

UNIVERSIDAD AUTÓNOMA
DEL ESTADO DE MÉXICO
CENTRO UNIVERSITARIO UAEM TEXCOCO



Segmentación de Imágenes de Color Inspirado en la
Percepción Humana del Color

Miguel Contreras Murillo

En este trabajo se propone imitar la percepción humana del reconocimiento del color empleando redes neuronales artificiales de tipo competitivas con el fin de evitar entrenar el algoritmo cada vez que una nueva escena se presenta.

TESIS

Que para obtener el Grado de
Maestro en Ciencias de la Computación

Presenta

Contreras Murillo Miguel

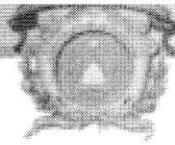
Tutor Académico

Dr. Farid García Lamont

Tutores Adjuntos

Dra. Alma Delia Cuevas Rasgado
Dr. José Sergio Ruiz Castilla

Texcoco, Estado de México, a 1 de agosto del 2016



DICTÁMEN DE AUTORIZACIÓN Y OBTENCIÓN DE GRADO DE MAESTRÍA

Texcoco, Méx., a 5 de septiembre del 2016

Título del proyecto:

Segmentación de Imágenes de Color Inspirado en la Percepción Humana del Color

Tesista:

Contreras Murillo Miguel

Dictamen:

No. de revisión: 1

- Rechazado
- Sujeto a modificaciones
- Aceptado, condicionado
- Aceptado

**POSGRADO
TITULACION
RECIBIDO**
Por Adriana Arevalo
Texcoco, Méx., a 6 de VII del 2016

Observaciones generales:

Tutor Adjunto	Tutor Académico	Tutor Adjunto
Dra. en C. Alma Delia Cuevas Rasgado	Dr. en C. Com. Farid García Lamont	Dr. en I.S. José Sergio Ruiz Castilla



RESUMEN

Usualmente la segmentación de imágenes se realiza considerando las características de textura y/o geométricas. Sin embargo, la segmentación de imágenes utilizando las características de color no es tan común. Los trabajos que abordan la segmentación de imágenes por sus características de color emplean o se basan en métodos no supervisados o técnicas de agrupamiento, principalmente fuzzy C-means. Los resultados que se reportan son buenos, la desventaja con dichas técnicas es que se requiere definir previamente la cantidad de grupos que en se desean agrupar los colores, pero esto puede limitar la cantidad de colores que existen en la imagen; por otra parte, el procesamiento de los algoritmos no pueden generalizarse para cualquier imagen ya que solo procesan los colores de cada imagen, si estos grupos se intentan emplear para segmentar imágenes diferentes, es muy probable que no funcionarán correctamente.

En este trabajo se propone imitar la percepción humana del reconocimiento del color empleando redes neuronales artificiales de tipo competitivas. Los seres humanos reconocen los colores primero por su cromaticidad y después por su intensidad; por otra parte, los seres humanos pueden reconocer áreas o secciones de imágenes dependiendo solamente de la cromaticidad de las partes que conforman la imagen. De aquí que, se propone entrenar una red neuronal que reconozca la cromaticidad de los colores. Una ventaja que tendrá la red neuronal con respecto a los métodos de agrupamiento es que no tendrá que ser entrenada por cada imagen, es decir, una vez entrenada la red neuronal a reconocer la cromaticidad del color, esta puede ser aplicada a cualquier imagen sin volver a ser entrenada.

DEDICATORIA

A mi familia, amigos, y quien sea esperaba su nombre en esta sección.

AGRADECIMIENTOS

Se agradece al Consejo Nacional de Ciencia y Tecnología, al Consejo Mexiquense de Ciencia y Tecnología, y a la Universidad Autónoma del Estado de México por las becas otorgada para la realización de mis estudios de maestría y culminar satisfactoriamente esta investigación. Además, al Centro de Aplicación de Tecnologías de Avanzada (CENATAV), La Habana, Cuba, donde todos me brindaron ayuda en todo lo que necesité en mi vida diaria y en parte de la investigación que en esta tesis se presenta.

A los profesores que contribuyeron en mi formación académica. A mis tutores Dra. Alma Delia Cuevas Rasgado y Dr. José Sergio Ruiz Castilla, quienes dieron guía a esta investigación, y al Dr. Farid García Lamont, quien además dio origen a lo que se presenta en esta tesis.

A mi familia que siempre me ha apoyado incondicionalmente en todo lo que he necesitado. Y a mis compañeros y amigos con quienes he convivido dentro y fuera de la escuela. La lista es tan amplia que va de la A a la Z.

Contenido

Resumen	iii
Dedicatoria	iv
Agradecimientos	v
Contenido	vi
Índice de tablas	viii
Índice de figuras	ix
1 Introducción	1
1.1 Planteamiento del problema	1
1.2 Justificación	1
1.3 Objetivo general	1
1.4 Objetivos específicos	1
1.5 Hipótesis	2
2 Trabajos previos y Antecedentes	3
2.1 Segmentación de imágenes de escala de grises	3
2.2 Métodos de segmentación de imágenes de color	4
2.2.1 Redes neuronales artificiales	4
2.2.2 Segmentación basada en Clustering	5
2.2.3 Superpixel	5
3 Espacios de color	6
3.1 Espacio de color RGB	6
3.2 Espacio de color HSV	7
3.3 Colores semánticos	8
3.4 Métodos comúnmente empleados	12
3.4.1 Sustracción de fondo	13
3.4.2 K-Nearest Neighbour (K-nn)	15
3.4.3 K-means	19
3.4.4 Redes neuronales competitivas	23
4 Método propuesto	25
4.1 Arquitecturas de redes neuronales	27
4.2 Entrenamiento de la red neuronal	28
4.3 Procesamiento de la imagen	28
5 Experimentos	30

5.1	Redes neuronales competitivas	31
5.2	Mapas auto-organizados	32
6	Discusión	34
6.1	Evaluación cuantitativa de la segmentación de imágenes de color	34
6.1.1	VOI (<i>Variation of Information</i>)	36
6.1.2	PRI (<i>Probabilistic Rand Index</i>)	40
6.1.3	GCE (<i>Global Consistency Error</i>)	43
6.2	Tamaño de la red neuronal	46
	Conclusiones	47
	Trabajo Futuro	49
	Anexo	50
	Bibliografía	61

Índice de tablas

Tabla 3.1 Valores RGB para los 11 colores semánticos estándar para internet.....	11
Tabla 2 Notación para el paradigma K-NN	16
Tabla 6.1 VOI (Promedio)	37
Tabla 6.2 VOI (Máximo).....	38
Tabla 6.3 VOI (Mínimo)	39
Tabla 6.4 PRI (Promedio).....	40
Tabla 6.5 PRI (Máximo)	41
Tabla 6.6 PRI (Mínimo).....	42
Tabla 6.7 GCE (Promedio)	43
Tabla 6.8 GCE (Máximo)	44
Tabla 6.9 GCE (Mínimo).....	45

Índice de figuras

Figura 3-1 Espacio de color RGB.....	6
Figura 3-2 Representación gráfica del espacio de color HSV.....	7
Figura 3-3 Espacio de color HSV.....	7
Figura 3-4. Paleta con muestras de colores para ser clasificados por personas.....	9
Figura 3-5 Evolución de los colores semánticos.....	9
Figura 3-6 Test Farnsworth D-15, consiste en ordenar los colores, un mal ordenamiento sería prueba de algún tipo de daltonismo.....	10
Figura 3-7 Cartas de Ishihara, contienen números de distintos colores para detectar algún tipo de daltonismo: (a) Carta de Ishihara No. 1(12); (b) Carta de Ishihara No. 11(6); (c) Carta de Ishihara No. 19(Nada, dígito oculto: los daltónicos ven el número 2); (d) Carta de Ishihara No. 23(42);.....	11
Figura 3-8 Distribución de los colores semánticos en distintas investigaciones.....	11
Figura 3-9 En un primer intento se utilizaba un mapeo de pixeles a colores semánticos, pero se tenía como resultado una fusión entre el primero y el segundo plano.....	13
Figura 3-10 Ejemplo de usos de cromas verdes. Todo pixel verde es clasificado como fondo.....	13
Figura 3-11 (a) y (b) son dos fotogramas consecutivos, (c) es el resultado de restar ambos fotogramas y (d) es el primer plano con valor en 1 y el fondo con valores en 0.....	14
Figura 3-12 Interfaz gráfica de la aplicación realizada para la recolección de muestras de colores semánticos.....	16
Figura 3-13 (a) es la representación de las muestras a color, (b) es el nuevo pixel a clasificar dentro las muestras, y (c) es el área de búsqueda limitada alrededor del pixel a clasificar.....	17
Figura 3-14 Cálculo de índices con posibles muestras válidas para la clasificación del pixel.....	18
Figura 3-15 Se limitan los índices a un radio específico, por eso se muestran en forma de esfera..	18
Figura 3-16 Los índices calculados se suman a las coordenadas cartesianas en el espacio RGB para su ubicación final.....	18
Figura 3-17 El segundo plano es segmentado por k-means, mientras el primer plano es segmentado con k-nn aunque presenta ruido.....	19
Figura 3-18 Segmentación resultante después de utilizar un umbral para los colores semánticos blanco, negro y gris.....	19
Figura 3-19 Dos vectores con diferentes magnitudes serán similares si apuntan hacia la misma dirección.....	20
Figura 3-20 La distancia más corta entre dos puntos es la línea recta, y se calcula como la raíz de la suma de los cuadrados de sus lados.....	21
Figura 3-21 En una ciudad con calles cuadradas no hay forma de llegar en línea recta, existen varios caminos más cortos, pero todos tendrán el mismo número de calles a recorrer.....	21
Figura 3-22 K-means con 3 clusters aplicado a escenas de video-vigilancia.....	22
Figura 3-23 a) imagen de un ojo humano, b) segmentación utilizando la distancia euclidiana con k-means y $k = 3$, c) segmentación utilizando la distancia city block con k-means y $k = 3$, b)	

segmentación utilizando la distancia perceptual con k-means y $k = 3$, b) segmentación utilizando la medida del coseno con k-means y $k = 3$	22
Figura 3-24 a) imagen original b) fondo generado con la media de 10 fotogramas consecutivos c) imagen segmentada con k-mean y la distancia euclidiana d) sustracción del primer y segundo plano, resultado de la resta entre a) y b) e) segmentación resultante de combinar K-means (en el segundo plano c)), y k-nn (en el primer plano).....	23
Figura 3-25 Estructura de una RNA de tipo competitiva	23
Figura 3-26 Ejemplo de la organización de las neuronas en la capa oculta de un MAO, y selección de las neuronas a entrenarse en un radio no mayor a 1, 2, o 3	24
Figura 4-1 Distribución de los conos R, G y B en un ojo sano (izquierda), y uno con daltonismo (derecha).....	25
Figura 4-2 Cítricos en color y blanco y negro. El color es importante para identificarlos.	25
Figura 4-3 a) Colores rojo, verde y azul con brillo y saturación máximo, b) Colores rojo, verde y azul con saturación disminuida en un 50%, c) Colores rojo, verde y azul con brillo disminuido un 20%.26	
Figura 4-4 a) En un cielo azul despejado, la saturación del color disminuye del punto más alto (cenit), hasta el horizonte donde se percibe casi blanco, el brillo y la cromaticidad son iguales en todos los puntos. B) al disminuir el brillo la imagen se torna oscura, pero mantiene la misma cromaticidad y saturación que la original.	26
Figura 4-5 a) y b) representan a la misma cromaticidad (verde) con iluminaciones distintas, c) y d) son cromaticidades distintas (rojo y rosa), con la misma iluminación.	27
Figura 5-1 Imágenes extraídas de la BSB, empleadas para los experimentos realizados	30
Figura 5-2 Imágenes obtenidas empleando mapas auto-organizados de 3x3, 4x4 y 5x5 neuronas 31	
Figura 5-3 Imágenes obtenidas empleando redes neuronales competitivas de 9, 16 y 25 neuronas	32
Figura 6-1 Ejemplo de imágenes segmentadas a mano de la BSB.....	35
Figura 6-2 Ejemplo de comparación de las imágenes: (a) y (c), segmentadas a mano y las segmentadas con nuestra propuesta, (b) y (d).....	36
Figura 6-3 VOI (Promedio).....	37
Figura 6-4 VOI (Máximo)	38
Figura 6-5 VOI (Mínimo).....	39
Figura 6-6 PRI (Promedio)	40
Figura 6-7 PRI (Máximo).....	41
Figura 6-8 PRI (Mínimo)	42
Figura 6-9 GCE (Promedio).....	43
Figura 6-10 GCE (Máximo).....	44
Figura 6-11 GCE (Mínimo)	45

1 Introducción

1.1 Planteamiento del problema

Los métodos que utilizan técnicas de clustering, o redes neuronales proporcionan muy buenos resultados dado que se adaptan a cada imagen que se presenta, pero necesitan de intervención humana, o de mucho tiempo para lograrlo.

Al utilizar clustering o agrupamiento es necesario definir a priori la cantidad de clusters en que se agruparan los colores de la imagen a procesar, en el caso de no proporcionar los resultados esperados, es necesario redefinir el número de clusters, ya sea aumentándolo o disminuyéndolos, para después procesar de nuevo la imagen, el proceso se puede repetir hasta llegar a resultados aceptables. Por tanto, los resultados de la segmentación son demasiado dependientes de la intervención humana.

Al usar redes neuronales se mejora la autonomía del proceso, ya que la intervención humana en la segmentación de la imagen es opcional; si así se desea, la computadora automáticamente determinará el mejor resultado. Pero presentan el problema de que necesitan ser entrenadas cada vez que se requiere de segmentar una nueva imagen, y aunque es un proceso que proporciona muy buenos resultados, consume demasiados recursos principalmente en durante el entrenamiento que no podrá ser reutilizado con cualquier otra imagen con características diferentes.

1.2 Justificación

Como se ha mencionado, de los métodos empleados para la segmentación, los basados en clustering requieren de una persona que indique la cantidad de grupos o de secciones en que una imagen debe dividirse. En cuanto a las redes neuronales, estas son entrenadas con los datos de cada imagen en particular, si se presenta una imagen nueva se requiere entrenar nuevamente la red neuronal; esto es, que la red neuronal no está aprendiendo, propiamente a reconocer los colores.

Estas dos situaciones hacen que la segmentación de imágenes de color no sea totalmente automática o inteligente. De aquí que es necesario que desarrollar un método que sea capaz de reconocer colores de cualquier imagen y de no necesitar de la ayuda de los seres humanos para decidir la cantidad de secciones que se debe dividir la imagen en base a sus características de color.

1.3 Objetivo general

Diseñar un algoritmo basado en redes neuronales para la segmentación de imágenes de color que requiera de poca intervención humana durante el procesamiento.

1.4 Objetivos específicos

- Obtener un modelo para representar la cromaticidad de los colores
- Construir el conjunto de entrenamiento adecuado de cromaticidades de los colores

- Diseñar la arquitectura de la red neuronal artificial
- Realizar pruebas para validación

1.5 Hipótesis

Los seres humanos pueden reconocer los colores de una escena al emplear su experiencia o conocimiento adquirido durante el tiempo. Al observar una nueva imagen o escena no se necesita volver a aprender cómo reconocer los colores, simplemente se emplea el conocimiento previo adquirido.

Por otra parte, los seres humanos pueden reconocer áreas en una escena que tengan las mismas o similares características de cromaticidad, aunque no necesariamente la misma intensidad.

De aquí que, para segmentar las imágenes de color, el algoritmo propuesto trata de imitar la forma en que los seres humanos parecen percibir el color, es decir, por cromaticidad. Esto es, desarrollar una red neuronal a la que se le entrene a reconocer la cromaticidad de los colores y una vez hecho esto, se pueda aplicar para cualquier imagen o escena sin necesidad de volverla a entrenar.

2 Trabajos previos y Antecedentes

La segmentación de imágenes es un tema ampliamente estudiado debido a que es indispensable para el reconocimiento de objetos. Este refiere al proceso de agrupar regiones donde los píxeles de una región comparten ciertas características, para que posteriormente se puedan extraer y reconocer objetos de forma automática. Existen diversas técnicas para la segmentación de imágenes de acuerdo a características de forma, textura, color, entre otras [1] [2] [3] [4].

2.1 Segmentación de imágenes de escala de grises

Por eso es importante segmentar imágenes, es decir, dividir la imagen en regiones más pequeñas de forma coherente para después extraer de la imagen las características locales de estas regiones obtenidas. Aunque no es una tarea sencilla dado el número de factores que afectan la calidad de la segmentación; por ejemplo, la intensidad, el enfoque, el contraste, el color, entre otros. Existen varias técnicas para segmentar imágenes, estas se clasifican principalmente en [1]:

- **Basadas en umbrales:** Todos los elementos que estén dentro un rango son agrupados en un mismo conjunto. Con esta técnica se obtienen buenos resultados cuando sólo se tienen dos regiones en la imagen, pero puede fallar si la imagen está compuesta de varias regiones, o si la frontera entre las dos regiones no está bien definida [5].
- **Basadas en bordes:** Se puede segmentar una imagen por medio de detectores de bordes, donde cada región está aislada de las demás, los detectores de bordes identifican aquellos puntos donde existe una gran diferencia entre varios píxeles contiguos. Pero puede presentar errores cuando los bordes encontrados dentro de una imagen no bordean totalmente una región. Un borde ideal debe representar una curva cerrada, mas no siempre es posible obtenerlo de esa manera, para mejorar los resultados se necesita de otras técnicas para mejorar la detección de bordes.
- **Basadas en *Clustering* (agrupamiento):** Agrupa a los píxeles en conjuntos que comparten características similares en tantos grupos como se hayan definido [6].
- **Basadas en texturas:** Una textura puede ser definida como un patrón repetitivo en una imagen, al usarse como base para segmentar una imagen se identifican desde figuras geométricas, hasta cualquier patrón asimétrico con distorsiones geométricas, por ejemplo: imágenes con distintas profundidades [7].
- **Basadas en regiones:** Se basan en la idea de que los píxeles dentro de una misma región tienen apariencias similares, como color o textura, y suelen ser adyacentes entre ellos.

La técnica basada en regiones ha sido la más empleada, las cuales se dividen en:

- **Seeded Region Growing (SRF):** En esta técnica se seleccionan al azar varios píxeles dentro de una imagen, a partir de ellos se evalúan a sus píxeles vecinos, si comparten características similares entonces se consideran como parte de la misma región [8].
- **Dividir y mezclar (Split and merge):** La imagen a segmentar es dividida en partes simétricas (generalmente cuatro), y se evalúa cada región resultante; si son poco homogéneas, es decir, si los píxeles que la componen son muy diferentes entre ellos en cada región, se vuelven a dividir hasta que todas las regiones sean homogéneas. Después se comparan las

regiones resultantes vecinas y si son parecidas, finalmente se mezclan para crear una nueva región más grande.

Estas técnicas han mostrado tener muy precisos, pero demandan alto costo computacional.

- Técnicas mixtas: Pueden ser la combinación de varias técnicas para obtener mejores resultados. Por ejemplo, mezclar detectores de bordes que es rápido, pero ineficiente, combinado con SRF, que es eficiente pero lento, se pueden tener buenos resultados de una manera más rápida.

2.2 Métodos de segmentación de imágenes de color

Es importante hacer notar que las técnicas presentadas anteriormente emplean imágenes en escala de grises. Para el caso de las imágenes de color, algunas de las técnicas anteriores pueden emplearse al extenderlo al caso de vectorial; sin embargo, el problema la segmentación de imágenes de color tiene características diferentes a las imágenes de escala de grises, por lo que se requiere un análisis más específico para este tipo de imágenes. En esta sección se muestran los trabajos relacionados a la segmentación de imágenes de color.

La segmentación de imágenes de color ha sido abordada empleando principalmente dos técnicas: redes neuronales artificiales [9] [10] y clustering [6] o agrupamiento ya que han mostrado ser muy eficientes, aunque no se delimita únicamente a estas dos técnicas. A continuación, se presentan los trabajos encontrados en el estado del arte para la segmentación de imágenes de color empleando: redes neuronales, métodos de clustering y se revisan otros enfoques menos empleados.

2.2.1 Redes neuronales artificiales

Existen tres tipos de redes neuronales artificiales (RNA), las que necesitan de supervisión, las que no la necesitan, y una mezcla de ambas. En el caso de la segmentación de imágenes no siempre es muy factible el uso RNA supervisadas debido a que necesitan de muchas muestras tomadas por humanos [11].

Las RNA son utilizadas en muchas otras áreas, y se pueden usar para clasificar o para agrupar datos (píxeles cuando se trata de imágenes), los métodos supervisados requieren de un conjunto de entrenamiento que consta de muestras y de un identificador que indica a cada muestra a cuál clase pertenece. Se tiene una capa de entrada, una o varias capas ocultas, y una capa de salida, cada capa puede contener varias neuronas que al ingresar un dato se generará uno o varios resultados en la capa de salida, durante el entrenamiento se calculará el error que se tuvo en la salida, con base en ese error se calcularán los pesos de las capas ocultas, el procedimiento se realizará varias veces de la misma forma hasta que la precisión llegue a un porcentaje adecuado o después de un número definido de iteraciones.

Las RNA no supervisadas también requieren de un conjunto de datos para el proceso de entrenamiento, pero no se necesita definir a cuál clase pertenecerá cada dato. A diferencia de las RNA supervisadas, este tipo de RNA primeramente agrupa los datos y después los clasifica de

forma aleatoria. Durante el entrenamiento, cada dato (pixel) excitará a una neurona más que a otras, y esta única neurona aprenderá a identificar este dato pareciéndose más al dato de entrada, disminuyendo su distancia con respecto a este. Al final del entrenamiento los datos (la imagen) resultarán agrupados de forma similar a k-means o fuzzy c-means, donde las neuronas de la RNA se agrupan de forma similar a los centroides de estos otros algoritmos.

2.2.2 Segmentación basada en Clustering

Es una forma de segmentar imágenes, se basa en el hecho de que varios pixeles en una imagen suelen ser similares entre ellos, aunque no necesariamente iguales, esta técnica agrupa los pixeles dependiendo de qué tan parecidos son con sus vecinos cercanos, ya sea en base a color, textura u otras características.

2.2.3 Superpixel

Convencionalmente cuando se segmenta una imagen por medio de clustering, ya sea en color o blanco y negro, se toman en cuenta todos los pixeles de la imagen, pero si se usan métodos como *split and merge*, *watershed*, entre otros que son muy iterativos o recursivos, combinados con imágenes en alta calidad, los datos y los recursos necesarios son muy altos. Para mejorar el rendimiento del proceso, se puede primero sobre-segmentar la imagen en clusters muy pequeños, pero que pueden ser considerados como nuevos pixeles de mayor tamaño y a partir de ellos se puede realizar una segmentación con mayor precisión que ocupe menos información para dar mejores resultados [12].

3 Espacios de color

Un espacio de color es cualquier organización de colores que permita reproducir los colores en una representación similar a la existente en la naturaleza. Los colores existen de forma difusa, esto es, no se puede saber dónde termina uno y comienza el siguiente [13]. En un espacio de color, esta gama de colores está acotada en un espacio discretizado generalmente en tuplas o vectores que contienen la información del color representado [14].

3.1 Espacio de color RGB

Los seres humanos perciben los colores gracias a células encargadas de detectar diferentes longitudes de onda de la luz, de esta forma se tienen células especializadas en detectar las tonalidades rojas, verdes y azules [13], a estos tres colores se les llama colores primarios y es partir de su mezcla en diferentes porcentajes que se obtienen los demás colores del espectro visibles, este espacio de color se llama RGB (*Red, Green, Blue*) [15]. Es ampliamente utilizado, desde la obtención hasta el despliegado de la imagen por dispositivos digitales.

El espacio RGB está basado en el sistema de coordenadas Cartesiano donde los colores son puntos definidos por vectores que se extienden desde el origen, el negro está en el origen [0,0,0] y el blanco está ubicado en la esquina opuesta al origen [1,1,1], [15], ver Figura 3-1 Espacio de color RGB Figura 3-1.

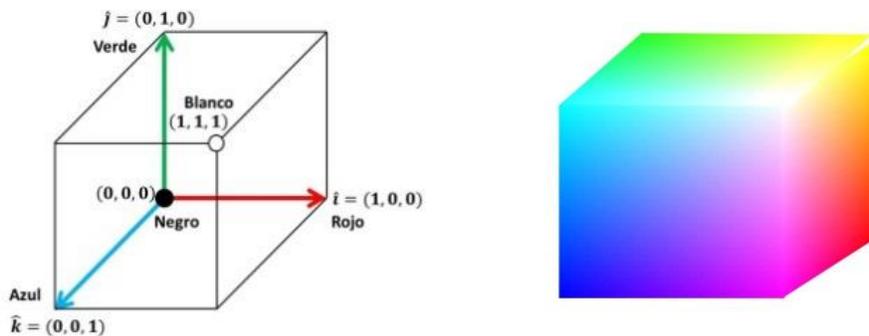


Figura 3-1 Espacio de color RGB

El color de un pixel p se escribe como una combinación lineal de los vectores base de verde, rojo y azul [15]:

$$\phi_p = r_p \hat{i} + g_p \hat{j} + b_p \hat{k} \quad (1)$$

En donde r_p , g_p y b_p son los componentes de rojo, verde y azul, respectivamente. La orientación y magnitud de un vector de color define la cromaticidad e intensidad del color, respectivamente [15]. Este espacio es sensible a la iluminación; es decir, a pesar de que dos vectores tengan la misma cromaticidad, estos representan diferentes colores si sus intensidades son diferentes.

Las cromaticidades correspondientes al rojo, verde y azul apuntan hacia las esquinas [1,0,0], [0,1,0] y [0,0,1], mientras que las cromaticidades cian, magenta y amarillo, también conocidos

como colores secundarios, apuntan hacia el resto de las esquinas que resultan al combinar dos colores primarios en igual medida, [1,1,0], [0,1,1] y [0,1,1].

3.2 Espacio de color HSV

Si bien los seres humanos cuentan con células distintas para cada color primario y la forma en la que se obtienen las imágenes digitales se basa en ello, no es precisamente de esta forma como el cerebro percibe los colores. Un color está formado de tres componentes importantes:

- Cromaticidad: la cualidad que se puede calificar como verde, amarillo, azul, etc.
- Saturación: todo color en su saturación más baja tiende a un tono pálido, a mayor saturación se tendrá un color más vivo.
- Luminiscencia: se refiere al brillo del color, a menor brillo el color tenderá a negro.



Figura 3-2 Representación gráfica del espacio de color HSV

La representación de color en el espacio HSV emula la percepción humana del color ya que la cromaticidad es desacoplada de la intensidad [16] [15]. En este espacio el color de un pixel p se representa por sus componentes de tono (h), saturación (s) e intensidad (v):

$$\varphi_p = [h_p, s_p, v_p] \quad (2)$$

El tono es la cromaticidad, la saturación es el nivel de blancura del color y la intensidad es el brillo del color; la Figura 3-3 Espacio de color HSV muestra la apariencia del espacio HSV. Los rangos de valores reales del tono, saturación e intensidad son $[0, 2\pi]$, $[0, 1]$ y $[0, 255]$, respectivamente.

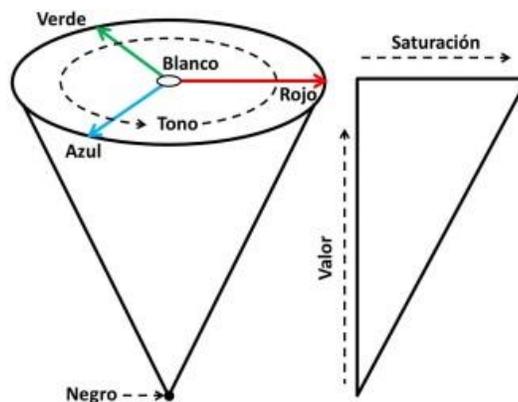


Figura 3-3 Espacio de color HSV

El espacio HSV tiene forma de cono, como se ve en la Figura 3-3 donde la cromaticidad está distribuida en la circunferencia de la base del cono, la cual se representa con valores en el rango $[0, 2\pi] \subset \mathbb{R}$. La saturación es la distancia al eje del brillo, y la intensidad es la altura en el eje que parte del centro de la circunferencia. La saturación y la intensidad están en el intervalo $[0, 1] \subset \mathbb{R}$.

La naturaleza de los colores no es lineal [11], de aquí no se pueda comparar los colores empleando la distancia Euclidiana. Por otra parte, en el espacio RGB los cambios de iluminación y saturación afectan considerablemente la representación de los colores ya que no los contempla explícitamente. Es por esto que en muchos trabajos de procesamiento de imágenes de color transforman las imágenes capturadas en el formato RGB a un espacio de color que no tengan las mismas debilidades del espacio RGB, como el espacio HSV, y así procesarlas exitosamente para que finalmente sean mapeadas de regreso al espacio RGB para su visualización.

3.3 Colores semánticos

Los colores semánticos se refieren a los nombres de los colores presentes en un lenguaje. Existen dos puntos de vista en la lingüística, por un lado, los relativistas defienden la idea de que cada lenguaje nombra los colores de forma arbitraria condiciona por la experiencia y cultura [17]; y por el otro, los universalistas, quienes defienden la idea de que la percepción humana del color está basada únicamente en la biología humana, depende directamente del funcionamiento neuropsicológico y es independiente de la cultura, por lo tanto, si todos los humanos perciben los colores de la misma forma, es posible que exista una semántica universal para los mismo colores [18], [19].

Aparentemente la naturaleza del color es difusa, y al no existir una manera natural para segmentar el espectro de luz visible no deberían existir coincidencias lingüísticas entre varios lenguajes si no comparten la misma raíz cultural. A pesar de esto, casualmente después una expedición a Chiapas, México, en la que se analizó el lenguaje Tzeltal y de otra expedición en Tahití, los científicos B. Berlin y P. Kay con investigaciones totalmente independientes [20] [21] [22], notaron que ambos lenguajes compartían la misma clasificación para los colores negro, blanco, rojo y amarillo con otros lenguajes más conocidos como el inglés, francés y chino, pero extrañamente ninguno de los lenguajes definía a los colores verde y azul como cromaticidades distintas.

En una investigación posterior realizada por los mismos autores, se realizó una encuesta a 20 personas con lenguas natales distintas, a cada una de ella se le presentó una paleta de colores distintos Figura 3-4, y se les pidió señalar los colores base, o los que mejor representaran a los existentes en su idioma, la información obtenida fue complementada con información obtenida de 78 textos literarios. Se encontró que los colores descritos en los 98 lenguajes analizados podían ser clasificados en 11 categorías básicas: blanco, negro, rojo, amarillo, verde, azul, café, púrpura, rosa, naranja, y gris, los demás colores resultan de combinar algunos de ellos.

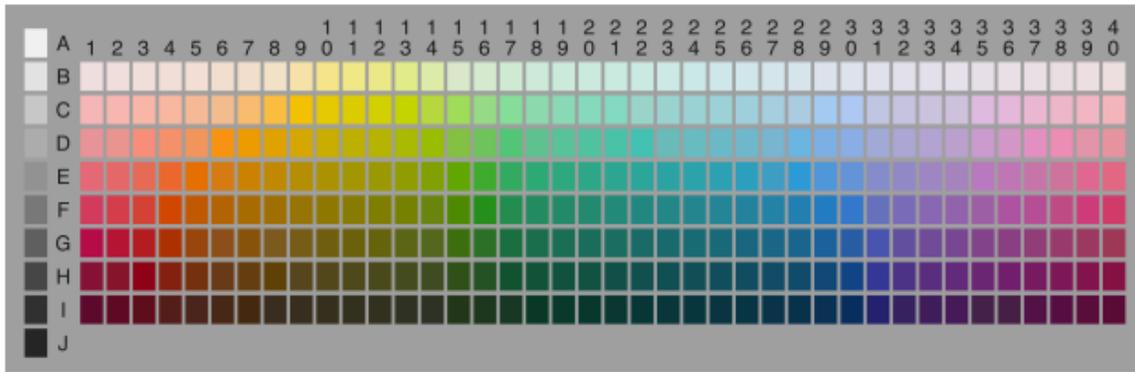


Figura 3-4. Paleta con muestras de colores para ser clasificados por personas

Aunque esta teoría parecía ser aplicable a la mayoría de los lenguajes, existían algunos en los que el número de colores existentes en el vocabulario eran inferiores a los 11 encontrados en la investigación, se notó que cuando esto ocurría, las personas hablantes de tal idioma tendían a clasificar varios colores como un mismo color, pero siempre de una forma similar. Por ejemplo, si existen sólo dos colores, estos son el blanco (colores claros) y negro (colores oscuros). Los colores claros incluyen al rojo, amarillo, u otros colores cálidos, al blanco, y otras tonalidades claras; mientras que los colores oscuros incluyen al negro, azul, verde, u otras tonalidades frías.

Además de esta observación, se notó que la aparición de los colores en un lenguaje solía tener un orden similar a pesar de no tener la misma raíz etimológica. Primeramente el blanco y el negro, después el rojo, seguido de ya sea el amarillo o el verde (no al mismo tiempo pero en este punto podía aparecer primero cualquiera de los dos seguido del otro), el azul y el verde solían clasificarse como la misma cromaticidad (color grue), y posterior a la aparición del color azul se iniciaba a diferencial al café, los colores púrpura, rosa, naranja y gris eran los últimos en aparecer, una investigación posterior llamó a estos últimos colores binarios y a los primeros como colores unarios [20].

$$\begin{bmatrix} \text{Blanco} \\ \text{Negro} \end{bmatrix} < [\text{Rojo}] < \begin{bmatrix} \text{Verde} \\ \text{Amarillo} \end{bmatrix} < \begin{bmatrix} \text{Amarillo} \\ \text{Verde} \end{bmatrix} < [\text{Azul}] < [\text{café}] < \begin{bmatrix} \text{Púrpura} \\ \text{Rosa} \\ \text{Naranja} \\ \text{Gris} \end{bmatrix}$$

Figura 3-5 Evolución de los colores semánticos

Un ejemplo muy claro de la aparición de los colores en el lenguaje existe en el español y el inglés, uno de los primeros lenguajes conocidos es el protoindoeuropeo del cual derivaron los lenguajes europeos y algunos asiáticos más utilizados hoy en día [23], en este idioma existía la palabra *bhel* que significaba brillo [24], [25], generalmente la fuentes emisoras de luz también emitían calor (esta palabra derivó en latín *flavus* que significaba amarillo o dorado, además de ser la misma raíz de la palabra flama), después de siglos de evolución lingüística en español ahora contamos con la palabra "blanco", mientras que en inglés desde sus inicios se tomó la palabra *bhel* con una perspectiva distinta, y aunque sí se mantuvo la su significado inicial en palabras como *bright* (brillo) o *bleach* (blanquear o descolorar), dada su relación con el fuego, también tomó el

significado de poner al fuego, de ahí surgió la palabra *black* (negro) [26], [27], y dado que el color azul se clasificaba inicialmente como, también la palabra *blue* (azul) [28].

Para definir el concepto de color se tienen las siguientes propiedades básicas [29], [30]:

- Es monoléxica, o sea que es definida por una sola palabra, por ejemplo, el color azul es un color semántico mientras que el azul marino es sólo una variante.
- Su significado no es aplicable a cromaticidades distintas.
- Puede ser aplicado a todo tipo de objetos.
- Es psicológicamente saliente, esto es, puede ser tomada como una de las principales características de un objeto.
- La palabra empleada para describirla no causa confusión y empleada por todas las personas que hablan el mismo idioma. Por ejemplo, el color fucsia no puede ser un color semántico debido a que no todas las personas lo identifican correctamente ni el uso de esa palabra es tan popular, en cambio el color rosa es conocido por todos los hispanoparlantes.

Aunque teóricamente los 11 colores semánticos pueden ser identificados por cualquier persona para clasificar los colores existentes, no es posible dividir el espectro de luz con alguna barrera fija, o sea, no existe un límite donde termina un color e inicia otro, esto porque la naturaleza del color es difusa. Para poder representarlos es necesario mapearlos a otro espacio de color con muestras clasificadas por distintos individuos, finalmente se calcula la media de los resultados para poder tener una base con la ubicación de los colores semánticos en otro espacio de color.

Para realizar estos experimentos es necesario realizarlo bajo condiciones iguales, y con individuos con una vista sana, por eso, primeramente, se realizan pruebas para detectar la sensibilidad de los individuos a los colores, por ejemplo, el test Farnsworth D-15 (consiste en ordenar colores con base en su similitud al contiguo) Figura 3-6, o el de Ishihara (se debe identificar números dentro un conjunto de puntos de diferentes colores) Figura 3-7. Después, dentro de un cuarto oscuro, se muestran cartas de colores a todos los voluntarios bajo condiciones iguales de iluminación, cromaticidades, número de muestras, y distancia entre el individuo y la muestra, para cada muestra se debe seleccionar en cuál de los 11 colores semánticos se puede clasificar [31].

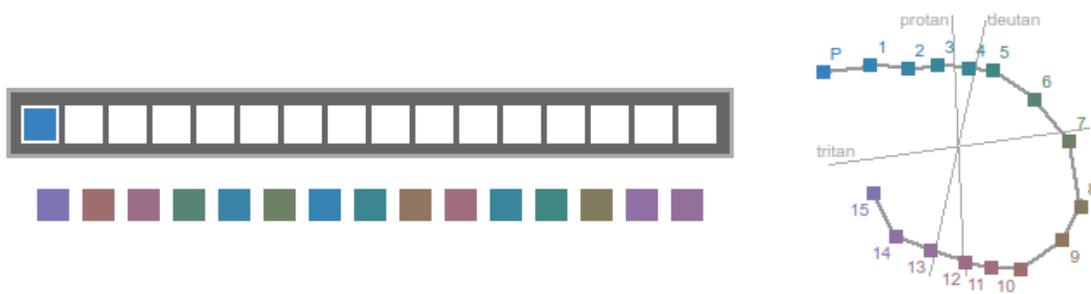


Figura 3-6 Test Farnsworth D-15, consiste en ordenar los colores, un mal ordenamiento sería prueba de algún tipo de daltonismo

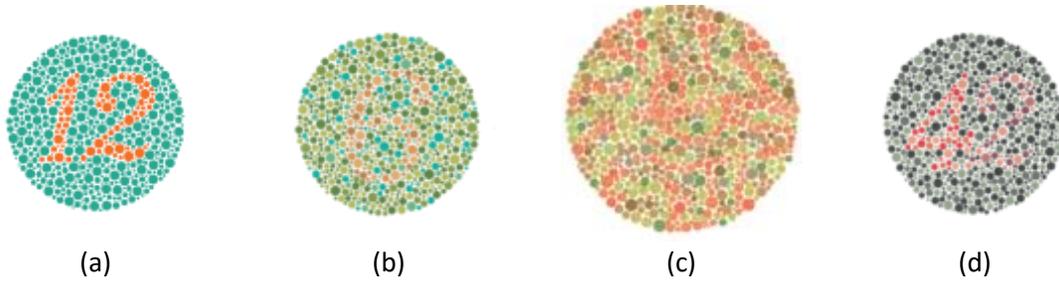


Figura 3-7 Cartas de Ishihara, contienen números de distintos colores para detectar algún tipo de daltonismo: (a) Carta de Ishihara No. 1(12); (b) Carta de Ishihara No. 11(6); (c) Carta de Ishihara No. 19(Nada, dígito oculto: los daltónicos ven el número 2); (d) Carta de Ishihara No. 23(42);

Los resultados suelen variar debido a que cada individuo tiene una percepción distinta a los colores, y aunque se suele tomar la media de varias muestras, los resultados finales suelen variar de investigación a investigación, Figura 3-8. Es por eso que para las pruebas con colores semánticos realizadas en el CENATAV se recurrió inicialmente a mapear directamente cada muestra a los colores semánticos estándar de internet, Tabla 3.1.

Tabla 3.1 Valores RGB para los 11 colores semánticos estándar para internet.

Colores	R	G	B
Negro	0	0	0
Azul	0	0	255
Marrón	165	42	42
Gris	181	181	181
Verde	0	255	0
Naranja	255	165	0
Rosa	255	20	147
Morado	155	48	255
Rojo	255	0	0
Blanco	255	255	255
Amarillo	255	255	0

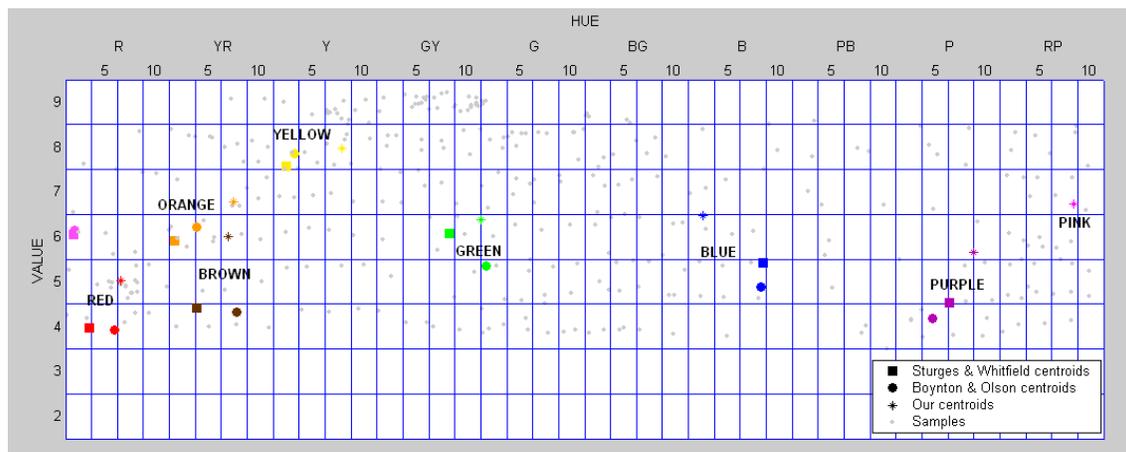


Figura 3-8 Distribución de los colores semánticos en distintas investigaciones

3.4 Métodos comúnmente empleados

La investigación “Segmentación de imágenes en el espacio de los colores semánticos” llevada a cabo por el CENATAV, consiste en identificar a personas y automóviles en escenas de video vigilancia, la búsqueda se realiza con base en el color de sus vestimentas o carrocerías (respectivamente), para poder identificar al sujeto en cuestión se necesita una descripción hablada, y para facilitar la descripción se limita la búsqueda a los colores semánticos descritos anteriormente.

En un intento anterior habían logrado segmentar las imágenes de muestra en colores semánticos, sin embargo, al segmentar las imágenes en su totalidad, si el área de interés (personas y automóviles) y el fondo de la imagen eran del mismo color, estos se fusionarían haciendo imposible identificar los objetos buscados. Así que se realizaron 3 tareas:

- Identificar los objetos que posiblemente correspondían a una persona o automóvil (discriminar el primer plano del segundo para evitar su fusión).
- Segmentar el primer plano con base en los colores semánticos.
- Segmentar el fondo de la imagen con base en su cromaticidad (no semántica) para mantener información posiblemente relevante, pero evitando su fusión con los objetos del primer plano.

Para realizar las pruebas del algoritmo fue utilizada una base de imágenes con 60,000 fotogramas consecutivos tomados por una cámara fija para video vigilancia con una resolución de 698 × 486 píxeles.





Figura 3-9 En un primer intento se utilizaba un mapeo de píxeles a colores semánticos, pero se tenía como resultado una fusión entre el primero y el segundo plano

3.4.1 Sustracción de fondo

Consiste en identificar cuáles objetos de la imagen corresponden al primer plano y cuáles al segundo plano o fondo, para lograrlo es necesario conocer el fondo para después eliminarlo de la escena en la están presentes los objetos de interés.

Existen varias formas para identificar el fondo de una imagen con base en su cromaticidad. Por ejemplo, para grabaciones de cine o televisión se utilizan cromas (muros o telones de un color específico), para poder identificar el fondo y después sustituirlo por uno distinto, generalmente son de una cromaticidad poco utilizada (verde o azul) para evitar que los objetos y el fondo sean confundidos. A pesar de ser una técnica bastante efectiva, no es factible su uso en escenarios no controlados debido a que los fondos de imagen varían de una locación a otra, suelen existir objetos de distintos colores, y la iluminación es distinta dependiendo de la hora del día (principalmente en exteriores).



Figura 3-10 Ejemplo de usos de cromas verdes. Todo píxel verde es clasificado como fondo

Dado que no existe un color específico para el fondo de cualquier imagen, se optó por utilizar una técnica distinta en la que los objetos clasificados como fondo son aquellos que no tuvieron un cambio significativo en varios fotogramas distintos, en cambio, si un objeto se movía a lo largo de la imagen, se podía considerar como una persona o automóvil, esta técnica tiene algunas

debilidades, por ejemplo, no detecta automóviles estacionados o peatones detenidos, además de que detecta todo tipo de objeto en movimiento ya sean personas, automóviles, animales, árboles, etc.

Esta técnica consiste en calcular la diferencia que existe entre un pixel en un fotograma contra el pixel correspondiente a las mismas coordenadas en otro fotograma. Si un pixel mantuvo el mismo valor, el resultado será 0, mientras que si hubo un cambio en sus valores la diferencia aumentará hasta un máximo de 1.

Para calcular la diferencia entre dos imágenes pixeles donde $p_{ij}(1)$ es el pixel en el primer fotograma, y $p_{ij}(2)$ el segundo fotograma:

$$\Delta I = |I(1) - I(2)|; \quad (3)$$

$$\Delta = \sqrt{(p_{rij}(1) - p_{rij}(2))^2 + (p_{gij}(1) - p_{gij}(2))^2 + (p_{bij}(1) - p_{bij}(2))^2} \quad (4)$$

Existen pixeles en los que se detecta una poca diferencia, por ejemplo, en las copas de los árboles, y otros en los que se detecta una mayor diferencia, para poder definir el umbral donde el movimiento es significativo se necesita de definir un umbral, puede ser fijo o dinámico, pero se optó por utilizar el método de Otsu [32] [33] que lo calcula en cada imagen, este método calcula el mejor umbral ajustándolo después de varias iteraciones donde la varianza es la mínima posible.

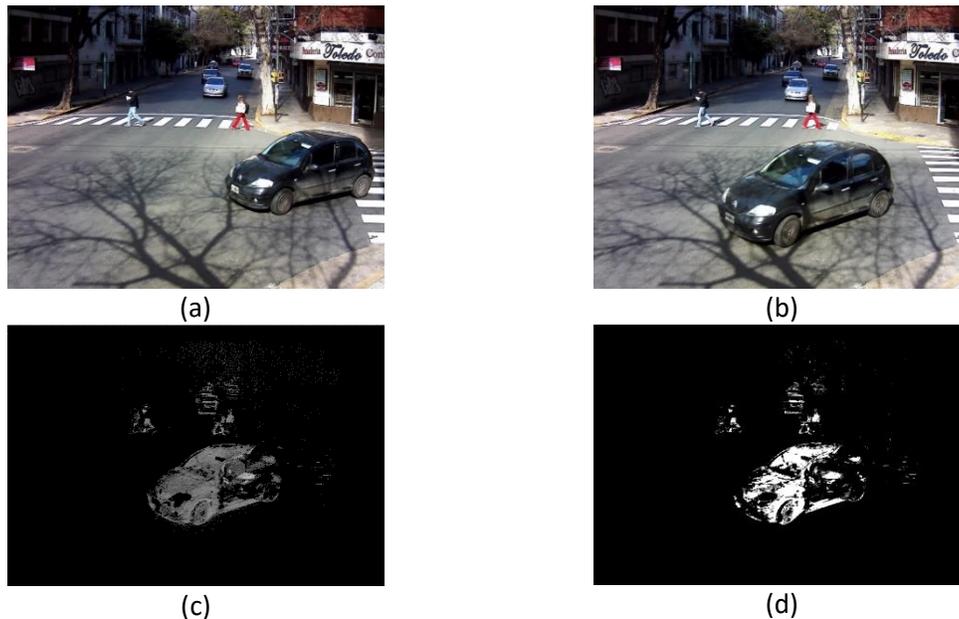


Figura 3-11 (a) y (b) son dos fotogramas consecutivos, (c) es el resultado de restar ambos fotogramas y (d) es el primer plano con valor en 1 y el fondo con valores en 0

En la sustracción de fondo se desea identificar únicamente la ubicación de los objetos en un fotograma, el otro fotograma sólo se usa como base, pero al utilizar dos fotogramas consecutivos los objetos en movimiento son detectados en ambos fotogramas como se puede observar en la Figura 3-11, debido a que no se conoce la dirección en la que se desplazan los objetos no se puede saber si el objeto en el fotograma analizado es el que se encuentra a la derecha o a la izquierda,

además de que sin otro tipo de algoritmo adicional no es posible identificar a los objetos y sus duplicados.

Para solucionar este problema, la resta de imágenes se realiza entre el fotograma a analizar y una imagen que sólo contiene el fondo, de esta forma, únicamente los objetos que no aparecen en la imagen de referencia son identificados como pertenecientes al primer plano. En un primer intento se buscaron fotogramas donde la calle apareciera vacía, pero era bastante concurrida así que de los 60,000 existentes en la base de imágenes, únicamente 5 estaban suficientemente libres de personas y automóviles para poderse tomar como referencia, pero a pesar de ser fondos en los que no aparecían objetos móviles, la iluminación de la calle no era constante así que mientras más tiempo transcurría mayor era la diferencia de iluminación y se producía ruido en los resultados.

Buscar entre todos los fotogramas una imagen que contenga únicamente el fondo de la calle es una tarea tediosa como para además aplicarse a varias tomas de distintas cámaras de video-vigilancia, además de que los resultados contienen demasiado ruido para ser útiles, así que se optó por generar el fondo dinámicamente. Calculando la media de varios fotogramas consecutivos, los píxeles que no tienen cambio alguno permanecen invariables, mientras que aquellos que presentan movimiento en algún fotograma tienden al estado en el que no presentan movimiento.

Esta técnica tampoco es infalible ya que tiene dificultades al momento identificar objetos en movimiento cuando los píxeles en movimiento son de una cromaticidad similar a la del fondo, este problema está presente en la mayoría de los algoritmos de segmentación de imágenes, e incluso los seres humanos tenemos dificultad para lograrlo cuando los objetos están camuflados; en algunos otros métodos se utiliza visión periférica, o sensores de profundidad para solucionar este problema, mas no es este el caso. En los resultados también se presenta ruido cuando partes del mismo u otro objeto están presentes en varios fotogramas, pero en la misma posición. A pesar de ello, los resultados son suficientemente buenos para identificar los objetos presentes en la imagen con técnicas posteriores a la segmentación de las imágenes de video-vigilancia, pero no son tratadas en esta investigación.

Una vez calculado el umbral, los valores cuyo movimiento fue superior a tal parámetro son considerados como automóviles o personas y son segmentados con los colores semánticos como base y con k-nn como clasificador, mientras que los píxeles con un cambio menos significativo son segmentados por medio de otra técnica de agrupamiento (k-means).

Estos algoritmos de clasificación y agrupación son ampliamente usados por su simplicidad y precisión, pero es posible utilizar otros métodos similares para los mismos fines.

3.4.2 K-Nearest Neighbour (K-nn)

En este paradigma de clasificación, cada nuevo caso es asignado a la clase más frecuente a la que pertenecen sus k-vecinos-más-cercanos. Existen distintas variantes del algoritmo básico, por ejemplo, en algunos casos se requiere que haya una mayoría de votos con un porcentaje mayor a un umbral predefinido, y en otros casos los vecinos más cercanos tienen un peso mayor.

La notación a utilizar en este tema es la siguiente [34]:

		X_1	...	X_j	...	X_n	C
(x_1, c_1)	1	x_{11}	...	X_{1j}	...	X_{1n}	c_1
\vdots	\vdots	\vdots		\vdots		\vdots	\vdots
(x_i, c_i)	I	x_{i1}	...	X_{ij}	...	X_{in}	c_i
\vdots	\vdots	\vdots		\vdots		\vdots	\vdots
(x_N, c_N)	N	x_{N1}	...	X_{Nj}	...	X_{Nn}	c_N
x	N + 1	$x_{N+1,1}$...	$X_{N+1,j}$...	$X_{N+1,n}$?

Tabla 2 Notación para el paradigma K-NN

Donde con N casos, cada uno está caracterizado por n variables predictoras, X_1, \dots, X_n y una variable a predecir, la clase C. Los N casos se denotan por $(x_1, c_1), \dots, (x_N, c_N)$ donde $x_i = (x_{i,1} \dots x_{i,n})$ para todo $i = 1, \dots, N$, $c_i \in \{c^1, \dots, c^m\}$ para todo $i = 1, \dots, N$, c^1, \dots, c^m denotan los m posibles valores de la variable clase C. El nuevo caso que se pretende clasificar se denota por $x = (x_1, \dots, x_n)$.

Este algoritmo de clasificación puede ser aplicado para la segmentación de imágenes, y para lograrlo necesita de conocimiento previo al igual que otros algoritmos de reconocimiento de patrones, aunque con la ventaja de que no necesita de un entrenamiento iterativo para poder clasificar los pixeles ya que trabaja directamente con los datos de muestra. Fue empleado para segmentar los objetos del primer plano, obtenidos por medio del procedimiento descrito en la sección anterior, y está enfocado a clasificar los pixeles en colores semánticos.

Manualmente se tomaron 10,000 muestras de color de pixeles pertenecientes únicamente a prendas de vestir y carrocerías, se creó una aplicación en matlab con la que se facilitó la tarea de recolección de muestras. Con ayuda de una interfaz gráfica se selecciona cada pixel y después por medio de botones se identifica el color al que pertenece, se guarda el dato y se continúa con el siguiente fotograma.

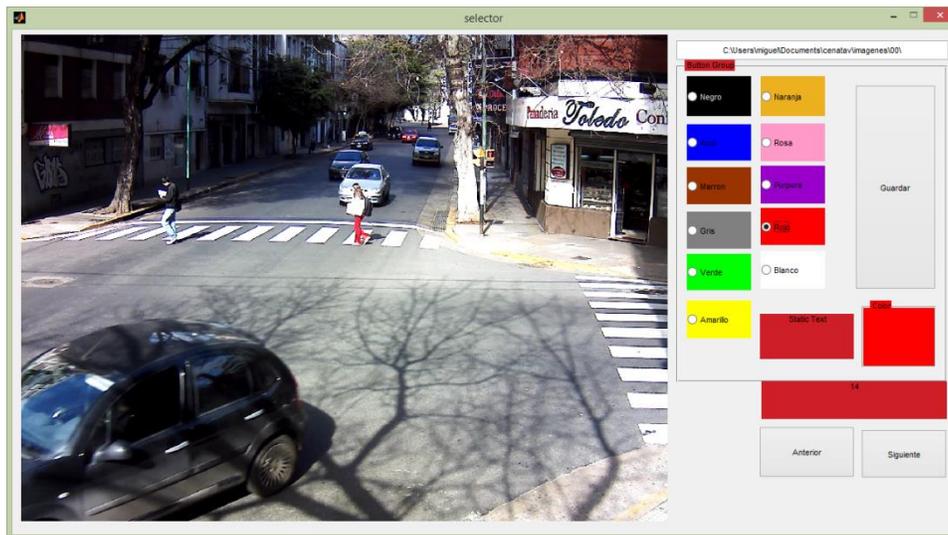


Figura 3-12 Interfaz gráfica de la aplicación realizada para la recolección de muestras de colores semánticos

El algoritmo básico de k-nn indica que se debe tomar en cuenta únicamente a los k-vecinos-más-cercanos cada que se clasifica un nuevo dato, pero no es posible identificar cuáles k-vecinos son los más cercanos sin antes calcular la distancia de todos los vecinos que existen contra el punto a clasificar, pero comparar cada punto de la imagen a clasificar contra todos los datos de entrenamiento se vuelve una tarea computacionalmente costosa. En algoritmos como k-means es factible realizar este tipo de comparaciones debido a que generalmente son pocos centroides por cada iteración, mientras que en este algoritmo se podría considerar cada muestra como un centroide.

Los datos seleccionados manualmente no estaban en orden, así que después de recolectar los datos, se eliminaron aquellos que se encontraban duplicados, y los datos restantes fueron almacenados en una matriz tridimensional de dimensión $256 \times 256 \times 256$ donde el dato almacenado era únicamente la clase a la que cada muestra pertenecía, para acceder a él se utilizan los datos $[r,g,b]$ del color como índice, a diferencia de una lista con 4 datos almacenados $[r,g,b]$ clase, el uso de una matriz tridimensional evita la necesidad de recorrer la lista completa para poder localizar la clase a la que pertenece un color ya que se accede directamente a las coordenadas deseadas.

Aún con los datos ordenados no es posible saber cuáles son los datos más cercanos a cada punto sin calcular la distancia que existe entre ellos, para disminuir la cantidad de operaciones que esto implicaría, se modificó el algoritmo básico, no se busca a k-vecinos-más-cercanos por una cantidad k específica sino cuantos de aquellas muestras existen dentro de un radio dado, es decir, que no rebase cierta distancia con respecto al pixel a clasificar, en caso de que no exista vecino alguno en aquella distancia, el radio se incrementa automáticamente hasta encontrar alguno.

Con una matriz de dimensión $256 \times 256 \times 256$ se tiene capacidad para almacenar hasta 16,777,216 datos, aunque sólo se cuenta con 10,000, razón por la cual existen más espacios nulos que aquellos con muestras. Primeramente, se seleccionan todos los que puedan contener algún valor dentro del radio definido, y después se asigna cada pixel a clasificar a la clase mayoritaria dentro ese radio.

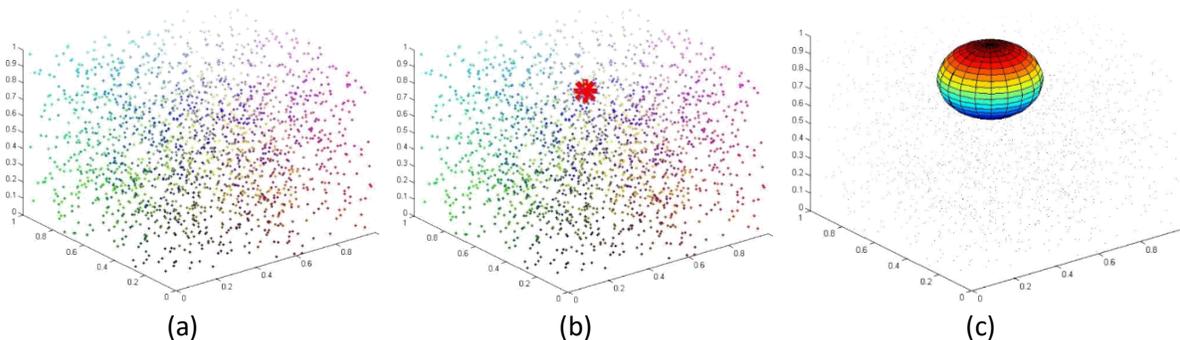


Figura 3-13 (a) es la representación de las muestras a color, (b) es el nuevo pixel a clasificar dentro las muestras, y (c) es el área de búsqueda limitada alrededor del pixel a clasificar

Para calcular los índices que podrían contener un dato válido se inicia con todos los puntos $[-\text{radio}, \text{radio}]$ en cada eje ($[x, y, z]$ o $[r, g, b]$).

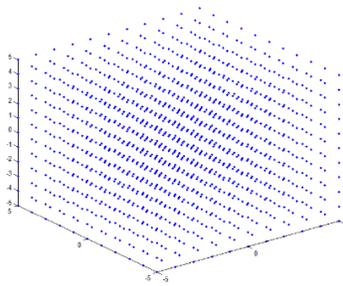


Figura 3-14 Cálculo de índices con posibles muestras válidas para la clasificación del pixel

Se descartan todos los que estén fuera del radio, o sea, que no cumplan con la condición $x^2 + y^2 + z^2 \leq r^2$.

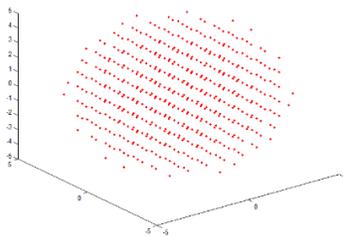


Figura 3-15 Se limitan los índices a un radio específico, por eso se muestran en forma de esfera

Y finalmente se suman a las coordenadas o valores RGB del pixel a clasificar y se verifica que contengan un valor distinto a nulo, $(r + x, g + y, b + z) \neq nulo$.

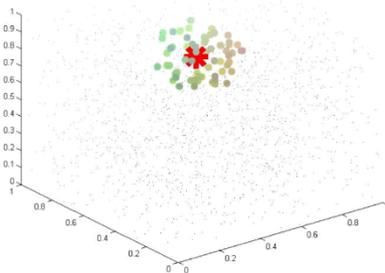


Figura 3-16 Los índices calculados se suman a las coordenadas cartesianas en el espacio RGB para su ubicación final

A pesar de contar con suficientes muestras de color, los colores semánticos blanco, negro y gris, no fueron clasificados correctamente en todos los casos, y solían ser confundidos con otros colores como el azul, verde, rojo o amarillo.



Figura 3-17 El segundo plano es segmentado por k-means, mientras el primer plano es segmentado con k-nn aunque presenta ruido

Esto se debe a que estos colores semánticos no representan a una cromaticidad específica sino a una saturación o brillo bajos, por ello se decidió no clasificarlos por medio de este algoritmo sino por un umbral, si la longitud del vector es pequeña, corresponde al color negro, si en el espacio HSV su saturación es baja se considera como gris, y si además la longitud es mayor que cierto umbral, se considera como un brillo alto y por lo tanto como blanco. Utilizando esta técnica se disminuyó el ruido y los resultados mejoraron. Cabe notar que en el algoritmo propuesto no se considera al color gris, únicamente al color blanco, que corresponde a una saturación baja, y al negro, que representa a un brillo bajo.



Figura 3-18 Segmentación resultante después de utilizar un umbral para los colores semánticos blanco, negro y gris

3.4.3 K-means

Como ya se mencionó, la segmentación de imágenes consiste en agrupar los píxeles que componen a la imagen en segmentos o grupos que comparten características similares, es por ello que es muy común el uso de técnicas existentes en la minería de datos para este fin. Para calcular la similitud entre un píxel y los demás de un grupo se pueden utilizar varias métricas: euclidiana, manhattan, coseno, etcétera.

Aplicado a la identificación de personas y automóviles con base en los colores semánticos, este algoritmo es aplicado únicamente al fondo sustraído con el fin de evitar fusiones con el primer plano, este método puede aplicarse en cualquier otro contexto, de hecho, es bastante común.

K-means suele utilizarse por ser simple y efectivo, y aunque es iterativo, generalmente no se necesita de muchas épocas para obtener buenos resultados. Al inicio del algoritmo se necesita indicar el número k de conjuntos presentes en la imagen, en el caso de una imagen este número indica la cantidad de colores presentes en la imagen o los que se espera que el algoritmo detecte, y la medida de similitud será calculada dependiendo de las características que se elijan, en este caso es el color, aunque no se limita únicamente a segmentación de imágenes.

Al inicio se seleccionan k-número de centroides al azar. Se calcula la distancia de cada punto a agrupar a cada uno de los centroides y se asigna al más cercano. Se recalcula cada centroide como la media de todos los puntos que están asignados a él. El algoritmo se repite n veces o hasta que los centroides no tengan una variación significativa.

El fin del algoritmo es minimizar la distancia que existe entre cada punto con respecto a uno de los centroides, existen varias medidas de similitud que se pueden aplicar al algoritmo, aunque dependerá de la forma en la que se representen los datos y el uso que se les dará.

Coseno: esta medida calcula la similitud en la dirección de dos vectores, a mayor apertura del ángulo menor será la similitud, aplicado al espacio RGB la magnitud del vector se refiere al brillo del pixel, mientras que la dirección apuntará a una cromaticidad específica con cierta saturación, en este aspecto, dos colores serán similares si apuntan a la misma cromaticidad y saturación, independientemente de la luminancia.

$$d = \frac{(x * x') + (y * y') + (z * z')}{|q| * |q'|} \quad (5)$$

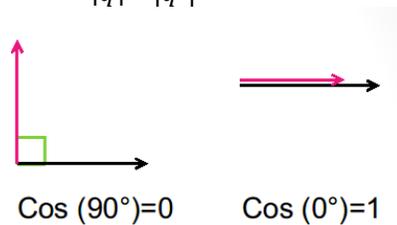


Figura 3-19 Dos vectores con diferentes magnitudes serán similares si apuntan hacia la misma dirección

Euclidiana: se puede aplicar en cualquier espacio euclidiano, y calcula la distancia más corta entre dos puntos, en el espacio RGB se puede utilizar y de hecho es quizás la medida más popular para este fin, aunque al calcular la distancia recta entre dos puntos, dos colores serán distintos si varían en su cromaticidad, saturación o brillo.

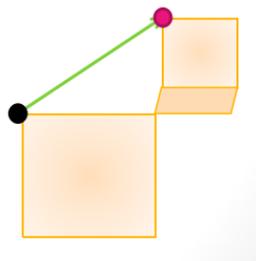


Figura 3-20 La distancia más corta entre dos puntos es la línea recta, y se calcula como la raíz de la suma de los cuadrados de sus lados

$$d = \sqrt{(x * x')^2 + (y * y')^2 + (z * z')^2} \quad (6)$$

City-block: también conocido como distancia Manhattan, del taxista, entre otras, es la suma de la diferencia absoluta de las coordenadas de dos puntos en un espacio euclidiano, su nombre lo obtiene por como un automóvil se desplaza de un punto a otro en una ciudad, mientras la distancia euclidiana calcula la distancia recta entre dos puntos, esta métrica calcula el número de calles que debe recorrer para llegar a su destino.

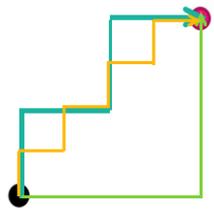


Figura 3-21 En una ciudad con calles cuadradas no hay forma de llegar en línea recta, existen varios caminos más cortos, pero todos tendrán el mismo número de calles a recorrer

$$d = |x - x'| + |y - y'| + |z - z'| \quad (7)$$

Perceptual: Fue creada en el CENATAV y es similar a la distancia euclidiana, pero desacopla la luminancia del matiz.

$$D(c, cs) = \sqrt{\alpha (\Delta|V(q)|)^2 + (1-\alpha (\Delta H)^2)} \quad (8)$$

Donde c es color del pixel de la imagen, cs son los colores semánticos, $\Delta|V(q)| = ||q| - |q' ||$ es diferencia de luminancia de los colores, $\Delta H = |H - H'|$ diferencia de matiz, y α es valor de ponderación.

Y para recalcular los centroides en cada iteración se debe calcular la media m de los n puntos asignados a cada grupo.

$$m = \sum_{a=1}^n \frac{R_{ai}}{n} + \sum_{a=1}^n \frac{G_{aj}}{n} + \sum_{a=1}^n \frac{B_{ak}}{n} \quad (9)$$

La función de similitud que se utilice en el algoritmo influye en los resultados de la segmentación y se debe seleccionar la correcta dependiendo de la aplicación que se le dé. Por ejemplo, en las siguientes figuras, las imágenes se ven mejor definidas al utilizar las funciones euclidiana y city-block, pero al realizar las mismas pruebas a un ojo humano, son las funciones del coseno y la perceptual las que logran detectar la pupila en la imagen, es fundamental en la biometría para el reconocimiento de iris.



Figura 3-22 K-means con 3 clusters aplicado a escenas de video-vigilancia

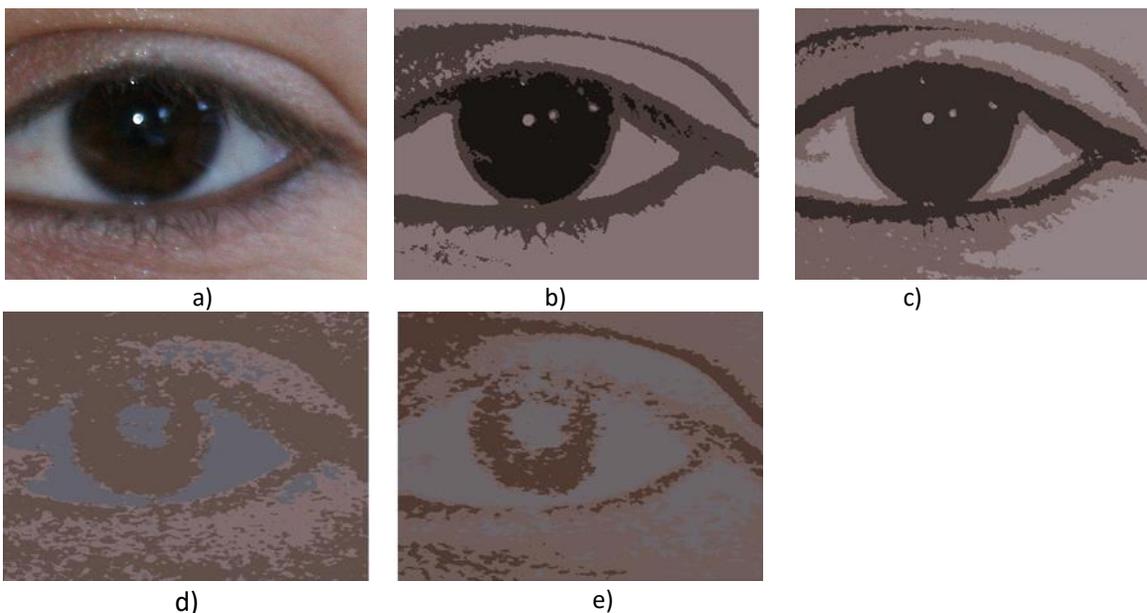


Figura 3-23 a) imagen de un ojo humano, b) segmentación utilizando la distancia euclidiana con k-means y $k = 3$, c) segmentación utilizando la distancia city block con k-means y $k = 3$, d) segmentación utilizando la distancia perceptual con k-means y $k = 3$, e) segmentación utilizando la medida del coseno con k-means y $k = 3$.

Finalmente, al combinar los resultados obtenidos con las distintas técnicas se llegó a la conclusión de que es factible segmentar imágenes con información semántica y evitar la fusión entre el primer y el segundo plano, pero debido a que este ejemplo sólo se enfoca en la segmentación de imágenes, será la investigación de la que este se desprende, la que se encargue de la identificación final de los automóviles y personas en escenas de video-vigilancia.



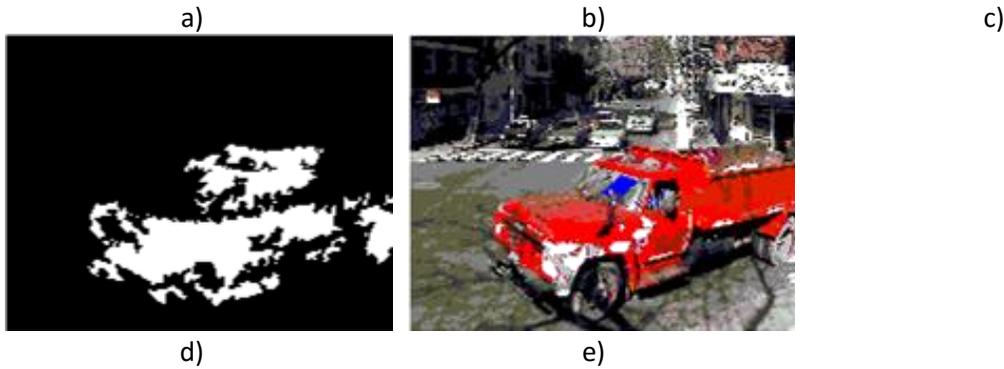


Figura 3-24 a) imagen original b) fondo generado con la media de 10 fotogramas consecutivos c) imagen segmentada con k-mean y la distancia euclidiana d) sustracción del primer y segundo plano, resultado de la resta entre a) y b) e) segmentación resultante de combinar K-means (en el segundo plano c)), y k-nn (en el primer plano).

3.4.4 Redes neuronales competitivas

Las redes neuronales artificiales están compuestas de una capa de entrada, una o varias capas ocultas y una capa salida. A partir de un conjunto de datos de entrada, se obtendrá un resultado en cada neurona de la capa de salida, durante el entrenamiento los pesos de cada neurona de la capa oculta son actualizados con respecto al estímulo obtenido en la capa de salida. Existen dos tipos de aprendizaje, el supervisado y el no supervisado, en el aprendizaje supervisado se cuenta con un conjunto de entrenamiento con sus respectivos resultados obtenidos, al ingresar estos valores se compara el resultado obtenido y el esperado, con base al error obtenido se recalculan los pesos de las capas ocultas.

En el caso del aprendizaje no supervisado no se requiere indicarle a la red el valor esperado, tal es el caso de las redes neuronales competitivas en las que simplemente se deja a las neuronas organizarse autónomamente tomando en cuenta a la neurona que se excitó más, durante el entrenamiento únicamente se actualiza el peso de esta neurona, así, cada neurona se especializa en reconocer un patrón específico mientras que cada una de las demás podrían o no especializarse en otro patrón distinto.

Una RNA competitiva tiene una estructura como la siguiente:

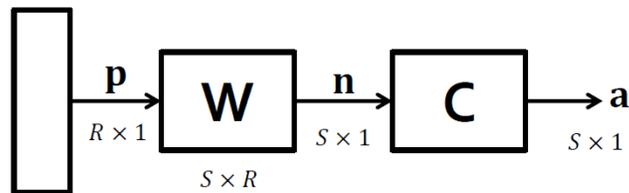


Figura 3-25 Estructura de una RNA de tipo competitiva

En una RNC de S neuronas, los datos de entrada de dimensión R están contenidos en un vector p , que al multiplicarse por una matriz de pesos W de dimensión $S \times R$ da como resultado un vector n con S número de salidas, al aplicársele la función *Compet* los datos de salida n se obtiene el

vector de salida a que contendrá varios ceros y un único valor 1 que indica cuál fue la neurona ganadora o con mayor excitación.

Los pesos en una RNA competitiva se actualizan únicamente si $a_i = 1$, esto es, sólo se modifican los pesos w_i de la neurona ganadora. Como se puede observar en la fórmula (10), para obtener el nuevo peso de la neurona se toma en cuenta el peso anterior de la misma neurona y los valores de entrada que excitaron a esa neurona, se tiene el parámetro α que indica la proporción en la que se tomarán estos valores, si α es un valor muy alto el nuevo valor será muy similar a la entrada p , si por el contrario el valor α es muy bajo, el peso de la neurona se verá poco afectado.

El comportamiento que tienen estas redes durante el entrenamiento es similar al de algunos métodos de agrupamiento, donde las neuronas de la red tienden a valores similares a los centroides del grupo después de varias iteraciones, de aquí que este algoritmo se pueda utilizar como algoritmo de agrupamiento, aunque de esta forma se desperdicia todo el aprendizaje obtenido [11].

Una variante a las redes neuronales competitivas es el mapa auto-organizado (MAO), donde al igual que en las redes competitivas, todas las neuronas compiten por especializarse en un patrón específico, pero en esta red, las neuronas vecinas dentro de un radio alrededor de la neurona ganadora también se actualizan. De esta forma se tienen zonas especializadas en un mismo patrón o, hasta cierto punto, en patrones similares.

$$\mathbf{w}_i(q) = (1 - \alpha)\mathbf{w}_i(q - 1) + \alpha p(q) \quad (10)$$

Para toda $i \in N_i^*(d)$ que contiene los índices de las neuronas dentro de un radio d .

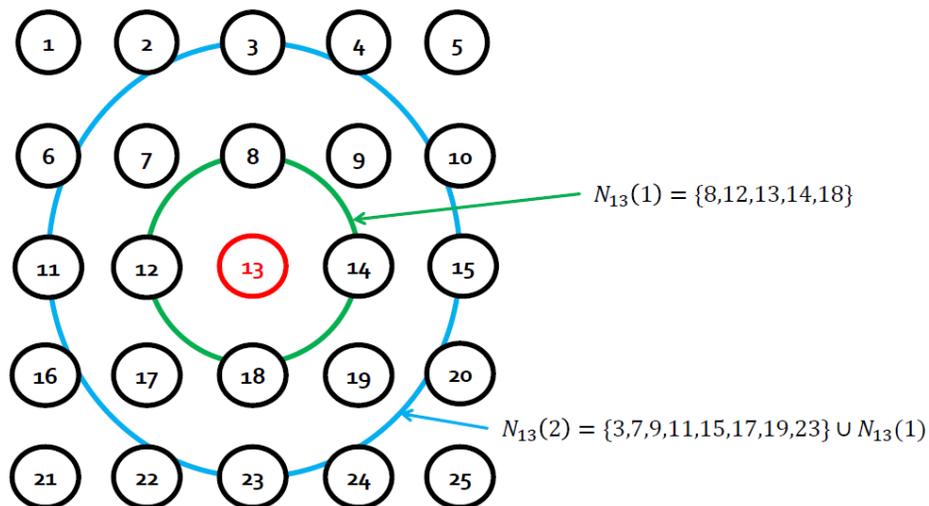


Figura 3-26 Ejemplo de la organización de las neuronas en la capa oculta de un MAO, y selección de las neuronas a entrenarse en un radio no mayor a 1, 2, o 3

4 Método propuesto

Los seres humanos dedicamos gran parte de nuestros primeros años de vida en aprender funciones básicas que utilizaremos a diario, entre ellas está la identificación de colores en cada escena que observamos. En la segmentación de imágenes a color, generalmente, se identifican los colores presentes en la imagen únicamente con base en el mismo conocimiento presente en ella, en el mejor de los casos, existe un aprendizaje en base a esa imagen, pero ese conocimiento adquirido es inútil al aplicarlo a cualquier otra escena, y todo el proceso de aprendizaje debe ser llevado a cabo desde cero para esta nueva imagen.

En los ojos humanos existen dos tipos de células sensibles a la luz: los conos y los bastones. Los bastones son células que no perciben cromaticidades, únicamente perciben distintas intensidades, estas se ubican alrededor del campo visual de tal forma que nos ayudan a ver en la oscuridad y a detectar movimiento en nuestro entorno (visión periférica), y aunque son mucho más sensibles y numerosas que los conos, no se encuentran en el centro de la retina (fóvea), lo que provoca que en condiciones de baja iluminación tengamos un punto ciego justo en el centro de nuestro campo visual, impidiéndonos ver a detalle y realizar actividades como leer.

En cambio, los conos se encuentran concentrados principalmente en la fóvea, permitiéndonos ver con gran detalle y a color si se cuenta con buenas condiciones de iluminación, estas células se especializan en identificar colores, ya sea el verde, rojo o azul. Inclusive cuando se presentan desórdenes como el daltonismo donde estas células no se desarrollan correctamente, la cromaticidad es la principal característica que tomamos en cuenta al observar una imagen.

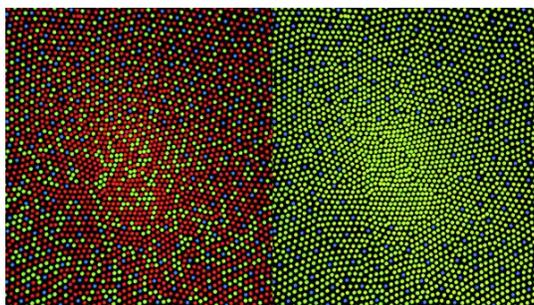


Figura 4-1 Distribución de los conos R, G y B en un ojo sano (izquierda), y uno con daltonismo (derecha)

La visión a color ha sido primordial para nuestra supervivencia como especie, nos permite identificar desde nuestros alimentos hasta qué peligros nos asechan. Por ejemplo, para saber si se trata de un limón, una lima, o una naranja, nos basamos principalmente en la cromaticidad, sin esta información identificar la tarea se dificulta.

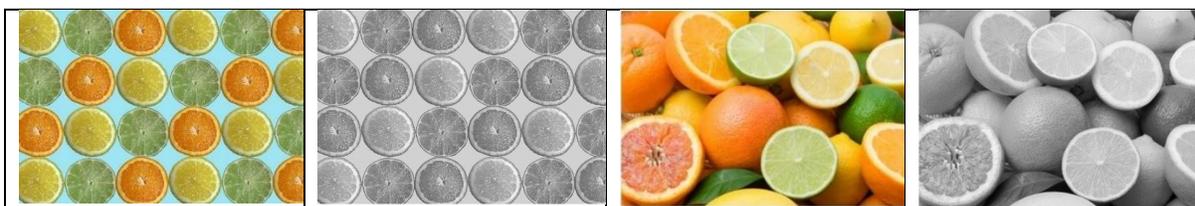


Figura 4-2 Cítricos en color y blanco y negro. El color es importante para identificarlos.

La luz visible está compuesta por ondas electromagnéticas de distintas frecuencias, cada una excitará más o menos dependiendo de la frecuencia que reciba, el cerebro califica esta información como rojo, verde, azul, o cualquier combinación de dos de estos componentes (la combinación de los tres colores resulta en blanco), a esta percepción se le conoce como cromaticidad.

Además de la cromaticidad, el color se compone por la saturación y la luminiscencia; la luminiscencia o brillo se refiere a la cantidad de luz que los ojos perciben, si esta cantidad es poca o nula se percibe como negro. Aunque el blanco y negro en algunos contextos pueden ser consideradas como palabras antónimas, este no es el caso, el ojo humano es capaz de adaptarse a distintas condiciones de iluminación, de tal forma que el color blanco no es producto de una iluminación excesiva sino a una saturación baja. La saturación se refiere a la pureza del color, o sea, la diferencia existente entre el color identificado y el blanco, siendo que este se presenta cuando los tres colores son percibidos simultáneamente.

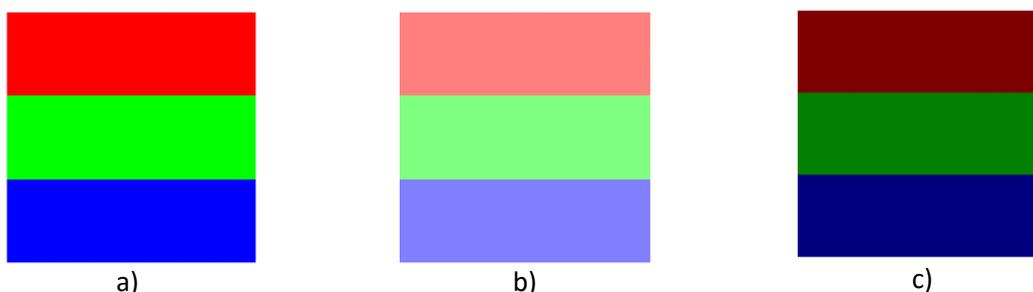


Figura 4-3 a) Colores rojo, verde y azul con brillo y saturación máximo, **b)** Colores rojo, verde y azul con saturación disminuida en un 50%, **c)** Colores rojo, verde y azul con brillo disminuido un 20%

Por ejemplo, durante un día despejado y soleado se puede observar que el tono o cromaticidad del cielo es azul, la iluminación es alta, pero se puede disminuir con lentes oscuros, aunque usarlos o no, no altera la saturación del color del cielo, en la parte más alta el azul será más puro que en la parte más baja, donde la saturación baja hasta convertirse casi completamente en blanco.



Figura 4-4 a) En un cielo azul despejado, la saturación del color disminuye del punto más alto (cenit), hasta el horizonte donde se percibe casi blanco, el brillo y la cromaticidad son iguales en todos los puntos. **B)** al disminuir el brillo la imagen se torna oscura, pero mantiene la misma cromaticidad y saturación que la original.

Hasta hace pocos años la mayoría de los enfoques de segmentación de imágenes consideraban únicamente la información de luminiscencia de las imágenes, desperdiciando toda la información cromática de las imágenes. En otros enfoques más recientes, la información a color ya se toma en

cuenta, pero utilizan el espacio RGB directamente que, aunque se obtienen buenos resultados, no contempla la diferencia que existe entre cromaticidad, saturación, e iluminación. De tal forma que no logra identificar si la diferencia entre dos colores se debe a un cambio en la cromaticidad (Figura 4-5) o, por ejemplo, a una sombra o un cambio de iluminación al momento de obtener la imagen, a pesar de ser la cromaticidad el componente principal en el que se basa la visión humana.

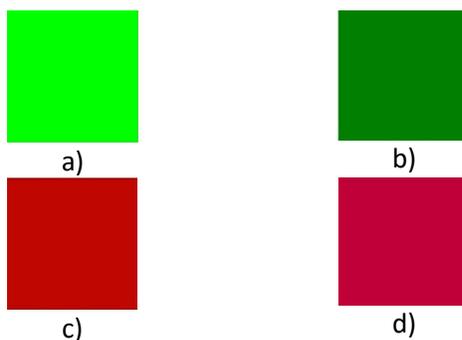


Figura 4-5 a) y b) representan a la misma cromaticidad (verde) con iluminaciones distintas, c) y d) son cromaticidades distintas (rojo y rosa), con la misma iluminación.

Es por ello que en la presente propuesta se pretende entrenar una red competitiva con muestras de colores en el espacio HSV con el fin de segmentar imágenes con base principalmente en su cromaticidad evitando reentrenar el algoritmo cada que una nueva imagen se presenta, además de ser menos sensible a cambios en la iluminación.

4.1 Arquitecturas de redes neuronales

El tipo de red neuronal adecuada para esta tarea es una del tipo no supervisada porque esta es capaz de encontrar la relación subyacente de los datos entre sí. Mientras que con una red neuronal supervisada el entrenamiento dependerá de la persona que construya su conjunto de entrenamiento; dado que el color es de naturaleza difusa, la percepción del color entre las personas es diferente y en consecuencia el conjunto de entrenamiento no será el mismo [35].

De las redes no supervisadas que se pueden emplear están las redes neuronales competitivas y los mapas auto-organizados. Aunque los mapas auto-organizados son una generalización de las redes competitivas, es posible que, con una red competitiva que requiere menos recursos computacionales, sea suficiente para esta tarea. La clave será definir la cantidad de neuronas que debe tener la red neuronal para reconocer varios colores, pero sin ser caro computacionalmente.

Aplicándolo a nuestro enfoque, después del entrenamiento, en la red neuronal competitiva las cromaticidades se organizarán al azar mientras que en el mapa auto-organizado por zonas con cromaticidades similares para que por cada pixel haya una neurona que tenga una mayor excitación, en base a esto se identifica la cromaticidad a la que pertenece y se etiqueta cada pixel de la imagen como pertenecientes a un mismo segmento.

En esta sección se presenta una nueva propuesta para la segmentación de imágenes de color, en donde se entrena una RN con muestras de cromaticidad de diferentes colores. Posteriormente se

extrae la cromaticidad de cada pixel de la imagen a segmentar al mapear el color al espacio HSV, la cromaticidad extraída es procesada por la RN y el nuevo tono del pixel es el tono de la neurona ganadora. Cuando todos los pixeles son procesados, la imagen es mapeada al espacio RGB.

4.2 Entrenamiento de la red neuronal

Debido a la naturaleza difusa del color, no es posible reconocer todos los colores del espectro; de aquí que, el espectro de color es “dividido” en un conjunto finito de colores. El número de colores que la RN puede reconocer depende de su tamaño; en este trabajo se realizan pruebas con una red neuronal competitiva (RNC) con 9, 16 y 25 neuronas, y con MAO de 3×3, 4×4 y 5×5 neuronas. Las RN son entrenadas con los elementos del conjunto Ψ construido con muestras de cromaticidad como sigue:

$$\Psi = \left\{ \psi_k = [\cos \theta_k, \sin \theta_k] \mid \theta_k = \frac{2\pi}{256} k: k = 0, 1, \dots, 255 \right\} \quad (11)$$

La cromaticidad es transformada en un vector debido al caso cuando el valor del tono es casi 0 o 2π . Considérese los cuadros (c) y (d) de la Figura 4-5, sus valores son $\pi/100$ y $19\pi/10$, respectivamente. Numéricamente ambos valores son muy distintos pero las cromaticidades de ambos cuadros son muy similares; si la cromaticidad de ambos cuadros es clasificada solamente por el valor escalar del tono, la cromaticidad es reconocida como si fueran muy diferentes.

Este problema se resuelve como sigue; siendo φ_p el color de un pixel representado en el espacio HSV como se muestra en la ec. (2), la cromaticidad es modelada como:

$$\psi_p = [\cos h_p, \sin h_p] \quad (12)$$

4.3 Procesamiento de la imagen

La segmentación de la imagen se hace al agrupar los colores empleando la cromaticidad de los colores de cada pixel de la imagen. Es importante mencionar que las RNA son entrenadas con información de la cromaticidad de los colores, por lo que no pueden reconocer el negro ni el blanco porque estos dos colores no tienen una cromaticidad definida. El blanco se obtiene cuando la saturación de un color es bajo, es decir, cuando $s \approx 0$; por otra parte, el negro se obtiene cuando la intensidad del color es baja, esto es, cuando $v \approx 0$.

Por lo tanto, antes de que un color sea procesado por la RN se debe evaluar su saturación e intensidad para clasificarlo como blanco o negro, respectivamente. Procesar el color de un pixel conlleva realizar los siguientes pasos. Sea el vector de color ϕ_p del pixel p representado en el espacio RGB:

1. El vector ϕ_p se mapea al espacio HSV obteniendo $\varphi_p = [h_p, s_p, v_p]$.
2. Se verifica si el color del pixel es negro; si $v_p \leq \delta_v$ entonces $v_p^* = 0$ y $s_p^* = 0$, ir al paso 5.
3. En caso contrario, se verifica si el color del pixel es blanco; si $s_p \leq \delta_s$ entonces $v_p^* = 191$ y $s_p^* = 0$, ir al paso 5.

4. En caso contrario, es decir, si $v_p > \delta_v$ y $s_p > \delta_s$ entonces:
 - a. Calcular el vector ψ_p y procesarlo con la RN.
 - b. Se obtiene el vector de peso de la neurona ganadora $\mathbf{w}_i = [w_{i,1}, w_{i,2}]$ y se etiqueta al pixel con el número i .
 - c. Calcular el tono con $h_p^* = \tan^{-1}(w_{i,2}/w_{i,1})$.
 - d. Se asignan los valores de saturación e intensidad: $v_p^* = 191$ y $s_p^* = 1$.
5. El nuevo vector $\varphi_p^* = [h_p^*, s_p^*, v_p^*]$ es mapeado al espacio RGB obteniendo el vector $\phi_p^* = [r_p^*, g_p^*, b_p^*]$.

En donde δ_s y δ_v son los umbrales para saturación e intensidad, respectivamente. Dada la naturaleza difusa del color, no hay valores específicos para decidir exactamente cuando un color es blanco o negro; de forma experimental encontramos que los mejores umbrales son $\delta_s = \mu_s - \sigma_s$ y $\delta_v = \mu_v - \sigma_v$; en donde μ_s y μ_v son la media de saturación e intensidad de la imagen, respectivamente; σ_s y σ_v son la desviación estándar de la saturación e intensidad de la imagen, respectivamente.

5 Experimentos

Para realizar las pruebas del algoritmo se utilizó como referencia la base de imágenes de la universidad de Berkeley (BSB), que en los últimos años se ha popularizado entre la comunidad investigadora, esta base de imágenes contiene 200 imágenes para entrenamiento y 100 más para realizar pruebas, entre 30 personas segmentaron manualmente todas las imágenes y se tienen 5 resultados para cada imagen. A partir de estos resultados se puede calificar la eficiencia del algoritmo propuesto.

Todos los experimentos fueron implementados en Matlab 2014a, y se seleccionó aleatoriamente un conjunto de 9 imágenes de las 300 imágenes que contiene la BSB para ser analizado en esta sección, ver Figura 5-1.

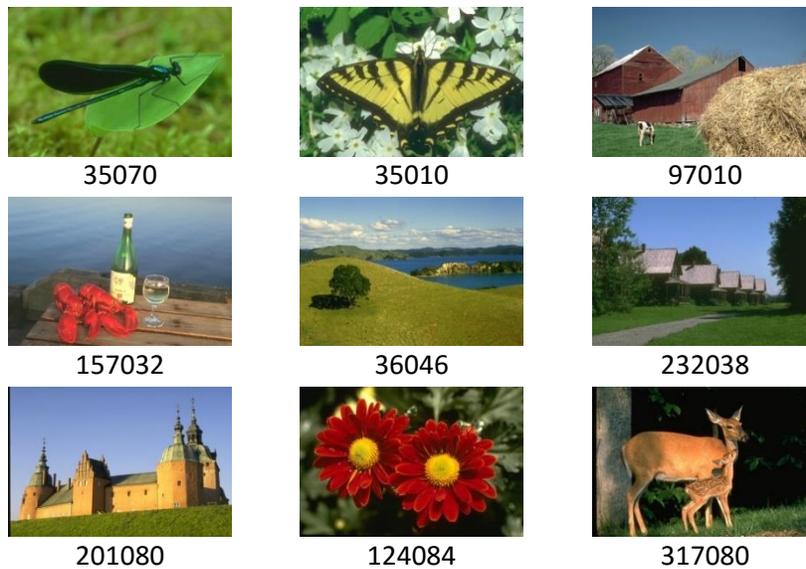


Figura 5-1 Imágenes extraídas de la BSB, empleadas para los experimentos realizados

5.1 Redes neuronales competitivas

En la Figura 5-3 se muestran las imágenes obtenidas al procesar las imágenes de la Figura 5-1 empleando las RNCs con los diferentes tamaños que se indican. A su vez en la Figura 5-2 se muestran las imágenes resultantes al ser procesadas las imágenes de la Figura 5-1 empleando los MAOs con los tamaños que se indican.

ID	MAO 3×3	MAO 4×4	MAO 5×5
35070			
35010			
97010			
157032			
36046			
232038			
201080			
124084			
317080			

Figura 5-2 Imágenes obtenidas empleando mapas auto-organizados de 3×3, 4×4 y 5×5 neuronas

Se puede observar fácilmente de las imágenes resultantes que estas pueden ser segmentadas solamente utilizando información de la cromaticidad; pero la segmentación también depende de la cantidad de neuronas de las RNs. Esto es, entre más grande sea la RN, mayor la cantidad de colores son reconocidos; de hecho, se puede apreciar que con los MAO se reconocen más colores que con las RNC, ya que se pueden observar más secciones dentro de la imagen empleando los MAO. Aunque también, por lo mismo, hay algunas partes de la imagen que no son segmentadas homogéneamente.

Al observar las imágenes se puede ver que estas tienen mejor segmentación empleando los MAO. Por ejemplo, en las imágenes obtenidas al procesar la imagen 35070 con las RNC, el fondo se combina con la hoja; mientras que la misma imagen procesada por los MAO se puede ver que la hoja es segmentada del fondo, a pesar de tener ambos tonos verdes pero los MAO son capaces de distinguir la diferencia de tonos.

5.2 Mapas auto-organizados

Otro ejemplo, en las imágenes obtenidas de la imagen 35010 utilizando las RNC, prácticamente el fondo de las hojas es segmentado en verde, mientras que con los MAO se pueden apreciar tonos en amarillo en los centros de las hojas. Las alas de la mariposa son segmentadas exitosamente con todas las redes neuronales, excepto utilizando la RNC de 25 neuronas, las cuales son segmentadas con el mismo tono de verde de las hojas del fondo. Una posible explicación es que esa red neuronal no “aprendió” correctamente a reconocer el tono amarillo durante el entrenamiento, por lo que la red neuronal debe ser entrenada nuevamente.

ID	RNC 9	RNC 16	RNC 25
35070			
35010			
97010			
157032			
36046			
232038			
201080			
124084			
317080			

Figura 5-3 Imágenes obtenidas empleando redes neuronales competitivas de 9, 16 y 25 neuronas

Dada la apariencia de las imágenes, la segmentación de las imágenes utilizando las RNC es mejor cuando la red neuronal es pequeña, mientras que con los MAO es lo contrario; es decir, el resultado de la segmentación de las imágenes utilizando los MAO es mejor cuando la red neuronal es grande. Por ejemplo, la imagen 124084 obtenida con la RNC de 9 neuronas es muy parecida a la obtenida utilizando el MAO de 5×5 neuronas. Ocurre de forma similar con las imágenes 97010, 157032, 201080 y 35010 si se utiliza un la RNC de 9 neuronas y el MAO de 5×5 neuronas.

6 Discusión

6.1 Evaluación cuantitativa de la segmentación de imágenes de color

Para evaluar qué tan buenos o malos son los resultados de una imagen segmentada existen varios métodos, aunque en su mayoría se basan en la clasificación de píxeles, donde separan las características texturales de las de color de cada píxel de la imagen original y de la imagen segmentada, para después compararse, y mientras más parecidos sean ambos conjuntos, mejor es la segmentación.

La evaluación cuantitativa de la segmentación de las imágenes de color se ha definido recientemente, donde se han desarrollado diferentes métricas, entre ellas están [36], [37], [38]:

Probabilistic Rand Index (PRI): Compara los resultados obtenidos con el algoritmo a evaluar con un conjunto de imágenes segmentadas a mano, los resultados pueden variar dependiendo de la persona que segmentó la imagen. La segmentación de la imagen será mejor, mientras mayor sea el número de píxeles que coincidan entre las imágenes. El rango es de 0 a 1, donde 1 significa que coincidieron totalmente.

$$R = \frac{a + b}{a + b + c + d} = \frac{a + b}{\frac{n}{2}} \quad (13)$$

Dado un conjunto de n elementos de dos particiones S a comparar, se define a a como el número de pares en S que están en el mismo conjunto en X , y en el mismo conjunto en Y , b es el número de pares de elemento de S que están en diferentes conjuntos en X y en diferentes conjuntos en Y , c se refiere al número de pares de elementos en S que están en el mismo conjunto en X pero en diferentes conjuntos en Y , y similarmente, d es el número de pares de elementos en S que están en distintos conjuntos en X pero en el mismo conjunto Y .

Variation of Information (VOI): Evalúa cuánta información de cada grupo se perdió o ganó con el algoritmo usado, en base la segmentación manual. El resultado será mejor mientras menor sea la pérdida o ganancia; esto es, la mejor segmentación se obtiene cuando el resultado sea 0.

Esta diferencia también puede ser definida como la entropía condicional promedio de una segmentación dada otra, así que se mide la cantidad de aleatoriedad en un segmento que no puede ser explicado por otro. Teniendo dos imágenes segmentadas X y Y , donde $X = \{X_1, X_2, \dots, X_k\}$, $p_i = |X_i|/n$, $n = \sum_k |X_i|$. Entonces la variación de información es:

$$VI(X, Y) = H(X) + H(Y) - 2I(X, Y) \quad (14)$$

Donde, $H(X)$ es la entropía de X , e $I(X, Y)$ es la información mutua entre X y Y .

Global Consistency Error (GCE): Evalúa hasta qué punto la segmentación puede ser vista como la refinación de la otra; esto es, cuántos cambios hubo entre las dos segmentaciones. El rango es de 0 a 1, los resultados son mejores cuando se aproximan a 0.

Los segmentos son conjuntos de píxeles, si un segmento es un subconjunto de otro, entonces el píxel se encuentra en un área de refinamiento y el error debe ser cero, si no hay relación entre subconjuntos, entonces las dos regiones se superponen inconsistentemente.

$$GCE = \frac{1}{n} \min \left\{ \sum_i E(S1, S2, pi), \sum_i E(S2, S1, pi) \right\} \quad (15)$$

Boundary Displacement Error (BDE): Evalúa la distancia que se desplazaron los bordes con respecto a la imagen de control, empleando la distancia euclidiana. Mejor serán los resultados mientras menor sea la distancia obtenida.

$$\mu_{LA}(u, v) = \begin{cases} \frac{u - v}{L - 1}, & 0 < u - v \leq L - 1 \\ 0, & u - v < 0 \end{cases} \quad (16)$$

Esta medición no será tomada en cuenta para evaluar el algoritmo propuesto debido a que este no calcula los bordes para realizar la segmentación, a diferencia de otros algoritmos que sí y en donde sí se puede aplicar esta medida.

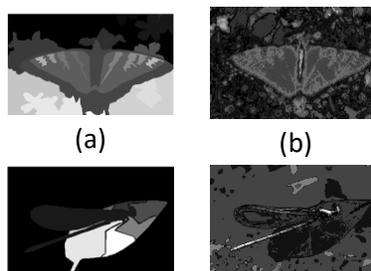
A pesar de que existen actualmente varias métricas, aun no hay algún conjunto de ellas que sean las estándar. Al revisar trabajos relacionados de segmentación de imágenes de color se ha observado que las métricas que más se emplean son la PRI y VOI [36]. De aquí que, para este trabajo, dichas métricas son las que se emplearan para evaluar el desempeño de nuestra propuesta, además de GCE que, aunque es menos popular, también es conocida.

La imagen segmentada con nuestra propuesta es comparada con las imágenes segmentadas a mano que se encuentran en la BSB que sirven como referencia. Cada imagen de la BSB tiene un conjunto de 5 imágenes segmentadas a mano, con las que se hace la comparación. Por ejemplo, en la se muestra las imágenes segmentadas a mano de la imagen 35010 de la BSB.



Figura 6-1 Ejemplo de imágenes segmentadas a mano de la BSB

Las imágenes segmentadas obtenidas con nuestra propuesta son comparadas con cada una de las imágenes segmentadas a mano de la BSB, en donde la similitud de la segmentación se mide con las métricas mencionadas anteriormente. En la Figura 6-2 se muestran dos ejemplos de imágenes segmentadas a mano, (a) y (c), y dos que se obtienen con nuestra propuesta, (b) y (d).



(c) (d)

Figura 6-2 Ejemplo de comparación de las imágenes: (a) y (c), segmentadas a mano y las segmentadas con nuestra propuesta, (b) y (d)

Para cada métrica se evaluaron las segmentaciones resultantes tomando como referencia a cada una de las segmentaciones manuales contenidas en la BSB. Se calculó el promedio, el valor máximo y el mínimo de los 5 resultados para cada prueba de cada imagen; Y finalmente se calculó el promedio de estos resultados en las 300 imágenes por cada algoritmo.

Se hicieron pruebas a los 4 algoritmos programados:

- K-nn: entrenado para identificar los 11 colores semánticos en una imagen.
- K-means: identifica k-número de colores que corresponden al k-número de centroides elegidos, para estas pruebas fueron 3, 5, 7, 10, 15, 20, 25, 30.
- RNA competitiva: contiene 9, 16, o 25 neuronas en la capa oculta, es el mismo número de colores que logra identificar.
- MAO: Contiene el mismo número de neuronas que la RNA competitiva, pero están organizadas matrices de 3×3 , 4×4 , y 5×5 .

6.1.1 VOI (*Variation of Information*)

Utilizando la fórmula de VOI, k-means parece empeorar a medida que identifica más colores, pasando de obtener los mejores resultados promedio a los peores. Se tienen un efecto similar utilizando los MAO donde aparece un ligero decremento en la precisión cuando se utilizan más neuronas, este crecimiento en la gráfica se puede deber a que hay más segmentos de los que identificaron las personas y, como esta fórmula calcula la entropía, es sensible a cualquier incremento en la cantidad de información existente en la segmentación final. Las RNA competitivas mejoran al momento de incrementar el número de colores que detecta, aunque no obtienen mejores que aquellos obtenidos por k-nnn.

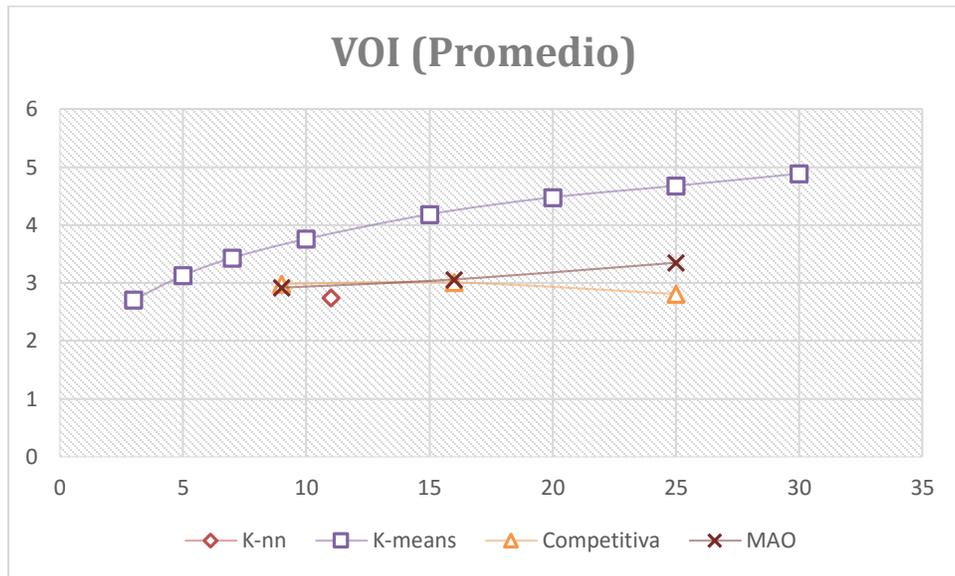


Figura 6-3 VOI (Promedio)

VOI (Promedio)		
	No. Colores	
K-nn	11	2.74753633
K-means	3	2.708228
	5	3.12702767
	7	3.43158533
	10	3.75782033
	15	4.183979
	20	4.47646533
	25	4.675028
	30	4.88443667
Competitiva	9	2.98105333
	16	3.00763667
	25	2.80660567
MAO	9	2.91481767
	16	3.05902433
	25	3.34996

Tabla 6.1 VOI (Promedio)

Observando las calificaciones más altas (los peores para esta fórmula), se observa un comportamiento similar al de los resultados promedio y los mejores resultados, siendo k-means el que obtiene los mejores y los peores resultados.

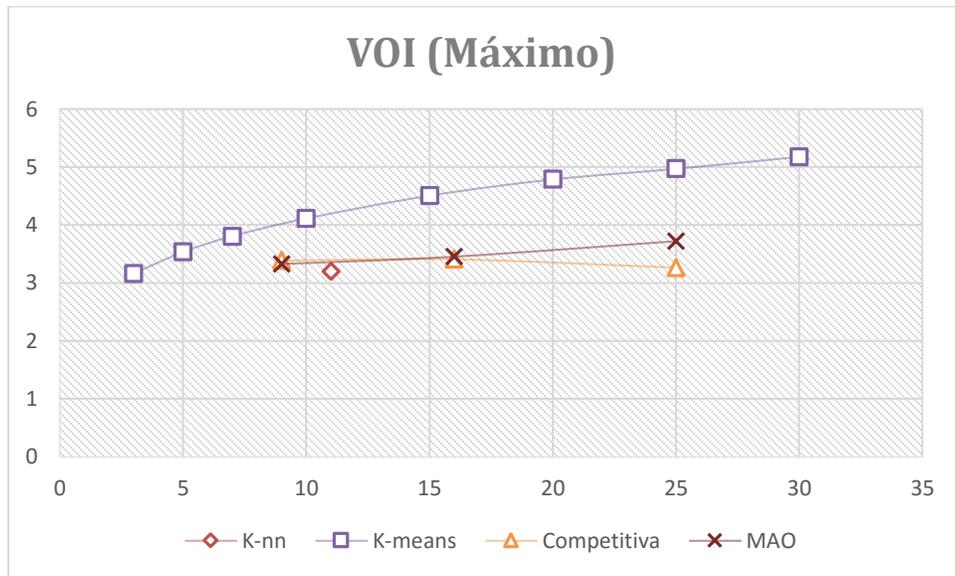


Figura 6-4 VOI (Máximo)

VOI (Máximo)		
	No. Colores	
K-nn	11	3.19894133
K-means	3	3.160587
	5	3.531693
	7	3.80586767
	10	4.11154767
	15	4.50620533
	20	4.78623733
	25	4.96945133
	30	5.176692
Competitiva	9	3.38512067
	16	3.40978033
	25	3.26297033
MAO	9	3.32663233
	16	3.452945
	25	3.71984433

Tabla 6.2 VOI (Máximo)

Las calificaciones más bajas son para aquellas segmentaciones con mejor desempeño, aunque k-means con 3 grupos resulta ser la mejor segmentación, rápidamente empeora conforme identifica más colores, k-nn se mantiene en segundo lugar con resultados similares a los obtenidos por las redes competitivas, el otro algoritmo propuesto obtiene buenos resultados, y aunque no son los mejores y empeoran mientras más neuronas tiene, la curva no es tan pronunciada como es el caso de k-means.

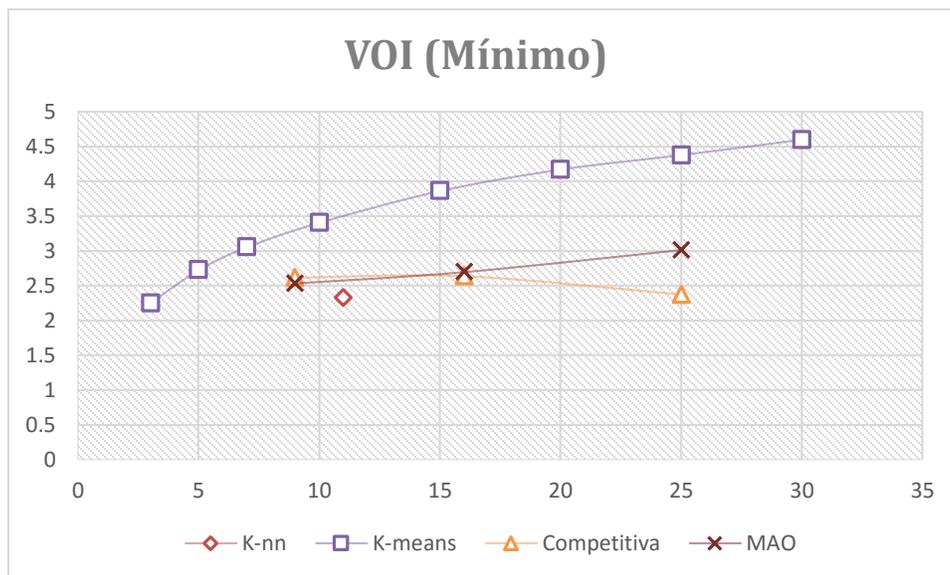


Figura 6-5 VOI (Mínimo)

VOI (Mínimo)		
	No. Colores	
K-nn	11	2.32925833
K-means	3	2.25390467
	5	2.72787933
	7	3.05595433
	10	3.407684
	15	3.86400267
	20	4.17227967
	25	4.380836
	30	4.60140433
Competitiva	9	2.61397833
	16	2.63800867
	25	2.375584
MAO	9	2.534409
	16	2.69749367
	25	3.015122

Tabla 6.3 VOI (Mínimo)

6.1.2 PRI (Probabilistic Rand Index)

Observando los datos correspondientes a la evaluación PRI, en promedio los MAO mejoran al incrementarse el número de neuronas, k-nn tiene una calificación baja, k-means en promedio se comporta bien, y aunque los mejores resultados, también tiene malos resultados, y las RNA competitivas tienden a empeorar con más neuronas.

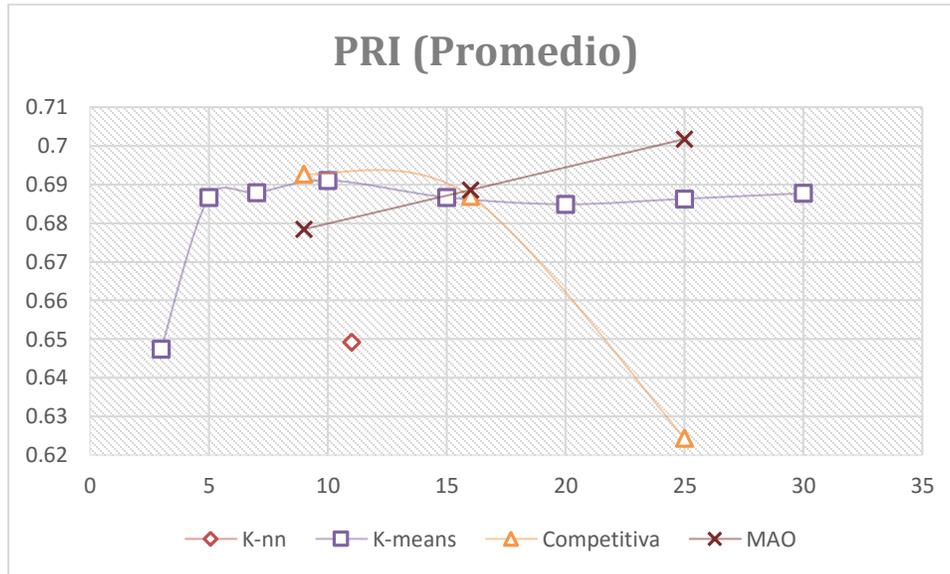


Figura 6-6 PRI (Promedio)

PRI (Promedio)		
	No. Colores	
K-nn	11	0.64927067
K-means	3	0.64738133
	5	0.686632
	7	0.687871
	10	0.69111633
	15	0.68669933
	20	0.68491233
	25	0.68632167
	30	0.68776467
Competitiva	9	0.69275333
	16	0.68695333
	25	0.62433
MAO	9	0.67849033
	16	0.68861167
	25	0.70177667

Tabla 6.4 PRI (Promedio)

El comportamiento del desempeño de los algoritmos es similar al observado cuando se consideró VOI, pero para este caso una calificación alta es mejor, por lo que k-means obtiene los mejores resultados cuando agrupa más colores al igual que cuando se utilizan los MAO, las redes competitivas tienen un comportamiento opuesto, empeorando al incrementar el número de neuronas. Finalmente, k-nn obtiene una calificación baja.

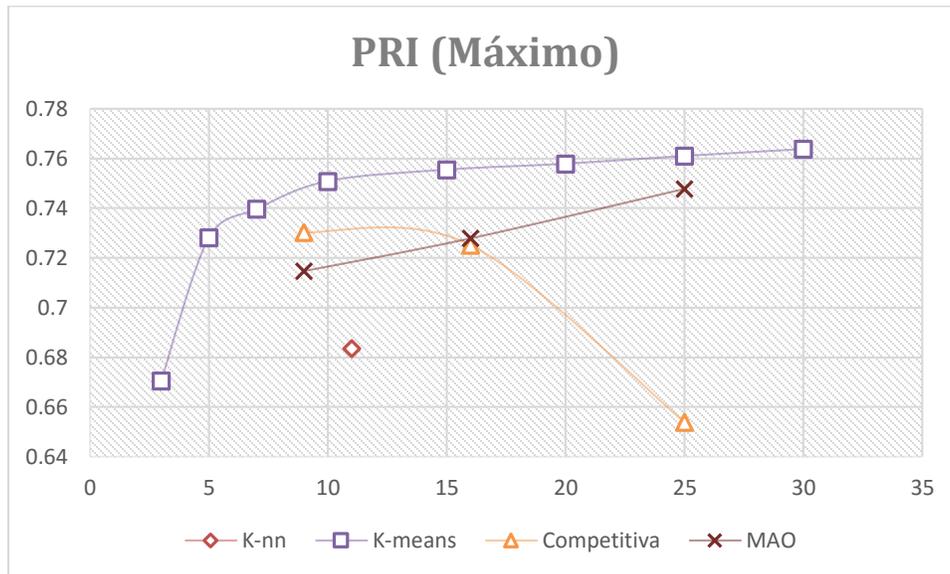


Figura 6-7 PRI (Máximo)

PRI (Máximo)		
	No. Colores	
K-nn	11	0.68343533
K-means	3	0.670491
	5	0.72806967
	7	0.73959433
	10	0.75086933
	15	0.75547733
	20	0.757853
	25	0.76099033
	30	0.76372167
Competitiva	9	0.73037667
	16	0.725257
	25	0.65389667
MAO	9	0.71461633
	16	0.727828
	25	0.74783933

Tabla 6.5 PRI (Máximo)

Tomando en cuenta los peores resultados las redes competitivas obtienen la mejor y la peor calificación, empeorando rápidamente conforme se incrementa el número de neuronas, k-means tiene un comportamiento similar a pesar de ser el que obtuvo los resultados máximos más altos en esta evaluación. K-nn y los MAO se comportan de igual forma que al considerar los valores promedio y máximos, donde los MAO mejoran su desempeño al identificar más colores.

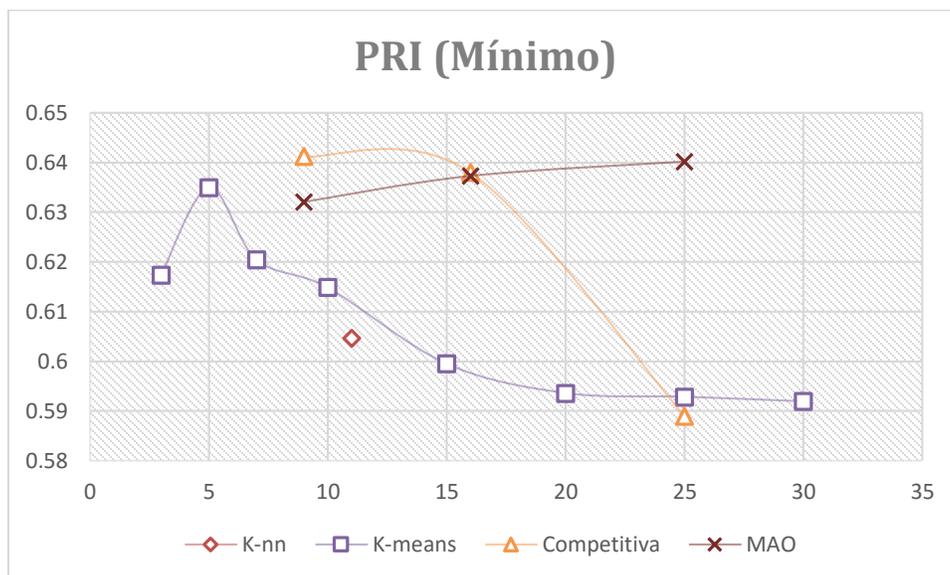


Figura 6-8 PRI (Mínimo)

PRI (Mínimo)		
	No. Colores	
K-nn	11	0.604688
K-means	3	0.617306
	5	0.63500733
	7	0.62049133
	10	0.61484733
	15	0.59947967
	20	0.59358367
	25	0.592836
	30	0.59190767
Competitiva	9	0.64120667
	16	0.63802967
	25	0.58894867
MAO	9	0.632052
	16	0.637268
	25	0.64014

Tabla 6.6 PRI (Mínimo)

6.1.3 GCE (Global Consistency Error)

Esta evaluación es sensible a los errores, por lo que una calificación baja es mejor que una alta. K-means obtiene buenos resultados, aunque rápidamente empeora para después mejorar de nuevo, k-nn tiene buenos resultados, quedando en segundo lugar. Las RNC mejoran al identificar más colores, contrario a los MAO que empeoran ligeramente con más neuronas.

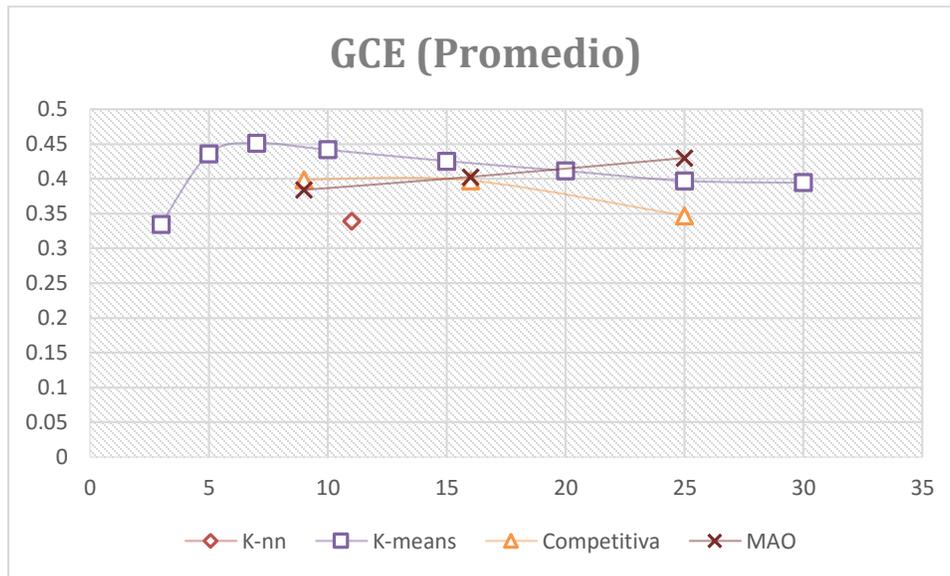


Figura 6-9 GCE (Promedio)

GCE (Promedio)		
	No. Colores	
K-nn	11	0.339218
K-means	3	0.33490733
	5	0.43577467
	7	0.451583
	10	0.441904
	15	0.42567633
	20	0.411665
	25	0.39728467
	30	0.39458767
Competitiva	9	0.39849367
	16	0.39730933
	25	0.34645933
MAO	9	0.38446633
	16	0.40280133
	25	0.42963067

Tabla 6.7 GCE (Promedio)

Las peores calificaciones corresponden a k-means aunque de nuevo es el de mejor desempeño cuando identifica sólo 3 colores, k-nn se mantiene en segunda posición, y el comportamiento de las redes neuronales es igual a cuando se considera el promedio.

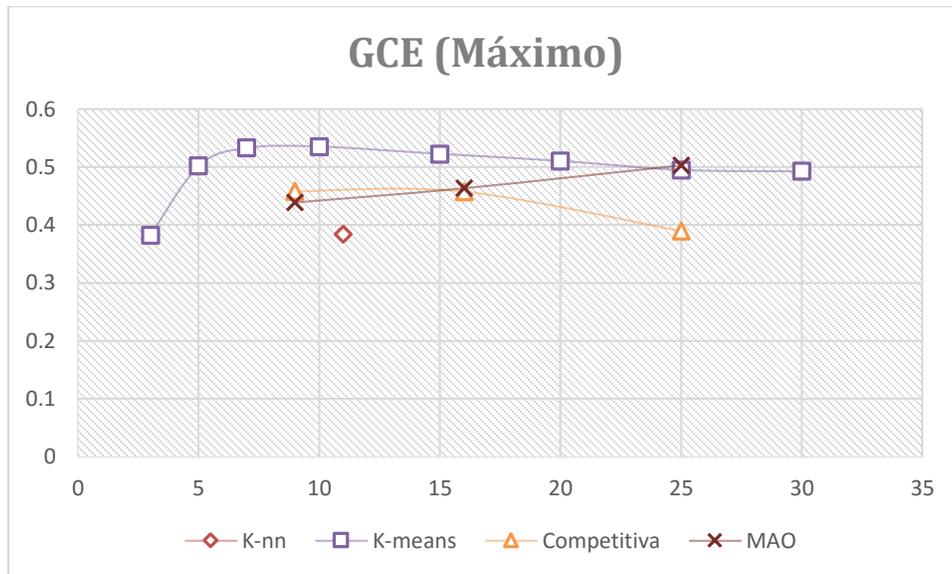


Figura 6-10 GCE (Máximo)

GCE (Máximo)		
	No. Colores	
K-nn	11	0.38338533
K-means	3	0.38201
	5	0.50162367
	7	0.53258533
	10	0.534932
	15	0.52234833
	20	0.51018733
	25	0.49466433
	30	0.49231867
Competitiva	9	0.45754467
	16	0.45717467
	25	0.38906467
MAO	9	0.43857967
	16	0.46325067
	25	0.50284767

Tabla 6.8 GCE (Máximo)

Por último, k-means continúa teniendo los mejores y los peores resultados, mientras k-nn continúa en segundo lugar, las RNC mejoran al contar con más neuronas, y los MAO empeoran ligeramente al identificar más colores.

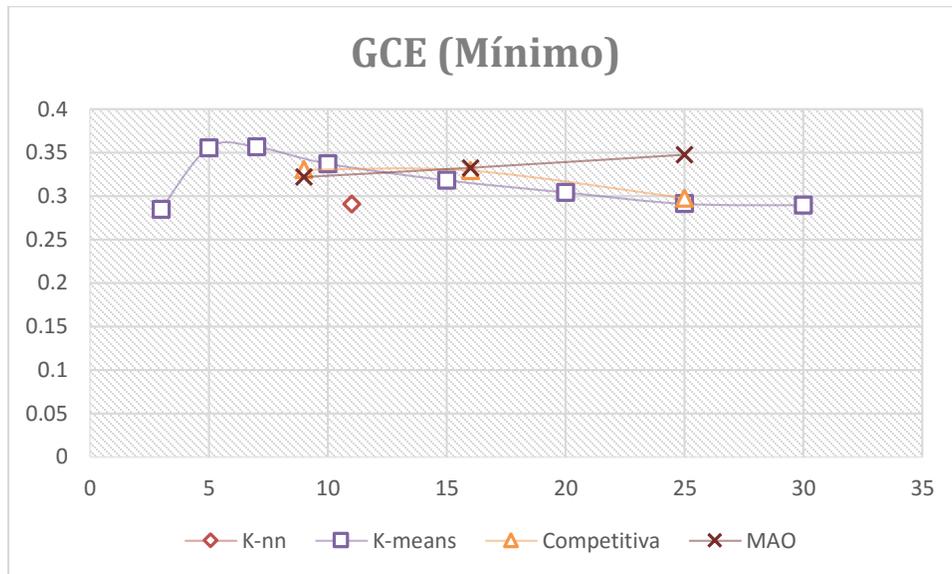


Figura 6-11 GCE (Mínimo)

GCE (Mínimo)		
	No. Colores	
K-nn	11	0.29070333
K-means	3	0.284832
	5	0.35547
	7	0.356361
	10	0.33709567
	15	0.31804633
	20	0.304159
	25	0.29108433
	30	0.28960333
Competitiva	9	0.33001967
	16	0.329085
	25	0.29714767
MAO	9	0.32145467
	16	0.33217633
	25	0.34723767

Tabla 6.9 GCE (Mínimo)

Como se puede observar en las tablas y gráficas, el uso de las redes neuronales supera a los demás algoritmos cuando se evalúa con PRI, y aunque no es la mejor con las demás evaluaciones los resultados no son malos, además, si se compara directamente con K-means.

6.2 Tamaño de la red neuronal

En este trabajo se ha presentado una propuesta para segmentar imágenes por características de color. Se presentan pruebas empleando redes neuronales competitivas y mapas auto-organizados de diferentes tamaños, que son entrenadas con muestras de cromaticidad de diferentes colores; posteriormente se procesan las imágenes extrayendo sólo la cromaticidad de los colores de las imágenes, al mapear previamente las imágenes al espacio HSV. Cada pixel es agrupado con el tono de la neurona ganadora de la red neuronal, finalmente la imagen resultante es mapeada al espacio RGB.

Los mapas auto-organizados mostraron, en cuanto a la apariencia de la imagen, tener mejor desempeño que las redes neuronales competitivas; es decir, aunque la cantidad y forma de las secciones obtenidas empleando ambas redes neuronales son parecidas, los tonos asignados por los mapas auto-organizados se asemejan más a los de las imágenes originales.

La cantidad de colores que pueden reconocer las redes neuronales depende de la cantidad de neuronas que tienen. Las redes neuronales con pocas neuronas funcionan mejor en imágenes que tienen pocos colores; mientras que las redes neuronales con un número grande de colores tienen mejor desempeño con las imágenes que tienen una mayor cantidad de colores.

Como se puede observar en los datos anteriores, al incrementarse el número de neuronas en la capa oculta de las RNC los resultados mejoran si se miden con VOI o con GCE, aunque empeoran al medirse con PRI. Los MAO tienen un comportamiento contrario, pues después de incrementarse el número de neuronas, los resultados mejoran con PRI y empeoran con VOI y GCE.

La diferencia que existe entre usar 9, 16 o 25 neuronas es más marcada en los resultados de las RNC que en los MAO, esto significa que el número de neuronas en la red afecta en mayor medida a las RNC que a los MAO, y que elegir un número de neuronas no adecuado de neuronas podría provocar una mala segmentación si se utiliza una RNC, aunque este problema afecta menos a los MAO.

Conclusiones

Como se puede apreciar en los resultados de la evaluación, la diferencia que existe entre los algoritmos más comunes y los propuestos es poca. Las RNC tienen buen desempeño, si se evalúa con VOI y GCE los resultados mejoran a medida que se incrementa el número de neuronas; mientras que los MAO no mejoran en estas evaluaciones, pero sí lo hacen con PRI, donde incluso superan a los otros algoritmos.

Los resultados de los MAO logran ser los mejores únicamente al ser evaluados con PRI, pero los resultados de las otras evaluaciones no son malos, así que, si se observan los resultados promedio, máximos y mínimos no existe una gran variación, lo cual indica que tiene un comportamiento más estable y confiable que los otros algoritmos. Por ejemplo, k-means obtuvo las mejores y las peores segmentaciones, haciendo sus resultados poco confiables.

Las RNC tienen un comportamiento distinto al de los MAO debido a la forma en la que se entrenan, que el número de neuronas se incremente no garantiza que el número de colores que identifica sea el mismo debido a que en las RNC es común que algunas neuronas no se activen, o se activen en muy pocos casos, mientras que los MAO esto es menos común debido a que las neuronas toman en consideración a las demás durante el entrenamiento.

Indicar el número de colores que identificará el algoritmo es un problema que existe en k-means. Utilizando una RNA existe un problema similar debido a que se debe indicar el número de neuronas a utilizar, pero a diferencia de k-means, al utilizar un MAO o una RNC se indica el número máximo de colores que podrá identificar y no se está forzado a segmentar las imágenes en ese número indicado; esta ventaja permite utilizar la misma red en varias escenas sin modificar ese parámetro para obtener buenos resultados (aunque modificarlo sí podría mejorarlos).

K-nn utilizado para identificar los 11 colores semánticos obtuvo muy buenos resultados debido a que su entrenamiento fue supervisado, pero para lograrlo requirió de al menos 10,000 muestras de color etiquetadas manualmente, además de que cada pixel de la imagen es comparado con un gran número de estas muestras, esto resulta en un entrenamiento muy tedioso y una ejecución lenta. Las RNA requirieron de un entrenamiento no supervisado con sólo 256 muestras, su ejecución requiere de multiplicaciones de matrices, que es proceso que requiere de menos operaciones que el número de comparaciones realizadas por k-nn, y además estas operaciones pueden realizarse en la GPU (Graphics Processor Unit o unidad de procesamiento gráfico) de forma paralela mejorando su rendimiento.

K-means no requiere de un aprendizaje previo a la ejecución, pero sí requiere de múltiples iteraciones para converger o llegar a un resultado aceptable. A pesar de que las RNA suelen requerir de más iteraciones para lograr una buena segmentación, estos ciclos se realizan durante el aprendizaje y no durante la ejecución del algoritmo, mejorando su velocidad (cabe destacar que se evaluó únicamente la eficacia del algoritmo y no la eficiencia, y que la velocidad de ejecución del algoritmo depende del hardware en el que sea implementado y los métodos de programación utilizados).

Que el espacio HSV tenga la información cromática está desacoplada de la saturación y el brillo, permite a los algoritmos ser menos sensibles a cambios de iluminación en comparación al espacio RGB. Durante el entrenamiento las neuronas están distribuidas aleatoriamente, pero después de varias iteraciones se llega a un ordenamiento más uniforme. Los colores semánticos no tienen una distribución tan uniforme, lo que provoca que haya colores que abarquen más tonalidades que otros y se puedan perder detalles importantes de imagen durante la segmentación, es por eso que se recomienda no utilizar el espacio de colores semánticos este fin a menos que el problema lo requiera. Dividir el espacio HSV de esta forma permite que se pueda aplicar a escenas en general.

Los resultados de la segmentación de imágenes dependerán del algoritmo que sea implementado y no es una tarea que se pueda aplicar de la misma forma a cualquier escena; inclusive cuando es realizada por seres humanos, los resultados dependerán de qué se requiere extraer de la imagen. Por ejemplo, en la base de imágenes utilizada se puede observar que cada individuo se enfocó en características distintas, de aquí que no exista un algoritmo capaz de segmentar perfectamente cualquier escena.

A pesar de lo anterior, los resultados obtenidos por nuestra propuesta son buenos y se demuestra que es posible implementar un algoritmo que con un solo entrenamiento sea capaz de segmentar cualquier imagen con resultados equiparables a otros algoritmos conocidos.

Trabajo Futuro

Aunque los resultados obtenidos fueron satisfactorios, se contempla combinar el algoritmo propuesto con k-means, de esta forma se espera que los resultados puedan mejorar.

Ya sea utilizando RNC o MAO, el algoritmo propuesto tiene como ventaja que no se necesita ser tan específico en el número de colores que la RN debe reconocer en cada imagen, si una imagen contiene menos colores que el número de neuronas en la red, simplemente no se activarán las neuronas correspondientes a los colores presentes, otros algoritmos de agrupamiento como k-means generalmente dividen la imagen en tantos conjuntos como les fue especificado aunque haya menos colores, resultando en una sobre-segmentación.

El número de neuronas activadas es igual al número de colores presentes en la imagen, este parámetro podría ser utilizado para volver a segmentar la imagen con k-means de forma más exacta. Aunque existirían casos donde una o más neuronas se active tan pocas veces que no sea útil tomarlas en cuenta, para resolver estos casos se necesitaría de un parámetro que regule a partir de cuántas veces una neurona se debe activar. Esta idea aún no ha sido probada y necesitará de ser evaluada de la misma forma que los algoritmos presentados en esta tesis.

Anexo

Además de esta tesis, se realizó un artículo que fue presentado en el 8° Congreso Mexicano de Inteligencia Artificial COMIA 2016, el cual se llevó a cabo en el Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), Puebla, México; y será publicado en la revista indexada Research in computing science en el volumen 144.



- ✔ Características cumplidas / cumpridas / Standards met: 19
- ✘ Características no cumplidas / Não cumpridas / Standards not met: 14

Título	Research in computing science
País	México
Situación	Vigente
Año de inicio	2000
Año de Terminación	9999
Frecuencia	Trimestral
Tipo de publicación	Publicación periódica
Soporte	Impreso en papel
Idioma	Inglés
ISSN	1665-9899
Temas	Ciencias Exactas y Naturales
Subtemas	Computación, ciencia y tecnología
Clasificación Dewey	005
Editorial	Instituto Politécnico Nacional, Centro de Investigación en Computación
Naturaleza de la publicación	Revista técnico-profesional
Naturaleza de la organización	Institución educativa
Notas	Fuente: V18,2006

• DATOS DE CONTACTO

Responsables	Juan Humberto Sossa (Edit. Resp.)
Calle y Número	Unidad Prof. Adolfo López Mateos, Av. Juan de Dios Batiz s/n casi esq. Miguel Othón de Mendizabal
Sector / Barrio / Colonia	Colonia Nueva Industrial Vallejo
Ciudad	México
Estado / Provincia / Departamento	Distrito Federal
Lugar	México, D.F.
Teléfono	(52-55) 5729-6000 ext. 56571
Sitio web de difusión	www.micai.org/racs/
Código Postal	07738

• ÍNDICES Y RESÚMENES

Latindex-Directorio
Periódica (Índice de Revistas Latinoamericanas en Ciencias)

• DISTRIBUCIÓN

Tiraje	500
Distribución (Vías)	Terrestre
Distribución Geográfica	Nacional

• DATOS DE REGISTRO

Folio	15377
Folio Único	12943
Centro de Acopio	México
Fecha de Alta	2006-05-04
Fecha de Modificación	2009-07-27

Segmentación de imágenes de color imitando la percepción humana del color

Miguel Contreras Murillo, Farid García Lamont, Alma Delia Cuevas Rasgado

Universidad Autónoma del Estado de México, Centro Universitario UAEM Texcoco,
Texcoco-Estado de México, México

miguelc297@gmail.com, fgarcial@uaemex.mx, almadeliacuevas@gmail.com

Resumen. Normalmente, en los trabajos de segmentación de imágenes de color entrenan redes neuronales con los colores de la imagen a segmentar, después se obtiene el número de colores dominantes dentro de la imagen para posteriormente procesar la imagen con fuzzy c-means, en donde los colores son representados en el espacio RGB. La desventaja que tienen estos trabajos es que, por un lado, deben entrenar las redes neuronales cada vez que se procesa una nueva imagen; por otro lado, en el espacio RGB la cromaticidad de un color puede ser modificada por su intensidad. En este trabajo proponemos segmentar las imágenes con información cromática de los colores, entrenando una red neuronal con muestras de cromaticidad de diferentes colores, que puede emplearse para segmentar cualquier imagen sin necesidad de volverla a entrenar; la cantidad de colores que reconoce la red neuronal depende de su tamaño. Se presentan experimentos con imágenes de la base de segmentación de Berkeley empleando redes neuronales competitivas y mapas auto-organizados.

Palabras clave: Segmentación, redes neuronales artificiales, espacios de color.

Color image segmentation by mimicking the human perception of color

Abstract. Usually, related works on color image segmentation train neural networks with the colors of the image to segment, then the number of dominant colors within the image is obtained in order to process the image using fuzzy c-means, where the colors are represented in the RGB space. The drawback with these methods is the neural networks must be trained every time a new image is given; but also, in the RGB space the color's chromaticity can be altered by its intensity. In this paper we propose to segment the images using chromatic data of colors, by training a neural network with chromaticity samples of different colors, which can be employed to segment any image just training it just once; the number of colors the neural network recognizes depends on its size. We show experiments with images of the Berkeley segmentation database using competitive neural networks and self-organizing maps.

Keywords: Segmentation, artificial neural networks, color spaces.

1. Introducción

La segmentación de imágenes es un tema ampliamente estudiado para la extracción y reconocimiento de objetos, de acuerdo a las características de textura, color, forma, entre otros. Dependiendo de la naturaleza del problema, las características de color de los objetos pueden proporcionar información relevante sobre ellos. Por ejemplo, la segmentación de imágenes de color ha sido aplicado en diferentes áreas como análisis de alimentos [1,2], geología [3], medicina [4,5] entre otras [6-9].

Los trabajos que abordan la segmentación de imágenes por características de color emplean diferentes técnicas [10,11], pero las más empleadas son las redes neuronales (RN) [12-14] y métodos basado en agrupamiento, específicamente, fuzzy c-means (FCM) [15-20]. Las RN son entrenadas para reconocer colores específicos, es decir, estas son entrenadas con los colores de la imagen a ser segmentada. Si se da una nueva imagen la RN debe ser entrenada nuevamente. Al emplear métodos basados en agrupamiento, se crean grupos de colores con características similares. La desventaja con tales métodos es que se requiere definir previamente la cantidad de grupos en que se divide la información; por lo tanto, el número de grupos se define dependiendo de la naturaleza de la escena.

Nuestra propuesta consiste en entrenar a la RN para reconocer diferentes colores, tratando de emular la percepción humana del color. Los seres humanos identifican principalmente los colores por su cromaticidad, después por su intensidad [21]. Por ejemplo, si se le pregunta a cualquier persona cual es el color de los cuadros (a) y (b) de la Fig. 1, lo más seguro es que responderá “verde”; nótese que el cuadro (a) es más brillante que el cuadro (b) pero la cromaticidad no cambia. Ahora, si se le vuelve a preguntar a esa misma persona cual es el color de los cuadros (c) y (d) de la Fig. 1, lo más seguro es que responda “rojo y rosa, respectivamente”; es importante mencionar que los cuadros (c) y (d) tienen la misma intensidad pero diferentes cromaticidades.

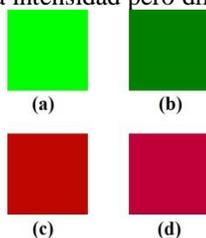


Fig. 1 Cuadros (a) y (b) con la misma cromaticidad pero con diferentes intensidades; cuadros (c) y (d) con diferentes cromaticidades pero con la misma intensidad

Los seres humanos tienen la capacidad innata de reconocer colores; al observar su entorno pueden reconocer, hasta cierto punto, regiones y/o objetos dentro de una escena por sus características cromáticas. Es importante mencionar que los humanos no necesitan aprender a identificar los colores cada vez que se les muestra una escena; ellos solo emplean su conocimiento previamente adquirido.

La contribución de este artículo es una propuesta para segmentar imágenes de color por características cromáticas, emulando la forma en que los seres humanos reconocen los colores. Con el fin de emular esta capacidad humana, proponemos entrenar un mapa auto-organizado (MAO), con muestras de cromaticidad de diferentes colores, una vez entrenada, el MAO procesa la imagen. En donde dependiendo de la cantidad de neuronas que tenga la RN es la cantidad de colores que puede reconocer la RN y en consecuencia el número de secciones que puede tener la imagen.

En la mayoría de los trabajos relacionados se emplea el espacio RGB para representar colores; sin embargo, este espacio es sensible a la iluminación por lo que la extracción de la cromaticidad de los colores no es precisa porque esta puede ser alterada por los cambios de intensidad. De aquí que, nosotros empleamos el espacio de color HSV, porque en este espacio la cromaticidad es separada de la intensidad [22].

El artículo está organizado de la siguiente forma: en la sección 2 se muestran las características de los espacios de color RGB y HSV. Presentamos nuestra propuesta para la segmentación de imágenes en la sección 3. En la sección 4 se muestran los experimentos realizados y se discuten los resultados obtenidos. Finalmente, el artículo termina con las conclusiones y trabajo futuro en la sección 5.

2. Espacios de color

Aunque el espacio RGB es ampliamente aceptado para representar colores por la comunidad de procesamiento de imágenes, los seres humanos no perciben el color como es representado en dicho espacio. La percepción humana del color es similar a la representación en el espacio HSV [21,22], de aquí que empleamos este espacio. En las secciones 2.1 y 0 se presentan las características de cada espacio.

2.1 Espacio de color RGB

El espacio RGB está basado en el sistema de coordenadas Cartesiano en donde los colores son puntos definidos por vectores que se extienden desde el origen, en donde el negro está en el origen y el blanco está ubicado en la esquina opuesta al origen [22], ver Figura 3-1.

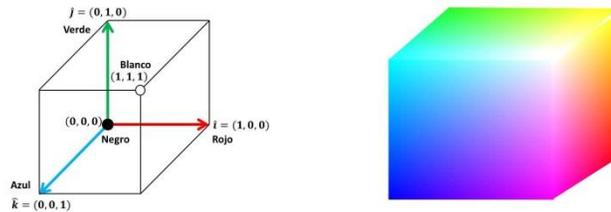


Fig. 2 Espacio de color RGB

El color de un pixel p se escribe como una combinación lineal de los vectores base de verde, rojo y azul [22]:

$$\phi_p = r_p \hat{i} + g_p \hat{j} + b_p \hat{k} \quad (1)$$

En donde r_p , g_p y b_p son los componentes de rojo, verde y azul, respectivamente. La orientación y magnitud de un vector de color define la cromaticidad e intensidad del color, respectivamente [22]. Como se ha mencionado antes, este espacio es sensible a la iluminación; es decir, a pesar de que dos vectores tengan la misma cromaticidad, estos representan diferentes colores si sus intensidades son diferentes.

2.2 Espacio de color HSV

La representación de color en el espacio HSV emula la percepción humana del color ya que la cromaticidad es desacoplada de la intensidad [21,22]. En este espacio el color de un pixel p se representa por sus componentes de tono (h), saturación (s) e intensidad (v):

$$\varphi_p = [h_p, s_p, v_p] \quad (2)$$

El tono es la cromaticidad, la saturación es el nivel de blancura del color y la intensidad es el brillo del color; la Figura 3-3 **Espacio de color HSV** muestra la apariencia del espacio HSV. Los rangos de valores reales del tono, saturación e intensidad son $[0,2\pi]$, $[0,1]$ y $[0,255]$, respectivamente.

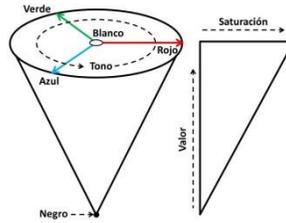


Fig. 3 Espacio de color HSV

3. Propuesta de segmentación

En esta sección presentamos nuestra propuesta para la segmentación de imágenes de color, en donde entrenamos una RN con muestras de cromaticidad de diferentes colores. Posteriormente se extrae la cromaticidad de cada pixel de la imagen a segmentar al mapear el color al espacio HSV, la cromaticidad extraída es procesada por la RN y el nuevo tono del pixel es el tono de la neurona ganadora. Cuando todos los pixeles son procesados, la imagen es mapeada al espacio RGB.

3.1 Entrenamiento de la red neuronal

Debido a la naturaleza difusa del color, no es posible reconocer todos los colores del espectro; de aquí que, el espectro de color es “dividido” en un conjunto finito de colores. El número de colores que la RN puede reconocer depende de su tamaño; en este artículo se realizan pruebas con una red neuronal competitiva (RNC) con 9, 16 y 25 neuronas, y con MAO de 3×3 , 4×4 y 5×5 neuronas. Las RN son entrenadas con los elementos del conjunto Ψ construido con muestras de cromaticidad como sigue:

$$\Psi = \left\{ \psi_k = [\cos \theta_k, \sin \theta_k] \mid \theta_k = \frac{2\pi}{256} k : k = 0, 1, \dots, 255 \right\} \quad (3)$$

La cromaticidad es transformada en un vector debido al caso cuando el valor del tono es casi 0 o 2π . Considérese los cuadros (c) y (d) de la Fig. 1, sus valores son $\pi/100$ y $19\pi/10$, respectivamente. Numéricamente ambos valores son muy distintos pero las cromaticidades de ambos cuadros son muy similares; si la cromaticidad de ambos cuadros es clasificada solamente por el valor escalar del tono, la cromaticidad es reconocida como si fueran muy diferentes.

Este problema se resuelve como sigue; sea φ_p el color de un pixel representado en el espacio HSV como se muestra en la ec. (2), la cromaticidad es modelada como:

$$\psi_p = [\cos h_p, \sin h_p] \quad (4)$$

3.2 Procesamiento de la imagen

La segmentación de la imagen se hace al agrupar los colores empleando la cromaticidad de los colores de cada pixel de la imagen. Es importante mencionar que las RNs son entrenadas con información de la cromaticidad de los colores, por lo que no pueden reconocer el negro ni el blanco porque estos dos colores no tienen una cromaticidad definida. El blanco se obtiene cuando la saturación de un color es bajo, es decir, cuando $s \approx 0$; por otra parte, el negro se obtiene cuando la intensidad del color es baja, esto es, cuando $v \approx 0$.

Por lo tanto, antes de que un color sea procesado por la RN se debe evaluar su saturación e intensidad para clasificarlo como blanco o negro, respectivamente. Procesar el color de un pixel conlleva realizar los siguientes pasos. Sea el vector de color ϕ_p del pixel p representado en el espacio RGB:

1. El vector ϕ_p se mapea al espacio HSV obteniendo $\varphi_p = [h_p, s_p, v_p]$.
2. Se verifica si el color del pixel es negro; si $v_p \leq \delta_v$ entonces $v_p^* = 0$ y $s_p^* = 0$, ir al paso 5.
3. En caso contrario, se verifica si el color del pixel es blanco; si $s_p \leq \delta_s$ entonces $v_p^* = 191$ y $s_p^* = 0$, ir al paso 5.
4. En caso contrario, es decir, si $v_p > \delta_v$ y $s_p > \delta_s$ entonces:
 - a. Calcular el vector ψ_p y procesarlo con la RN.
 - b. Se obtiene el vector de peso de la neurona ganadora $\mathbf{w}_i = [w_{i,1}, w_{i,2}]$ y se etiqueta al pixel con el número i .
 - c. Calcular el tono con $h_p^* = \tan^{-1}(w_{i,2}/w_{i,1})$.
 - d. Se asignan los valores de saturación e intensidad: $v_p^* = 191$ y $s_p^* = 1$.
5. El nuevo vector $\varphi_p^* = [h_p^*, s_p^*, v_p^*]$ es mapeado al espacio RGB obteniendo el vector $\phi_p^* = [r_p^*, g_p^*, b_p^*]$.

En donde δ_s y δ_v son los umbrales para saturación e intensidad, respectivamente. Dada la naturaleza difusa del color, no hay valores específicos para decidir exactamente cuando un color es blanco o negro; de forma experimental encontramos que los mejores umbrales son $\delta_s = \mu_s - \sigma_s$ y $\delta_v = \mu_v - \sigma_v$; en donde μ_s y μ_v son la media de saturación e intensidad de la imagen, respectivamente; σ_s y σ_v son la desviación estándar de la saturación e intensidad de la imagen, respectivamente.

4. Experimentos y discusión

Recientemente la base de segmentación de Berkeley¹ (BSB) se está convirtiendo en la referencia para probar algoritmos de segmentación de imágenes de color [16]. Para los experimentos, implementados en Matlab 2014a, se seleccionó aleatoriamente un conjunto de 9 imágenes de las 300 imágenes que contiene la BSB, ver .

¹
<http://www.eecs.berkeley.edu/Research/Projects/CS/vision/bds/>

En la se muestran las imágenes obtenidas al procesar las imágenes de la empleando las RNCs con los diferentes tamaños que se indican. A su vez en la se muestran las imágenes resultantes al ser procesadas las imágenes de la empleando los MAOs con los tamaños que se indican.

Se puede observar fácilmente de las imágenes resultantes que estas pueden ser segmentadas solamente utilizando información de la cromaticidad; pero la segmentación también depende de la cantidad de neuronas de las RNs. Esto es, entre más grande sea la RN, mayor la cantidad de colores son reconocidos; de hecho, se puede apreciar que con los MAO se reconocen más colores que con las RNC, ya que se pueden observar más secciones o dentro de la imagen empleando los MAO. Aunque también, por lo mismo, hay algunas partes de la imagen que no son segmentadas homogéneamente.

Al observar las imágenes se puede ver que estas tienen mejor segmentación empleando los MAOs. Por ejemplo, en las imágenes obtenidas al procesar la imagen 35070 con las RNCs, el fondo se combina con la hoja; mientras que la misma imagen procesada por los MAOs se puede ver que la hoja es segmentada del fondo, a pesar de tener ambos tonos verdes pero los MAOs son capaces de distinguir la diferencia de tonos.

Otro ejemplo, en las imágenes obtenidas de la imagen 35010 utilizando las RNCs, prácticamente los fondos de las hojas son segmentados en verde, mientras que con los MAOs se pueden apreciar tonos en amarillo en los centros de las hojas. Las alas de la mariposa son segmentadas exitosamente con todas las redes neuronales, excepto utilizando la RNC de 25 neuronas, las cuales son segmentadas con el mismo tono de verde de las hojas del fondo. Una posible explicación es que esa red neuronal no “aprendió” correctamente a reconocer el tono amarillo durante el entrenamiento, por lo que la red neuronal debe ser entrenada nuevamente.

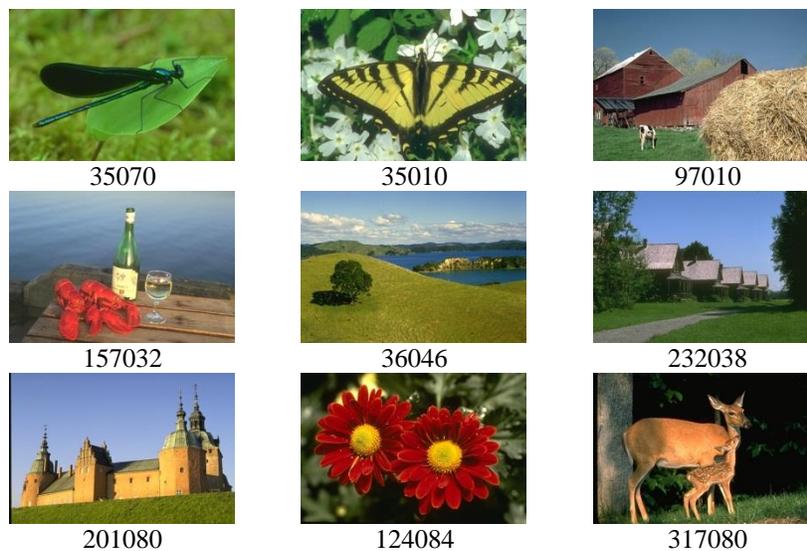


Fig. 4 Imágenes extraídas de la BSB, empleadas para los experimentos realizados



Fig. 5 Imágenes obtenidas empleando redes neuronales competitivas de 9, 16 y 25 neuronas

Dada la apariencia de las imágenes, la segmentación de las imágenes utilizando las RNCs es mejor cuando la red neuronal es pequeña, mientras que con los MAOs es lo contrario; es decir, el resultado de la segmentación de las imágenes utilizando los MAOs es mejor cuando la red neuronal es grande. Por ejemplo, la imagen 124084 obtenida con la RNC de 9 neuronas es muy parecida a la obtenida utilizando el MAO de 5×5 neuronas. Ocurre de forma similar con las imágenes 97010, 157032, 201080 y 35010 si se utiliza un la RNC de 9 neuronas y el MAO de 5×5 neuronas.

5. Conclusiones y trabajo futuro



Fig. 6 Imágenes obtenidas empleando mapas auto-organizados de 3×3, 4×4 y 5×5 neuronas

En este trabajo se ha presentado una propuesta para segmentar imágenes por características de color. Se presentan pruebas empleando redes neuronales competitivas y mapas auto-organizados de diferentes tamaños, que son entrenadas con muestras de cromaticidad de diferentes colores; posteriormente se procesan las imágenes extrayendo sólo la cromaticidad de los colores de las imágenes, al mapear previamente las imágenes al espacio HSV. Cada pixel es agrupado con el tono de la neurona ganadora de la red neuronal, finalmente la imagen resultante es mapeada al espacio RGB.

Los mapas auto-organizados mostraron, en cuanto a la apariencia de la imagen, tener mejor desempeño que las redes neuronales competitivas; es decir, aunque la cantidad y forma de las secciones obtenidas empleando ambas redes neuronales son parecidas, los tonos asignados por los mapas auto-organizados se asemejan más a los de las imágenes originales.

La cantidad de colores que pueden reconocer las redes neuronales depende de la cantidad de neuronas que tienen. Las redes neuronales con pocas neuronas funcionan mejor en imágenes que tienen pocos colores; mientras que las redes neuronales con un número grande de colores tienen mejor desempeño con las imágenes que tienen varios colores.

Como trabajo futuro se contempla hacer una evaluación cuantitativa de la segmentación de las imágenes obtenidas con nuestra propuesta. Empleando las métricas de índice aleatorio probabilístico y de variación de información, que se están volviendo últimamente en las métricas estándar para medir el desempeño de los algoritmos de segmentación de color [16]. La imagen segmentada con nuestra propuesta es comparada con las imágenes segmentadas a mano que se encuentran en la BSB que sirven como referencia. Cada imagen de la BSB tiene un conjunto de 5 imágenes segmentadas a mano, con las que se hace la comparación. Por ejemplo, en la se muestra las imágenes segmentadas a mano de la imagen 35010 de la BSB.



Fig. 7 Ejemplo de imágenes segmentadas a mano de la BSB

Las imágenes segmentadas obtenidas con nuestra propuesta son comparadas con cada una de las imágenes segmentadas a mano de la BSB, en donde la similitud de la segmentación se mide con las métricas mencionadas anteriormente. En la Fig. 8 se muestran dos ejemplos de imágenes segmentadas a mano, (a) y (c), y dos que se obtienen con nuestra propuesta, (b) y (d).

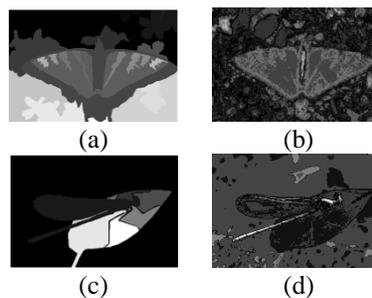


Fig. 8 Ejemplo de comparación de las imágenes: (a) y (c), segmentadas a mano y las segmentadas con nuestra propuesta, (b) y (d)

Por otra parte, se contempla hacer pruebas con imágenes con ruido para observar el comportamiento de nuestra propuesta.

Agradecimientos

El primer autor agradece al CONACyT la beca para realizar estudios de maestría, con el número de registro 634201.

Referencias

1. Gökmen, V., Sügüt, I.: A non-computer vision based analysis of color in foods. Int. J. Food Eng., 3(5), article 5 (2007).

2. Lopez, J.J., Cobos, M., Aguilera, E.: Computer-based detection and classification of flaws in citrus fruits. *Neural Comput. Appl.*, 20(7), 975-981 (2011).
3. Lespitö, L., Kuntuu, I., Visa, A.: Rock image classification using color features in Gabor space. *J. Electron. Imaging*, 14(4), 1-3 (2005).
4. Ghoeneim, D.M.: Optimizing automated characterization of liver fibrosis histological images by investigating color spaces at different resolutions. *Theor. Biol. Med. Model.*, 8, 25 (2011).
5. Harrabi, R., Braiek, E.B.: Color image segmentation using multi-level thresholding approach and data fusion techniques: application in the breast cancer cells images. *EURASIP J. Image Video Process.*, 11, (2012).
6. Wang, G., Man, L., Wang, B., Xiao, Y., Pan, W., Lu, X.: Fuzzy-based algorithm for color recognition of license plates. *Pattern Recognit. Lett.*, 29(7), 1007-1020 (2008).
7. del Fresno, M., Macchi, A., Marti, Z., Dick, A., Clausse, A.: Application of color image segmentation to estrus detection. *J. Vis.*, 9(2), 171-178 (2006).
8. Rotaru, C., Graf, T., Zhang, J.: Color image segmentation in HSI space for automotive applications. *J. Real-Time Image Process.*, 3(4), 311-322 (2008).
9. Bianconi, F., Fernandez, A., Gonzalez, E., Saetta, S.A.: Performance analysis of color descriptors for parquet sorting. *Expert Syst. Appl.*, 40(5), 1636-1644 (2013).
10. Aghbarii, Z.A., Haj, R.A.: Hill-manipulation: an effective algorithm for color image segmentation. *Image Vis. Comput.*, 24(8), 894-903 (2006).
11. Mignotte, M.: A non-stationary MRF model for image segmentation from a soft boundary map. *Pattern Anal. Appl.*, 17(1), 129-139 (2014).
12. Mousavi, B.S., Soleymani, F., Razmjoo, N.: color image segmentation using neuro-fuzzy system in a novel optimized color space. *Neural Comput. Appl.*, 23(5), 1513-1520 (2013).
13. Ong, S., Yeo, N., Lee, K., Venkatesh, Y., Cao, D.: Segmentation of color images using a two-stage self-organizing network. *Image Vis. Comput.*, 20(4), 279-289 (2002).
14. Jiang, Y., Zhou, Z.H.: SOM ensemble-based image segmentation. *Neural Process. Lett.*, 20(3), 171-178 (2004).
15. Wang, L., Dong, M.: Multi-level low-rank approximation-based spectral clustering for image segmentation. *Pattern Recognit. Lett.*, 33(16), 2206-2215 (2012).
16. Mújica-Vargas, D., Gallegos-Funes, F.J., Rosales-Silva, A.J.: A fuzzy clustering algorithm with spatial robust estimation constraint for noisy color image segmentation. *Pattern Recognit. Lett.*, 34(4), 400-413 (2013).
17. Huang, R., Sang, N., Luo, D., Tang, Q.: Image segmentation via coherent clustering in $L^a * L^b$ color space. *Pattern Recognit. Lett.*, 32(7), 891-902 (2011).
18. Nadernejad, E., Sharifzadeh, S.: A new method for image segmentation based on fuzzy c-means algorithm on pixonal images formed by bilateral filtering. *Signal Image Video Process.*, 7(5), 855-863 (2013).
19. Guo, Y., Sengur, A.: A novel color image segmentation approach based on neutrosophic set and modified fuzzy c-means. *Circuits Syst. Signal Process.*, 32(4), 1699-1723 (2013).
20. Kim, J.Y.: Segmentation of lip region in color images by fuzzy clustering. *Int. J. Control Autom. Sys.*, 12(3):652-661 (2014).
21. Ito S., Yoshioka, M., Omatu, S., Kita, K., Kugo, K.: An image segmentation method using histograms and the human characteristics of HSI color space for a scene image. *Artif. Life and Robot.*, 10(1), pp. 6-10 (2006).
22. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*, 2nd edn. Prentice Hall (2002).

Bibliografía

- [1] R. A.-H. Zaher Al Aghbari, «Hill-manipulation: An effective algorithm for color image segmentation,» *Elsevier*, pp. 894-903, 2006.
- [2] E. Navon, O. Miller y A. Averbuch, «Color image segmentation based on adaptive local thresholds,» *Image and Vision Computing*, nº 23, pp. 69-85, 2005.
- [3] H.-Y. Yang, X.-Y. Wang, X.-Y. Zhang y J. Bu, «Color texture segmentation based on image pixel classification,» *Engineering Applications of Artificial Intelligence*, nº 25, pp. 1656-1669, 2012.
- [4] C.-C. Kang, W.-J. Wang y C.-H. Kang, «Image segmentation with complicated background by using seeded region,» *Int. J. Electron. Commun.*, nº 66, pp. 767-771, 2012.
- [5] X.-Y. W. X.-Y. Z. J. B. Hong-Ying Yang, «Color texture segmentation based on image pixel classification,» *Elsevier*, pp. 1656-1669, 2012.
- [6] M. D. Lijun Wang, «Multi-level Low-rank Aproximation-based Spectral Clustering for image segmentation,» *Elsivier*, pp. 2206-2215, 2012.
- [7] G. Pajares y J. de la Cruz, *Visión por computador, imágenes digitales y aplicaciones*, México: Alfaomega, 2008.
- [8] W.-J. W. C.-H. K. Chung-Chia Kanga, «Image segmentation with complicated background by using seeded region,» *Elsevier*, pp. 767-771, 2012.
- [9] . W. Pascal, L. Michael, B. Marc, B. Tanya, D. Adam y H. Nicholas, *An Introduction to Scientific Computing in MATLAB*, Academic Press, 2014.
- [10] . G. Erkam y K. Gulgun, «Definition of artificial neural networks with comparison to other,» *Procedia Computer Science*, p. 426–433, 2011 .
- [11] N. Y. K. L. Y. V. D. C. S.H. Ong, «Segmentation of color images using a tow-stage self-organizing network,» *Elsevier*, pp. 279-289, 2002.
- [12] R. Huang, N. Sang, D. Luo y T. Qiling, «Image segmentation via coherent clustering in L/a/b/ color space,» *Pattern Recognition Letters*, pp. 891-902, 2011.
- [13] C. Pérez Vega, *VISION, LUZ Y COLOR*, UNIVERSIDAD DE CANTABRIA, Dpto. de Ingeniería de Comunicaciones, 2006.

- [14] J. Cortés Parejo, *La Percepción del Color*, 2000.
- [15] R. C. González y R. E. Woods, *Digital Image Processing*, 2nd ed., Prentice Hall, 2002.
- [16] S. Ito, M. Yoshioka, S. Omatu, K. Kita y K. Kugo, «An image segmentation method using histograms and the human characteristics of HSI color space for a scene image.,» *Artif. Life and Robot.*, pp. 6-10, 2006.
- [17] V. F. Ray, «Techniques and problems in the study of human color perception.,» *Southwestern Journal of Anthropology.*, nº 8, pp. 251-259, 1952.
- [18] R. Benavente, J. Van de Weijer., M. Vanrell, C. Schmid, R. Baldrich, J. Verbeek y D. Larlus, «Color Names,» *The Wiley-IS&T Series in Imaging Science and Technology*, 2011.
- [19] A. Maerz y M. Paul., *A dictionary of color*, 1st. ed., McGraw-Hill, 1930.
- [20] B. Berlin y P. Kay, *Basic color terms: their universality and evolution.*, Berkeley: University of California, 1969.
- [21] P. Kay, «Universality of Color Categorization».
- [22] A. J. Elliot, M. D. Fairchild y A. Franklin, *Handbook of Color Psychology (Cambridge Handbooks in Psychology)*, Cambridge: Cambridge University Press, 2016.
- [23] «Proto-Indo-European_language,» [En línea]. Available: https://en.wikipedia.org/wiki/Proto-Indo-European_language.
- [24] «The American Heritage Dictionary,» [En línea]. Available: <https://www.ahdictionary.com/word/indoeurop.html>.
- [25] «bhel- / Indo-European roots,» [En línea]. Available: http://www.thefreedictionary.com/_/roots.aspx?type=Indo-European&root=bhel-.
- [26] «Black,» [En línea]. Available: <http://www.etymonline.com/index.php?term=black> .
- [27] «Black,» [En línea]. Available: <https://en.wiktionary.org/wiki/black>.
- [28] «Bleu,» [En línea]. Available: <https://en.wiktionary.org/wiki/bleu>.
- [29] L. R. Ronchi y S. Rizzo, *La ricerca di avanguardia vista dall'AIC nel terzo millennio*, 2004.
- [30] H. Moreira Villegas y J. Lillo Jover, *Percepción del color y daltonismos*, Perámide, 2013.

- [31] R. Benavente, M. Vanrell y R. Baldrich, «A Data Set for Fuzzy Colour Naming,» *COLOR research and application*, pp. 48-56, 2006.
- [32] N. Otsu, «A Threshold Selection Method from Gray-Level Histograms,» *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, nº 1, pp. 62-66, 1979.
- [33] . H. J. Vala y A. Baxi , «A Review on Otsu Image Segmentation,» *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 2, nº 2, 2013.
- [34] A. Moujahid, I. Inza y P. Larrañaga, «Clasificadores K-NN,» *Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad del País Vasco–Euskal Herriko Unibertsitatea*.
- [35] H. Sébastien, C. Sylvain y P. Robert, «Are unsupervised neural networks ignorant? Sizing the effect of environmental distributions on unsupervised learning,» *Cognitive Systems Research*, p. 357–371, 2006.
- [36] D. Mújica-Vargas, F. Gallegos-Funes y A. Rosales-Silva, «A fuzzy clustering algorithm with spatial robust estimation constraint,» *Pattern Recognition Letters, Elsevier*, pp. 400-413, 2013.
- [37] V. C. Manisha Sharma, «Objective Evaluation Parameters of Image Segmentation Algorithms,» *International Journal of Engineering and Advanced Technology (IJEAT)*, 2012.
- [38] B. Sathya y R. Manavalan, «Image Segmentation by Clustering Methods: Performance Analysis,» *International Journal of Computer Applications*, 2011.
- [39] G. Paschos, «Perceptually Uniform Color Spaces for Color Texture Analysis: An Empirical Evaluation,» *IEEE Transactions on Image Processing vol.10 No. 6*, 2001.
- [40] H. Zhang, J. Fritts y S. Goldman , « Image segmentation evaluation: A survey of unsupervised methods.,» *Comput. Vis. Image Underst.*, nº 110, pp. 260-280, 2008.