVM).	37
3.3. Funciones de Kernel comunes para SVM. a) Lineal. b) Polinomial. c) Base Radial (RBF). d) Sigmoidal	38
3.4. Modelo de Mezcla de Gaussianas (GMM). a) Gráfico de un histograma. b) Gráfico con la aproximación del histograma utilizando 8 componentes gaussianas para un espectro de audio.	40
3.5. Elementos de la arquitectura general de un Modelo de Oculto de Markov (HMM).	42
3.6. Esquema de una neurona biológica (izquierda) y una neurona artificial (derecha)	45
3.7. Funciones de activación de neuronas artificiales. a) identidad. b) paso binario. c) unidad lineal rectificada (ReLU). d) tangente hiperbólica. e) softsig. f) logística.	45
3.8. Arquitectura de una red de perceptrones multicapa.	46
3.9. Esquema de arquitectura de Mapa auto-organizado (SOM).	48
3.10. Capa competitiva de redes SOM con diferentes dimensiones y vecindades. a) Red unidimensional con vecindad 2. b) Red bi-dimensional de vecindad 4. c) Red bi-dimensional de vecindad 6. d) Red tri-dimensional de vecindad 6.	49
3.11. Arquitectura de una red neuronal de Energía de Coulomb Restringida (RCE).	51
4.1. Distribución de la piel en el espacio de color RGB según base de datos [BD12].	57
4.2. Distribución de la piel en el espacio de color RGB normalizado según base de datos [BD12].	58
4.3. Distribución de la piel en el espacio de color HSV según base de datos [BD12].	59
4.4. Distribución de la piel en el espacio de color HSL según base de datos [BD12].	59
4.5. Distribución de la piel en el espacio de color HSI según base de datos [BD12].	60
4.6. Distribución de la piel en el espacio de color YUV según base de datos [BD12].	61
4.7. Distribución de la piel en el espacio de color YIQ según base de datos [BD12].	61
4.8. Distribución de la piel en el espacio de color YCbCr según base de datos [BD12].	62
4.9. Distribución de la piel en el espacio de color YCgCr según base de datos [BD12].	63
4.10. Distribución de la piel en el espacio de color CIE-LAB según base de datos [BD12].	64
6.1. a) Categorización de algoritmos en grupos de relevancia. b) Desglose de los tipos de algoritmos utilizados en los trabajos estudiados.	81
6.2. a) Experimentos agrupados por cantidad de modelos de color que se utilizan para segmentar la piel. b) Experimentos que usan 2 o más modelos de color agrupados por modelo. c) Experimentos que usan 1 modelo de color agrupados por modelo. d) Experimentos agrupados por modelos de color sin distinguir la cantidad que utilizan.	83
6.3. a) Promedio de píxeles segmentados para cada sistema. b) Cantidad de Neuronas requeridas para la segmentación de las imágenes para cada sistema.	89

ÍNDICE DE CUADROS

5.1. Bases de datos utilizadas por los artículos de estudio para segmentación de piel.	70
6.1. Algoritmos de segmentación de la piel implementados por los artículos de estudio	80
6.2. Sistemas de representación del color utilizados por los artículos de estudio.	82
6.3. Efectividad de la segmentación reportada en los artículos de estudio	85

LISTA DE ABREVIATURAS Y SÍMBOLOS

ANN	Artificial Neural Net, Red Neuronal Artificial
BMU	Best Matching Unit, neurona ganadora de una red SOM
CIE	sigla en francés de Commission Internationale de l'Éclairage o Comisión Internacional Iluminación
CIE-LAB	Modelo de color que representa los colores más allá del espectro visible
CIE-XYZ	Modelo de color que representa la totalidad de los colores del espectro visible
EM	Expectation Maximization, Maximización de Expectativa
GMM	Gaussian Mixture Model, Modelo de Mezclas Gaussianas
HMM	Hidden Markov Model, Modelo Oculto de Markov
HSI	siglas del inglés Hue, Saturation, Intensity que define un modelo de color basado en el modelo HSV que mantiene la misma representación de la tonalidad pero difiere en la representación de la saturación y de la luminosidad
HSL	siglas del inglés Hue, Saturation, Lightness que define un modelo de color basado en el modelo HSV que mantiene la misma representación de la tonalidad pero difiere en la representación de la saturación y de la luminosidad
HSV	siglas del inglés Hue, Saturation, Value que define un modelo de color basado en tonalidad, saturación, y un valor de luminosidad
K-Mean	sigla del inglés K-Mean, K-Medias
KNN	sigla del inglés K-Nearest Neighbor, K Vecinos mas Cercanos
MLP	sigla del inglés Multi-layer Perceptron, Red Neuronal Multicapa de Perceptrones
NBC	sigla del inglés Naive Bayes Classifier, Clasificador de Bayes Ingenuo
RCE	Restricted Coulomb Energy, Red Neuronal de Energía de Coulomb Restringida
RGB	Red-Green-Blue, Modelo de color basado en la síntesis aditiva que tiene por colores primarios al rojo, verde y azul
RGB Norm	Modelo de color basado en el modelo RGB el cual normaliza los componentes para reducir la variabilidad de la luz
SOFM	Self-organizing Feature Map, Mapa Auto-organizado
SOM	Self-organizing Map, Mapa Auto-organizado

SVM	Support Vector Machine, Máquina de Soporte Vectorial
VQ	Vector Quantization, Cuantificación de Vectores
YCbCr	Modelo de color basado en un componente de luminosidad (Y) y dos componentes de cromaticidad (CbCr)
YCgCr	Modelo de color basado en un componente de luminosidad (Y) y dos componentes de cromaticidad (CgCr)
YIQ	Modelo de color basado en un componente de luminosidad (Y) y dos componentes de cromaticidad (IQ)
YUV	Modelo de color basado en un componente de luminosidad (Y) y dos componentes de cromaticidad (UV)

BIBLIOGRAFÍA

- [AM07] M. M. Abdelwahab and W. B. Mikhael. Efficient feature representation employing pca and vq in the transform domain for facial recognition. In *2007 50th Midwest Symposium on Circuits and Systems*, pages 281–284, Aug 2007.
- [BAO17] Emir Buza, Amila Akagic, and Samir Omanovic. Skin detection based on image color segmentation with histogram and k-means clustering. *2017 10th International Conference on Electrical and Electronics Engineering (ELECO)*, pages 1181–1186, 2017.
- [Bau72] Leonard E. Baum. An equality and associated maximization technique in statistical estimation for probabilistic functions of markov processes. *Inequalities*, 3:1–8, 1972.
- [BCL01] David A. Brown, Ian Craw, and Julian Lewthwaite. A som based approach to skin detection with application in real time systems. In *British Machine Vision Conference*, 2001.
- [BD94] Michael R. Berthold and Jay Diamond. Boosting the performance of rbf networks with dynamic decay adjustment. In *Proceedings of the 7th International Conference on Neural Information Processing Systems*, pages 521–528, Cambridge, MA, USA, 1994. MIT Press.
- [BD12] Rajen Bhatt and Abhinav Dhall. UCI machine learning repository: Skin segmentation dataset, 2012.
- [BK10] Kishor Bhoyar and Omprakash Kakde. Skin color detection model using neural networks and its performance evaluation. *Journal of Computer Science*, 6:963–968, 12 2010.
- [BP66] Leonard E. Baum and Ted Petrie. Statistical inference for probabilistic functions of finite state markov chains. *The Annals of Mathematical Statistics*, 37(6):1554–1563, Diciembre 1966.
- [BV15] Sneha Tiwari Birendra and Nagaraj K. Vernekar. Robust color-based multi-face detection using skin color segmentation. *Journal of Computer Engineering (IOSR-JCE)*, pages 54–60, 2015.
- [CEV12] Z. Chen, T. Ellis, and S. A. Velastin. Vehicle detection, tracking and classification in urban traffic. In *2012 15th International IEEE Conference on Intelligent Transportation Systems*, pages 951–956, Sep. 2012.

- [CGVRGPSP10] Jose M. Chaves-González, Miguel A. Vega-Rodríguez, Juan A. Gómez-Pulido, and Juan M. Sánchez-Pérez. Detecting skin in face recognition systems: A colour spaces study. *Digit. Signal Process.*, 20(3):806–823, May 2010.
- [CKLO08] Qiu Chen, Koji Kotani, Fei-Fei Lee, and Tadahiro Ohmi. A vq-based fast face recognition algorithm using optimized codebook. In *2008 International Conference on Wavelet Analysis and Pattern Recognition*, volume 1, pages 298–303, Aug 2008.
- [CMR13] Paulo Casati, Diego Moraes, and Evandro Luis Linhari Rodrigues. Sfa: A human skin image database based on feret and ar facial images. In *IX Workshop de Visão Computacional*, Rio de Janeiro, Brasil, 2013.
- [Com99] Computer Vision Laboratory. Ljubljana University, Slovenia. Cvl face database, 1999. <http://www.lrv.fri.uni-lj.si/facedb.html>, Last accessed on 2019-02-20.
- [CPB03] D. Chai, S. L. Phung, and A. Bouzerdoum. A bayesian skin/non-skin color classifier using non-parametric density estimation. In *Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on*, volume 2, pages II–464–II–467 vol.2, May 2003.
- [CSR06] W. M. Campbell, D. E. Sturim, and D. A. Reynolds. Support vector machines using gmm supervectors for speaker verification. *IEEE Signal Processing Letters*, 13(5):308–311, May 2006.
- [CV95] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Mach. Learn.*, 20(3):273–297, Septiembre 1995.
- [CZL⁺02] Li Chen, Jiliu Zhou, Zhiming Liu, Wei Chen, and Guoqing Xiong. A skin detector based on neural network. In *IEEE 2002 International Conference on Communications, Circuits and Systems and West Sino Expositions*, volume 1, pages 615–619 vol.1, June 2002.
- [dDG03] J. J. de Dios and N. Garcia. Face detection based on a new color space ycgcr. In *Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429)*, volume 3, pages III–909–12 vol.2, Sept 2003.
- [DKT17] Dua Dheeru and Efi Karra Taniskidou. UCI machine learning repository, 2017.
- [DZSD08] Z. Dan, S. Zheng, S. Sun, and R. Dong. Speaker recognition based on ls-svm. *3rd International Conference on Innovative Computing Information and Control*, pages 25–28, 2008.
- [GCRÁCP⁺09] A. Guerrero-Curieses, J. L. Rojo-Álvarez, P. Conde-Pardo, I. Landesa-Vázquez, J. Ramos-López, and J. L. Alba-Castro. On the performance of kernel methods for skin color segmentation. *EURASIP Journal on Advances in Signal Processing*, 2009(1):856039, Jun 2009.

- [GEL17] Carla Luna Gennari, César Armando Estrebou, and Laura Cristina Lanzarini. Reconocimiento de gestos aplicado al control de dispositivos. In *Actas del XXIII Congreso Argentino de Ciencias de la Computación (CACIC 2017)*, pages 1040–1049, 2017.
- [Gen17] Carla Luna Gennari. Segmentación de manos basada en redes neuronales. su uso en un sistema embebido para control de tv. Master's thesis, Facultad de Informática. UNLP, 2 2017. 10.
- [GG07] Anthony T.C. Goh and S.H. Goh. Support vector machines: Their use in geotechnical engineering as illustrated using seismic liquefaction data. *Computers and Geotechnics*, 34(5):410 – 421, 2007. Special Issue on Biologically Inspired and Other Novel Computing Techniques in Geomechanics.
- [GM05] Guo Dong and Ming Xie. Color clustering and learning for image segmentation based on neural networks. *IEEE Transactions on Neural Networks*, 16(4):925–936, July 2005.
- [GM12] J. Greenhalgh and M. Mirmehdi. Real-time detection and recognition of road traffic signs. *IEEE Transactions on Intelligent Transportation Systems*, 13(4):1498–1506, Dec 2012.
- [GMX11] Kamarul Hawari Bin Ghazali, Jie Ma, and Rui Xiao. Article: An innovative face detection based on skin color segmentation. *International Journal of Computer Applications*, 34(2):6–10, November 2011. Full text available.
- [GRG06] H. Greenspan, A. Ruf, and J. Goldberger. Constrained gaussian mixture model framework for automatic segmentation of mr brain images. *IEEE Transactions on Medical Imaging*, 25(9):1233–1245, Sep. 2006.
- [HAAB⁺15] Ahmad Hassanat, Mouhammd Al-Awadi, Eman Btoush, Amani Al-Btoush, EsraaTMa Alhasanat, and Ghada Altarawneh. New mobile phone and webcam hand images databases for personal authentication and identification. *Procedia Manufacturing*, 3:4060 – 4067, 2015. 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015.
- [HATWSS14] Zaher Hamid Al-Tairi, Rahmita Wirza, M Iqbal Saripan, and Puteri Sulaiman. Skin segmentation using yuv and rgb color spaces. *Journal of Information Processing Systems*, 10:283–299, 06 2014.
- [HBGR09] Murad Al Haj, Andrew D. Bagdanov, Jordi González, and Xavier F. Roca. Robust and efficient multipose face detection using skin color segmentation. In Helder Araujo, Ana Maria Mendonça, Armando J. Pinho, and María Inés Torres, editors, *Pattern Recognition and Image Analysis*, pages 152–159, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.
- [HRBLM07] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

- [HSW08] Reza Hassanpour, Asadollah Shahbahrami, and Stephan Wong. Adaptive gaussian mixture model for skin color segmentation. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 2(5):1348 – 1353, 2008.
- [HTMK02] Quan Huynh-Thu, Mitsuhiko Meguro, and Masahide Kaneko. Skin-color-based image segmentation and its application in face detection. In *IAPR Workshop on Machine Vision Application*, pages 48–51, December 2002.
- [IK13] Noor Ibraheem and Rafiqul Zaman Khan. Multiple histogram technique for robust skin color based segmentation. *Journal of Engineering Research*, 2:50–54, 05 2013.
- [KGMS09] Dakshina Ranjan Kisku, Phalguni Gupta, Hunny Mehrotra, and Jamuna Kanta Sing. Multimodal belief fusion for face and ear biometrics. *Intelligent Information Management*, 1(3):166–171, 2009.
- [KK12] Amanpreet Kaur and B.v Kranthi. Article: Comparison between ycbcr color space and cielab color space for skin color segmentation. *International Journal of Applied Information Systems*, 3(4):30–33, July 2012. Published by Foundation of Computer Science, New York, USA.
- [KKNS14] Michal Kawulok, Jolanta Kawulok, Jakub Nalepa, and Bogdan Smolka. Self-adaptive algorithm for segmenting skin regions. *EURASIP Journal on Advances in Signal Processing*, 2014(170):1–22, 2014.
- [KKS⁺17] S. Kolkur, D. Kalbande, P. Shimpi, C. Bapat, and J. Jatakia. Human skin detection using rgb, hsv and ycbcr color models. *International Conference on Communication and Signal Processing*, abs/1708.02694, 2017.
- [KMB07] P. Kakumanu, S. Makrogiannis, and N. Bourbakis. A survey of skin-color modeling and detection methods. *Pattern Recogn.*, 40(3):1106–1122, March 2007.
- [KO02] K. Kotani and T. Ohmi. Face recognition using vector quantization histogram method. In *Proceedings. International Conference on Image Processing*, volume 2, pages II–II, Sep. 2002.
- [Koh82] Teuvo Kohonen. Self-organizing formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69, 1982.
- [KRTG08] Dakshina Ranjan Kisku, Ajita Rattani, Massimo Tistarelli, and Phalguni Gupta. Graph application on face for personal authentication and recognition. In *10th International Conference on Control, Automation, Robotics and Vision, ICARCV 2008, Hanoi, Vietnam, 17-20 December 2008, Proceedings*, pages 1150–1155, 2008.
- [Lei13] Shi Lei. A face detection algorithm in color image based on skin color segmentation. In *Modern Materials and Technologies of Industrial Production*, volume 811 of *Advanced Materials Research*, pages 417–421. Trans Tech Publications, 12 2013.

- [LLX08] Ming Li, Ruiling Luo, and Yujuan Xing. A novel multi-reduced svm approach for speaker recognition. In *Fuzzy Systems and Knowledge Discovery, 2008. FSKD '08. Fifth International Conference on*, volume 4, pages 462–466, Octubre 2008.
- [LzP10] Qiong Liu and Guang zheng Peng. A robust skin color based face detection algorithm. In *2010 2nd International Asia Conference on Informatics in Control, Automation and Robotics (CAR 2010)*, volume 2, pages 525–528, March 2010.
- [MB04] F. Melgani and L. Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8):1778–1790, Aug 2004.
- [MP43] Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, Dec 1943.
- [NKN16] M. Naveena, G. H. Kumar, and P. Navya. Detection of a person in a crowd based on skin color segmentation. In *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 175–179, Sept 2016.
- [OKH15] S. Oh, Y. Kim, and J. Hong. Urban traffic flow prediction system using a multifactor pattern recognition model. *IEEE Transactions on Intelligent Transportation Systems*, 16(5):2744–2755, Oct 2015.
- [Ots79] Nobuyuki Otsu. A Threshold Selection Method from Gray-level Histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62–66, January 1979.
- [PBC05] Son Lam Phung, Abdesselam Bouzerdoum, and Douglas Chai. Skin segmentation using color pixel classification: Analysis and comparison. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(1):148–154, jan 2005.
- [PCB01] Son Lam Phung, D. Chai, and A. Bouzerdoum. A universal and robust human skin color model using neural networks. In *Neural Networks, 2001. Proceedings. IJCNN '01. International Joint Conference on*, volume 4, pages 2844–2849 vol.4, 2001.
- [PFS⁺07] I. Parra Alonso, D. Fernandez Llorca, M. A. Sotelo, L. M. Bergasa, P. Revenga de Toro, J. Nuevo, M. Ocana, and M. A. Garcia Garrido. Combination of feature extraction methods for svm pedestrian detection. *IEEE Transactions on Intelligent Transportation Systems*, 8(2):292–307, June 2007.
- [Pla98] John C. Platt. Fast training of support vector machines using sequential minimal optimization. In *Advances in kernel methods: support vector learning*, pages 185–208. MIT Press, Cambridge, MA, USA, 1998.
- [PM47] Walter Pitts and Warren S. McCulloch. How we know universals the perception of auditory and visual forms. *The bulletin of mathematical biophysics*, 9(3):127–147, Sep 1947.
- [Psy17] Psychological Image Collection at Stirling. Aberdeen pilot database, 2017. <http://pics.stir.ac.uk/>, Last accessed on 2019-02-20.

- [RCE82] Douglas L. Reilly, Leon N. Cooper, and Charles Elbaum. A neural model for category learning. *Biol. Cybern.*, 45(1):35–41, August 1982.
- [RCGS16] R. F. Rahmat, T. Chairunnisa, D. Gunawan, and O. S. Sitompul. Skin color segmentation using multi-color space threshold. In *International Conference on Computer and Information Sciences*, pages 391–396, Aug 2016.
- [RHW86] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–, October 1986.
- [RKP14] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi. Dream: Diabetic retinopathy analysis using machine learning. *IEEE Journal of Biomedical and Health Informatics*, 18(5):1717–1728, Sep. 2014.
- [RKRSRVK12] P. Rama Koteswara Rao, Y. Srinivasa Rao, and D. Vijaya Kumar. Principal factor analysis and svm based effective speaker recognition. In *Computing Communication Networking Technologies (ICCCNT), 2012 Third International Conference on*, pages 1–7, Julio 2012.
- [RQE⁺16a] Franco Ronchetti, Facundo Quiroga, Cesar Estrebou, Laura Lanzarini, and Alejandro Rosete. Lsa64: An argentinian sign language dataset. *XXII Congreso Argentino de Ciencias de la Computación (CACIC 2016)*, pages 794–803, 2016.
- [RQE⁺16b] Franco Ronchetti, Facundo Quiroga, César Estrebou, Laura Lanzarini, and Alejandro Rosete. Sign language recognition without frame-sequencing constraints: A proof of concept on the argentinian sign language. In Manuel Montes y Gómez, Hugo Jair Escalante, Alberto Segura, and Juan de Dios Murillo, editors, *Advances in Artificial Intelligence - IBERAMIA 2016: 15th Ibero-American Conference on AI, San José, Costa Rica, November 23-25, 2016, Proceedings*, pages 338–349. Springer International Publishing, 2016.
- [RQEL16] Franco Ronchetti, Facundo Quiroga, César Estrebou, and Laura Lanzarini. Hands-hape recognition for argentinian sign language using probsom. *Journal of Computer Science and Technology*, 16(1):1–5, 2016.
- [RQLE15] Franco Ronchetti, Facundo Quiroga, Laura Lanzarini, and Cesar Estrebou. Distribution of action movements (dam): a descriptor for human action recognition. *Frontiers of Computer Science*, 9(6):956–965, 2015.
- [RTBA12] Devendra Singh Raghuvanshi, M. Tech, MANIT Bhopal, and Dheeraj Agrawal. Human face detection by using skin color segmentation , face features and regions properties. *International Journal of Computer Application*, 38(9), January 2012.
- [SCVS03] S.Kr Singh, D.S. Chauhan, Mayank Vatsa, and Richa Singh. A robust skin color based face detection algorithm. *Tamkang Journal of Science and Engineerin*, 6(4):227–234, 12 2003.

- [SGK⁺15] Khamar Basha Shaik, P. Ganesan, V. Kalist, B.S. Sathish, and J. Merlin Mary Jenitha. Comparative study of skin color detection and segmentation in hsv and ycbcr color space. *Procedia Computer Science*, 57:41 – 48, 2015. 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015).
- [SKR11] C. Sui, N. M. Kwok, and T. Ren. A restricted coulomb energy (rce) neural network system for hand image segmentation. In *2011 Canadian Conference on Computer and Robot Vision*, pages 270–277, May 2011.
- [SREC87] Christopher L. Scofield, Douglas L. Reilly, Charles Elbaum, and Leon N. Cooper. Pattern class degeneracy in an unrestricted storage density memory. In *Proceedings of the 1987 International Conference on Neural Information Processing Systems, NIPS'87*, pages 674–682, Cambridge, MA, USA, 1987. MIT Press.
- [TBTL08] D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet. Semantic annotation and retrieval of music and sound effects. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(2):467–476, Feb 2008.
- [TCYC12] W. R. Tan, C. S. Chan, P. Yogarajah, and J. Condell. A fusion approach for efficient human skin detection. *IEEE Transactions on Industrial Informatics*, 8(1):138–147, Feb 2012.
- [TKSY17] Eman Thabet, Fatimah Khalid, Puteri Suhaiza Sulaiman, and Razali Yaakob. Low cost skin segmentation scheme in videos using two alternative methods for dynamic hand gesture detection method. *Adv. in MM*, 2017:7645189:1–7645189:9, 2017.
- [TM07] Z. Tang and Z. Miao. Fast background subtraction and shadow elimination using improved gaussian mixture model. In *2007 IEEE International Workshop on Haptic, Audio and Visual Environments and Games*, pages 38–41, Oct 2007.
- [TPM⁺11] S. Thakur, S. Paul, A. Mondal, S. Das, and A. Abraham. Face detection using skin tone segmentation. In *2011 World Congress on Information and Communication Technologies*, pages 53–60, Dec 2011.
- [TtSx13] Liu Ting-ting and Guan Sheng-xiao. On text-independent speaker recognition via improved vector quantization method. In *Control Conference (CCC), 2013 32nd Chinese*, pages 3912–3916, Julio 2013.
- [VSA03] Vladimir Vezhnevets, Vasilij Sazonov, and Alla Andreeva. A survey on pixel-based skin color detection techniques. In *GraphiCon*, pages 85–92, 2003.
- [Wer74] P. J. Werbos. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. PhD thesis, Harvard University, 1974.
- [XP17] Y Xu and Gouchol Pok. Identification of hand region based on ycgcr color representation. *International Journal of Applied Engineering Research*, 12(6):1031–1034, 2017.

- [XS13] F. Xiang and Shahrel Azmin Suandi. Fusion of multi color space for human skin region segmentation. *International Journal of Information and Electronics Engineering*, 3(2):963–968, 3 2013.
- [XZ06] Zhanwu Xu and Miaoliang Zhu. Color-based skin detection: survey and evaluation. In *2006 12th International Multi-Media Modelling Conference*, pages 10 pp.–, 2006.
- [YGX01] Xiaoming Yin, Dong Guo, and Ming Xie. Hand image segmentation using color and rce neural network. *Robotics and Autonomous Systems*, 34(4):235 – 250, 2001.
- [YMS17] Abdulganiyu Abdu Yusuf, Fatma Susilawati Mohamad, and Zahraddeen Sufyanu. Human face detection using skin color segmentation and watershed algorithm. *American Journal of Artificial Intelligence*, 1:29–35, 2017.
- [YW96] Jie Yang and A. Waibel. A real-time face tracker. In *Applications of Computer Vision, 1996. WACV '96., Proceedings 3rd IEEE Workshop on*, pages 142–147, Dec 1996.
- [ZME09] A. Zulfiqar, A. Muhammad, and A.M.M. Enriquez. A speaker identification system using mfcc features with vq technique. In *Intelligent Information Technology Application, 2009. IITA 2009. Third International Symposium on*, volume 3, pages 115–118, Noviembre 2009.
- [ZMXNJS06] Lu Zhe-Ming, Xu Xiu-Na, and Pan Jeng-Shyang. Face detection based on vector quantization in color images. *International Journal of Innovative Computing, Information and Control*, 2(3):667–672, 2006.
- [ZZGL08] Yanling Zhao, Xiaoshi Zheng, Huixian Gao, and Na Li. A speaker recognition system based on vq. In *Industrial Electronics and Applications, 2008. ICIEA 2008. 3rd IEEE Conference on*, pages 1988–1990, Junio 2008.