

**UNIVERSIDAD POLITÉCNICA
SALESIANA**

SEDE CUENCA

**CARRERA DE INGENIERÍA
DE SISTEMAS**

TEMA:

“ESQUEMAS DE VOTACIÓN BORDA
APLICADOS A LA CLASIFICACIÓN DE
IMÁGENES UTILIZANDO LOS HISTO-
GRAMAS DE COLOR RGB, HSV Y EL
DESCRIPTOR DE DISTRIBUCIÓN DE CO-
LOR MPEG-7”

Tesis previa a la obtención del Título de:
Ingeniero de Sistemas

AUTOR:

Javier Ramiro Poveda Figueroa

DIRECTOR:

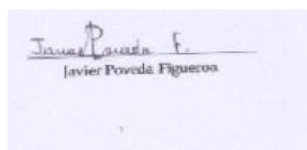
Ing. Vladimir Robles.

Cuenca, Enero del 2013

DECLARATORIA DE RESPONSABILIDAD:

Yo, Javier Ramiro Poveda Figueroa portador de la cédula de ciudadanía 0105839542, estudiante de Ingeniería de Sistemas , certifico que los conceptos desarrollados, análisis realizados, así como los criterios vertidos en la totalidad del presente trabajo, son de mi exclusiva responsabilidad, y autorizo a la Universidad Politécnica Salesiana el uso de la misma con fines académicos.

Cuenca, enero del 2013



Javier Poveda F.
Javier Poveda Figueroa

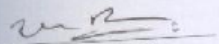
CERTIFICACIÓN:

Ing. Vladimir Robles.

Certifica:

Haber dirigido y revisado prolijamente cada uno de los capítulos del informe de monografía realizada por el Señor Javier Ramiro Poveda Figueroa.

Cuenca, enero del 2013



Ing. Vladimir Robles

DEDICATORIA

" Las ideas dan forma a la sociedad "

John Maynard Keynes. Economista Británico.

Este trabajo de investigación esta dedicado a muchas personas a quienes he considerado que han formado mi espíritu.

A mis abuelos Luis Poveda Orellana y Enrique Figueroa Polo, a mi Papá Marcial Poveda García y a mi bisabuelo materno Luis Figueroa Espejo, quienes me han inculcado esta facinación por la investigación .

A mis abuelas Raquel García Heras y Enriqueta Mendieta Riofrio, a mi bisabuela materna Amelia Polo Palacio y a mi madre Nancy Figueroa Mendieta quienes me han enseñado la importancia del amor y el coraje para ser una persona de bien.

Finalmente, quiero dedicar esta tesis a todos los profesores que iluminan el día a día el camino largo y fascinante de la vida.

AGRADECIMIENTOS

Quisiera agradecer en primer lugar a mi familia que me dio un gran apoyo al redactar el presente trabajo de investigación con sus consejos e ideas.

A todos mis amigos que me supieron dar apoyo incondicional durante todo este tiempo y en mi vida en general. Bernardo, Mauricio, Andrea y a los demás les agradezco muchas por su apoyo.

A mis compañeros del centro de investigación que siempre estuvieron presentes brindándome su apoyo ya sea directa o indirectamente.

A mi director de tesis y gran amigo Vladimir Robles quien me supo guiar con sus consejos, e ideas tanto a nivel académico como a nivel social. De verdad muchas gracias.

PUBLICACIONES

Algunas ideas, imágenes y tablas han aparecido previamente en las siguientes publicaciones:

- Poveda, J. y Robles, V. (2012). "Image retrieval based on the combination of RGB and HSV's histograms and Colour Layout Descriptor". INGENIUS. N7, (enero/junio). pp 3-10. ISSN: 1390-650X
- Poveda, J. y Robles, V. (2012). "Combination of HSV and RGB histograms and MPEG-7 Descriptor: Preliminar Results and Future Work". INGENIUS. N8 , (julio/diciembre), pp. 3-10, ISSN: 1390-650X.

ÍNDICE GENERAL

I INTRODUCCIÓN	1
1 INTRODUCCIÓN	3
1.1 Antecedentes	3
1.2 Objetivos	4
1.2.1 Objetivo General	4
1.2.2 Objetivos específicos	4
1.3 Alcance	5
II EL CEREBRO, EL OJO Y EVOLUCIÓN HUMANA	7
2 EL CEREBRO, EL OJO Y EVOLUCIÓN HUMANA	9
2.1 Evolución Humana	9
2.1.1 Técnicas de Selección	9
2.1.2 Teoría del gen egoísta	11
2.1.3 Evolución humana.	12
2.1.4 Evolución del ojo	13
2.2 Funcionamiento del cerebro humano.	13
2.2.1 Sustancia blanca y gris.	14
2.2.2 Lóbulos del córtex.	14
2.3 Funcionamiento del ojo humano.	15
2.3.1 Principios físicos de las lentes.	16
2.3.2 El ojo como cámara fotográfica	17
2.3.3 Profundidad	21
2.3.4 Anatomía y función de los elementos estructurales de la retina	24
2.3.5 Función de corteza visual primaria.	25
III CONTENT-BASED IMAGE RETRIEVAL	29
3 CONTENT-BASED IMAGE RETRIEVAL	31
3.1 Historia de la recuperación de imágenes	31
3.1.1 Visión por computador	31
3.1.2 Aplicaciones de visión por computador	31
3.1.3 Historia de la visión por computador	32
3.1.4 Problemas de Visión por Computador y la Visión Humana	33
3.2 Proceso de recuperación de imágenes	34
3.3 Características de las imágenes	34
3.4 Métodos y aplicaciones de la recuperación de imágenes.	35
3.4.1 Métodos de recuperación de imágenes	35
3.4.2 Aplicaciones de la recuperación de imágenes	35
IV HISTOGRAMAS DE COLOR Y DESCRIPTORES DE COLOR	37
4 HISTOGRAMAS DE COLOR Y DESCRIPTORES DE COLOR	39
4.1 Introducción	39
4.2 Tipos de histogramas de color	39
4.2.1 Comparación entre dos histogramas	39
4.3 Tipos de Descriptores de Color del estándar MPEG-7	43
4.3.1 Características del estándar MPEG-7	43
4.3.2 Descripción técnica del estándar MPEG-7	44

4.3.3	Tipos de descriptores para imágenes	45
4.3.4	Otros descriptores	50
4.4	Reducción de la dimensionalidad	51
V TÉCNICAS DE MACHINE LEARNING 53		
5	TÉCNICAS DE MACHINE LEARNING	55
5.1	Introducción	55
5.1.1	Naturaleza del conocimiento	55
5.1.2	Machine Learning y la naturaleza del conocimiento	56
5.2	Métodos de Machine Learning	56
5.2.1	Tipos de aprendizaje	56
5.2.2	Aprendizaje Supervizado	56
5.2.3	Aprendizaje no supervisado	56
5.2.4	Conjunto de entrenamiento y prueba para aprendizaje supervisado .	56
5.2.5	Conjunto de entrenamiento para aprendizaje no supervisado	58
5.2.6	Algoritmos de Machine Learning	58
5.2.7	Cómo elegir un algoritmo de Machine Learning para resolver un problema puntual	60
5.3	K-Nearest Neighbors	60
5.4	Técnicas de combinación de características: Esquemas de votación Borda	63
5.4.1	Sistemas de votación	63
5.4.2	Esquemas de Votación Borda	63
5.4.3	Método de Condorcet	66
VI PLAN DE EXPERIMENTACIÓN 69		
6	PLAN DE EXPERIMENTACIÓN	71
6.1	Selección y preparación del corpus	71
6.2	Selección de las características a emplear	71
6.2.1	Extracción de las características de las imágenes	71
6.3	Diseño de la técnica de combinación de características a emplear	75
6.3.1	Parte uno: extracción de características de las imágenes, cuantización de las imágenes, mezcla y subconjuntos de prueba y entrenamiento	76
6.3.2	Parte dos: Aplicación de los K-Vecinos más cercanos y esquemas de Votación Borda	77
6.3.3	Diseño del prototipo	78
6.3.4	Desarrollo de algoritmos en este sistema	81
6.4	Diseño del plan de experimentación	84
6.4.1	Objetivo del experimento	84
6.4.2	Descriptores de color elegidos	84
6.4.3	Unidades experimentales	86
6.4.4	Elección del algoritmo de Machine Learning y objetivo de aplicación de los esquemas de votación Borda.	86
VII ANÁLISIS DE RESULTADOS 87		
7	ANÁLISIS DE RESULTADOS	89
7.1	Análisis de la Precisión, Cobertura y F-Measure	89
7.1.1	Ejecución del plan de experimentación	89

7.1.2	Análisis de resultados	93
7.2	Comparativa con el Estado del Arte	97
7.2.1	Experimentación	97
7.3	Propuesta de mejoras y trabajo futuro	103
VIII CONCLUSIONES Y RECOMENDACIONES		105
IX GLOSARIO		111
BIBLIOGRAFÍA		114
X ANEXOS		121
A	ANEXOS	123

ÍNDICE DE FIGURAS

Figura 1	Conjunto de escarabajos de color verde y café [77] 10	
Figura 2	Los escarabajos de color verde tienen una desventaja y desaparecen a manos de un depredador.[77]	10
Figura 3	Los escarabajos café se empiezan a reproducir por su ventaja natural frente a los depredadores [77] 10	
Figura 4	Los escarabajos café han exparcido sus genes para garantizar su supervivencia. [77] 11	
Figura 5	Lóbulos del cerebro humano[24] 14	
Figura 6	Funciones ejecutivas del lóbulo frontal[28] 15	
Figura 7	Corte vertical del ojo derecho [67] . 16	
Figura 8	Lente convexa [34] 16	
Figura 9	Lente cóncava [34] 17	
Figura 10	Formación de una imagen a partir de una lente convexa[34] 17	
Figura 11	Efecto de la fuerza de las lentes sobre distancia focal.[34] 18	
Figura 12	El modelo de Pin-hole [9] 18	
Figura 13	Elementos de la visión humana [34] . 20	
Figura 14	Mecanismo de acomodación [34] . 21	
Figura 15	Apertura de la pupila. [34] 21	
Figura 16	Triangulación de una imagen [9] 23	
Figura 17	Profundidad del ojo humano por: 1) Las dimensiones de la imagen en la retina. 2)Por estereopsis[34]	23
Figura 18	Capas de la retina [34] . 24	
Figura 19	Ojo humano [37] . 25	
Figura 20	córtex visual primario [34] . 26	
Figura 21	Fechas y eventos en el desarrollo de visión por computador [74] 33	
Figura 22	Precisión de los histogramas con las bases de datos[19] . 33	
Figura 23	Explicación del proceso de Content-Based Image Retrieval [44] 34	
Figura 24	Imagen de la cual se va a extraer el histograma . 40	
Figura 25	Histograma de la imagen 24. 40	
Figura 26	Representación de la gama de colores RGB en el plano cartesiano [41]. 42	
Figura 27	Representación del histograma HSV. En donde V es representado por el principal eje ortogonal del plano. El ángulo representa el valor cromático, mientras que el radio representa el nivel de saturación (Pureza de color) [36]. 43	
Figura 28	Ejemplo de extracción de MPEG-7 [54]. 44	
Figura 29	Principales elementos de MPEG-7 [54]. 44	
Figura 30	Resultado de aplicar la selección del color representativo de la imagen en el espacio de color RGB ₂₉ . 48	

- Figura 31 Recorrido en zig-zag aplicado a Y, Cb y Cr del color de la imagen representativo[75]. 48
- Figura 32 Clasificación genética de la población en donde se etiqueraron los datos $\{(x^1, y^1), (x^2, y^2), (x^3, y^3), \dots, (x^n, y^n)\}$ de tal forma que puedan ser clasificados.[7]. 57
- Figura 33 Aprendizaje no supervisado en donde se busca clasificar los datos no etiquetados [1]. 57
- Figura 34 Árboles de decisión de un sistema de información de un bando [18]. 59
- Figura 35 Etiquetas, número de besos y número de peleas [35]. 62
- Figura 36 Conjunto de datos de las películas que van a ser clasificadas [35]. 63
- Figura 37 Cálculo de las distancias [35]. 63
- Figura 38 Imágenes de la base de datos Wang que está dividida en 10 categorías, puede ser descargada de <http://www.cs.washington.edu/research/imagetdatabase/groundtruth>. 72
- Figura 39 Imágenes de la base de datos ZuBud que está dividida 201 categorías. En esta imagen solamente se muestran las primeras 20 categorías. Puede ser descargada de <http://www.vision.ee.ethz.ch/showroom/zubud/> 73
- Figura 40 Imágenes de la base de datos UW que está dividida 22 categorías. Cada categoría se tiene un número de imágenes aleatorio. Puede ser descargada de: wang.ist.psu.edu/docs/related. 74
- Figura 41 Primera parte del experimento [29]. 76
- Figura 42 Parte dos del experimento [29]. 77
- Figura 43 Diagrama que explica las relaciones entre los módulos del sistema. 79
- Figura 44 Módulo para procesar el histograma RGB que muestra las distintas etapas para generar el archivo con los datos del histograma RGB. 79
- Figura 45 Módulo para procesar el histograma HSV que muestra las distintas etapas para generar el archivo con los datos del histograma HSV 80
- Figura 46 Módulo para procesar el histograma CLD que muestra las distintas etapas para generar el archivo con los datos del histograma HSV . 81
- Figura 47 Esquema que muestra los pasos que permiten generar los resultados del algoritmo RGB. 82
- Figura 48 Esquema que muestra los pasos que permiten generar los resultados del algoritmo HSV . 82
- Figura 49 Esquema que muestra los pasos que permiten generar los resultados del algoritmo CLD . 83
- Figura 50 Proceso mediante el cual se ejecuta el módulo de los esquemas de votación Borda. 83
- Figura 51 Procedimiento la extracción de características del espacio de color RGB. 90
- Figura 52 Procedimiento la extracción de características del espacio de color HSV. 91
- Figura 53 Procedimiento la extracción de características del estándar de color MPEG-7 Color Layout Descriptor. 92

Figura 54	Procedimiento para la ejecución del algoritmo de losK-Vecinos más cercanos para RGB Y HSV . 94
Figura 55	Procedimiento para la ejecución del algoritmo de losK-Vecinos más cercanos para CLD 95
Figura 56	Procedimiento para la ejecución del algoritmo de los esquemas de votación Borda 96
Figura 57	Resultados con los diversos descriptores de color [19]. 98
Figura 58	Resultados del experimento con la base de datos Wang . 99
Figura 59	Resultados del experimento comenzando desde la izquierda: Precisión del 82.66 %, Cobertura del 100 % y F-Measure de 90.5 % . 100
Figura 60	Resultados del experimento con la base de datos ZuBud. 100
Figura 61	Resultados del experimento comenzando desde la izquierda: Precisión del 50 %, Cobertura del 100 % y F-Measure de 66.66 % . 101
Figura 62	Resultados del experimento con la base de datos UW . 102
Figura 63	Resultados del experimento comenzando desde la izquierda: Precisión del 50 %, Cobertura del 100 % y F-Measure de 66.66 % . 102

ÍNDICE DE CUADROS

Cuadro 1	Porcentaje de generaciones de zorros que evolucionaron. 9
Cuadro 2	Votos del jurado donde en el grupo 1 hay mayor preferencia por el candidato 1, en el grupo 2 por el candidato 1 y en el grupo 3 por el candidato 2 . 65
Cuadro 3	Precisión en la base de datos Wang. 98
Cuadro 4	Precisión en la base de datos ZuBud. 100
Cuadro 5	Precisión en la base de datos UW. 101

Parte I

INTRODUCCIÓN

INTRODUCCIÓN

1.1 ANTECEDENTES

La visión por computador es una área dentro de la inteligencia artificial que tiene como objetivo detectar los patrones dentro de las imágenes mediante la aplicación de principios geométricos [64], para que de esta forma el computador pueda detectarlos y tomar alguna decisión dentro del contexto que esta trabajando un agente inteligente[68]. Para llegar a cumplir este objetivo los científicos de diversas áreas del conocimiento se han unido con la finalidad de cumplirlo. Entre las áreas del conocimiento se encuentran [40, 64]:

- El dominio de la relación espacial.
- La medición del espacio tridimensional.
- La propagación de la luz a través de un lente.
- El modelo matemático de formas y tamaños de objetos.

Para lograr este objetivo se han buscado varios modelos de percepción visual para tener aproximaciones al modelo de la visión humana. Desde tiempos inmemoriales se han planteado varias teorías sobre como el ser humano puede observar el mundo a su alrededor. Los primeros en plantear que el ojo sigue un conjunto de reglas sistemáticas para observar el mundo fueron los árabes durante la edad media europea [39]. Luego, se realizaron varios estudios sobre el área de la visión tal como lo explica [64] *“Kepler realiza estudios sobre el análisis geométrico de la formación de la imagen en el ojo. Consecuentemente, Isaac Newton realiza estudios sobre la visión de color, mientras que en 1910 Helmholtz estudia la óptica fisiológica, y en 1912 Wertheimer realiza estudios acerca del movimiento aparente de agrupaciones de puntos o campos. Estos estudios se han convertido en los pilares fundamentales para las teorías actuales de la percepción visual. Dichos estudios tomaron fuerza en los años sesenta cuando empezaron aparecer las primeras computadoras, que se requerían para el tratamiento de información espacial donde involucraba el análisis de propiedades del sistema de visión humana”*.

Debido a que la visión humana es un sistema muy complejo ya que el mismo puede detectar cambios en los colores de las imágenes, sus bordes y textura, fue muy difícil desarrollar un conjunto de algoritmos y teorías que expliquen su funcionamiento, para ello se utilizaron los siguientes tres enfoques [40, 64]:

- El Profesor Azriel Rosenfeld desarrolló un conjunto de técnicas empíricas basadas en ciertos criterios matemáticos que permitían la estimación de bordes y líneas utilizando diferentes criterios.
- Otro grupo que estaba estudiando el problema de visión por computador restringió su estudio a un mundo de bloques blancos mate iluminados sobre fondo negro, en donde los bloques podían tener cualquier forma siempre que sus superficies fueran planas y sus bordes rectos. Se buscó que este trabajo generará un cierto

conocimiento que podría ser extrapolado a otros mundos mucho más complejos. Pero no fue así.

- Finalmente, en 1986 se consigue establecer un conjunto de modelos de cálculo con la finalidad de expresar la formación de la imagen a través de ecuaciones diferenciales relacionando los valores de la intensidad con los de la superficie. Para ello se estudió a gran detalle la iluminación, la geometría, la reflectancia de la superficie y el punto de vista del observador que actúan entre sí para establecer los valores de intensidad de la imagen.

David Marr durante la década de 1970 y principios de 1980 estudia a profundidad los modelos de visión en donde se centra en investigar los siguientes elementos que le llevaron a definir un modelo que permite aproximar la percepción visual y cómo estudiar los sistemas de conocimiento: percepción del movimiento, textura, distancia, entre otras. Sin embargo, los investigadores Gibson y Aloimonos aseguraron que la teoría de Marr no se puede determinar como una teoría de percepción visual debido a que no consideró a todos los sistemas visuales existentes, lo que no permitiría obtener información de imágenes en 3 dimensiones [64]. Desde ese entonces se desarrollaron varias técnicas para resolver el problema de la visión por computador tales como: Filtro Gaussiano, Filtro Laplaciano, renderización de imágenes, entre otras [74].

La técnica de Content-Based Image Retrieval, también conocida como CBIR, fue definida por T. Kato en 1992 para describir en uno de sus experimentos la recuperación automática de una base de datos de imágenes a partir de características básicas, como color, textura y forma, que pueden ser extraídas automáticamente de las imágenes con la finalidad de poderlas clasificar. Las características para la recuperación pueden ser primitivas o semánticas, pero el proceso de extracción es automático [25].

Mediante CBIR se busca comparar imágenes con otras para recuperar las que sean parecidas. Para lograr este objetivo se aplican algoritmos de Machine Learning con la finalidad de obtener las características de las mismas para poder detectar sus patrones, analizarlos y determinar que imágenes son parecidas a la imagen de búsqueda. Las técnicas que se utilizan para obtener las características de las imágenes son histogramas y descriptores. Dependiendo de la necesidad que se tenga se utilizan histogramas y/o descriptores de color, forma y textura. Los histogramas representan el número de bins que tiene la imagen, mientras que los descriptores describen el contenido de la imagen [25].

1.2 OBJETIVOS

1.2.1 *Objetivo General*

Mejorar los niveles de precisión y cobertura en la recuperación de imágenes a través de la utilización de técnicas de fusión de características de color.

1.2.2 *Objetivos específicos*

- Explicar que es recuperación de imágenes basada en contenidos.

- Estudiar y describir los descriptores de Color RGB, HSV y el descriptor de distribución de color del estándar MPEG-7.
- Estudiar y describir en qué consiste la técnica de los K-Vecinos más cercanos (K-Nearest Neighbors).
- Estudiar y describir los esquemas de Votación Borda como técnica de fusión de características.
- Seleccionar y preparar 3 corpus de imágenes usados en el estado del arte para aplicar la propuesta.
- Diseñar e implementar un plan de experimentación, a través de un conjunto de scripts que permitan validar la propuesta.
- Comparar los resultados obtenidos con el estado de arte actual.
- Plantear un plan de mejora que posibilite mejorar el proceso como trabajo futuro.

1.3 ALCANCE

El presente trabajo de investigación se centra en la utilización de los histogramas de color RGB, HSV y el descriptor de color MPEG-7 Color Layout Descriptor para extraer sus características. Luego, se las utilizará con el algoritmo de Machine Learning los K-Vecinos más cercanos con la finalidad de obtener las valoraciones de cada descriptor. Finalmente, ellas serán aplicadas al algoritmo de los esquemas de votación Borda para clasificar los datos.

Se utilizarán herramientas para la extracción de las características de las imágenes a partir del cual se definirá un sistema para clasificarlas. Además, el sistema permitirá realizar pruebas y otras otras operaciones con las herramientas.

Parte II

EL CEREBRO, EL OJO Y EVOLUCIÓN
HUMANA

Para iniciar el estudio de el ojo humano es necesario tocar dos tópicos muy importantes: La evolución humana y el cerebro humano. A partir del primer punto, se explicará la evolución del ser humano a lo largo del tiempo, para ello se describirá la teoría de la evolución de las especies planteada por el naturalista inglés Charles Darwin y continuada por los biólogos y psicólogos evolucionistas. Para el cerebro humano se utilizará la anatomía y fisiología humana con la finalidad de explicar su estructura y sus funciones principales. Finalmente, se revisará la anatomía y fisiología del ojo humano y su relación con la visión por computador.

2.1 EVOLUCIÓN HUMANA

El naturalista inglés Charles Darwin describe a la evolución como un proceso de transformación progresiva de las especies para adaptarse al medio ambiente [31]. Por ejemplo, en la década de 1950 en la antigua Unión Soviética el científico Dimitri Belyaev empezó a realizar un experimento con zorros con la finalidad de poder domesticarlos durante un tiempo para que se puedan adaptar con el ser humano. Para ello empezó a separar a los zorros que no eran muy agresivos con el ser humano y al cabo de 6 generaciones los zorros disminuyeron su nivel general de agresividad. Al paso del tiempo los zorros ya eran amigables con el ser humano. Ello demuestra como a través del paso del tiempo las especies se transforman para adaptarse al medio ambiente. El cuadro 1 muestra los resultados del experimento [16].

Evolución: Proceso de transformación progresiva de las especies para adaptarse al medio ambiente

2.1.1 Técnicas de Selección

Existen diversas teorías sobre cómo los seres vivos son seleccionados para permitir la supervivencia de la especie. En el presente tema de tesis mencionaré cuatro de ellos: Selección Natural, Selección Sexual y Selección Artificial.

Selección Natural

La selección natural es el proceso mediante el cual la supervivencia o reproducción de diferentes genotipos en una población permiten los cambios en la frecuencias de una población [78]. Es decir, un individuo que posee ciertas características que permitirán la supervivencia de la

Nota: Un individuo que posee ciertas características que permitirá la supervivencia de la especie mediante la transmisión de sus genes

Generación de zorros.	Porcentaje de adaptación al ser humano
10	18
20	35
30-35	70-80

Cuadro 1: Porcentaje de generaciones de zorros que evolucionaron.



Figura 1: Conjunto de escarabajos de color verde y café [77]



Figura 2: Los escarabajos de color verde tienen una desventaja y desaparecen a manos de un depredador.[77]

especie mediante la transmisión de sus genes, mas adelante se describirá la teoría del gen egoísta que explica estas características. En este tipo de selección la naturaleza es el agente que selecciona las características de los individuos, tal como Charles Darwin lo explicó en su libro de la teoría de las especies [16].

La figura 1 muestra un grupo de escarabajos café y verde en un determinado escenario de la naturaleza. La figura 2 muestra como los escarabajos verdes son presa de un águila debido a que no se pueden ocultar en la naturaleza. La figura 3 muestra a la selección natural en acción; debido a que los escarabajos verdes son una presa fácil para el águila, los escarabajos solo se reproducen con los escarabajos café. La figura 4 muestra a los escarabajos que pueden sobrevivir frente al ataque de un águila[77].

Selección Sexual

Selección Sexual: La selección sexual nos indica que quien puede reproducirse es aquel que es más sensual para el resto de la especie

Según anota [57]: “La selección sexual nos indica que quien puede reproducirse es aquel que es más sensual para el resto de la especie, a diferencia de la

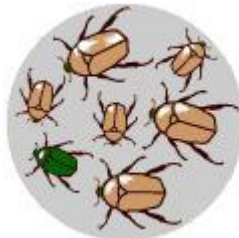


Figura 3: Los escarabajos café se empiezan a reproducir por su ventaja natural frente a los depredadores [77]



Figura 4: Los escarabajos café han exparcido sus genes para garantizar su supervivencia. [77]

selección natural donde el más fuerte es quien sobrevive". En este tipo de selección se compete para la reproducción, es decir, la capacidad de atraer a la mayor cantidad de compañeros/as sexuales con la finalidad de permitir una nueva generación con características que mejoren sus posibilidades de reproducirse [57].

Esta teoría surge debido a que Charles Darwin después de la publicación de su libro "Origen de las Especies" se dio cuenta que la teoría de la selección natural no podía explicar debido a ciertos comportamientos en el reino animal, tal como la cola de plumas del pavo real. Por ello el planteó la posibilidad de que la descendencia de los individuos deseables heredarán sus características maximizando sus posibilidades de reproducirse y sobrevivir. En este tipo de forma de supervivencia Darwin le da una gran importancia a las hembras, ya que ellas son las que tienen la capacidad de hacer que la especie sobreviva [56].

Selección Artificial

Según se explica en Yamasaki et al. [86]: "La Selección Artificial es el proceso intencional o no intencional de modificación de individuos de una población a través de acción humana, resultando en evolución fenotípica de plantas y animales". Esto quiere decir que los seres humanos seleccionan ciertas características que tienen las plantas o animales para mezclarlas, con el objetivo que su descendencia pueda adaptarse a las necesidades humanas. Un ejemplo de ello son los vegetales como el Brócoli y la Coliflor que provienen de la familia de los repollos Brassica Oleracea que a través de cientos de años mediante la selección de determinados repollos con técnicas de horticultura se obtuvo los repollos mencionados anteriormente [15].

2.1.2 Teoría del gen egoísta

La teoría del gen egoísta fue propuesta por el biólogo evolucionista inglés Richard Dawkins en el año de 1976. En esta teoría se propone que los genes y no los individuos son la unidad de selección [33]. Según esta teoría "Los genes que promueven su propia supervivencia y replicación se esparirán a expensas de otros genes" [81]; ello quiere decir que solamente los individuos que posean ciertos genes podrán reproducirse lo que garantizará que sus genes pasarán a su descendencia y con ello garantizará la inmortalidad de sus genes [14].

Para Richard Dawkins "Los genes son entidades egoístas que quieren inmortalizar su supervivencia"[81] para ello los genes moldean el cuerpo

y utilizan ciertos indicadores externos que muestran cierta “garantía genética” de que la especie podrá sobrevivir [14].

En la sección 2.1.1 se mencionó que los individuos que tienen ciertas características físicas pueden asegurar la supervivencia de la especie, ello muestra que los individuos tienen ciertos genes que garantizarán que la especie pueda sobrevivir. Por ejemplo, en África los cazadores que cazan elefantes buscan los que tienen colmillos más grandes debido a que tienen mayor cantidad de marfil. Por ello, los elefantes hembra han empezado a elegir a los elefantes macho que tienen colmillos pequeños para que la siguiente generación no posea colmillos grandes con la finalidad de que la especie no desaparezca en manos de los cazadores. Ello muestra que los individuos de colmillos pequeños poseen genes que se perpetuarán en la población debido a la amenaza de los cazadores [16].

En la sección 2.1.1 se explicó que la selección está dada por la elección para mezclarse con el más sexy. Tal como [58] lo explica, la selección sexual es aquella que permitió la evolución humana, ya que se seleccionó a aquellos que tenían mejor humor y un mayor nivel de inteligencia, lo que permitió que el cerebro humano se desarrolle. Ello garantizó que los genes de estos individuos se perpetuaran y permitan la aparición de la especie humana.

2.1.3 *Evolución humana.*

En esta parte del capítulo se describirán algunos aspectos relacionados con la evolución humana, tales como cooperación e inteligencia.

Cooperación

La [27] define a cooperar como el conjunto de acciones que nos permiten obrar conjuntamente con otro u otros para un mismo fin. La cooperación tiene dos bases: una biológica y la otra social. En la parte biológica me referiré a la parte genética, en la parte social acudiré a la filosofía, política y economía.

En la sección 2.1.2 se explicó que un gen que se esparce a través de la población para garantizar la supervivencia de la misma. La función de la cooperación es ayudar a la supervivencia de gen mediante la cooperación entre grupos. Un ejemplo de ello es la protección que los padres dan a sus hijos y se sacrifican por ellos debido a que ellos portan los genes de sus padres, sin la ayuda de los padres los hijos no podrían sobrevivir, lo que no permitiría que sus genes se esparzan. Desafortunadamente, la cooperación como tal solamente se da con aquellos que están mucho más relacionados con nosotros, nuestros padres, abuelos, tíos y aquellos que tengan un mayor grado de consanguinidad, debido a que compartimos genes y mientras más sea el número de genes que compartamos, mayor será la cooperación que exista entre nosotros [81].

A pesar de la parte biológica, tenemos una parte social, que quizá es lo más importante para garantizar la supervivencia como especie. Al comienzo vivíamos en pequeños grupos, los cuales cooperaban entre sí, con la finalidad de garantizar la supervivencia del grupo. Actualmente esa tendencia de cooperar a la forma como se realizaba en las etapas primigenias del ser humano ha desaparecido [81]. Sin embargo, han aparecido otras estructuras sociales que permiten que los seres humanos podamos vivir en grandes grupos, la más importante de todas y una

construcción moderna: El estado. El estado moderno surge a partir de la crisis de feudalismo europeo, debido a las conjeturas sociales de la época [60], en ello se produce la aparición del capitalismo como nuevo sistema económico, las ciencias sociales que permiten el control del estado y el manejo de las ciencias naturales que permiten la creación de nuevos productos con la finalidad de garantizar la producción. En el nuevo régimen empiezan a aparecer nuevas ideas, la más importante quizá es la del filósofo alemán Hegel, cuya idea más importante es que el estado garantice tanto las necesidades naturales como las necesidades sociales de sus habitantes [53]. Ello quiere decir que el altruismo actualmente depende de las políticas estatales que se den, ya no mas en el plano individual.

Inteligencia y evolución del cerebro

El cerebro humano es el órgano más complejo que existe dentro del cuerpo humano, dentro de el se almacenan recuerdos, ideas, pensamientos, etc. Este órgano complejo posee un conjunto de patrones programados que permiten la supervivencia del individuo [81], tal como lo explica Geoffrey Miller siguiendo la tradición de Charles Darwin, fue moldeado por selección sexual[58]. Mediante esta teoría se puede explicar el desarrollo del cerebro humano ya que como lo explica Charles Darwin en una de sus cartas con su colega Alfred Russell Wallace:

“La selección natural sólo podría haber dotado al hombre salvaje de un cerebro mínimo grado superior al de un mono, mientras que por el contrario posee uno que es poco inferior al de un filósofo”[33]

En la sección 2.1.2 se explicó que los parámetros que permitieron la evolución del ser humano fueron humor e inteligencia, para ello se necesitó un cerebro mucho más grande capaz de comunicarse, cooperar y poder trabajar en equipo [82]. Mediante la selección de los individuos que tenían dichas características a través del paso del tiempo, se produjo la transición de los homínidos al género homo.

La selección del más inteligente sobre el más fuerte ha sido uno de los pilares para la explicación de la evolución del cerebro humano.

2.1.4 *Evolución del ojo*

El ojo es un órgano que permite captar señales del exterior las cuales son procesadas con la finalidad de obtener una imagen del escenario en donde se desenvuelve el ser vivo. La evolución del ojo se produjo por selección natural debido a las adaptaciones que se produjeron por las necesidades de las especies de poder detectar la luz para garantizar su supervivencia en el medioambiente en que se encontraban. Por ello, la evolución de los diversos ojos de varias especies dependen mucho de las necesidades de supervivencia de las mismas [47]. Por ejemplo: el ojo humano tiene la precisión de color de una lente Vario/Tessar mientras que las salamandras no pueden percibir muy bien los colores[16].

2.2 FUNCIONAMIENTO DEL CEREBRO HUMANO.

Según [31] “ el cerebro es la materia nerviosa que ocupa el cráneo de los vertebrados y es el asiento de las sensaciones, así como el principio de los movimientos voluntarios”. El cerebro humano forma parte del encéfalo, el

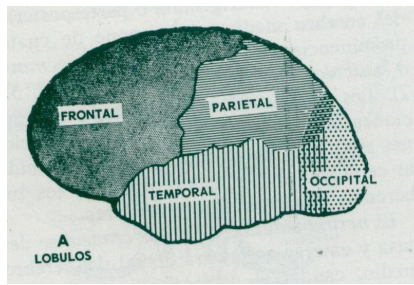


Figura 5: Lóbulos del cerebro humano[24]

cual está formado por una sustancia blanca y una sustancia gris. Además, el cerebro humano está formado por 5 lóbulos: Frontal, Parietal, Occipital, Temporal y de la Ínsula [84]. En esta sección se explicará la sustancia blanca y gris. Además, se analizará los 5 lóbulos del cerebro humano. En la figura 5 se muestra los 5 lóbulos del cerebro humano.

2.2.1 Sustancia blanca y gris.

La sustancia blanca y gris forman el cerebro humano. Cada una de ellas tiene una función específica. En la presente subsección se explicará la sustancia blanca y gris.

Sustancia Gris.

La sustancia gris comprende a las neuronas, las cuales se disponen fundamentalmente en la corteza y núcleos grises de la base. [84]

Sustancia blanca.

Son fibras que establecen conexiones con estructuras subyacentes, las que ponen en conexión diferentes regiones cerebrales y las que ascienden hacia núcleos basales y corteza a partir de formaciones más bajas [84].

2.2.2 Lóbulos del córtex.

Lóbulo Frontal

El lóbulo frontal es la zona más evolucionada del ser humano. En ella se encuentran las siguientes funciones: el lenguaje expresivo o habla, el movimiento voluntario, los procesos necesarios para el cálculo, la atención y memoria, las zonas ejecutivas, las cuales se muestran en la figura 6 [28]:

Las zonas en las que se divide el lóbulo frontal son: el córtex motor o área primaria motora, el córtex premotor, el operculum frontal, la zona pre-frontal [28]. El córtex motor es el encargado del control de los movimientos aprendidos, el córtex premotor tiene funciones motoras y cognitivas más complejas, como los movimientos oculares, el operculum frontal controla el área del lenguaje como la semántica, la

Atención focalizada sobre estímulos relevantes e inhibición de los irrelevantes (paradigma estándar: paradigma de Stroop)
Flexibilidad de pasar de una tarea a otra; flexibilidad de respuesta apropiada a las demandas de un contexto variable que no admite respuestas de rutina
Planificación de tareas dirigidas a un objetivo
Previsión
Monitorizar
La información (<i>working memory</i>)
La ejecución
Codificación para el tiempo y lugar. Organizar la conducta temporalmente para alcanzar los objetivos
Resolución de problemas
Localizar recursos, curiosidad-motivación
Formular conceptos abstractos
Autoconsciencia, conducta moral
Discurso social como habilidad para interactuar productivamente con otros en discusiones y conversaciones

Figura 6: Funciones ejecutivas del lóbulo frontal[28]

zona pre-frontal es la zona más evolucionada del ser humano ya que tiene funciones como el razonamiento y la formación de conceptos, la generación de acciones voluntarias, etc[28].

Lóbulo Parietal

El lóbulo parietal es aquel que nos permite movilidad, ya que integra la información sensorial introspectiva, como los músculos y articulaciones y exteroceptiva, del exterior. Se le atribuyen básicamente funciones sensitivas, asociativas, así como de reconocimiento del espacio [32].

Lóbulo Occipital

Este lóbulo se encarga de la visión, elabora la información visual aunque está trasciende a los lóbulos parietales y temporales [32].

Lóbulo Temporal

En esta área se procesa el audio, asientan importantes estructuras de la memoria (hipocampo), y del sistema emotivo inconsciente (sistema límbico) [32].

Lóbulo de la Ínsula

También conocido como lóbulo central. Esta parte del cerebro humano está relacionado con el ritmo cardíaco y el control de la presión arterial, control visceromotor y de las funciones sensitivas viscerales, sensibilidad intraabdominal y algunas funciones del habla[23].

2.3 FUNCIONAMIENTO DEL OJO HUMANO.

Según lo explica la [27]: "El ojo es el órgano de la vista en el hombre y en los animales". Mediante el ojo las especies pueden detectar cambios de luz, detectar objetos y sus texturas, entre otras funciones. En la presente sección se explicará el funcionamiento del ojo humano, para ello se describirá los principios físicos de las lentes, anatomía y función de los

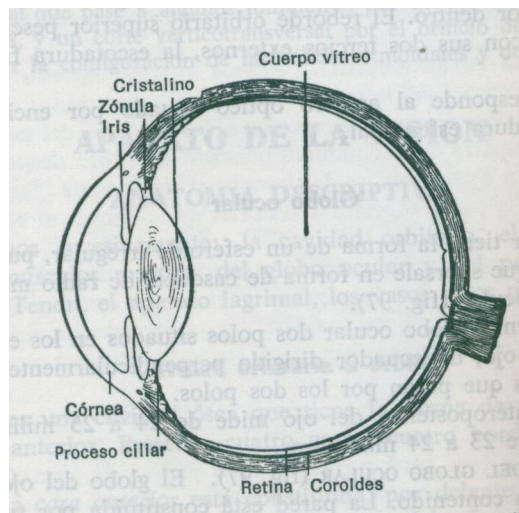


Figura 7: Corte vertical del ojo derecho [67].

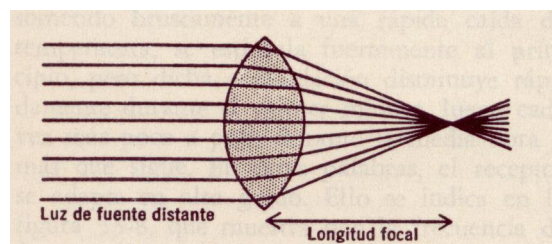


Figura 8: Lente convexa [34]

elementos estructurales de la retina y la fisiología de la visión. En la figura 7 se muestra las partes del ojo humano.

2.3.1 Principios físicos de las lentes.

Antes de empezar el estudio del ojo es necesario conocer los principios que rigen las lentes, para luego relacionarlos con el ojo humano. En este apartado se estudiará las propiedades de las lentes.

Lente cóncavo y convexo

Lente convexa: Tal como [65] lo anota: "Un lente convexo es más grueso en el centro que en sus bordes". Para este tipo de lentes la distancia focal es mayor a 0 [83]. En este tipo de elementos los rayos de luz que atraviesan el centro de la lente son exactamente perpendiculares a la misma, por lo que la atraviesan sin refracción alguna. Sin embargo, al irlos acercando al borde de la lente, los rayos chocan contra ella con ángulos progresivamente mayores. Los rayos que penetraron la lente salen convergiendo en un solo punto [34].

Lente cóncava: Tal como [65] lo anota: "Un lente convexo es más delgado en el centro que en sus bordes". Para este tipo de lentes la distancia focal es menor a 0 [83]. En este tipo de lentes los rayos no sufren refracción alguna por lo que al chocar con el centro de la lente no se refractan, de hecho permanecen perpendiculares al centro [34].

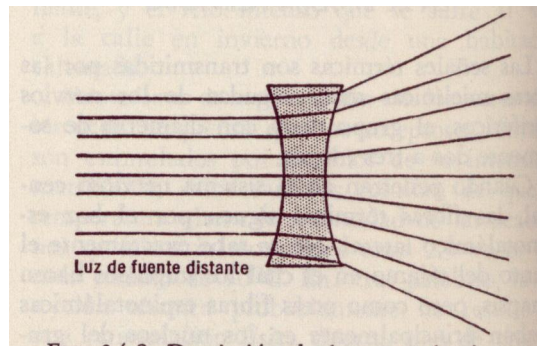


Figura 9: Lente cóncava [34]

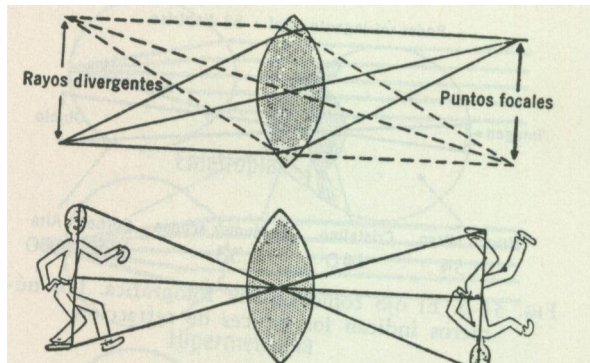


Figura 10: Formación de una imagen a partir de una lente convexa[34]

Formación de una imagen por una lente convexa

En la sección 2.3.1 se explicó las características de las lentes convexas. A partir de ello se explicará la formación de imágenes en una lente convexa, que es muy parecida al ojo humano.

Debido a que los rayos luminosos que atraviesan la lente por el centro no son desviados, todos los rayos luminosos coinciden en un mismo punto focal. La imagen que se obtiene como resultado es totalmente invertida con respecto al objeto original [34].

Medición del poder de refracción de una lente

[34] describe a este fenómeno de la siguiente manera: "Cuanto mas desvía la luz una lente, mayor es su "poder de refracción. Este poder de refracción se mide en dioptrías. Este poder equivale a un metro dividido por su distancia focal".

2.3.2 El ojo como cámara fotográfica

En la presente sección explicaré el funcionamiento de las cámaras y su relación con el ojo humano.

El modelo de Pin-hole

El modelo de pin-hole es un modelo imaginario que nos permite entender el funcionamiento de una cámara. En este modelo una porción luz pasa por un pequeño orificio, el resto de luz es detenida por una pared. El resultado de la obtención de luz es una imagen que tiene un tamaño de su distancia focal [9].

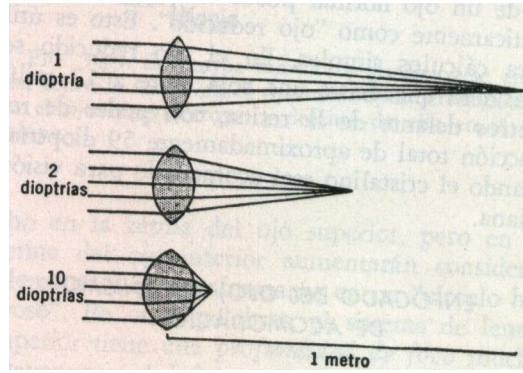


Figura 11: Efecto de la fuerza de las lentes sobre distancia focal.[34]

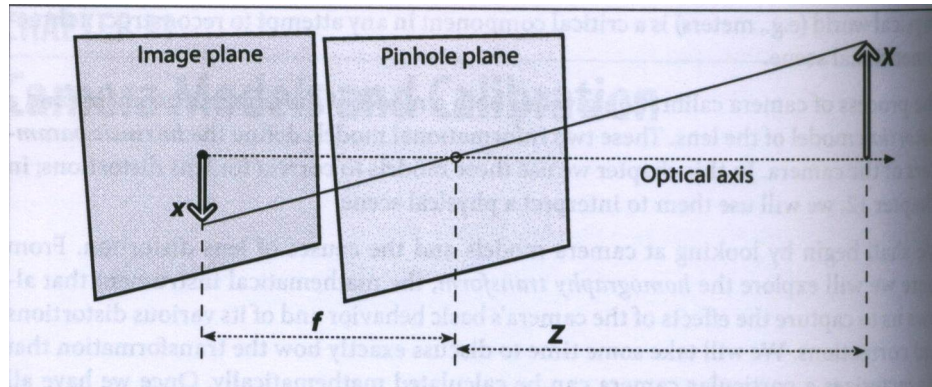


Figura 12: El modelo de Pin-hole [9]

En la figura 12 se muestra la representación de este modelo.

La fórmula 2.1 describe la relación entre la distancia focal f , el tamaño del sensor W y el campo de vista θ [74].

$$f = \frac{W}{2} * (\tan \frac{\theta}{2})^{-1} \quad (2.1)$$

Para ejemplificar este modelo se utiliza una relación de triángulos semejantes mediante la cual podemos realizar una representación en el plano del modelo [9]. La ecuaciones 2.2 y 2.3 describen los puntos de proyección en una pantalla, donde (X, Y, Z) son los puntos en el plano real y los puntos (x, y) son los puntos en la pantalla y (c_x, c_y) son los posibles desplazamientos que se den en la imagen[9].

$$x = f_x \left(\frac{X}{Z} \right) + c_x \quad (2.2)$$

$$y = f_y \left(\frac{Y}{Z} \right) + c_y \quad (2.3)$$

Proyección geométrica básica

Es la relación entre el conjunto de puntos del mundo físico (X_i, Y_i, Z_i) y los puntos en la pantalla x_i, y_i . Cuando se trabaja con este tipo de

coordenadas, es conveniente usar lo que es conocido como coordenadas homogéneas. Este tipo de coordenadas asociadas con un punto en un espacio de dimensiones n son expresadas en un vector $(n+1)$. Los puntos de proyección en el mundo físico a una cámara pueden ser resumidos de la siguiente forma en donde $w=Z$ [9]:

$$q = MQ \quad (2.4)$$

$$q = \begin{bmatrix} x \\ y \\ w \end{bmatrix} \quad (2.5)$$

$$M = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (2.6)$$

$$Q = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (2.7)$$

Distorsión de los lentes

Debido a que el modelo de Pin-Hole solo permite que un poco de luz pase a la cámara, es necesario utilizar una lente que permita enfocar una mayor cantidad de luz para obtener más rápidamente la imagen. Sin embargo, no hay lente perfecta debido a que se manufacturan lentes esféricos que son mucho más baratos y fáciles de desarrollar, ya que los lentes parabólicos, que son matemáticamente ideales, son mucho más difíciles y costosos de desarrollar [9]. En esta sección se explicarán dos distorsiones de los lentes: Radial y Tangencial.

Distorsión Radial: Este tipo de distorsión es resultado de la forma de los lentes. En las cámaras reales se produce distorsión de los píxeles cerca de los bordes de la imagen. Esto se debe a que los rayos más lejanos del centro de la lente se doblan más de los que son cercanos a él [9]. Para corregir este tipo de distorsión utilizamos las siguientes ecuaciones [9] donde k^n son los coeficientes de distorsión y $r^n = x^n + y^n$ son los coeficientes de distorsión radial [74, 9]:

$$x_{\text{corregido}} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \quad (2.8)$$

$$y_{\text{corregido}} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \quad (2.9)$$

Distorsión Tangencial: Surge en el proceso de procesamiento de la imagen [9]. Los parámetros de estas ecuaciones son similares a los de

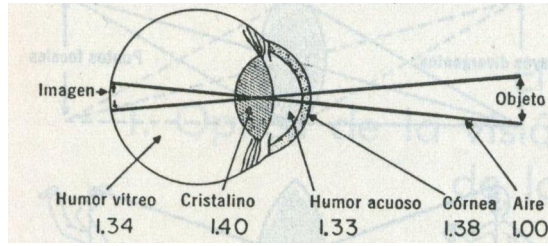


Figura 13: Elementos de la visión humana [34] .

las ecuaciones 2.8 y 2.9, con la diferencia de que p_n son los parámetros de entrada que alimentan a la ecuación [9, 76].

$$x_{\text{corregido}} = x + [2p_1 y + p_2 (r^2 + 2x^2)] \quad (2.10)$$

$$y_{\text{corregido}} = y + [p_1 (r^2 + 2y^2) + 2p_2 x] \quad (2.11)$$

El ojo como cámara fotográfica

Previamente, en las secciones 2.3.2, 2.3.2, 2.3.2, se explicó como las cámaras pueden adquirir las imágenes de los objetos del exterior. En esta sección se explicará algunos elementos de la visión del ojo.

Generalidades: [34] describe al sistema de visión del ojo como: "el sistema de lentes del ojo está formado por: 1) la interfase entre el aire y superficie anterior a la córnea; 2) la interfase entre la superficie posterior de la córnea y humor acuoso; 3) la interfase entre el humor acuoso y superficie anterior al cristalino, y 4) la interfase entre la superficie posterior del cristalino y el humor vítreo. En cada una de estas interfases los rayos luminosos se desvían; el resultado final es un sistema de lentes convergentes capaz de formar imágenes en la retina". En la figura 13 se muestran los elementos de la visión con sus índices de refracción.

Enfocado del ojo: Mecanismo de acomodación:

[34] explica que este mecanismo es: "El poder refringente del cristalino ocular tiene 18 dioptrías, del total de 59 dioptrías del sistema de lentes del ojo; las 41 dioptrías restantes resultan casi totalmente de la interfase anterior de la córnea con el aire. Normalmente el cristalino está formado de una fuerte cápsula elástica llena de fibras viscosas de tipo proteínico, pero transparentes. Cuando el cristalino se halla relajado, sin tensión sobre la cápsula, adopta forma esférica por la elasticidad de la cápsula que lo rodea. [...] Hay 70 ligamentos que se fijan radialmente alrededor de la lente y tiran de los bordes de la misma hacia el cuerpo ciliar. Estos ligamentos se hallan constantemente en tensión por la tracción elástica de sus fijaciones en el cuerpo ciliar, y la tensión de estos ligamentos hace que el cristalino se conserve relativamente plano en condiciones normales de reposo del ojo. A nivel de las inserciones en los tendones en el cuerpo ciliar se halla el músculo ciliar. Cuando estas fibras musculares entran en contracción, las inserciones de los ligamentos son estiradas hacia adelante, con lo cual disminuye algo de tensión sobre el cristalino, y éste adopta forma más esférica, a modo de pelota, gracias a la elasticidad de su cápsula. Cuando el músculo ciliar se halla totalmente relajado, el cristalino está lo más plano posible y su fuerza en dioptrías es la mínima. Por otra

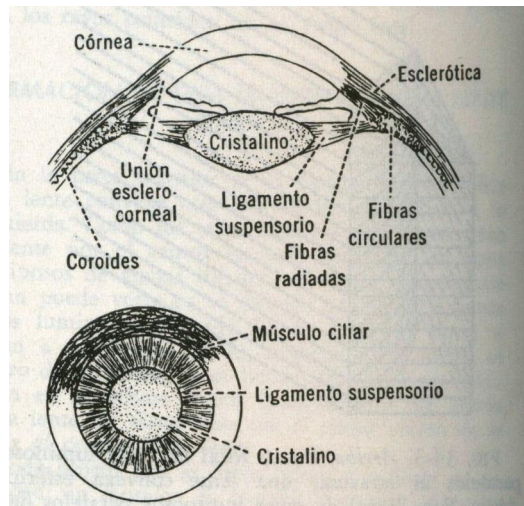


Figura 14: Mecanismo de acomodación [34].

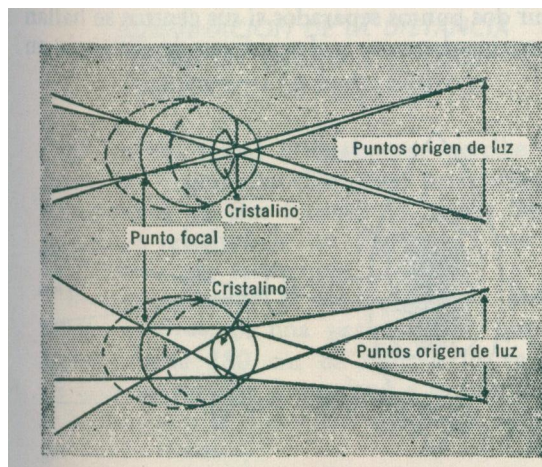


Figura 15: Apertura de la pupila. [34]

parte, cuando el músculo ciliar se contrae lo más fuerte que puede, la fuerza en dioptrías de la lente es máxima". En la figura 14 se puede apreciar el mecanismo de acomodación del ojo humano.

Abertura pupilar: [34] describe a este sistema como: "Una función principal del iris estriba en aumentar la cantidad de luz que penetra en el ojo en la obscuridad y disminuir la cantidad de luz que penetra en el ojo cuando la misma es intensa. [...]La cantidad de luz que penetra en el ojo es proporcional al área de la pupila, o bien al cuadrado del diámetro de la pupila. La pupila del ojo humano puede llegar a ser tan pequeña como aproximadamente 1.5mm de diámetro, y tan grande como casi 8mm de de diámetro. Por lo tanto, la cantidad de luz que penetra en el ojo puede variar unas 30 veces a consecuencia de los cambios en las dimensiones en la abertura pupilar". En la figura 15 se muestra la abertura de la pupila.

2.3.3 Profundidad

En esta sección se tocarán algunos conceptos sobre la percepción de la profundidad del ojo humano. Para ello se describirá la visión este-

reo, triangulación y para finalmente describir la profundidad del ojo humano.

Visión Estereo

Se trata de un sistema que permite obtener distintas imágenes del objeto o vista observado simultáneamente [76]. Para obtener una imagen 3D utilizamos esta técnica con dos cámaras que apuntan a un mismo objeto o área. Esta técnica utiliza los siguientes pasos [9]:

1. Remover las distorsiones radiales y tangenciales de los lentes. Con ello obtenemos imágenes no distorsionadas.
2. Ajustar los ángulos y distancias entre cámaras.
3. Encontrar las mismas características tanto en la cámara izquierda como la derecha, éste proceso es conocido como correspondencia. La salida de este paso es un mapa de disparidades, donde las disparidades son las diferencias en las coordenadas x en planos de imagen de las mismas características vistas en las cámaras izquierda y derecha: $x^l - x^r$
4. Si conocemos la configuración geométrica de las cámaras, entonces nosotros podemos cambiar el mapa en distancias por triangulación. Esto es conocido como reproyección, y su salida es un mapa de profundidad.

Triangulación

Este es el proceso mediante el cual se calcula un determinado punto P de un espacio físico, a partir de dos imágenes distintas en donde se pueda apreciar el mismo punto [76].

[9] explica el sistema de triangulación de la siguiente forma: “Suponiendo que tenemos dos cámaras cuyos planos son exactamente coplanares con cada uno, con exactamente los mismos ejes que están a una determinada distancia, y con los mismos tamaños de los puntos focales $f_l = f_r$. También se asume que los puntos principales izquierda y derecha han sido calibrados para tener las mismas coordenadas de pixel en sus imágenes izquierda y derecha. [...] Asumiendo que las imágenes están alineadas a las filas y que cada pixel de la fila de la cámara de la izquierda está alineado con cada pixel de la cámara de la derecha. También asumiremos que podemos encontrar un punto P en el mundo físico en las vistas de las imágenes de izquierda y derecha, las cuales corresponden a las coordenadas $x^l - x^r$ ” [9]. En la figura 16 se muestra lo explicado en este apartado.

La fórmula 2.12 explica como obtener la profundidad en donde f es la distancia focal, T es la distancia entre las dos cámaras y $x^l - x^r$ es la disparidad en el punto x izquierda y derecha [9]:

$$Z = \frac{fT}{x^l - x^r} \quad (2.12)$$

Percepción de la profundidad del ojo humano.

El ojo humano percibe la profundidad de tres formas [34]:

1. Según las dimensiones relativas de los objetos.

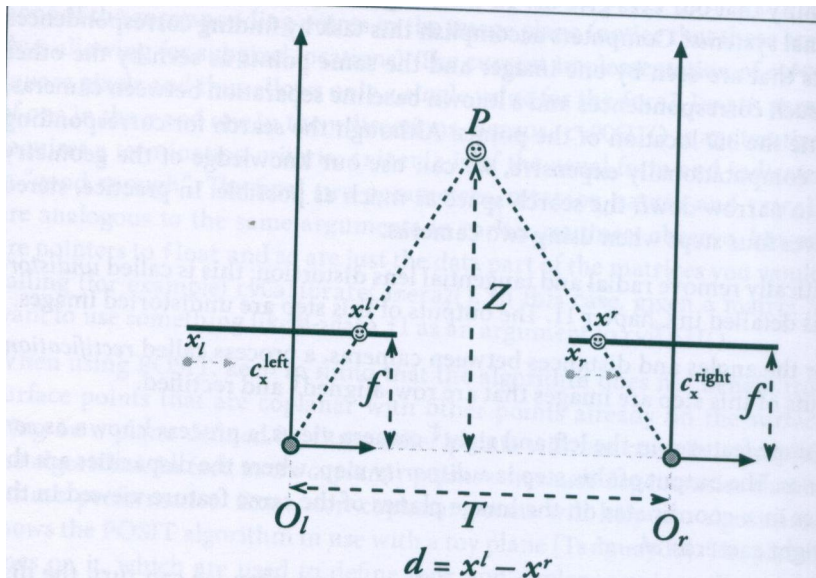


Figura 16: Triangulación de una imagen [9]

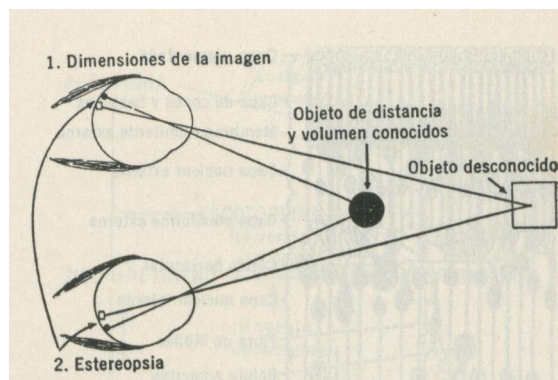


Figura 17: Profundidad del ojo humano por: 1) Las dimensiones de la imagen en la retina. 2) Por estereopsis[34]

2. Por paralaje.
3. Por estereopsis.

En la figura 17 se muestra la percepción del ojo humano.

Determinación de la distancia según las dimensiones relativas de los objetos: Si una persona puede observar un determinado objeto y lo puede ver con un solo ojo, puede determinar a que distancia se halla según la dimensión que tenga su imagen en la retina [34].

Determinación de la distancia por paralaje de movimiento: Cuando una persona mueve su cabeza de un lado a otro las imágenes de los objetos cercanos se mueven rápidamente por la retina, mientras que las imágenes de los objetos distantes quedan casi estacionarias [34].

Determinación de la distancia por estereopsis: Permite determinar las distancias relativas cuando el objeto es cercano. Debido a que la distancia de los ojos se encuentra a poco menos de 5cm cada uno, las imágenes de las dos retinas son diferentes, la unión de las dos imágenes me permiten determinar la profundidad de la misma[34], tal como se lograba con la triangulación de la imagen.

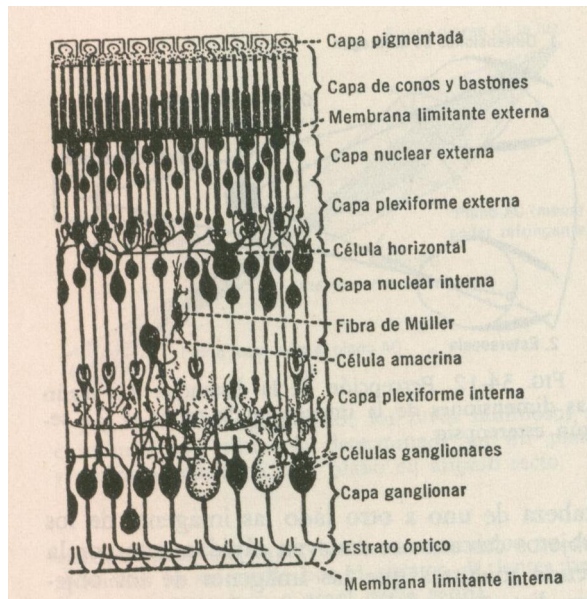


Figura 18: Capas de la retina [34] .

2.3.4 Anatomía y función de los elementos estructurales de la retina

La retina es una parte del ojo que es sensible a la luz, ella contiene los conos y los bastones que son elementos que detectan la luz [34]. En la presente sección se describirán los elementos estructurales de la retina.

Capas de la retina

Las capas de la retina son las siguientes[34]:

1. Capa pigmentada.
2. Capa de conos y bastones.
3. Capa limitante externa.
4. Capa nuclear externa.
5. Capa plexiforme externa.
6. Capa nuclear interna.
7. Capa Ganglionar.
8. Capa de fibras del nervio óptico.
9. Membrana limitante interna

La luz que ingresa en la retina atraviesa de abajo a arriba todas las capas de la misma como se muestra en la figura 18. La luz debe atravesar todas las capas antes de alcanzar las células excitables [34].

Conos y bastones

Los conos y bastones son células que se encuentran en la retina, estas células permiten descubrir determinados colores. Los conos descubren colores específicos, mientras que los bastones pueden distinguir cualquier color que no sea rojo intenso. Cuando los conos y los bastones son

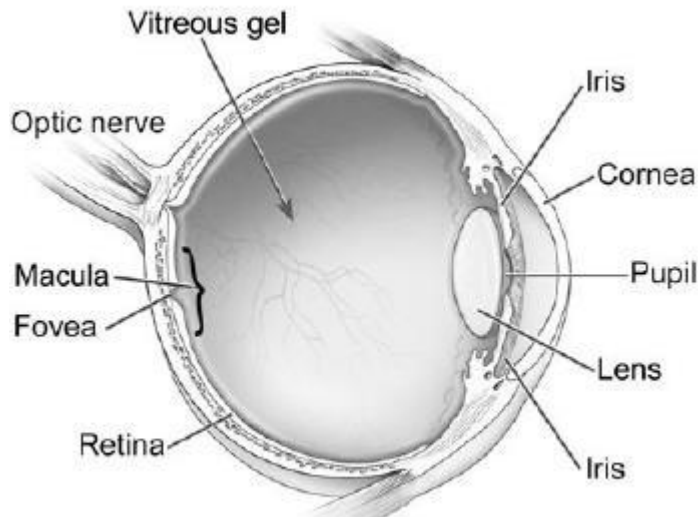


Figura 19: Ojo humano [37].

excitados, los impulsos nerviosos se transmiten por neuronas sucesivas en la propia retina, y, finalmente, por las fibras del nervio óptico hasta la corteza cerebral [34].

Los bastones son muy delgados ya que poseen un diámetro de unas 2 micras en las porciones más centrales de la retina y de 4 a 5 micras en las periféricas. Los conos tienen diámetros de 1.5 micras en la fovea (porción más central de la retina) y de 5 a 8 micras en la periferia [34].

[34] explica el fenómeno de la formación de la visión a nivel químico de la siguiente forma: “Las porciones sensibles de los conos y bastones contienen productos químicos sensibles a la luz que se descomponen cuando quedan expuestos, aunque sólo sea a mínimas cantidades de luz. Los productos de descomposición estimulan las membranas celulares de los conos y bastones, desencadenando impulsos nerviosos que luego son transmitidos al sistema nervioso. Los conos y bastones son neuronas modificadas y pueden considerarse en realidad como formando parte del sistema nervioso”. Los conos se encuentran en la fovea, y los bastones en su gran mayoría se encuentran en la mácula [17]. En la figura 19 se muestra una imagen del ojo con la fovea y mácula.

2.3.5 Función de corteza visual primaria.

El sistema visual puede descubrir la organización especial de la escena visual gracias a la corteza visual primaria [34]. Tal como lo describe [34]: “Puntos específicos de la retina establecen conexión con puntos específicos de la corteza visual; las mitades derechas de las dos retinas se conectan con la corteza visual derecha, mientras que las mitades izquierdas con la corteza visual izquierda. La mácula se halla representada en el polo occipital de corteza visual y las áreas periféricas de la retina se representan por círculos concéntricos cada vez más lejos del polo occipital. [...]Es en el centro de esta región donde se halla representada la fovea, que proporciona el grado máximo de agudeza visual”. En la figura 20 se puede apreciar el córtex visual primario.

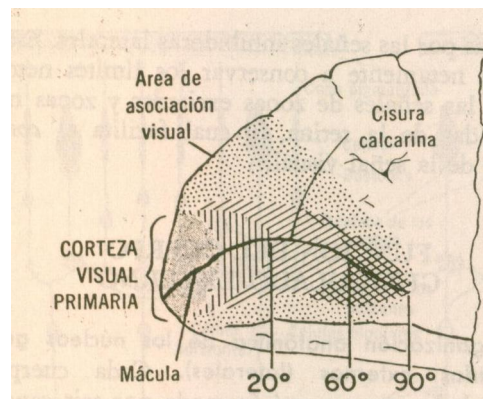


Figura 20: córtex visual primario [34].

Descubrimiento de líneas y contornos por la corteza visual primaria

Cuando una persona ve a un determinado objeto el córtex visual empieza a relacionar los contrastes de la escena en la que está observando dicho objeto. Debido a que los objetos tienen un cambio de oscuro a claro y viceversa, el córtex visual detecta estos cambios. La intensidad de estimulación que recibe el córtex visual depende del gradiente de contraste, es decir, cuanto más neto sea el borde del contraste, y a mayor diferencia de intensidad entre las zonas claras y oscuras, mayor será el grado de estimulación del córtex visual primario [34].

Descubrimiento de la orientación de líneas y bordes: Tal como lo describe [34]: “La corteza visual ha de descubrir no sólo la existencia de líneas y bordes en las diferentes zonas de la imagen retiana, sino también la orientación de cada línea o borde. [...] El córtex visual primario lo hace de la siguiente manera. La corteza visual está compuesta de miles de pequeñas columnas de neuronas, cada una de un diámetro aproximado de 0.5mm, que se extienden desde la superficie de la corteza hasta la profundidad a través de sus seis capas. Cada columna de células es excitada por un borde o una línea orientada en una dirección particular.[...] Así, pues, cada pequeña zona del córtex visual primario tiene múltiples columnas, cada una de las cuales responde específicamente a un borde orientado en una dirección determinada”

Análisis nervioso del color

Tal como lo describe [34]: “Algunas células ganglionarias transmiten señales que proviene de conos sensibles al azul, al rojo y al verde, quedando así representados los tres colores primarios”. Las combinaciones de las células ganglionarias son las siguientes [34]:

1. Algunas células ganglionarias combinan información originada en tres variedades de conos .
2. Dependiendo de las combinaciones las células ganglionarias responden de manera diferente a los colores. Unas responden positivamente al azul, otras al verde, otras al amarillo, etc.

Percepción de la intensidad luminosa

Según [34] “La corteza visual primaria es excitada por: 1) los cambios bruscos de la intensidad luminosa, en particular cuando estos cambios tienen lugar en bordes netos, y 2) el movimiento de la imagen a través de la retina. Ambos

estímulos producen una rápida descarga de las células ganglionares. [...] Estos efectos muestran otra vez que la corteza visual primaria está muy bien adaptada a la percepción de las señales de contraste, a diferencia de la percepción de las señales constantes"

Parte III

CONTENT-BASED IMAGE RETRIEVAL

Content-based image retrieval, también conocido como CBIR, es la aplicación de técnicas de visión por computador para resolver el problema de recuperación de imágenes, ya que analiza los contenidos de la misma. Para realizar este proceso buscamos en una base de datos de imágenes, la cual es dividida en dos grupos: un grupo de entrenamiento, que permite obtener las características de las imágenes, y el grupo de verificación, que permite comparar las características extraídas de ambos grupos a fin de verificarlas (en el capítulo 6 del plan de experimentación se profundizará más en el tema).

En el presente capítulo se explicará algunas generalidades sobre visión por computador, el problema de la visión por computador en relación con el cerebro y el ojo humano, debido a las limitaciones del presente método de detección de imágenes con las características de interpretación del ojo humano.

3.1 HISTORIA DE LA RECUPERACIÓN DE IMÁGENES

3.1.1 *Visión por computador*

La visión por computador es una rama de la Inteligencia Artificial que permite percibir los objetos de una escena [68], para ello se diseñan un conjunto de algoritmos que posibilitan extraer sus componentes tales como texturas, colores, entre otras. Mediante esta técnica se intenta diseñar algoritmos capaces de detectar las características de las imágenes, tales como [74]:

1. Modelado en 3 dimensiones.
2. Detección de rostros.
3. Autenticación visual.

Para lograr la imitación de el ojo humano los investigadores en esta área han estado desarrollando técnicas matemáticas para la recuperación de imágenes [74], tales como filtros matemáticos para el análisis del contorno de la imagen, Content-Based Image Retrieval, entre otros.

Los modelos utilizados en visión por computador son usualmente desarrollados en física y en gráficos por computador. Mediante ellas se intenta describir el mundo como lo miramos, tal como la reconstrucción de las propiedades como iluminación, forma y distribución del color, entre otras[74].

3.1.2 *Aplicaciones de visión por computador*

Algunas aplicaciones de está técnica son [74]:

1. Detección de caracteres ópticos: Lectura de códigos postales escritos, cartas y detección automática del número de placa.
2. Inspección de máquina: Rápida inspección de las partes para aseguramiento de la calidad usando visión estereo con iluminación

especializada, para medir la tolerancia en alas de aviones o partes de cuerpo del automóvil o buscar los defectos en la construcción acerera utilizando rayos X.

3. Imágenes médicas: Registrar imágenes pre-operativas e intra-operativas o realizar estudios de largo plazo de la morfología del cerebro de las personas mientras envejecen.
4. Seguridad en automóviles: Detectar obstáculos inesperados tales como peatones en las calles, bajo condiciones donde las técnicas de visión activa como el radar no funcionan.

3.1.3 Historia de la visión por computador

Década de 1970: En esta época se inicia la investigación en el área de visión por computador, ya que se da la necesidad de recuperar la estructura tridimensional del mundo y usarla, como una forma de entender el mundo a su alrededor, debido a la aparición de nuevas áreas del conocimiento como la robótica. En esta área los primeros intentos de entender la escena incluyeron las técnicas de extracción de bordes e inferir en la estructura 3D el objeto a partir de la estructura de las líneas en 2D [74].

Se introdujeron tres niveles de descripción de un sistema de procesamiento visual [74]:

1. Teoría computacional: Cuál es la meta de una tarea computacional y cuáles son los límites para resolver dicho problema.
2. Representaciones y algoritmos: Cuáles son las entradas, salidas e información intermedia representada, y cuáles son los algoritmos que son usados para calcular el resultado deseado.
3. Implementación de Hardware: Cómo los algoritmos son representados en el Hardware actual.

Década de 1980: Durante esta década se puso más atención en las técnicas matemáticas para realizar análisis sobre las imágenes, en la cual las más importantes fueron las pirámides de imagen, ya que nos permiten realizar tareas como mezcla de imágenes y procesamiento a escala espacial de las imágenes [74]. Luego se introdujo la técnica conocida como wavelets, que como [20] lo explica:

"Wavelets son familias de funciones que se encuentran en el espacio y se emplean como funciones de análisis, examinan a la señal de interés para obtener sus características de espacio, tamaño y dirección". Además, se investigó mejores técnicas para detección de bordes y contorno [74].

Década de 1990: Durante esta época se realizó un gran esfuerzo para resolver los problemas relacionados a la detección de imágenes en movimiento. Después, se modelaron sistemas en 3D que fueron construidos con tales técnicas. Surgió un nuevo campo llamado visión basada en física, en donde se combinaba las medidas de color e intensidad de las imágenes con modelos físicos de transporte de la radiación y formación de imagen [74].

Década del 2000: Esa década se centró en la renderización de imágenes y otras investigaciones como punto de cruz, captura de luz y renderización de luz [74].

En la figura 21 se puede apreciar el cuadro cronológico con los eventos más importantes que se mencionaron anteriormente.

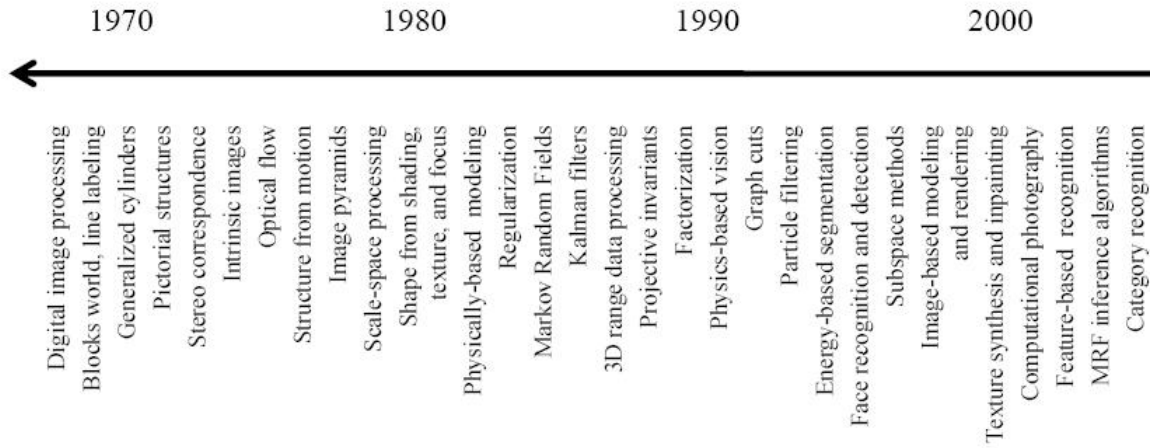


Figura 21: Fechas y eventos en el desarrollo de visión por computador [74]

feature	wang	uw	irma	ucid	zubud	average
color histogram	16.9	12.3	-	51.5	7.8	22.1
LF SIFT global search	37.2	31.5	27.7	31.7	7.0	27.0
LF patches histogram	17.9	14.6	24.9	58.0	24.4	28.0
LF SIFT histogram	25.6	21.4	30.8	50.4	18.3	29.3
inv. feature histogram (monomial)	19.2	12.9	55.8	53.8	7.8	29.9
MPEG7: scalable color	25.1	13.0	-	60.7	32.2	32.7
LF patches signature	24.3	17.6	42.7	68.7	36.5	38.0
Gabor histogram	30.5	20.5	44.9	74.1	24.4	38.9
32x32 image	47.2	26.4	22.8	82.8	27.0	41.2
MPEG7: color layout	35.4	21.2	47.7	75.2	27.0	41.3
Xx32 image	55.9	26.7	21.4	83.2	20.9	41.6
Tamura texture histogram	28.4	16.8	33.0	63.4	84.4	45.2
LF SIFT signature	35.1	20.9	99.3	58.4	20.0	46.7
gray value histogram	45.3	23.0	42.6	86.6	47.0	48.9
LF patches global	42.9	42.7	48.2	63.4	47.8	49.0
MPEG7: edge histogram	32.8	22.9	99.3	69.9	23.5	49.7
inv. feature histogram (relational)	38.3	23.6	39.2	83.2	93.9	55.6
Gabor vector	65.5	37.9	42.5	95.8	73.0	62.9
global texture feature	51.4	32.4	67.7	95.4	98.3	69.0

Figura 22: Precisión de los histogramas con las bases de datos[19] .

3.1.4 Problemas de Visión por Computador y la Visión Humana

En el capítulo 2 se explicó que tanto el cerebro humano como el ojo humano han surgido mediante los procesos de selección sexual y natural respectivamente. Lo que quiere decir que ambos desarrollaron procesos adaptativos durante billones de años, desde la aparición de la primera especie que pudo distinguir el blanco y negro hasta el homo sapiens actual. Uno de los mayores retos del área de la visión por computador es igualar o superar la capacidad de reconocimiento de los objetos en un sistema de computador, debido a que actualmente no se ha logrado obtener la precisión con la cual los seres humanos podemos reconocer objetos. En la figura 22 se puede apreciar la precisión de los histogramas en diferentes bases de datos. Más información en el capítulo 4 sobre histogramas y 6 de experimentación.

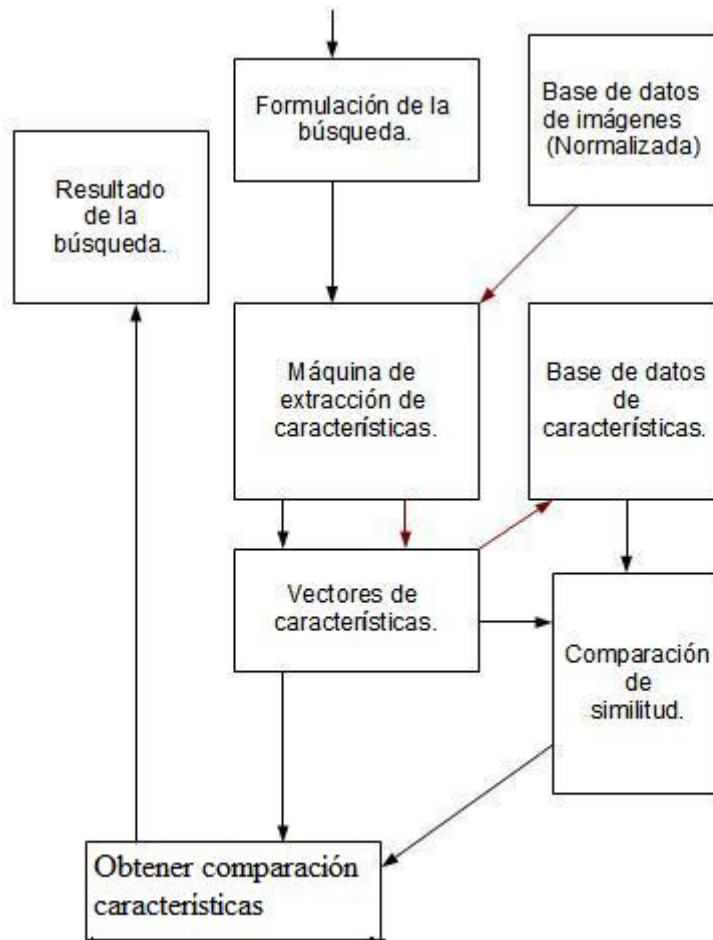


Figura 23: Explicación del proceso de Content-Based Image Retrieval [44]

3.2 PROCESO DE RECUPERACIÓN DE IMÁGENES

El proceso de CBIR consta de los siguientes pasos [44, 21]:

1. Adquisición de imágenes: Obtención de una imagen de una base de datos .
2. Extracción de Características: Extraer las características de las imágenes para poder distinguirlas .
3. Coincidir: Se refiere a que las características coincidan con el de las imágenes que sean parecidas .

En la figura 23 se muestran los pasos de CBIR. Para mayor información sobre los pasos que se van a seguir en este experimento referirse al capítulo 6 de experimentación.

3.3 CARACTERISTICAS DE LAS IMÁGENES

Las imágenes están formadas por tres elementos [26]:

- Color: Representa el conjunto de valores que permiten formar los colores de las imágenes en la luz visible. Los colores de ellas

son producto de la combinación de tres colores: rojo, verde y azul. Mediante su variación se puede formar cualquier color de la naturaleza.

- Bordes: El contorno de los objetos que están dentro de las imágenes.
- Texturas: Es el conjunto de propiedades que tiene una imagen. Entre ellas tenemos: contraste, dirección, polarización, escala, etc.

3.4 MÉTODOS Y APLICACIONES DE LA RECUPERACIÓN DE IMÁGENES.

3.4.1 *Métodos de recuperación de imágenes*

Los métodos de recuperación de imágenes son los siguientes [44]:

- Recuperación a partir del color de las imágenes: Mediante este procedimiento se busca la similitud entre los elementos gráficos del corpus de datos con la finalidad de determinar la similitud entre ellos. Para lograr este objetivo se comparan todas las imágenes entre si para determinar los elementos de cada categoría del corpus.
- Recuperación a partir de la textura de las imágenes: Similar al método anterior, pero con la diferencia que se compara las texturas de las imágenes.
- Recuperación a partir de la forma de las imágenes: Similar a los métodos anteriores, pero con la diferencia que se compara las formas de las imágenes.

3.4.2 *Aplicaciones de la recuperación de imágenes*

Existen diversas aplicaciones de Content-Based Image Retrieval, las más relevantes son [25]:

1. Diagnóstico Médico: Permite el análisis de imágenes de las enfermedades y demás dolencias de los pacientes, a partir de tomografías, radiología e histopatología.
2. Búsqueda web: Permite encontrar imágenes en la web mediante la formulación de una búsqueda, con ello a partir de una imagen se compara con las características que se encuentran en la base de datos de imágenes en el buscador web y se obtiene una respuesta.
3. Sistemas de Información Geográfica: Mediante esta técnica se pueden obtener las características de las imágenes del suelo y mediante algoritmos de Machine Learning se pueden reconocer sus propiedades como calidad de cultivo, nivel de clorofila, entre otras.

Parte IV

HISTOGRAMAS DE COLOR Y
DESCRIPTORES DE COLOR

HISTOGRAMAS DE COLOR Y DESCRIPTORES DE COLOR

4.1 INTRODUCCIÓN

En el capítulo anterior se explicó los conceptos sobre visión por computador y el Content-Based Image Retrieval. Se describió el proceso de CBIR, uno de los pasos es la obtención de las características de las imágenes, para ello utilizamos Histogramas de Color y Descriptores de Color. En el presente capítulo se explicará los conceptos relacionados con el tema.

4.2 TIPOS DE HISTOGRAMAS DE COLOR

Los histogramas son el conteo de datos coleccionados organizados en un conjunto de bins, que son puntos que describen la intensidad del color. Los histogramas pueden representar la distribución de color de una imagen o video. Esta técnica es muy aplicada en el área de visión por computador, tales como Detectar escenas de la transiciones de video mediante el marcado cuando un eje y estadísticas de color cambian de marco en marco[9], o la representación de la distribución de color de una imagen. En la figura 24 y 25 se muestra un ejemplo de un histograma de color. Como se puede apreciar en la figura 25, se ha calculado los valores del histograma de la figura 24, en donde 0 representa el valor del pixel más oscuro, mientras que 255 es el pixel más claro.

4.2.1 Comparación entre dos histogramas

Es la habilidad de comparar dos imágenes a partir de un criterio específico para buscar características similares. Hay que tomar en cuenta que los histogramas tienen el mismo tamaño [9].

Métodos de comparación

Método de correlación: Este método nos permite determinar si las imágenes son similares. Un resultado de 1 es un resultado perfecto, -1 es el peor resultado y 0 no indica relación entre los histogramas. En la ecuación 4.1 se muestra este método en donde $H'_1(i)$ es el valor del primer histograma en la posición i y $H'_2(i)$ es el valor del segundo histograma en la posición i [9].

$$d_{\text{correl}}(H_1, H_2) = \frac{\sum_i H'_1(i) * H'_2(i)}{\sqrt{\sum_i H'_1(i) * H'_2(i)}} \quad (4.1)$$

Método de Chi-Cuadrada: En este caso un valor bajo representa una mayor similitud con el histograma, un valor alto representa una menor



Figura 24: Imagen de la cual se va a extraer el histograma .

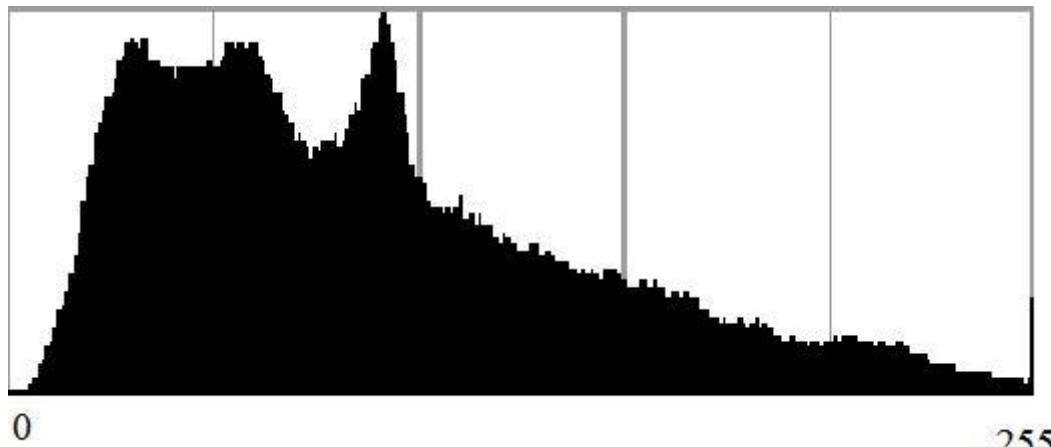


Figura 25: Histograma de la imagen 24.

similitud con el histograma y o es un resultado perfecto [9]. En la ecuación 4.2 $H_1(i)$ es el valor del primer histograma en la posición i y $H_2(i)$ es el valor del segundo histograma en la posición i [9].

$$d_{\text{chi-cuadrada}}(H_1, H_2) = \sum_i \frac{(H_1(i) - H_2(i))^2}{H_1(i) + H_2(i)} \quad (4.2)$$

Método de Intersección: En este caso un valor alto indica una mayor similitud, mientras que un valor bajo indica una menor similitud. En la ecuación 4.3 $H_1(i)$ es el valor del primer histograma en la posición i y $H_2(i)$ es el valor del segundo histograma en la posición i [9].

$$d_{\text{interseccion}}(H_1, H_2) = \sum \text{mín}(H_1(i), H_2(i)) \quad (4.3)$$

Distancia Bhattacharyya: En este caso un valor bajo indica una mayor similitud, mientras que un valor alto indica una menor similitud. Un valor perfecto es 0, mientras que un valor 1 es malo [9].

En la ecuación 4.4 $H_1(i)$ es el valor del primer histograma en la posición i y $H_2(i)$ es el valor del segundo histograma en la posición i [9].

$$d_{\text{Bhattacharyya}}(H_1, H_2) = \sqrt{1 - \frac{\sum_i \frac{\sqrt{H_1(i) * H_2(i)}}{\sqrt{\sum_i H_1(i) * \sum_i H_2(i)}}}{\sum_i H_1(i) * \sum_i H_2(i)}} \quad (4.4)$$

Distancia Euclídea: Este método me permite comparar un histograma con otro. En este caso si es un valor bajo indica un acierto, mientras si indica un valor alto es un falso negativo. En la ecuación 5.4 $H_1(i)$ es el valor del primer histograma en la posición i y $H_2(i)$ es el valor del segundo histograma en la posición i [8].

$$d_{\text{euclídea}}(H_1, H_2) = \sqrt{\sum_{i=1}^n (H_1(i) - H_2(i))^2} \quad (4.5)$$

Tipos de Histogramas

En esta subsección se explicará el histograma de Color RGB, HSV, YUV.

Histograma RGB: El histograma de Color RGB es el acrónimo de Red, Green and Blue: Rojo, Verde y Azul, valores que permiten construir colores a partir de ellos, ya que su combinación permite obtener cualquier tono que se desee [70]. Es el histograma más utilizado en gráficos por computador y se aumenta o disminuye los valores de los componentes en RGB para obtener un nuevo valor. Para la representación de un color se lo describe como se se estuviese trabajando en coordenadas cartesianas. [55].

Debido a que cada componente de RGB consta de 256 colores y cada uno se puede combinar con cualquiera de los otros dos componentes para obtener un determinado color, en total tenemos un valor de 16581375 colores que pueden ser creados a partir de la combinación de colores en RGB [70]. En la figura 26 se puede apreciar la representación

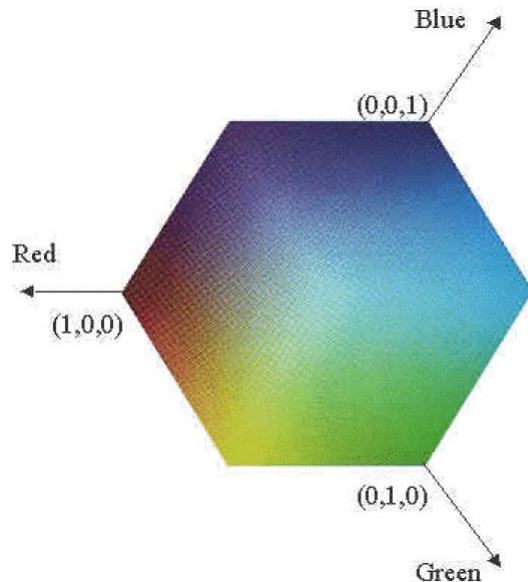


Figura 26: Representación de la gama de colores RGB en el plano cartesiano [41].

la distribución de color en donde se puede ver la formación de colores mediante la combinación de rojo, verde y azul [55].

Histograma HSV: El histograma HSV es el acrónimo de Hue, Saturation and Value. Este histograma es el más cercano a la percepción del ojo humano. Tal como [73] lo explica: “El ojo humano percibe los colores por excitación de los mediante las células del iris: conos y bastones. La excitación de los conos lleva a la percepción del color, mientras que las células de los bastones ayudan a la percepción de varios niveles de gris. [...]El histograma de color HSV separa el componente de luminancia(Intensidad) de un pixel de color para formar sus componentes de crominancia(tono y saturación)”. Esta representación es ideal para el ojo humano, debido a que trabaja como la separación de los conos y los bastones[29].Por ello, cuando la luz ingresa por la pupila mediante un canal, se encarga de dividir en tres señales que estimulan a los conos y los bastones que envían impulsos al cerebro para que puedan analizar el color real [73].

En la representación de los componentes del histograma el tono específica el componente cromático, la saturación indica el valor de predominancia del componente cromático y la luminancia es la intensidad del color [73, 29].

La formación de colores se representa de la siguiente forma: Tal como [73], lo explica el componente cromático define el color por cambiar su ángulo en un rango de $[0, 2\pi]$. Saturación es la profundidad o pureza del color y es medido como la distancia radial del eje central entre 0 en el centro a 1 en el espacio. Finalmente, la intensidad es representada por el eje vertical[29]. En la figura 27 se puede apreciar el espacio de color hsv.

Histograma YUV: El espacio de color YUV es utilizado en los estándares de color para video análogo PAL, NTSC, SECAM [72]. Este espacio de color separa la información del espacio de color RGB en información de luminancia y crominancia que funcionan en las aplicaciones de compresión [30].

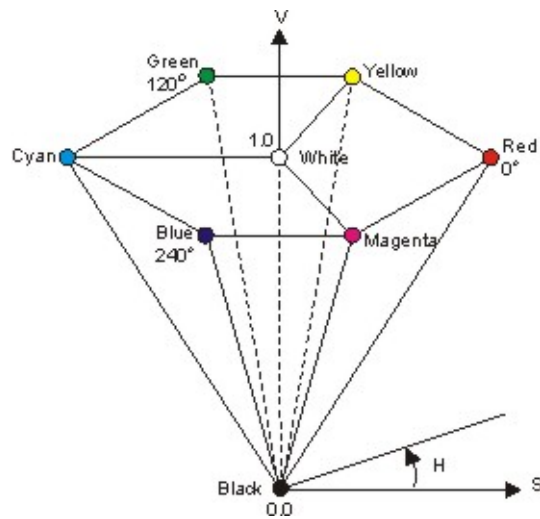


Figura 27: Representación del histograma HSV. En donde V es representado por el principal eje ortogonal del plano. El ángulo representa el valor cromático, mientras que el radio representa el nivel de saturación (Pureza de color) [36].

Los componentes de este espacio de color son Y para representar la luminancia, o intensidad, la cual es adecuada para los artefactos a blanco y negro. Los componentes UV representan la crominancia de este espacio de color [22].

4.3 TIPOS DE DESCRIPTORES DE COLOR DEL ESTÁNDAR MPEG-7

La mayor parte de los descriptores de color se basan en el estándar MPEG-7, el cual fue creado por el grupo de expertos en movimiento de imágenes -MPEG en inglés- que se encuentra dentro de la ISO. Este descriptor se convirtió en el estándar multimedia por excelencia en el año 2001 debido a que describe contenido multimedia al adjuntar metadatos al mismo [54].

4.3.1 Características del estándar MPEG-7

Este estándar se encarga de describir las características del contenido multimedia, de esta forma los usuarios pueden buscar, consultar y obtener el mismo de una forma muy efectiva. El proceso MPEG-7 incluye extracción de características, la descripción del objeto y la aplicación utilizando la descripción del contenido multimedia que se encuentra dentro del mismo [54].

Los ejemplos de búsqueda de MPEG-7 son [54]:

1. Audio
2. Gráficos
3. Imagen
4. Video

En la figura 28 se puede apreciar el proceso mencionado anteriormente.

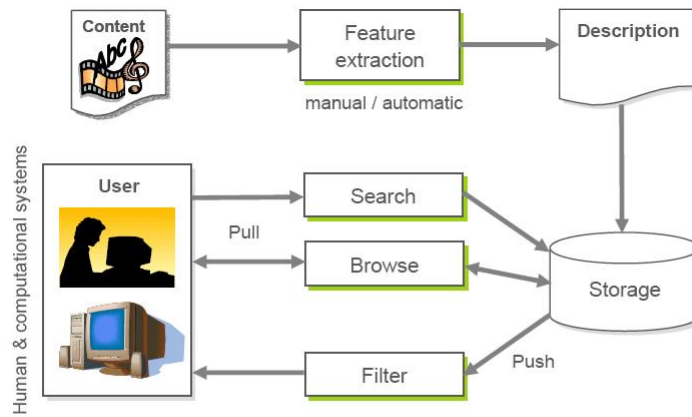


Figura 28: Ejemplo de extracción de MPEG-7 [54].

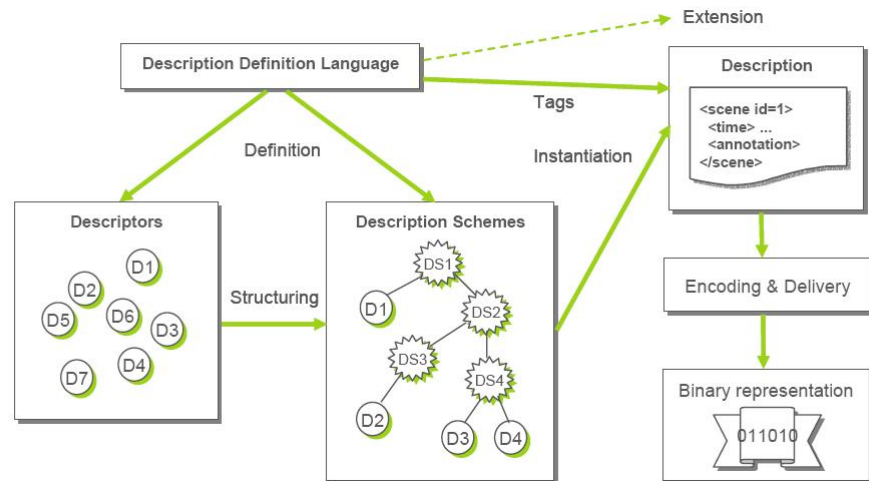


Figura 29: Principales elementos de MPEG-7 [54].

4.3.2 Descripción técnica del estándar MPEG-7

Las principales herramientas para implementar el estándar MPEG-7 son [54]:

1. Descriptores
2. Esquemas de descripción.
3. Lenguaje de definición de descripción

En la figura 29 se muestra los principales elementos de MPEG-7

Descriptores y esquemas de descripción

Un descriptor define la sintaxis y semántica de cada característica, como el color de un histograma, mientras que un esquema de descripción especifica la estructura y semántica de las relaciones entre sus componentes que pueden ser descriptores y esquemas de descripción [54].

Lenguaje de definición de descripción

Este lenguaje permite la creación de nuevos descriptores y esquemas de descripción MPEG-7. Permite la extensión y modificación de los esquemas de descripción existentes. Este lenguaje utiliza el formato XML para ser definido [54].

Formato binario para el formato MPEG-7

Como [54] lo explica: "Las descripciones textuales MPEG-7(Archivos XML) tienden a hacerse muy grandes y en consecuencia muy ineficientes. Esa es la razón el estándar MPEG-7 define un formato binario, BiM, que permite la transmisión y compresión de las descripciones. Actualmente, BiM puede ser utilizado para comprimir cualquier documento XML".

Aplicaciones MPEG-7

Las aplicaciones son las siguientes [54]:

1. Vigilancia de tráfico
2. Compras.
3. Edición multimedia
4. Periodismo
5. Entretenimiento en casa
6. Comercio electrónico
7. Selección de transmisión de medios

4.3.3 *Tipos de descriptores para imágenes*

En esta sección se describirá los tipos de descriptores para imágenes: los descriptores de color, los descriptores de textura, los descriptores de forma, descriptores de movimiento y descriptores de localización [5].

Descriptor de color

Color Space: "Color space consiste en un tipo de datos que especifica el espacio de color en el cual se expresan o trabajan los otros descriptores de color. Los espacios de color que contemplan son los siguientes: "[5]

1. RGB
2. YCbCr
3. HSV
4. HMMD
5. Matriz de transformación lineal con referencia a RGB
6. Monocromático

Color Quantization: Como lo describe [5]: *“Este descriptor es el más conveniente para ser utilizado en imágenes o zonas de ellas, en las cuales un pequeño número de colores es suficiente para caracterizar la información cromática de la región determinada. Sería aplicable por ejemplo en imágenes de banderas o marcas determinadas. En este caso la cuantificación se usa para extraer un reducido número de colores que sean suficientes como para caracterizar la imagen o región. También se calcula una coherencia espacial entre estos colores y dónde están situados, cosa se utilizará en algoritmos de similitud.”*

Color Scalable color: Como lo describe [5] : *“Este descriptor consiste en un histograma de color en el espacio HSV, codificado con una medida de Haar. Su representación se puede escalar de manera que se adecue lo máximo al tamaño de datos con el que se quiere trabajar. Este descriptor es útil en comparaciones imagen-imagen o en búsquedas basadas en características de color. La fiabilidad de la búsqueda aumenta proporcionalmente al número de colores distintos que se tengan en cuenta en el histograma.”*

Color Layout: Este descriptor de color representa la distribución de color de las imágenes en una forma muy compacta [4]. Para calcular el presente descriptor de color se realiza los siguientes pasos [29]:

- Calcular los bloques CLD: En este paso la imagen es dividida en 64 bloques de píxeles. La altura y el ancho de cada bloque son calculados al dividir el alto y el ancho de la imagen en 8. Este proceso es desarrollado por los tres componentes del espacio de color en RGB y de esta forma obtendremos 3 matrices de 64 bloques.
- Representación de la selección de color: Cada bloque debe tener un solo color. Para seleccionar el color representativo, los valores de los píxeles en los bloques son promediados, como se puede apreciar en la figura 30. Luego de ello, cada matriz de 8x8 es transformada al espacio de color YCbCr. En el capítulo 6 se explica como transformar un espacio de color de RGB a YCbCr¹
- Transformada discreta del coseno: Antes de revisar la transformada discreta del coseno, revisaremos algunos conceptos previamente.

Transformada de Fourier: Como lo describe [66]: *“Las series de Fourier pueden describir cualquier señal útil para la ingeniería en un tiempo finito y cualquier señal periódica para cualquier tiempo como una combinación lineal de senoides. Pero no tiene la posibilidad de describir señales aperiódicas para todo el tiempo”*. Por esa razón la transformada de Fourier nos permite describir señales periódicas en tiempo discreto como en tiempo continuo. En este caso se explicará la transformada de Fourier en tiempo discreto.

Transformada de Fourier en tiempo discreto: La transformada de Fourier en tiempo discreto trabaja con un muestreo de la señal, a diferencia de la transformada de Fourier en tiempo continuo, que trabaja con la frecuencia de la función. La fórmula 4.6 muestra la descripción matemática de la transformada de Fourier en tiempo Discreto, en donde

¹ YCbCr es un espacio de color que tiene las mismas características de YUV, la diferencia es que se aplica al estándar digital 72.

$x[n]$ es el valor de la señal original, j es $\sqrt{-1}$ y n es el contador del sumatorio [66].

$$X(F) = \sum_{n=-\infty}^{\infty} x[n] \exp(-j2\pi n) \quad (4.6)$$

Transformada discreta de Fourier: La transformada discreta de Fourier es una transformada utilizada en el análisis de Fourier. A diferencia de la transformada de Fourier en tiempo discreto, esta transformada es finita. La transformada discreta de Fourier está definida por la fórmula 4.7 en donde N es el número de muestras obtenidas de una señal, i representa $\sqrt{-1}$ y k representa el conjunto de valores que puede tomar n [9].

$$f_k = \sum_{n=0}^{N-1} x_n * \exp\left(-\frac{2\pi i}{N} kn\right) \quad (4.7)$$

Transformada discreta del coseno: Está transformada es una variante de la transformada discreta de Fourier, debido a que la transformada discreta del coseno nos proporciona valores reales los cuales nos permiten representar la codificación de datos[9].

Para el Color Layout Descriptor utilizamos las tres componentes de la imagen (Y, Cb y Cr) que son transformados a tres matrices de coeficientes DCT de 8x8 [4]. La transformada discreta del coseno está dada por la siguiente fórmula 4.8 [29].

$$c_k = \alpha_p \alpha_q \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A_{mn} * \cos \frac{\pi(2m+1)p}{2M} * \cos \frac{\pi(2n+1)q}{2N} \quad (4.8)$$

Donde

1. A_{mn} es el coeficiente para las posiciones p, q de la matriz DCT.
2. M es el número de filas de la matriz.
3. N es el número de columnas de la matriz.
4. α_p y α_q son constantes donde los valores se calculan de la siguiente forma.

$$\alpha_p = \begin{cases} \frac{1}{\sqrt{M}} & \text{Si } p = 0 \\ \sqrt{\frac{2}{M}} & \text{Si } 1 \leq p \leq M-1 \end{cases}$$

$$\alpha_q = \begin{cases} \frac{1}{\sqrt{N}} & \text{Si } q = 0 \\ \sqrt{\frac{2}{N}} & \text{Si } 1 \leq q \leq N-1 \end{cases}$$

- Recorrido en zig-zag: Este proceso nos sirve para localizar los componentes de baja frecuencia de la imagen que se encuentran en la parte superior izquierda de la matriz.[8]. Usualmente, el CLD emplea los primeros 6 componentes del color Y, y tres componentes de Cb y Cr [29]. En la figura 31 se puede apreciar el recorrido en zig-zag de la imagen.

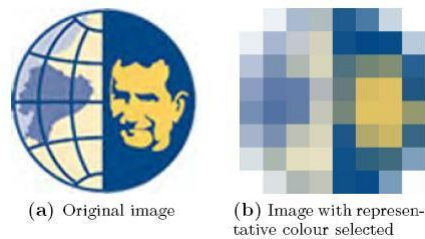


Figura 30: Resultado de aplicar la selección del color representativo de la imagen en el espacio de color RGB₂₉.

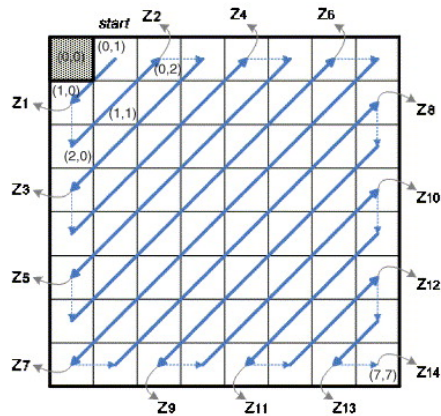


Figura 31: Recorrido en zig-zag aplicado a Y, Cb y Cr del color de la imagen representativo[75].

Color Structure Descriptor: Como lo describe [5]: “Este descriptor caracteriza la distribución de los colores en una imagen. Construye una especie de histograma de color, en el cual tendrán mayor importancia los colores que más se reparten por la imagen. El descriptor divide la imagen en bloques de 8x8 píxeles y analiza dentro de estos bloques los distintos colores que aparecen, incrementándolos así en el histograma. A diferencia de un histograma de color, permite distinguir entre dos imágenes que tengan la misma cantidad de píxeles de un color pero con distinta distribución de estos píxeles. Este descriptor es útil para comparaciones imagen-imagen y añade funcionalidades distintas a las del histograma de color que permiten mejorar la búsqueda de similitud en determinados tipos de imágenes, como por ejemplo las imágenes de naturaleza.”

GoF/GoP Color (Group of Frames/Group of Pictures): Este descriptor se aplica a secuencias de video o de imágenes fijas. Este descriptor da la posibilidad de calcular tres formas distintas este histograma de color [5]:

1. Histograma promedio: “Toma de cada imagen de la secuencia el promedio de los valores del histograma”.
2. Histograma de mediana: “Toma de cada imagen de la secuencia el valor central del conjunto de valores del histograma”.
3. Histograma de intersección: “Toma de cada imagen de la secuencia el mínimo del conjunto de valores del histograma, para así ver cuál es el color “menos común” en el conjunto de imágenes”.

Descriptores de textura

Textura Homogénea: Como lo describe [5] : “Este descriptor emergió como una importante herramienta a la hora de buscar y escoger dentro de grandes colecciones de imágenes de gran similitud visual. Este descriptor utiliza un banco de 30 filtros que permite obtener una afinada descripción de las distintas texturas de la imagen para poder comparar de esta manera con las de otras.”

Búsqueda de texturas: Como lo describe [5] : “Este descriptor especifica la caracterización perceptiva de una textura, la cual es similar a la caracterización de ella que hace un ojo humano, en cuanto a términos de regularidad, tosquedad y direccionalidad. Es útil para búsquedas y clasificaciones a “grosso modo” de texturas.”

Histograma de bordes: Como lo describe [5]: “El Edge Histogram es un descriptor que nos facilita información sobre el tipo de contornos o bordes que aparecen en la imagen. Trabaja dividiendo la imagen en 16 sub-imágenes y es capaz de analizar en ellas el tipo de bordes existentes con el uso de distintos filtros que le permiten diferenciar si son bordes horizontales, verticales, oblicuos o aleatorios. Su utilización principal es la comparación imagen-imagen, especialmente en imágenes de naturaleza con una gran nonuniformidad de contornos. Su uso es muy útil también en combinación con el de otros descriptores como por ejemplo el histograma de color.”

Descriptores de forma

Forma de la región : Este descriptor permite clasificar las imágenes de acuerdo a la región de la misma [5].

Forma del contorno : Este descriptor permite clasificar las imágenes de acuerdo al contorno de la misma [5].

Forma 3D : Este descriptor permite clasificar las imágenes de acuerdo a la malla de las imágenes en 3D [5].

Descriptores de movimiento [5]

Movimiento de cámara : “Es un descriptor que da información sobre los movimientos que efectúa la hipotética cámara que toma la secuencia de imágenes.”

Movimiento de trayectoria : “Este descriptor nos permite analizar la trayectoria de un objeto en una secuencia de imágenes, la cual se consigue con la localización en tiempo y espacio de un punto representativo del objeto determinado.”

Movimiento parabolico : “Consiste en describir el movimiento de ciertos objetos en una cadena de imágenes. Estos objetos se definen como regiones en la imagen, y su movimiento se registra de una manera compacta como un conjunto de parámetros. Este descriptor permite diferenciar numerosos tipos de movimiento elementales como translaciones, rotaciones, ‘zooms’, de manera que cualquier otro movimiento se puede especificar como una combinación de estos.”

Movimiento de actividad : “Se trata de un descriptor que intenta evaluar la ‘intensidad de la acción’ en una secuencia de imágenes, de manera parecida a como lo percibiríamos nosotros los humanos. Es decir, aportar información que permite diferenciar escenas lentas, rápidas, a cámara lenta, etc.”

Descriptores de Localización [5]

Localización de región : “Este descriptor permite la localización de determinadas regiones en una imagen.”

Localización temporal espacial : “Es como el de localización de región pero aplicado a secuencias de imágenes de manera que localiza determinadas regiones analizando tiempo y espacio.”

4.3.4 *Otros descriptores*

En esta sección se explicará otro tipo de descriptores distintos a MPEG-7: SIFT, SURF y ASIFT.

SIFT

SIFT describe la forma local de la región mediante el histograma de bordes [79]. En este descriptor a una imagen se le transforma la información en coordenadas invariantes de escala y rotación [63].

Este descriptor de color permite detectar [43]:

1. Detección extrema de espacio escala.
2. Localización de puntos clave.
3. Asignación de orientación.
4. Descriptor de puntos clave.

La fórmula 4.9 es para el cálculo de éste histograma en donde $g\sigma$ es un kernel Gaussiano isotrópico $\sigma^2 I$ en donde I representa la imagen y σ^2 representa la varianza de la imagen, x es la coordenada espacial y σ es la coordenada escalada [80].

$$G(x; \sigma) \triangleq (g\sigma * I)(x) \quad (4.9)$$

SURF

Surf está inspirado en SIFT y usa la suma de las respuestas Wavelet Haar y la matriz Hessian [6, 63]. Comparado con otros descriptores, el desempeño de SURF es muy superior debido a que usa cálculo de gradientes, las mismas le ayudan a reducir el efecto de los cambios fotométricos [6].

SURF sigue los siguientes pasos [63]:

- Detección de los puntos de interés.
- Localización de los puntos de interés
- Descripción de los puntos de interés.
- Matching.

ASIFT

ASIFT es muy parecido a SIFT. Con la diferencia de que éste método simula tres parametros [63]:

1. Zoom

2. Ángulo de la cámara en latitud.
3. Ángulo de la cámara en longitud.

Para ello siguen estos pasos[63, 61]:

- Afinación de la aproximación local.
- Simulación de distorsiones.
- Matching.

4.4 REDUCCIÓN DE LA DIMENSIONALIDAD

La reducción de la dimensionalidad de los datos se la realiza con la finalidad de reducir el tiempo y los recursos computacionales requeridos para procesar los datos de los descriptores y/o de los histogramas. Debido a que cuando se leen los datos de las imágenes se obtienen n dimensiones que van a ser procesados con la finalidad de resolver el problema de clasificación de las mismas, el tiempo y los recursos que se necesitan para procesarlas es demasiado alto. Por ello se reducen sus dimensiones para que el problema del procesamiento al momento de utilizar los recursos y el tiempo sea disminuido al mínimo posible.

Parte V

TÉCNICAS DE MACHINE LEARNING

5.1 INTRODUCCIÓN

Machine Learning es una combinación entre ingeniería, ciencias de la computación y estadística que nos permite clasificar datos en bruto [35]. La definición formal según el profesor de la Universidad de Carnegie Mellon, Tom Mitchell dice: *"Nosotros decimos que una máquina aprende con respecto a una determinada tarea T, con una métrica de desempeño P, y un tipo de experiencia E, si la fiabilidad del sistema mejora su desempeño P, en la tarea T, siguiendo la experiencia E. Dependiendo como especificamos T, P y E, la tarea de aprendizaje puede ser minería de datos, descubrimiento autónomo, etc. [...]Machine Learning busca las mejores arquitecturas computacionales y algoritmos que pueden ser utilizados eficientemente para capturar, almacenar, indexar, retornar y mezclar estos datos, como múltiples subtareas de aprendizaje pueden ser orquestadas en sistemas más grandes"*. Las áreas en donde puede ser aplicado Machine Learning son entre otras[59]:

1. Psicología.
2. Economía.
3. Biología.
4. Teoría de control.

En el presente capítulo se describirá la naturaleza del conocimiento, los tipos de aprendizaje, como elegir un algoritmo adecuado para resolver un problema puntual y los esquemas de votación Borda.

5.1.1 Naturaleza del conocimiento

Según [85]: *"Conocimiento viene de la acción de conocer: Tener la idea o la noción de una persona o cosa"*, para llegar a ello debemos seguir este proceso:

1. Adquirir datos: Proceso mediante el cual se obtienen datos mediante impulsos externos: Vista, oído, gusto, tacto.
2. Proceso de generación del conocimiento: Procesamiento de los datos en el cerebro humano.
3. Conocimiento generado: Una vez que se procesan los datos, se empiezan a relacionar patrones de los mismos, los cuales producen un entendimiento de los datos y su posterior interpretación.

El proceso que se mostró anteriormente explica como se genera el conocimiento dentro del cerebro humano, el mismo nos posibilita interpretar datos de diversa índole, lo cual nos permite tomar decisiones. Sin embargo, el cerebro humano es muy frágil debido a que sus conexiones nerviosas se suelen afectar con el tiempo. Además, el criterio humano para el análisis de datos es totalmente subjetivo, debido a que los analiza de acuerdo con su experiencia.

5.1.2 *Machine Learning y la naturaleza del conocimiento*

Debido a que el ser humano no puede analizar objetivamente los datos y el conocimiento se puede perder con el tiempo, se ha buscado la forma de mantenerlo e interpretarlo de modo mucho más objetivo. Para ello se crearon algoritmos inteligentes para la adquisición automática del conocimiento y su posterior interpretación, por ello se desarrollaron los algoritmos de machine learning [3]. Para la creación de estos algoritmos utilizamos estadística, debido a que nos permite identificar patrones de la información coleccionada [35].

5.2 MÉTODOS DE MACHINE LEARNING

5.2.1 *Tipos de aprendizaje*

Los tipos de aprendizaje de Machine Learning son: Aprendizaje supervisado y no supervisado.

5.2.2 *Aprendizaje Supervizado*

Aprendizaje supervisado es una técnica que utiliza datos etiquetados para predecir que valores aparecerán de un conjunto de datos no clasificados. En este tipo de aprendizaje nosotros definimos que tipo de valores tienen que ser etiquetados antes de pasar por este proceso. Tenemos que permitir que nuestro algoritmo de aprendizaje supervisado aprenda estos valores para reconocer los patrones de cada categoría. Cuando aparece un nuevo valor el algoritmo de aprendizaje trata de reconocer su etiqueta [68]. En la figura 32 se muestra un conjunto de vinos de cava y champagne que fueron clasificados mediante este algoritmo, para ello se trabajó en pares, lo que permitió determinar a que categoría pertenece cada uno [42].

5.2.3 *Aprendizaje no supervisado*

Es una técnica que trabaja con datos que no están etiquetados. Mediante esta técnica podemos clasificar los datos que no han sido etiquetados, para ello buscamos los patrones del conjunto del sistema [68]. En la figura 33 se puede apreciar este tipo de aprendizaje ya que utilizando un determinado tipo de algoritmo de esta área puede buscar los patrones de cada una de las etiquetas de los datos.

5.2.4 *Conjunto de entrenamiento y prueba para aprendizaje supervisado .*

Cuando trabajamos con Machine Learning utilizamos un conjunto de datos para poder clasificarlo de alguna manera. A este conjunto de datos se lo divide en dos subconjuntos, el uno de entrenamiento y el otro para pruebas [9].

El conjunto de entenamiento nos sirve para que el algoritmo pueda aprender los patrones del conjunto de datos, para que de esta forma el sistema tenga una referencia para poder clasificar los datos.

El conjunto de prueba nos permite comparar este subconjunto de datos con los datos del conjunto de entrenamiento con la finalidad de verificar la eficiencia del algoritmo, para ello se compara cada

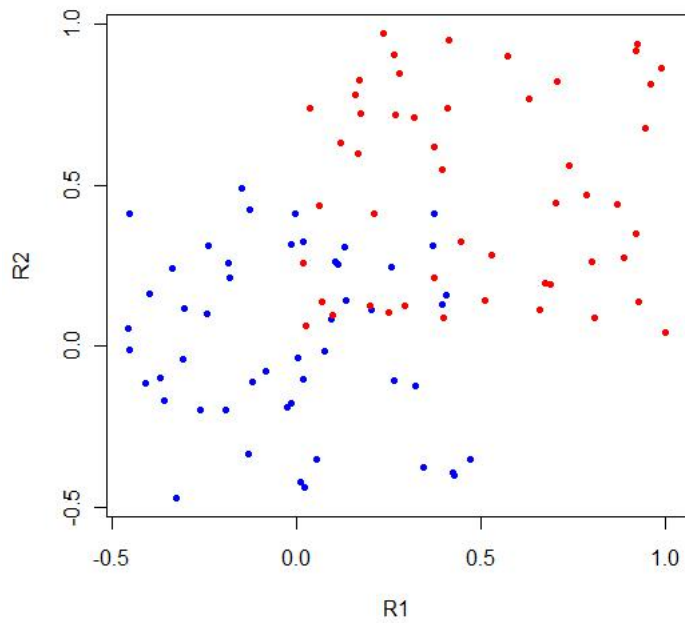


Figura 32: Clasificación genética de la población en donde se etiqueraron los datos $\{(x^1, y^1), (x^2, y^2), (x^3, y^3), \dots, (x^n, y^n)\}$ de tal forma que puedan ser clasificados.[7].

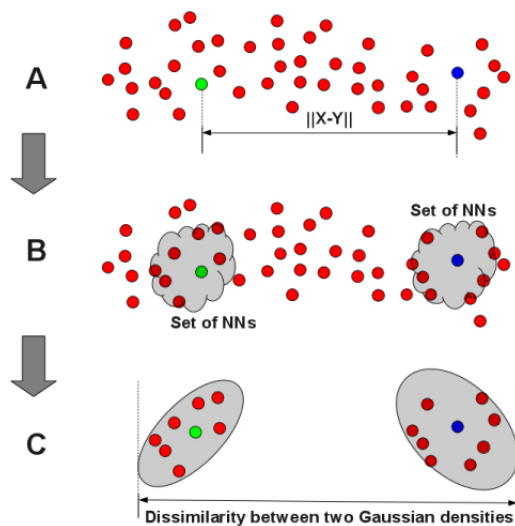


Figura 33: Aprendizaje no supervisado en donde se busca clasificar los datos no etiquetados [1].

elemento del conjunto de prueba con todos los elementos del conjunto de entrenamiento. Esta relación se puede definir mediante la siguiente ecuación matemática que representa el producto cartesiano de dos elementos tales que $A \times B = \{(a, b) | a \in A \wedge b \in B\}$ [52] 5.1 :

$$\{a, b\} \times \{b, c, d\} = \{(a, b), (a, c), (a, d), (b, b), (b, c), (b, d)\} \quad (5.1)$$

5.2.5 Conjunto de entrenamiento para aprendizaje no supervisado

El conjunto de entrenamiento de aprendizaje supervisado es muy parecido a la definición del conjunto de entrenamiento para aprendizaje no supervisado con algunas diferencias. La primera es que los datos no son etiquetados como en el subconjunto de entrenamiento del aprendizaje supervisado. Segundo, los datos son clasificados de acuerdo a patrones de comportamiento comunes. La tercera diferencia es que no se divide en los dos subconjuntos de datos test y train.

Debido a que el conjunto de datos posee elementos que aún no han sido etiquetados, se requiere un método que permita determinar a que categoría pertenece cada uno de ellos. Por esa razón, se utiliza una técnica conocida como clustering, que permite clasificar los datos no etiquetados. Dependiendo del algoritmo que vaya a ser utilizado se manejarán las siguientes técnicas tales como reglas de asociación en el algoritmo a-priori, clustering en kmedias, etc [35]. En el presente capítulo se describirá la técnica de clustering.

5.2.6 Algoritmos de Machine Learning

Naive Bayes

Naive Bayes es un método de clasificación supervisada que se basa en la teoría de decisión bayesiana, la cual escoge los datos del subconjunto en base a la mayor probabilidad de los elementos que se tenga, para ello utilizamos la regla de Bayes que nos permite determinar la probabilidad entre varios elementos del corpus. A partir de ello se puede determinar a que categoría pertenece el elemento.

La fórmula 5.2 permite determinar a qué categoría pertenece el dato, en donde $p(x, y | c_i)$ es la probabilidad de que se produzca la ocurrencia de c_i dados los puntos x, y , $p(c_i)$ es la probabilidad de ocurrencia de c_i [35]:

$$p(c_i | x, y) = \frac{p(x, y | c_i) p(c_i)}{p(x, y)} \quad (5.2)$$

Arboles de decisión

Es un método de clasificación supervisada que como lo explica [9]: “El árbol encuentra una característica en los datos y los mezcla (threshold) en el nodo actual para luego dividir los datos en clases separadas. Los datos son divididos y se repite el procedimiento recursivamente tanto en las ramas derechas como en las izquierdas del árbol. A pesar de que no tiene el mejor desempeño, es el más rápido y tiene la más alta funcionalidad”

Como lo describe [45]: “Para el proceso de clasificación de este tipo de aprendizaje se formulan preguntas, las cuales serán buscadas dentro de cada

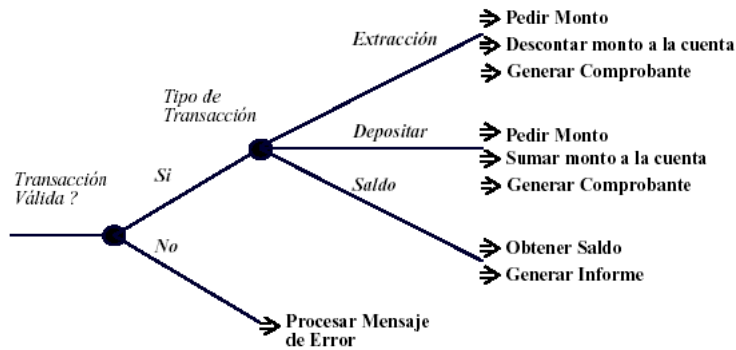


Figura 34: Árboles de decisión de un sistema de información de un banco [18].

nodo. Ya que las preguntas se encuentran contenidas en cada nodo, y cada nodo interno apunta a un nodo hijo para intentar responder a la pregunta formulada. Las preguntas son tratadas como si fuese un árbol donde el camino escogido dependerá de las respuestas elegidas. Las preguntas son muy sencillas, tan solo se pueden responder con un si o un no". En la figura 34 se puede apreciar un ejemplo sobre los árboles de decisión.

K-Medias

K-Medias es un algoritmo de aprendizaje no supervisado creado por McQueen en 1967 que utiliza el método de clustering para clasificar varios elementos que no pueden ser divididos en varios conjuntos de datos, debido a que no pueden ser etiquetados [12]. Tal como lo explica [12] "Es el proceso de agrupar datos en clases o clusters de tal forma que los objetos de un cluster tengan una similitud alta entre ellos y baja con objetos de otros clusters", para ello utiliza varios patrones de comportamiento en los datos que determinarán a qué categoría pertenecen los datos [2]. Los centroides se relacionan con este algoritmo y su relación será explicada a continuación.

Tal como lo explica [12] "El nombre de K-Medias viene porque representa cada uno de los clusters por la media (o media ponderada) de sus puntos, es decir, por su centroide. La representación mediante centroides tiene la ventaja de que tiene un significado gráfico y estadístico inmediato. Cada cluster por tanto es caracterizado por su centro o centroide que se encuentra en el centro o el medio de los elementos que componen el cluster".

El algoritmo de K-Medias se ejecuta de la siguiente forma [9]:

1. "Las entradas son un conjunto de datos y el número de centroides K"
2. "Asignar aleatoriamente las posiciones de los centroides"
3. "Mover los centroides a los datos que son parecidos"
4. "Si no se han clasificado los elementos regresar al paso 3"

Los problemas que se producen con este algoritmo son [9]:

- "K-Medias no garantiza encontrar la mejor solución para posicionar los centroides. Sin embargo, garantiza hallar alguna solución"
- "K-Medias no garantiza cuantos centroides se deban usar. La respuesta varía de acuerdo a su número"

- “K-Medias presume que los datos han sido normalizados”

Las soluciones propuestas para minimizar el error son las siguientes [9]:

- “Ejecutar K-Medias varias veces, cada una con diferentes posicionamientos de los centroides”
- “Comenzar con un centroide y luego aumentar el número de centroides hasta que la solución sea convergente”

5.2.7 Cómo elegir un algoritmo de Machine Learning para resolver un problema puntual

Todo depende de la necesidad puntual que tengamos. Por ejemplo, si deseamos clasificar un conjunto de datos desde cero, entonces necesitaremos un algoritmo de clasificación no-supervizada, pero si deseamos clasificar en base a los datos que se encuentran en un corpus, lo mejor es utilizar clasificación supervisada [35].

5.3 K-NEAREST NEIGHBORS

Los K-Vecinos más cercanos es la técnica de aprendizaje supervisado más simple, debido a que utiliza un sistema de votación para clasificar los datos[9], este es el primer algoritmo diseñado para Machine Learning que trabaja con valores numéricos y nominales que definirán a que categorías pertenecen cada uno de los datos [35]. Para ello utiliza el conjunto de datos de prueba etiquetados que serán utilizados en el proceso de clasificación. Para lograr este objetivo se utilizan las distancias Euclídea, Manhattan, etc. que permitirán definir a qué categorías pertenecen datos que han sido analizados. Estas distancias buscan la menor distancia entre dos elementos. Para ello se compara el elemento elegido del conjunto de prueba con todos los elementos del conjunto de entrenamiento para que de esta forma se pueda determinar la menor distancia entre categorías y así saber a cual pertenece.

Selección de datos

Previamente, en la sección 3.3, se explicó cómo funcionan tanto el conjunto de entrenamiento como el conjunto de prueba para aprendizaje supervisado, para ello se utiliza un corpus de datos que será dividido en dos partes: una para entrenamiento y otra para pruebas. Mediante estos dos conjuntos de datos se aplicará el algoritmo de los K-Vecinos más cercanos [35]. El conjunto de entrenamiento etiqueta a sus elementos para que de esta forma puedan ser clasificados, debido a que cada etiqueta del conjunto de prueba determinará la categoría a la cual pertenecen las imágenes.

Aplicación de K-Vecinos al conjunto de entrenamiento y de pruebas.

Una vez que ha seleccionado los datos se procede a aplicar el algoritmo de los k-vecinos más cercanos, el cual compara todos los elementos del subconjunto de test con todos los elementos del subconjunto de train con la finalidad de obtener un conjunto de votaciones, las cuales nos permitirán determinar a que categoría del subconjunto de datos pertenece el dato que está siendo analizado [35]. Para lograr ello utilizamos

las siguientes fórmulas matemáticas: la fórmula 5.3[46] es la medida de distancia absoluta, la fórmula 5.4 representa a la distancia euclídea.

$$d_A(x, y) = \sum_{i=1}^N |x_i - y_i| \quad (5.3)$$

$$d_E(x, y) = \sum_{i=1}^N \sqrt{(x_i - y_i)^2} \quad (5.4)$$

Una vez que se ha aplicado alguna de las fórmulas expuestas con anterioridad, se busca la distancia más pequeña, ella nos muestra a que categoría pertenece el dato.

Procedimiento que se utiliza en éste método

Una vez que se ha elegido el conjunto de datos y la media de la distancia que se utilizará en este proceso, se procede a aplicar el algoritmo al conjunto de datos que serán analizados. Una vez que se haya dividido el corpus en subconjuntos de test y train se les aplica la distancia para buscar el punto más cercano entre los dos puntos. Recordemos que cada distancia tiene un resultado diferente, por lo que hay que ser muy cauteloso en la forma en como se realiza el experimento.

Un corpus de datos puede tener subgrupos de elementos, los cuales tienen características comunes que permiten descubrir que patrones tiene cada uno, con la finalidad de obtener un determinado valor que nos permitirá definir a que categoría del subconjunto de datos pertenece.

Los pasos descritos por [46] para calcular los K-Vecinos más cercanos son los siguientes:

- “Almacenar los valores de salida del vecino más cercano M para consultar el escenario q en el vector $r = \{r^1, r^2, r^3, \dots, r^M\}$ al repetir el siguiente bucle M veces:”
- 1. “Ir al siguiente escenario s^i en el conjunto de datos, donde i es la actual iteración dentro del dominio $\{1, \dots, P\}$ ”
- 2. “Si q no está en el conjunto de datos o $q < d(q, s^i)$; $q \leftarrow d(q, s^i)$; $t \leftarrow s^i$ ”
- 3. “Iterar hasta que hayamos alcanzado el final del conjunto de datos (i.e. $i = P$)”
- 4. “Almacenar q en el vector c y t en el vector r ”
- “Calcular la media aritmética alrededor de r como muestra la ecuación 5.5:”

$$r = \frac{1}{M} \sum_{i=1}^M r_i \quad (5.5)$$

- “Retornar el valor r del escenario q ”

Movie title	# of kicks	# of kisses	Type of movie
California Man	3	104	Romance
He's Not Really into Dudes	2	100	Romance
Beautiful Woman	1	81	Romance
Kevin Longblade	101	10	Action
Robo Slayer 3000	99	5	Action
Amped II	98	2	Action
?	18	90	Unknown

Figura 35: Etiquetas, número de besos y número de peleas [35].

Ventajas y desventajas de K-Vecinos más cercanos

Las ventajas de este algoritmo son [35]:

1. Alta precisión.
2. Insensible a presiones externas.
3. No se asume ideas previas a los datos

Las desventajas de este método son [35]:

1. Computacionalmente caro.
2. Requiere mucha memoria.

Ejemplo de K-Nearest Neighbors

El siguiente ejemplo intenta clasificar un conjunto de películas en dos categorías, las cuales son romance y acción, para ello se utilizan dos variables que determinaran a qué categoría pertenecen: número de patadas y de besos. Tal como se puede apreciar en la figura 35 se han definido un conjunto de elementos que muestran el número de patadas y besos de cada película, así como la categoría a la cual pertenecen.

En la figura 36 ha aparecido una película que deseamos que sea clasificada en alguna de las mismas que fueron mencionadas anteriormente. Los únicos datos que permiten la clasificación de la misma son el número de patadas y el de besos tal como se puede apreciar en la figura 36 [35].

Finalmente, se comparan los datos de la película que aun no ha sido clasificada con los datos que previamente fueron aprendidos por el algoritmo. Para ello calculamos la distancia euclídea con los elementos que los datos nos han proporcionado. La figura 37 muestra el cálculo de las distancias entre los elementos. Según los resultados obtenidos la película pertenece a la categoría de las películas románticas ya que las distancias con el resto de las categorías obtenidas son: 20.5, 18.7 y 19.2 respectivamente.

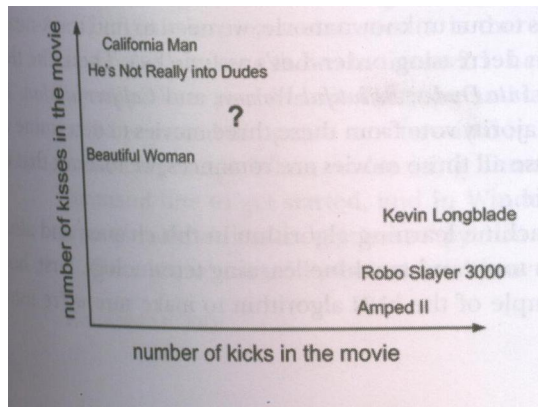


Figura 36: Conjunto de datos de las películas que van a ser clasificadas [35].

Movie title	Distance to movie “?”
California Man	20.5
He's Not Really into Dudes	18.7
Beautiful Woman	19.2
Kevin Longblade	115.3
Robo Slayer 3000	117.4
Amped II	118.9

Figura 37: Cálculo de las distancias [35].

5.4 TÉCNICAS DE COMBINACIÓN DE CARACTERÍSTICAS: ESQUEMAS DE VOTACIÓN BORDA

5.4.1 *Sistemas de votación*

Los sistemas de votación datan del siglo XVIII y han sido aplicados en varias áreas como política, informática, etc [69]. El marqués de Condorcet y el matemático francés Jean-Charles Borda diseñaron dos procedimientos de votación, con la finalidad de desarrollar un sistema que permita que se puntuen a todos los candidatos de mayor a menor, y según el mérito del candidato que sea el ganador, ya que anteriormente los ganadores eran las personas que tenían mayor popularidad[10]. En la presente sección se explicará estos dos métodos.

5.4.2 *Esquemas de Votación Borda*

Los esquemas de votación Borda han sido aplicados a varias áreas de las ciencias de la computación, sobre todo en el área de procesamiento del lenguaje natural [10]. En la presente sección se explicará las características de los esquemas de votación Borda.

Regla de Borda

La regla de Borda fue desarrollada por el matemático francés Jean-Charles Borda en el siglo XVIII, debido a que se dio cuenta que el método utilizado por la Real Academia de Ciencia de París para la

elección de sus miembros no era la más adecuada, ya que los mismos eran elegidos por su popularidad más no por sus méritos. Para solucionar este problema, Borda propuso un método mediante el cual los candidatos puedan ser votados usando un modelo de puntuación de mayor a menor, según el mérito del candidato [10].

En este método se utiliza un conjunto de candidatos, a los cuales se los va a calificar de acuerdo a la puntuación asignada por parte de los miembros del jurado. De acuerdo con [13]: a los valores del conjunto R se le asigna un determinado valor numérico r_i , lo que determinará el peso que tenga cada uno de los elementos. Luego, sumamos estos valores y obtendremos al ganador. Tal como [10] lo define, la regla de votación es:

“Sea $X = \{x_1, x_2, x_3, \dots, x_n\}$ un conjunto finito de candidatos, con $n \geq 3^2$ y m votantes, debe cumplirse la igualdad de $m \geq 3$. Se dice que P es una relación de preferencia sobre X si P es una relación binaria asimétrica, esto es, si $x_i P x_j$, entonces no puede ocurrir $x_j P x_i$. La relación de indiferencia I asociada a una relación de preferencia P recoge la ausencia de preferencia: $x_i I x_j$, lo que significa que ni $x_i P x_j$ ni $x_j P x_i$. La relación de preferencia débil, $P \cup I$ contempla tanto la preferencia como indiferencia: $x_i (P \cup I) x_j$ quiere decir que $x_i P x_j$ o $x_i I x_j$. Sea P^k la relación de preferencia del votante k , ($k = 1, 2, \dots, m$), sobre el conjunto de n alternativas X . A tal relación de preferencia se le puede asociar la matriz 5.6:”

$$\begin{pmatrix} r_{11}^k & r_{12}^k & \dots & r_{1n}^k \\ r_{21}^k & r_{22}^k & \dots & r_{2n}^k \\ \dots & \dots & \dots & \dots \\ r_{n1}^k & r_{n2}^k & \dots & r_{nn}^k \end{pmatrix} \quad (5.6)$$

donde

$$r_{ij}^k = \begin{cases} 1 & \text{Si } x_i P x_j \\ 0 & \text{Encaso contrario} \end{cases}$$

Dependiendo del número de candidatos es posible definir la forma de calificación que se acople a lo que se busca obtener de este método, por ejemplo, la clasificación de imágenes, el etiquetado de roles semánticos, etc. Tal como lo explica [13]: *“Suponiendo que tenemos m alternativas para que el miembro del jurado tenga un rango de valores enteros que van desde m (su alternativa preferida) hasta 1 (su alternativa menos preferida)”*. Para ello se siguen estos pasos [13]:

- Cada miembro del jurado asigna un distinto valor numérico para cada alternativa, de esta forma se define un perfil local de preferencias sin ataduras alrededor de las alternativas.
- Las n categorías son mezcladas en un perfil de preferencia global o calificación.

La mezcla de las categorías se obtiene sumando los valores asignados a cada alternativa, de esta forma se puede obtener el valor de los votos para cada uno de los candidatos [13]. La regla de Borda puede ser resumida en la siguiente fórmula matemática[10]:

$$r_k = \sum_{i=1}^N \sum_{j=1}^M r_{ij} \quad (5.7)$$

	Grupo 1	Grupo 2	Grupo 3
Candidato 1	3	1	2
Candidato 2	2	2	3
Candidato 3	1	3	1

Cuadro 2: Votos del jurado donde en el grupo 1 hay mayor preferencia por el candidato 1, en el grupo 2 por el candidato 1 y en el grupo 3 por el candidato 2 .

Ejemplo de la regla de Borda

Tres candidatos se han presentado como postulantes a la presidencia ejecutiva de una empresa de desarrollo de software, los cuales van a ser calificados por los tres grupos más importantes de la junta directiva de la empresa. En la tabla 2 se puede mostrar el orden de preferencia de los tres grupos.

Una vez que cada uno de los miembros definió el orden mediante el cual cada uno desea un determinado postulante, donde 3 es el concursante con mayor preferencia, 2 el que tiene preferencia intermedia y 1 el candidato con menor preferencia, se procede a definir tres matrices de datos, en donde cada una representa la elección de cada miembro del jurado. En ellas se colocarán valores de 0 donde no exista una elección, caso contrario se colocará un 1.

$$M_{\text{grupo1}} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} \quad (5.8)$$

$$M_{\text{grupo2}} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad (5.9)$$

$$M_{\text{grupo3}} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad (5.10)$$

Una vez que se han calculado las matrices se suman los valores de las columnas de cada uno de los valores de las matrices, de esta forma obtenemos los siguientes resultados:

$$M_{\text{suma1}} = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \quad (5.11)$$

$$M_{\text{suma2}} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} \quad (5.12)$$

$$M_{\text{suma3}} = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} \quad (5.13)$$

Finalmente se suman las matrices que fueron obtenidas en el procedimiento anterior, de esta forma podemos observar los siguientes resultados en donde el candidato 2 obtuvo 4 votos, el candidato 3 obtuvo 3 votos, y finalmente el candidato 1 con 2 votos. Por lo tanto, el candidato 2 es el ganador.

$$\begin{bmatrix} 2 \\ 4 \\ 3 \end{bmatrix} \quad (5.14)$$

5.4.3 Método de Condorcet

Fue planteado por el marqués de Condorcet Marie Jean Antoine Nicolas Caritat el cual es muy parecido al sistema de Votación planteado por Charles Borda, pero tiene algunas diferencias significativas. A diferencia del sistema de Borda donde se elige al candidato por la mayoría de votos, la votación según Condorcet es: *“Si hay una elección que es preferida en toda comparación con otras elecciones, la elegida debe ser la ganadora”* [50]. En esta sección se describirá el presente método y sus características.

La regla básica de Condorcet nos dice: *“Dado un conjunto A de alternativas, este método se basa en la ejecución de comparaciones entre pares alrededor de las alternativas m del conjunto A. [...] Se deben realizar m(m-1)/2 comparaciones para cada miembro que decide, de esta forma podemos evaluar sus preferencias por alguna alternativa”* [13]. Para realizar estas comparaciones debemos seguir estas reglas: Si tenemos n votantes para el par (a,b) tendremos[13]:

- n_1 votantes tienen un ranking $a \geq b$
- n_2 votantes tienen un ranking

Si $n = n_1 + n_2$ se obtiene las siguientes reglas [13]:

- Si $n_1 > n_2$ podemos decir que $a \geq b$ es la elección de n miembros.
- Si $n_1 = n_2$ podemos decir que $a \sim b$ es la elección de n miembros.
- Si $n_1 < n_2$ podemos decir que $b < a$ es la elección de n miembros.

En resumen, el método de Condorcet sigue estas reglas [13]:

- Dada una alternativa $a_i \in A$ la comparamos con las alternativas restantes $m - 1$ alternativas a_j de A con $i \neq j$

- Contamos el número n_1 de veces donde a_i es debilmente preferida por a_j .
- Contamos el número n_2 de veces donde a_j es debilmente preferida por a_i .
- La alternativa de a_i es preferida a la alternativa de a_j si $n_1 > n_2$, alternativamente a_j es preferida a la alternativa de a_i si $n_2 > n_1$ caso contrario (si $n_1 = n_2$) las dos alternativas son indiferentes.

Parte VI

PLAN DE EXPERIMENTACIÓN

PLAN DE EXPERIMENTACIÓN

En el presente capítulo se diseñará el plan de experimentación. En el mismo se describirán las razones porque fueron elegidas las bases de datos UW, Wang y ZuBud. Además, porqué se eligió los histogramas de color RGB, HSV y el descriptor de color Color Layout Descriptor. Finalmente, se explicará la elección del algoritmo de Machine Learning: los K-Vecinos más cercanos y las herramientas utilizadas.

6.1 SELECCIÓN Y PREPARACIÓN DEL CORPUS

Los corpus que fueron elegidos para el presente experimento fueron UW, Wang y ZuBud debido a que estas bases de datos fueron organizadas por varios investigadores de visión por computador con la finalidad de que se puedan realizar experimentos con algoritmos de machine learning sobre ellas. El corpus UW que contiene 1.109 imágenes que representan 18 sitios vacacionales; en este corpus la relevancia de las imágenes se considera de la siguiente forma: *“En el caso de Content Based Image Retrieval, las relevancias son definidas por clases, dada una imagen de consulta de una cierta clase, todas las imágenes de la base de datos de la misma clase son consideradas relevantes. [...] El corpus Wang está formado por 1000 imágenes de la base de datos de fotos Corel las cuales han sido seleccionadas manualmente formando 10 clases de 100 imágenes cada una [19]”*. Las clases que conforman la base de datos Wang son: Africa, playa, monumentos, buses, dinosaurios, elefantes, flores, caballos, montañas y comida. Finalmente, el corpus ZuBud, que es el acrónimo de Zurich Buildings Database for Image Based Recognition está formado por 1005 imágenes de 201 edificaciones, en donde 5 imágenes pertenecen a la misma edificación, pero cada imagen fue tomada en distintas condiciones tales como: clima, perspectiva y la cámara que fue utilizada. Para realizar una comparación de imágenes se utiliza el siguiente procedimiento: dada una imagen de consulta será comparada con el resto de 1005 imágenes, solamente las imágenes que son parecidas a lo edificios serán consideradas relevantes[19]. La figura 38 representa el corpus Wang, la figura 39 representa al corpus ZuBud y la figura 40 representa el corpus UW.

6.2 SELECCIÓN DE LAS CARACTERÍSTICAS A EMPLEAR

6.2.1 Extracción de las características de las imágenes

Primero, se deben procesar las imágenes con la finalidad de obtener sus características antes de realizar el experimento. Para ello extraen las características de las mismas en el espacio de color en RGB. A partir de este espacio de color se convertirá a otros histogramas de color HSV y YCbCr con la finalidad de poder realizar el experimento.



(a) Categoría 1



(b) Categoría 2



(c) Categoría 3



(d) Categoría 4



(e) Categoría 5



(f) Categoría 6



(g) Categoría 7



(h) Categoría 8



(i) Categoría 9



(j) Categoría 10

Figura 38: Imágenes de la base de datos Wang que está dividida en 10 categorías, puede ser descargada de <http://www.cs.washington.edu/research/imagetatabase/groundtruth>



(a) Objeto 1



(b) Objeto 2



(c) Objeto 3



(d) Objeto 4



(e) Objeto 5



(f) Objeto 6



(g) Objeto 7



(h) Objeto 8



(i) Objeto 9



(j) Objeto 10



(k) Objeto 11



(l) Objeto 12



(m) Objeto 13



(n) Objeto 14



(o) Objeto 15



(p) Objeto 16



(q) Objeto 17



(r) Objeto 18

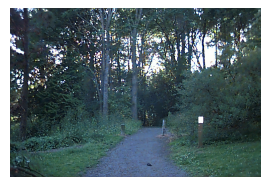


(s) Objeto 19



(t) Objeto 20

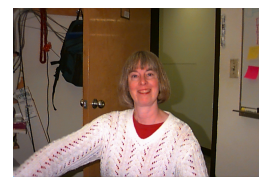
Figura 39: Imágenes de la base de datos ZuBud que está dividida en 201 categorías. En esta imagen solamente se muestran las primeras 20 categorías. Puede ser descargada de <http://www.vision.ee.ethz.ch/showroom/zubud/>



(a) Arboregreens



(b) Australia



(c) Barcelona



(d) Cambridge



(e) Campus in fall



(f) Cannon Beach



(g) Cherries



(h) Columbia George.



(i) Geneva



(j) Green Lake



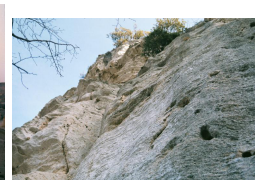
(k) Green Lake



(l) Indonesia



(m) Iran



(n) Italy



(o) Japan



(p) San Juans



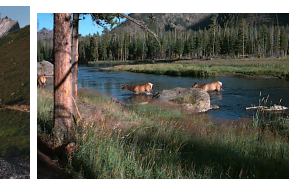
(q) Leaf less trees



(r) Spring Flowers



(s) Swiss Mountains



(t) Yellowstone

Figura 40: Imágenes de la base de datos UW que está dividida 22 categorías. Cada categoría se tiene un número de imágenes aleatorio. Puede ser descargada de: wang.ist.psu.edu/docs/related.

Conversion del espacio de color RGB a HSV y YCbCr

RGB a HSV: Los histogramas de color RGB y HSV representan el conteo de pixeles que tienen las imágenes que están formados por diferentes valores, cada uno con su propia magnitud [8]. Debido a que RGB tiene una estructura por los colores básicos, rojo, verde y azul, pueden formar una gran gama de colores. Sin embargo, el histograma RGB no tiene el nivel de precisión del ojo humano al detectar los colores de su medio ambiente, por ello se desarrolló el histograma HSV que se aproxima a la precisión del ojo humano. Por esa razón es necesario realizar la conversión de RGB a HSV, para ello se utilizan las siguientes fórmulas[55]: La fórmula 6.1 muestra la conversión a Hue, 6.2 muestra la conversión a Saturation, 6.3 ,y representa a Value.

$$H = \cos^{-1} \left\{ \frac{0,5[(R - G) + (R - B)]}{\sqrt{(R - G)^2 + (R - B)(R - G)}} \right\} \quad (6.1)$$

$$S = 1 - \frac{3}{R + G + B} [\min(R, G, B)] \quad (6.2)$$

$$V = \frac{1}{3}(R + G + V) \quad (6.3)$$

RGB a YCbCr: Como se mencionó anteriormente RGB puede formar una gran variedad de colores a partir de los colores Rojo, Verde y Azul. Sin embargo, RGB no es muy bueno en detectar cambios de colores debido a que su estructura no tiene una buena representación del cambio de luz [49]. Por ello era necesario buscar un nuevo sistema de representación de color que permita detectar estos cambios. YCbCr permite detectar estos cambios. Además, de que es el mejor en detectar los cambios de color en las imágenes [49]. La ecuación 7.4 representa el valor de la luminancia y las ecuaciones 6.5,6.6 representan los componentes de cromancia [38].

$$Y = 0,299R + 0,587G + 0,114B \quad (6.4)$$

$$Cb = -0,172R - 0,339G + 0,511B + 128 \quad (6.5)$$

$$Cr = 0,511R - 0,428G - 0,083B + 128 \quad (6.6)$$

6.3 DISEÑO DE LA TÉCNICA DE COMBINACIÓN DE CARACTERÍSTICAS A EMPLEAR

El procedimiento del presente experimento fue realizado en dos partes:

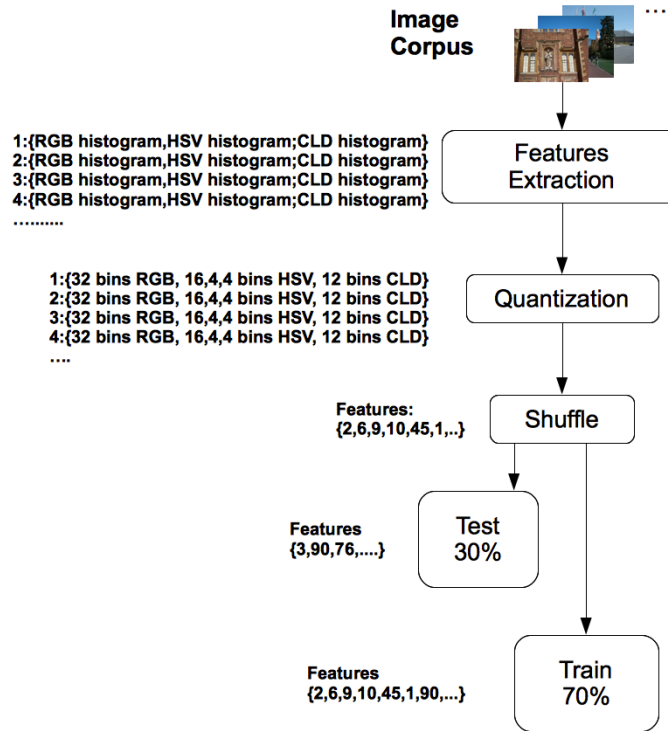


Figura 41: Primera parte del experimento [29].

6.3.1 Parte uno: extracción de características de las imágenes, cuantización de las imágenes, mezcla y subconjuntos de prueba y entrenamiento

En esta sección se mostrará la primera parte del experimento que se puede apreciar en la figura 41.

Extracción de las características de las imágenes

Para la extracción de características de las imágenes utilizamos los histogramas de color a los que nos referimos anteriormente. Antes de comenzar se explicará como se realizó el procedimiento de conversión de RGB a los otros histogramas, ya que como se estudió previamente es la base para la conversión a todos los histogramas.

Cuantización de las imágenes

Una vez que hemos obtenido los histogramas en RGB, HSV y el descriptor de color Color Layout Descriptor, cuantizamos los histogramas, Color Layout Descriptor ya viene cuantizado en 6 valores para Y, 3 para Cb y 3 para Cr. En el caso de RGB y HSV se cuantizará a 256 valores cada uno.

Mezcla de las imágenes

La mezcla de imágenes se la realizará una vez que las hayamos cuantizado. Este proceso se lo realiza para no sesgar los datos cuando se aplica el algoritmo de K-Vecinos más cercanos.

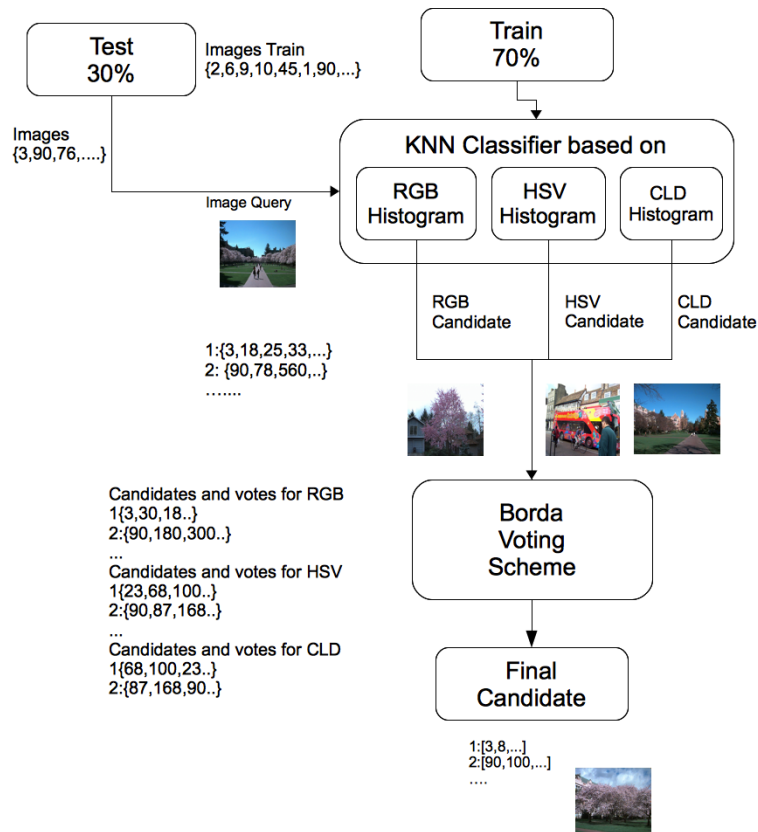


Figura 42: Parte dos del experimento [29].

División del conjunto de entrenamiento y de prueba

En este caso se divide el conjunto de datos en dos subconjuntos de datos: Entrenamiento y prueba. Se realiza este proceso debido a los algoritmos elegidos en Machine Learning. K-Vecinos más cercanos necesitan los dos subconjuntos en entrenamiento. El subconjunto de entrenamiento tendrá el 70% de los valores del conjunto de datos, el resto pertenecerá al subconjunto de prueba.

6.3.2 Parte dos: Aplicación de los K-Vecinos más cercanos y esquemas de Votación Borda

En esta sección se explicará la aplicación de los K-Vecinos más cercanos y esquemas de votación Borda. En la figura 42 se muestra la parte dos del experimento.

Aplicación de los K-Vecinos más cercanos

Una vez que hemos mezclado y dividido el conjunto de datos en entrenamiento y prueba aplicamos el algoritmo de los K-Vecinos más cercanos para obtener un conjunto de votos, los cuales permitirán que se pueda aplicar los esquemas de votación Borda al conjunto de datos.

Esquemas de votación Borda

Mediante este método clasificaremos al conjunto de datos que obtuvimos en el paso anterior para obtener el resultado final. Al aplicar este

proceso obtendremos una matriz de $n \times n$ en donde el candidato que resulte con el mayor número de votos será el ganador.

Herramientas utilizadas en el experimento

Las herramientas utilizadas para este experimento fueron:

- **OpenCV** versión 2.4.0: Es una herramienta de uso libre que permite programar funciones para visión por computador. Esta herramienta permite el uso de más de 2500 algoritmos optimizados que permiten la manipulación de imágenes. OpenCV puede ser ejecutado tanto en Mac, Windows y GNU/Linux, y puede ser utilizado en C, C++, Python y próximamente en Java [62].

OpenCV utiliza las siguientes librerías[11] [64]:

1. **CxCore**: “Es la librería que posee las estructuras y algoritmos básicos que usan las demás funciones.”
 2. **CV**: “Implementa las funciones principales de procesamiento”.
 3. **HighGui**: “Sirve para realizar la interfaz gráfica e interacción con archivos de imágenes y video”.
 4. **ML**: “Posee algoritmos de aprendizaje y clasificadores”.
 5. **CVAux**: “Admiten algoritmos experimentales BG/FG”.
- La versión 4.6 del lenguaje de programación de alto nivel C++ para Unix.
 - La versión 2008 express edition de el lenguaje de programación de Microsoft Visual C++.
 - El editor de texto Microsoft Word 2007.
 - El lenguaje de programación de alto nivel Java 1.6.1

6.3.3 *Diseño del prototipo*

En esta subsección se describirá el diseño del prototipo para el desarrollo del experimento. Para ello se definieron 5 módulos en el sistema tal como se puede apreciar en la figura 43.

Módulo para procesar el histograma RGB

Este módulo fue programado en C++ en el sistema operativo Ubuntu Hardy en OpenCv 2.4.0 y realiza las siguientes operaciones. Primero extrae características de las imágenes son extraídas. Segundo, se realiza un conteo general de los pixeles, ello permite la construcción del histograma en RGB. Luego, las características de las imágenes son cuantizadas a 256 bins. Finalmente, los bins cuantizados son escritos en un archivo de word con la finalidad de que puedan ser utilizados con el algoritmo de los K-Vecinos más cercanos tal como se puede apreciar en la figura 44.

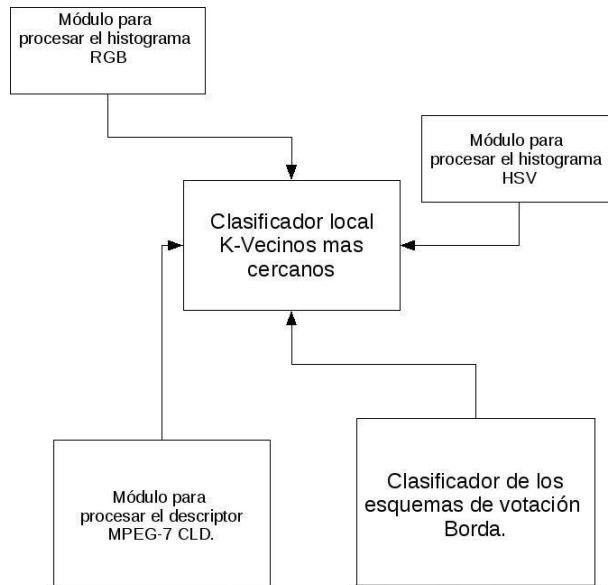


Figura 43: Diagrama que explica las relaciones entre los módulos del sistema.



Figura 44: Módulo para procesar el histograma RGB que muestra las distintas etapas para generar el archivo con los datos del histograma RGB.

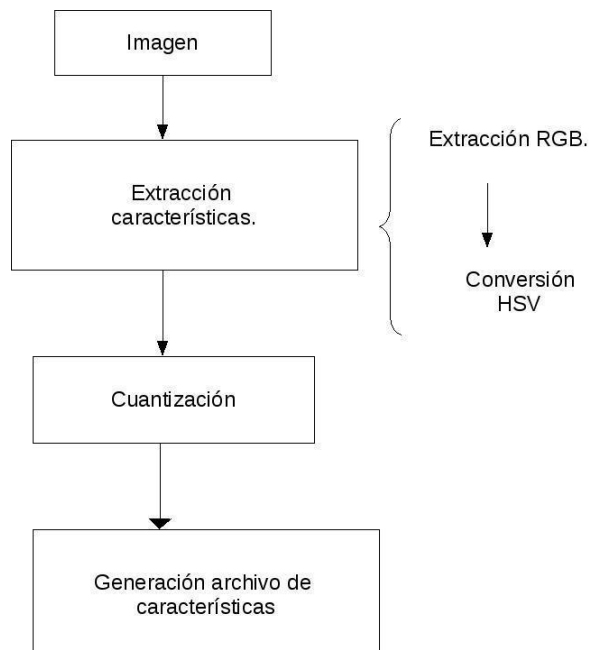


Figura 45: Módulo para procesar el histograma HSV que muestra las distintas etapas para generar el archivo con los datos del histograma HSV

Módulo para procesar el histograma HSV

El módulo que permite procesar el espacio de color HSV es muy parecido al de RGB. Sin embargo, tiene dos diferencias que serán mencionadas a continuación. Primero, se realiza una conversión del espacio de color RGB a HSV. La segunda diferencia radica en la forma que fue cuantizada la imagen, ya que el estándar de OpenCV define el valor para H de 180 grados y para S y V de 256. Los valores que son generados en este módulo es escrito en un archivo de word, el cual será usado por el algoritmo de los K-Vecinos más cercanos tal como se puede apreciar en la figura 45.

Módulo para procesar el descriptor de color en MPEG-7 Color Layout Descriptor

Este módulo fue programado en el lenguaje de programación C++ con OpenCV en el sistema operativo Ubuntu Hardy siguiendo los pasos que se detallan a continuación. Primero, se extraen las características de color en RGB. Segundo, se transforman las características de color en RGB a YCbCr. Luego, se divide la imagen en una matriz de 8x8. Después, la transformada discreta del coseno es aplicada para disminuir las frecuencias de baja frecuencia. Finalmente, los coeficientes de CLD son escritos en un archivo de word que será utilizado por los K-Vecinos más cercanos tal como está descrito en la figura 46.

Clasificador local de los K-Vecinos más cercanos

El módulo de los K-Vecinos más cercanos fue programado en Visual Studio 2008 Express Edition y sigue estos pasos. Primero, un modelo para extraer de las características de color de los archivos de texto que fueron mencionados anteriormente. Segundo, las características de color son mezcladas con la finalidad de no sesgar los resultados.

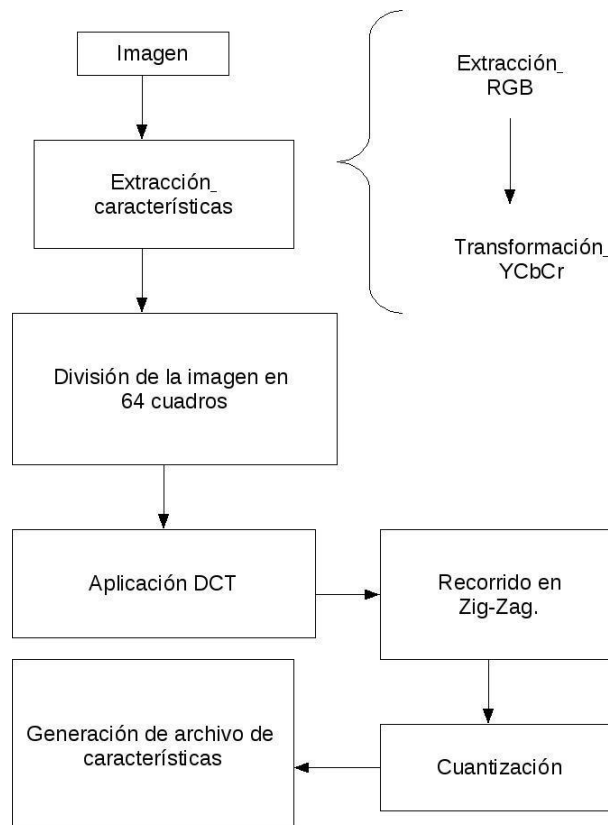


Figura 46: Módulo para procesar el histograma CLD que muestra las distintas etapas para generar el archivo con los datos del histograma HSV .

Después de ello, las características son separadas en dos subgrupos: entrenamiento que tendrá 70% de los elementos y el subgrupo de prueba con 30% de los elementos. Luego, los elementos del set de prueba son comparados uno a uno con los elementos de la sección de entrenamiento. Finalmente, los resultados son escritos en un archivo de word para que sean utilizados por los esquemas de votación Borda. En la figura 47 se puede apreciar el proceso de generación de los resultados KNN-RGB, mientras que en la figura 48 se muestra la generación de datos por parte de KNN para el espacio de color HSV, y finalmente en la figura 49 se observa la generación de datos por parte de KNN.

Clasificador de los esquemas de votación Borda

Este módulo fue programado en Java y sigue los pasos que se detallan a continuación. Primero, lee los archivos de word en donde se encuentran los resultados de los esquemas de votación Borda. Segundo, obtiene los resultados de cada uno de los algoritmos de KNN y los usa para formar matrices que representan la decisión de los votantes. Finalmente las suma y obtienen los resultados tal como se muestra en la figura 50.

6.3.4 *Desarrollo de algoritmos en este sistema*

Para el desarrollo de este sistema se obtuvo los valores de RGB y HSV y se los cuantizó mediante OpenCV en GNU/Linux. Mediante el algoritmo 6.1 se puede apreciar como cuantizar y obtener los valores en RGB . En el algoritmo 6.2 se puede apreciar como cuantizar y obtener

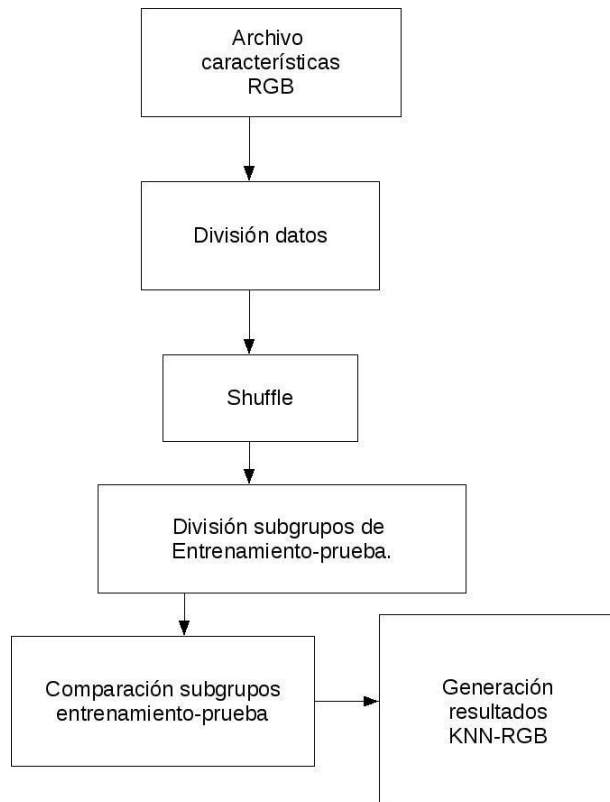


Figura 47: Esquema que muestra los pasos que permiten generar los resultados del algoritmo RGB.

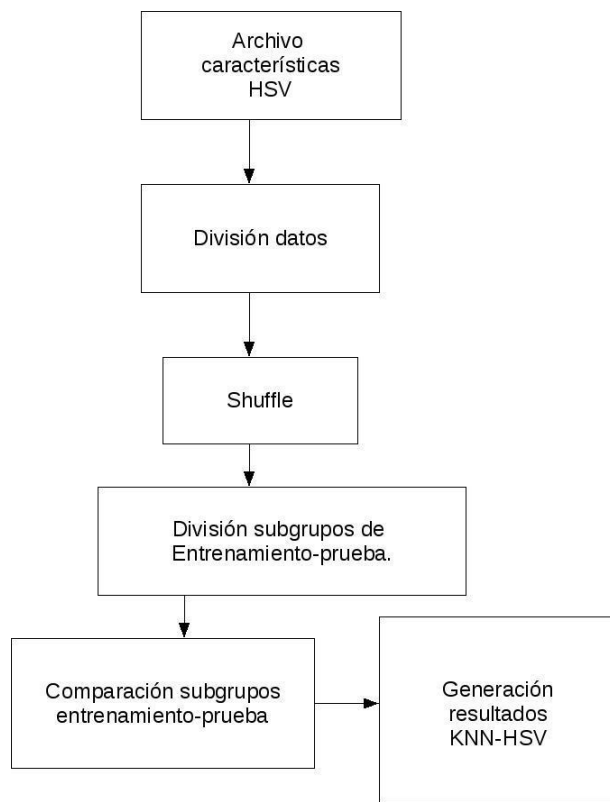


Figura 48: Esquema que muestra los pasos que permiten generar los resultados del algoritmo HSV .

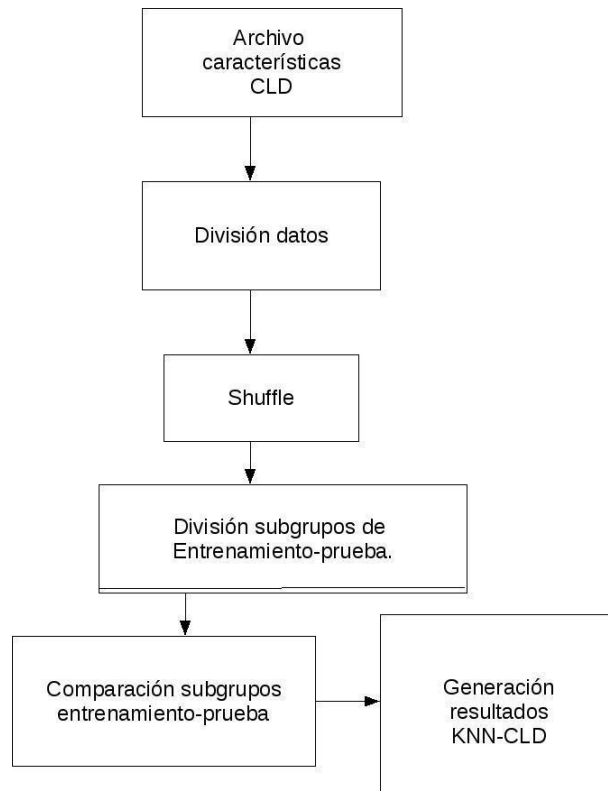


Figura 49: Esquema que muestra los pasos que permiten generar los resultados del algoritmo CLD .

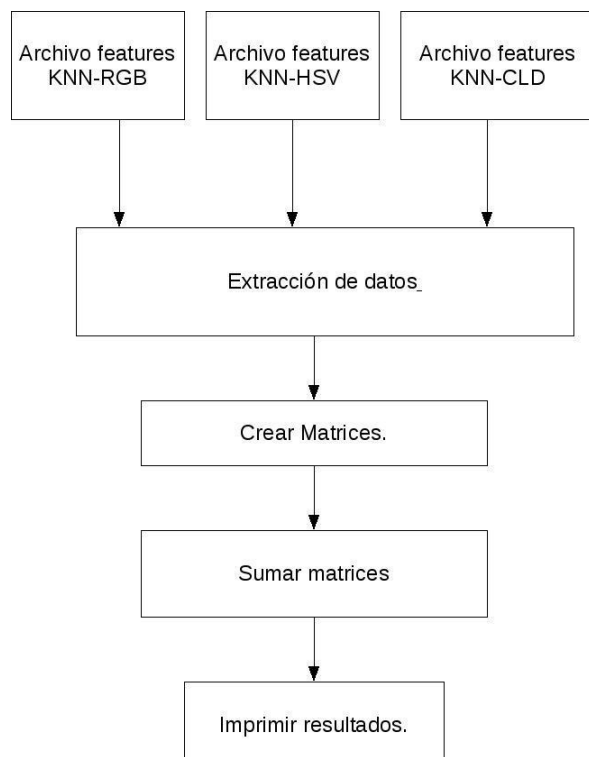


Figura 50: Proceso mediante el cual se ejecuta el módulo de los esquemas de votación Borda.

Algoritmo 6.1 Obtencion de los valores del histograma en RGB y su respectiva cuantización .

```

/* Librerias de Open Cv para el procesamiento de imagenes */
#include "LeerArchivo.h"
#include <cv.h>
#include <highgui.h>
#include <opencv.hpp>
using namespace std;
/* Metodo principal */
int main() {
cv::Mat rgb; /* Definicion de una matriz mediante la cual obtendran
los valores de RGB */
/* Leo la imagen, si no le leyo correctamente el sistema retorna -1 */
if( !(rgb=cv::imread("hola.jpg", 1)).data )
return -1;
int histSize[] = {hbins, sbins,vbins}; /* Defino el tamaño del histograma
*/
float rgranges[] = { 0, 256 }; /* Valores que puede obtener el histograma
en RGB, como se vio 0-255 */
const float* ranges[] = { rgranges, rgranges , rgranges }; /* Rangos
de 0-255 */
cv::MatND hist; /* Matriz que tendra los valores cuantizados */
int channels[] = {0, 1,2}; /* Canales a ser leidos R, G, B */
cv::calcHist(&hsv, 1, channels, cv::Mat(),hist, 3, histSize, ran-
ges,true,false); /* Cuantizacion del histograma */
}

```

los valores en HSV. No se podrá incluir los algoritmos de Color Layout Descriptor, K-Vecinos más cercanos y los esquemas de votación Borda debido a que son demasiado extensos.

6.4 DISEÑO DEL PLAN DE EXPERIMENTACIÓN

6.4.1 *Objetivo del experimento*

El objetivo el presente trabajo de experimentación consiste en la combinación de los histogramas de color RGB, HSV y el estándar MPEG-7 Color Layout Descriptor mediante el uso del algoritmo de aprendizaje supervisado K-Vecinos más cercanos y los esquemas de votación Borda con la finalidad de mejorar la precisión y cobertura en el proceso de recuperación de imágenes. Para ello se explicará los métodos, las variables y las funciones aplicadas dentro del mismo al momento de extraer las características de las imágenes, aplicar los algoritmos sobre ellas y los resultados que fueron obtenidos.

6.4.2 *Descriptores de color elegidos*

Los descriptores elegidos para este experimento fueron los histogramas de color RGB, HSV y CLD. Las razones para su elección fueron las siguientes:

1. RGB: Es el histograma más utilizado en el área de visión por computador [55], y constituye el histograma base del conjunto de histogramas de Color y en el es el más fácil de formar cualquier

Algoritmo 6.2 Obtención de los valores del histograma en HSV y su respectiva cuantización

```
/* Librerias para la lectura de imagenes */
#include "LeerArchivo.h"
#include <cv.h>
#include <highgui.h>
#include <opencv.hpp>
using namespace std;
/* Metodo principal */
int main( int argc, char** argv ) {
cv::Mat src; /* Matriz que obtendra los valores en RGB */
int hbins = 16, sbins = 4 , vbins = 4; /* Valores que seran utilizados
para cuantizar el histograma */
/* Si la imagen no puede ser leida */
if( !(src=cv::imread("hola.jpg", 1)).data )
return -1;
cvtColor(src, hsv, CV_BGR2HSV); /* Conversion de RGB a HSV */
int histSize[] = {hbins, sbins,vbins}; /* Tamaño del histograma */
float hranges[] = { 0, 180 }; /* Valores que obtendra h */
float svranges[] = { 0, 256 }; /* Valores que obtendra sv que son los
mismos en openCv */
const float* ranges[] = { hranges, svranges , svranges }; /* Defino los
rangos de la matriz */
cv::MatND hist; /* Histograma con los valores cuantizados */
int channels[] = {0, 1,2}; /* Tres canales para la lectura */
cv::calcHist(&hsv, 1, channels, cv::Mat(),hist, 3, histSize, ran-
ges,true,false); /* Calculo del histigrama cuantizado */
}
```

tipo de color de la naturaleza, a partir de el se puede convertir a cualquier histograma de color. Por esa razón, se ha elegido el histograma en RGB para combinarlo con los otros dos histogramas [51].

2. HSV: Es el histograma con mayor precisión de todos debido a que intenta emular la capacidad de detectar el color de la misma forma que lo haría el ojo humano y es el más eficiente ya que como lo menciona [73] es el más adecuado en segmentación de imagen y generación de aplicaciones que son muy importantes en la aplicación de CBIR.
3. CLD: Según [71]: “Este descriptor está designado para describir la distribución de color en una región arbitrariamente marcada”. Mediante ella se puede representar la distribución de color, lo que permitirá un mejor resultado en el análisis de color.

Cuantización de los histogramas

Este proceso permite reducir la dimensionalidad de los bins de las imágenes que fueron extraídos con la finalidad de mejorar su manejo por parte del algoritmo de los K-Vecinos más cercanos y por eficiencia en el computador. Como se pudo apreciar en el capítulo anterior los K-Vecinos más cercanos son diseñados de tal forma que pueden procesar los datos al 100% gracias a la forma en como ha sido definido el algoritmo. Sin embargo, la cantidad de memoria que se requiere para procesar el número de bins de las imágenes es demasiado grande, lo que consumiría la memoria del sistema, los vectores de características serían demasiado grandes para ser procesadas, se utilizaría demasiado tiempo en manejar los vectores. Por ello se reduce la dimensionalidad de los elementos.

6.4.3 *Unidades experimentales*

Para este experimento se utilizarán 1000 imágenes de la base de datos Wang, 1005 imágenes de la base de datos ZuBud y 1219 imágenes de la base de datos UW. Las imágenes fueron divididas en dos subgrupos de prueba y entrenamiento. El grupo de train tuvo los siguientes valores: UW tuvo 844 imágenes, 700 imágenes para Wang y 603 para ZuBud. Para test se tuvieron los siguientes valores: 375 para UW, 300 para Wang y 402 para ZuBud.

6.4.4 *Elección del algoritmo de Machine Learning y objetivo de aplicación de los esquemas de votación Borda.*

El algoritmo elegido para realizar el experimento es el de los K-Vecinos más cercanos por las siguientes razones. Primero, el experimento que se va a realizar requiere un algoritmo de aprendizaje supervisado, debido a que los datos han sido pre-etiquetados y en consecuencia se puede determinar a qué categoría pertenecen con el manejo de las mismas. Segundo, este algoritmo garantiza una cobertura del 100% gracias a su estructura que no permite que se produzcan pérdidas durante su ejecución. Finalmente, los K-Vecinos más cercanos generan un conjunto de votaciones que sirven para clasificar las imágenes. En este caso se utilizarán los votos en el algoritmo de los esquemas de votación Borda.

Parte VII

ANÁLISIS DE RESULTADOS

ANÁLISIS DE RESULTADOS

7.1 ANÁLISIS DE LA PRECISIÓN, COBERTURA Y F-MEASURE

7.1.1 Ejecución del plan de experimentación

El plan de experimentación se ejecutó en cinco etapas que serán descritas a continuación. Primero, se pone en marcha el módulo para procesar el diagrama RGB, una vez que este diagrama obtuvo los bins cuantizados en RGB se procede a ejecutar el módulo para procesar el diagrama HSV. Luego, se procede a ejecutar el módulo para procesar el descriptor MPEG-7 Color Layout Descriptor. Una vez que se hayan obtenido los elementos de las imágenes se procede a utilizar sus resultados en el clasificador local de los K-Vecinos más cercanos. Finalmente, con los resultados del módulo anterior se procede a ejecutar el clasificador de los esquemas de votación Borda.

Ejecución del módulo para procesar el diagrama RGB

La ejecución fue realizada en Ubuntu Hardy. Para ello se llevo a cabo el siguiente procedimiento:

1. Se ingresó al directorio donde se encuentra el archivo ejecutable generado en C++.
2. Si es necesario se cambia el nombre del formato de las imágenes que se desea procesar.
3. Se definió en qué lugar y con qué nombre se desean guardar las características extraídas de la imagen.
4. Colocar el identificador de la categoría, el mismo será usado en el algoritmo de los K-Vecinos más cercanos.
5. Compilar y ejecutar el archivo binario para extraer las características de la imagen.

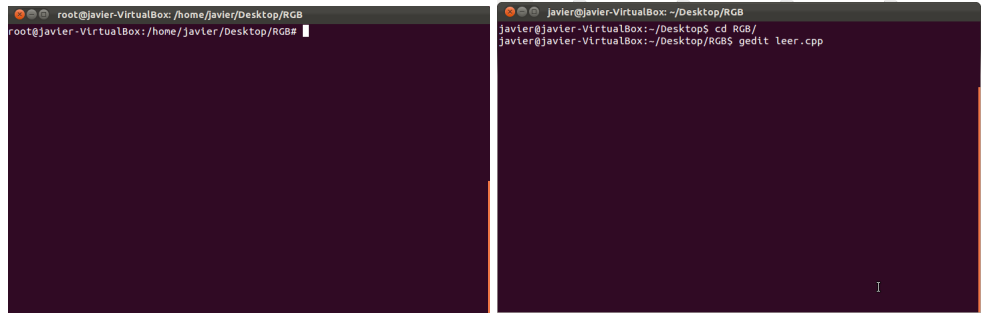
En la figura 51 se puede apreciar el proceso mencionado anteriormente.

Ejecución del módulo para procesar el diagrama HSV

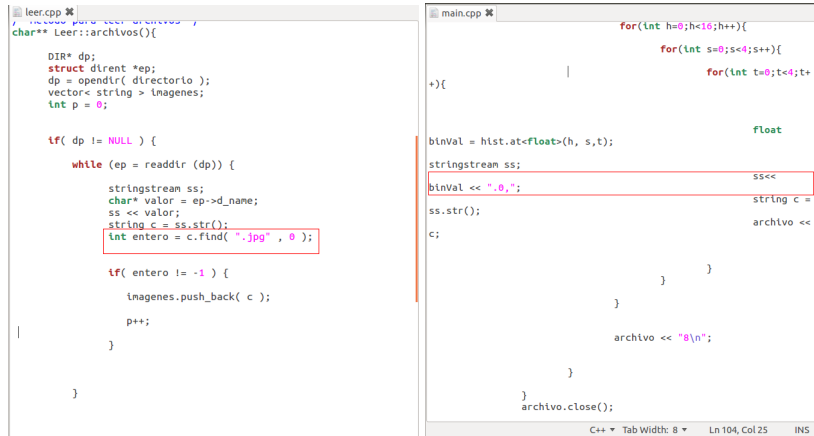
La ejecución de este módulo es muy similar a la del módulo para procesar el diagrama en RGB. En la figura 52 se puede apreciar el procedimiento.

Ejecución del módulo para procesar el estándar de color MPEG-7 Color Layout Descriptor

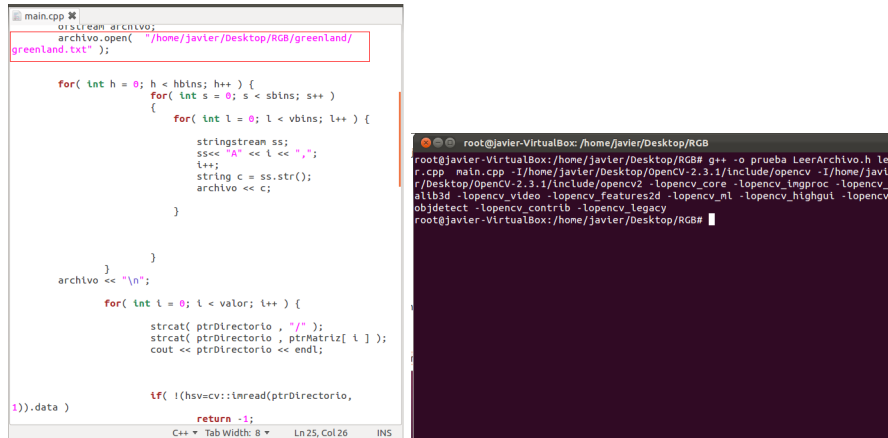
La ejecución de este módulo es muy similar a la del módulo para procesar el diagrama en RGB. En la figura 53 se puede apreciar el procedimiento.



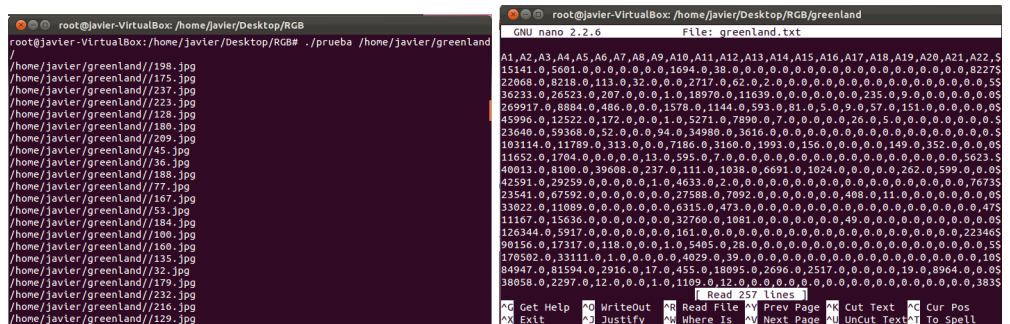
(a) Ingreso al directorio donde se encuentran los archivos de C++ (b) Ejecutamos un editor de texto para abrir el lector de imágenes .



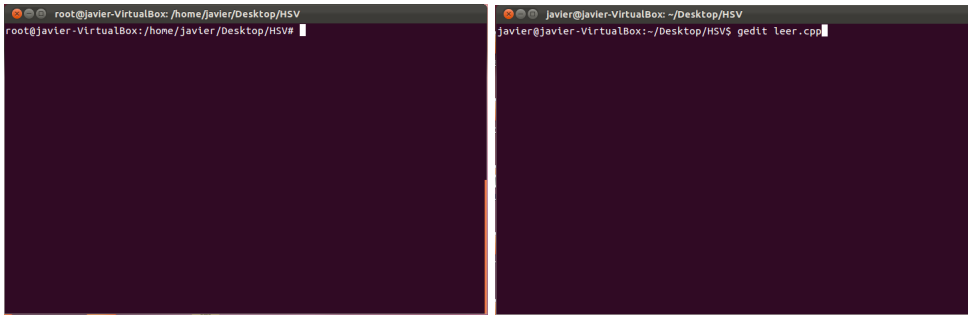
(c) Comprobar que la extensión del archivo este en formato jpg o png debido a que las imágenes de estos corpus están en estos formatos . (d) Abrir main.cpp y colocamos el identificador para que sea usado en los K-Vecinos más cercanos.



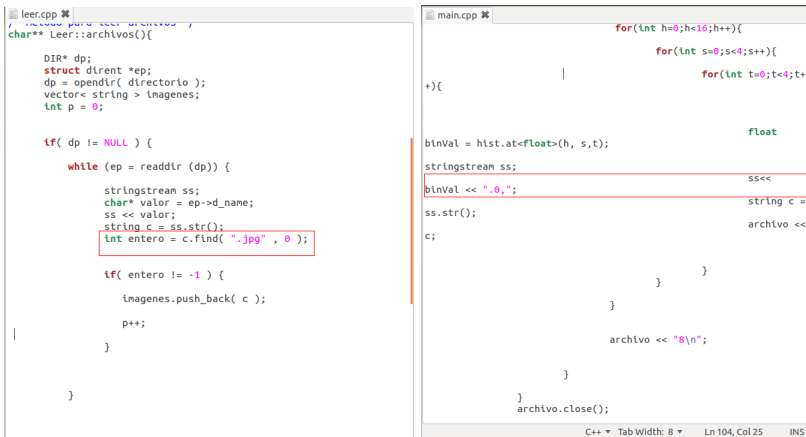
(e) Elegir el directorio en donde se van a escribir los bins de RGB del archivo main.cpp. (f) Compilamos los archivos leer.cpp y main.cpp.



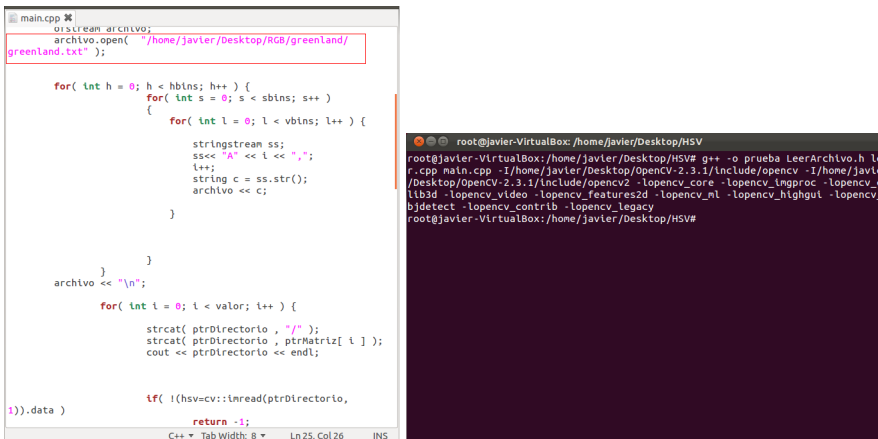
(g) Ejecutamos el binario del programa. (h) Archivo generado.



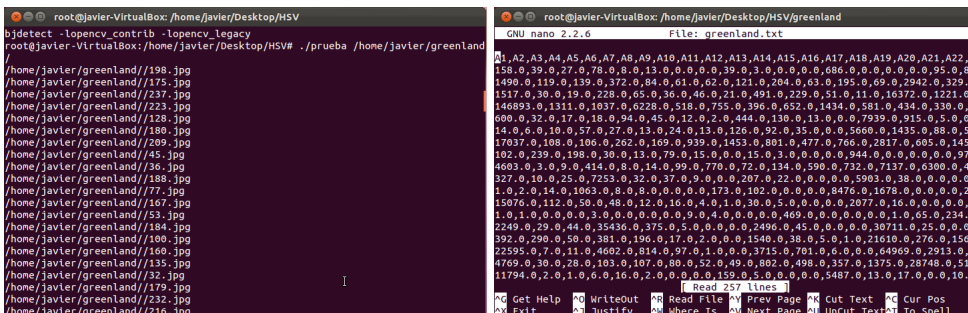
- (a) Ingreso al directorio donde se encuentran los archivos de C++ (b) Ejecutamos un editor de texto para abrir el lector de imágenes .



- (c) Comprobar que la extensión del archivo este en formato jpg o png debido a que las imágenes de estos corpus están en estos formatos . (d) Abrir main.cpp y colocamos el identificador para que sea usado en los Vecinos más cercanos.

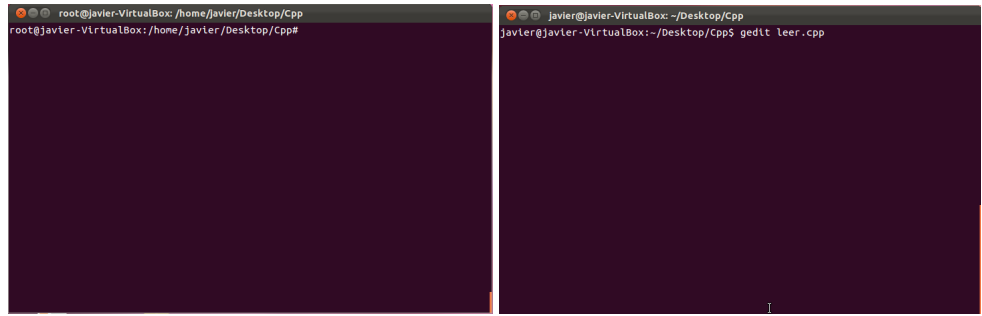


- (e) Elegir el directorio en donde se van a escribir los bins de HSV del archivo main.cpp (f) Compilamos los archivos leer.cpp y main.cpp

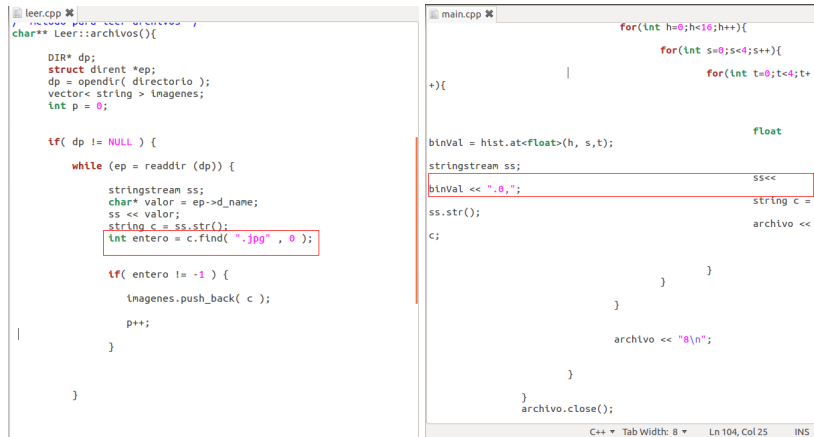


- (g) Ejecutamos el binario del programa. (h) Archivo generado.

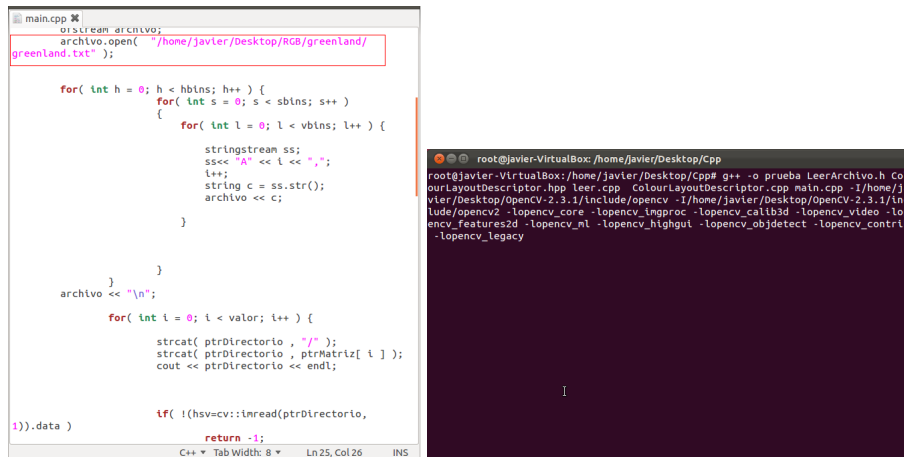
Figura 52: Procedimiento la extracción de características del espacio de color HSV.



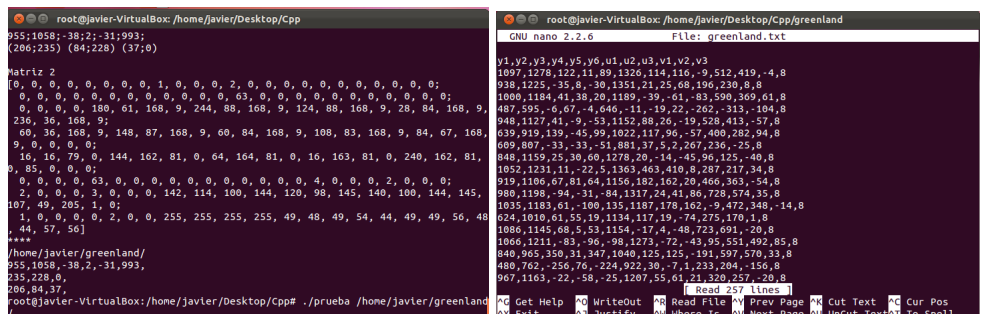
(a) Ingreso al directorio donde se encuentran los archivos de C++ (b) Ejecutamos un editor de texto para abrir el lector de imágenes .



(c) Comprobar que la extensión del archivo este en formato jpg o png debido a que las imágenes de estos corpus están en estos formatos . (d) Abrir main.cpp y colocamos el identificador para que sea usado en los K-Vecinos más cercanos.



(e) Elegir el directorio en donde se van a escribir los bins de RGB del archivo main.cpp. (f) Compilamos los archivos leer.cpp colourlayoutdescriptor.cpp y main.cpp



(g) Ejecutamos el binario del programa. (h) Ejecutamos el binario del programa.

Figura 53: Procedimiento la extracción de características del estándar de color MPEG-7 Color Layout Descriptor.

Ejecución del módulo del clasificador de los K-Vecinos más cercanos

La ejecución de este módulo se realizó en Windows XP y siguió el siguiente procedimiento.

1. Definimos el número de categorías que serán leídas con sus respectivas imágenes.
2. Escribimos el nombre de las categorías que serán utilizadas tanto en train como en test.
3. Definimos la forma en como será medida la distancia entre los elementos
4. Definimos la forma en como será escrito el archivo.
5. Ejecutamos el algoritmo.
6. Volver a repetir los pasos anteriores con los otros dos descriptores.

En la figura 54 se puede apreciar los pasos mencionados con anterioridad para los histogramas RGB y HSV, en la figura 55 se muestra el procedimiento para el estándar de color MPEG-7 Color Layout Descriptor. Para este ejemplo se trabajó con la base de datos Wang.

Ejecución del módulo del clasificador de los esquemas de votación Borda

La ejecución de este módulo se realizó en Windows XP y se ejecutaron los siguientes pasos:

1. Abrimos el programa EsquemasVotacionBorda en Java.
2. Colocamos los tres archivos que tienen información sobre los esquemas de votación Borda
3. Ejecutamos el archivo.
4. Abrimos el programa conteo
5. Variamos el parámetro de la matriz de valores
6. Obtenemos los votos con mayor rango.

7.1.2 *Análisis de resultados*

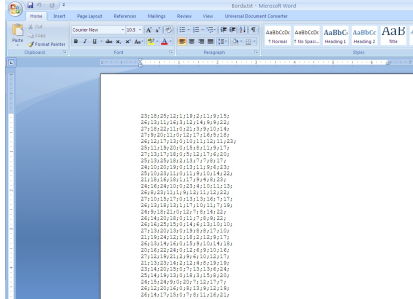
En esta sección se realizará el análisis de los resultados del experimento realizado en este trabajo de investigación. El análisis se lo realiza al comparar los resultados de los K-Vecinos más cercanos con los resultados obtenidos con los esquemas de votación Borda. Este proceso consiste en el porcentaje de precisión en la extracción de los objetos que se ha conseguido [64], para ello se dividió los tres conjuntos de imágenes que fueron mencionados con anterioridad de la siguiente forma. Para la base de datos UW se tomaron 844 imágenes para el conjunto de entrenamiento y 375 para prueba. Para la base de datos Corel se tomaron 700 imágenes para el conjunto de entrenamiento y 300 para el conjunto de prueba. Finalmente, se utilizaron 603 imágenes para el conjunto de entrenamiento y 402 imágenes para la base de datos ZuBud.


```

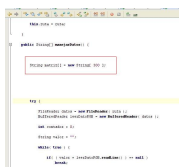
public static void main(String[] args) {
    new Main( "C:\\RGB\\ordenamiento.txt" , "C:\\CLD\\ordenamiento.txt" , "C
}

```

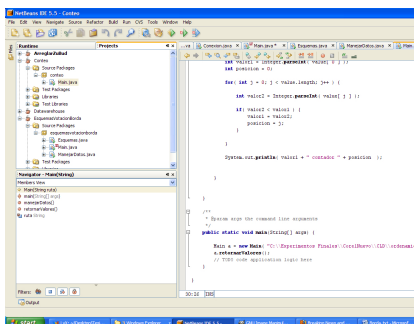
(a) Colocación de las rutas de los votos de KNN.



(b) Resultados de la ejecución del módulo



(c) Parametro que especifica el número de imágenes que pertenecen a la categoría .



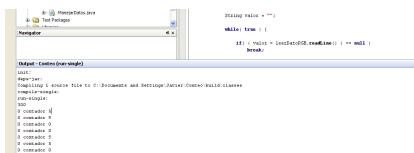
(d) Abrimos el proyecto conteo para ordenar los votos.

```

/**
 * @param args the command line arguments
 */
public static void main(String[] args) {
    Main a = new Main( "C:\\ordenamiento.txt" );
    a.retornarValores();
    // TODO code application logic here
}
}

```

(e) Elegimos el archivo generado por el módulo del clasificador de los esquemas de votación Borda para obtener la clasificación



(f) Resultados finales de los esquemas de votación Borda.

Figura 56: Procedimiento para la ejecución del algoritmo de los esquemas de votación Borda

Precisión

Esta medida permite medir la proporción de las imágenes que son relevantes para la búsqueda, ella permite determinar que tan preciso es el sistema que ha sido diseñado. Para ello utilizamos la siguiente fórmula matemática [41] en donde a representa imágenes retornadas exitosamente, mientras que b indica las imágenes recuperadas no relevantes.

$$P = \frac{a}{a + b} \quad (7.1)$$

Cobertura

Indica el número de imágenes relevantes en una base de datos que son recuperadas en respuesta a una búsqueda. Esta variable está representada por la siguiente fórmula matemática [41] en donde a representa imágenes retornadas exitosamente, mientras que c indica las imágenes relevantes no recuperadas.

$$C = \frac{a}{a + c} \quad (7.2)$$

F-Measure

Según [48] “Se considera la media armónica entre la precisión y la cobertura y sirve para medir la exactitud de una prueba”. La fórmula 7.3 representa cómo calcular la presente medida.

$$F = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (7.3)$$

7.2 COMPARATIVA CON EL ESTADO DEL ARTE

7.2.1 Experimentación

En el capítulo anterior se explicó cómo se iba a realizar el presente experimento, mientras que en esta parte se explicará los resultados. Hay que tomar en cuenta que el descriptor de color Color Layout Descriptor utilizó pesos para mejorar su precisión, para ello se utilizó las siguientes fórmulas [8]: La fórmula 7.4 es el peso que le asignamos a Y , en donde j es la posición actual del subconjunto de prueba y k es la posición actual del subconjunto de entrenamiento. La fórmula 7.5 es el peso que le asignamos a Y , en donde j es la posición actual del subconjunto de prueba y k es la posición actual del subconjunto de entrenamiento. La fórmula 7.6 es el peso que le asignamos a Y , en donde j es la posición actual del subconjunto de prueba y k es la posición actual del subconjunto de entrenamiento.

$$w_i^Y = \frac{0,2}{(j + k)^2} \quad (7.4)$$

feature	wang	uw	irma	ucid	zubud	average
color histogram	50.5	63.0	–	43.3	75.6	58.1
LF SIFT global search	38.3	63.6	20.9	62.5	81.3	53.3
LF patches histogram	48.3	62.0	31.4	37.5	64.7	48.8
LF SIFT histogram	48.2	62.3	32.7	44.7	68.0	51.2
inv. feature histogram (monomial)	47.6	62.6	24.4	41.6	71.0	49.5
MPEG7: scalable color	46.7	63.9	–	37.9	54.3	50.7
LF patches signature	40.4	59.9	23.0	27.6	42.6	38.7
Gabor histogram	41.3	59.7	25.2	22.3	48.7	39.4
32x32 image	37.6	60.1	40.9	14.0	41.9	38.9
MPEG7: color layout	41.8	61.0	29.8	21.7	47.7	40.4
Xx32 image	24.3	57.0	35.0	13.9	47.0	35.4
Tamura texture histogram	38.2	60.8	30.4	33.2	15.8	35.7
LF SIFT signature	36.7	61.2	10.9	34.1	62.7	41.1
gray value histogram	31.7	59.4	26.1	11.8	36.5	33.1
LF patches global	30.5	55.7	17.6	30.3	38.5	34.5
MPEG7: edge histogram	40.8	61.4	10.9	25.2	46.3	36.9
inv. feature histogram (relational)	34.9	59.7	24.1	14.4	6.3	27.9
Gabor vector	23.7	56.3	27.7	4.7	15.9	25.7
global texture feature	26.3	56.5	16.4	6.7	2.6	21.7

Figura 57: Resultados con los diversos descriptores de color [19].

Feature	Precisión
RGB	79.33
HSV	72.33
CLD	53
Borda	82.66

Cuadro 3: Precisión en la base de datos Wang.

$$w_i^{Cb} = \frac{0,4}{(j+k)^2} \quad (7.5)$$

$$w_i^{Cr} = \frac{0,4}{(j+k)^2} \quad (7.6)$$

Para comparar los datos se utilizó la referencia al paper científico de Tomás Desealers [19]. Los resultados a comparar son los siguientes 57:

Experimentación con la base de datos Wang

Precisión

En esta base de datos se dividió el corpus de datos en 300 imágenes para el conjunto de prueba y 700 para el conjunto de entrenamiento de las 1000 imágenes del corpus. En la tabla 3 se pueden apreciar los resultados obtenidos del experimento con esta base de datos:

A pesar de que la precisión de Color Layout Descriptor fue baja, la fusión superó al resto de descriptores debido a la distribución de color dentro del corpus de datos. En la figura 58 se puede apreciar los resultados del experimento, en ella se puede ver que los resultados de las dos primeras imágenes de búsqueda muestran aciertos, mientras que las dos últimas muestran fallos.

Resultados	Imagen Búsqueda	Primer Resultado	Segundo Resultado	Tercer Resultado	Cuarto Resultado
Imagen Búsqueda: Africa Resultados De Borda: Africa Comida Buses Flores					
Imagen Búsqueda: Playa Resultados De Borda: Playa Museos Montañas Buses Flores					
Imagen Búsqueda: Comida. Resultados de Borda: Africa Museos Caballos Buses					
Imagen Búsqueda: Montañas Resultados de Borda: Africa Playas Montañas Museos					

Figura 58: Resultados del experimento con la base de datos Wang .

Cobertura

La cobertura de esta base de datos fue del 100% debido a que el algoritmo de los K-Vecinos más cercanos procesa todos los datos.

F-Measure

El valor de F-Measure es 90.5% gracias a los valores altos que fueron obtenidos en Borda, pero el valor disminuyó en relación con la cobertura porque el valor de la precisión.

Análisis de resultados.

En la figura 59 se puede apreciar la precisión, cobertura y f-measure. La precisión fue del 82.66% debido a que el descriptor MPEG-7 Color Layout Descriptor fue muy baja frente a los histogramas RGB y HSV, mientras que la cobertura fue del 100% por la estructura del algoritmo de los K-Vecinos más cercanos. La media armónica entre precisión y cobertura fue baja debido a que el descriptor Color Layout Descriptor tuvo malos resultados.

Experimentación con la base de datos ZuBuD

Precisión

En esta base de datos se dividió el corpus de datos en 402 imágenes para el conjunto de prueba y 603 para el conjunto de entrenamiento de las 1000 imágenes del corpus. En la tabla 4 se pueden apreciar los resultados obtenidos del experimento con esta base de datos y en la figura 60 se puede apreciar los resultados del experimento, en ella se puede ver que los resultados de las dos primeras imágenes de búsqueda muestran aciertos, mientras que las dos últimas muestran fallos:

Cobertura



Figura 59: Resultados del experimento comenzando desde la izquierda: Precisión del 82.66 %, Cobertura del 100 % y F-Measure de 90.5 % .

Feature	Precisión
RGB	87.31
HSV	94.72
CLD	31.11
Borda	50

Cuadro 4: Precisión en la base de datos ZuBud.

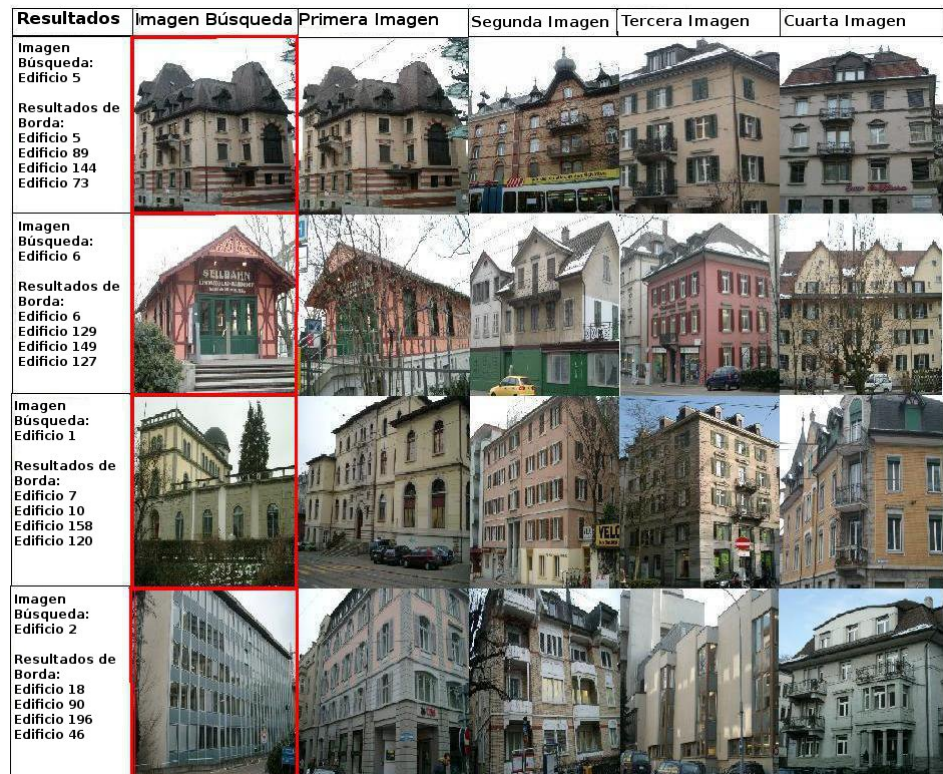


Figura 60: Resultados del experimento con la base de datos ZuBud.

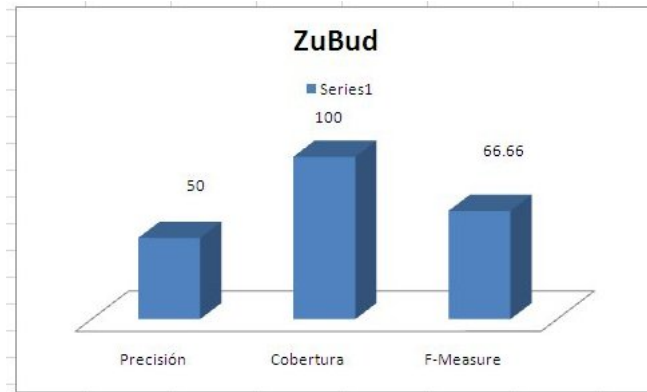


Figura 61: Resultados del experimento comenzando desde la izquierda: Precisión del 50%, Cobertura del 100% y F-Measure de 66.66% .

Feature	Precisión
RGB	77.18
HSV	82.08
CLD	38.39
Borda	66.42

Cuadro 5: Precisión en la base de datos UW.

La cobertura de esta base de datos fue del 100% debido a que el algoritmo de los K-Vecinos más cercanos procesa todos los datos.

F-Measure

El valor de F-Measure es 66.66%, el valor es bajo debido a que el estándar de color MPEG-7 Color Layout Descriptor es demasiado bajo.

Análisis de resultados.

En la figura 61 se puede apreciar la precisión, cobertura y f-measure. La precisión fue del 50% debido a que el descriptor MPEG-7 Color Layout Descriptor fue muy baja frente a los histogramas RGB y HSV, mientras que la cobertura fue del 100% por la estructura del algoritmo de los K-Vecinos más cercanos. La media armónica entre precisión y cobertura fue baja debido a que el descriptor Color Layout Descriptor tuvo malos resultados.

Experimentación con la base de datos UW

Precisión

En esta base de datos se dividió el corpus de datos en 375 imágenes para el conjunto de prueba y 844 para el conjunto de entrenamiento de las 1000 imágenes del corpus. En la tabla 5 se pueden apreciar los resultados obtenidos del experimento con esta base de datos, y en la figura 62 se puede apreciar los resultados del experimento, en ella se puede ver que los resultados de las dos primeras imágenes de búsqueda muestran aciertos, mientras que las dos últimas muestran fallos::

Cobertura

La cobertura de esta base de datos fue del 100% debido a que el algoritmo de los K-Vecinos más cercanos procesa todos los datos.

F-Measure

Resultados	Imagen Búsqueda	Primera Imagen	Segunda Imagen	Tercera Imagen	Cuarta Imagen
Imagen Búsqueda: Arboregreens Resultados de Borda: Arboregreens Spring flowers Campus in fall Yellowstone					
Imagen Búsqueda: Yellowstone Resultados de Borda: Yellowstone San Juans Cannon Beach Spring Flowers					
Imagen Búsqueda: Spring Flowers Resultados de Borda: Campus in fall Arboregreens Spring Flowers Indonesia					
Imagen Búsqueda: Australia Resultados de Borda: Cambridge Australia Cherries Indonesia					

Figura 62: Resultados del experimento con la base de datos UW .

El valor de F-Measure es 79.82 %, el valor es bajo debido a que el estándar de color MPEG-7 Color Layout Descriptor es demasiado bajo.

Análisis de resultados.

En la figura se puede apreciar la precisión, cobertura y f-measure. La precisión fue del 82.66 % debido a que el descriptor MPEG-7 Color Layout Descriptor fue muy baja frente a los histogramas RGB y HSV, mientras que la cobertura fue del 100 % por la estructura del algoritmo de los K-Vecinos más cercanos. La media armónica entre precisión y cobertura fue baja debido a que el descriptor Color Layout Descriptor tuvo malos resultados.

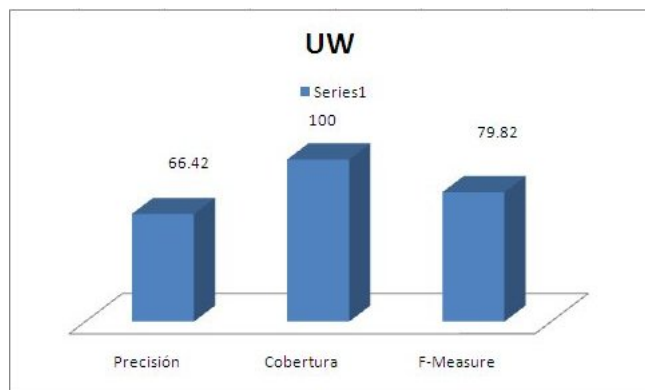


Figura 63: Resultados del experimento comenzando desde la izquierda: Precisión del 50 %, Cobertura del 100 % y F-Measure de 66.66 % .

7.3 PROPUESTA DE MEJORAS Y TRABAJO FUTURO

Se propone para mejorar los resultados la técnica de clasificación de boosting Adaboost. Debido a que esta técnica combina varios clasificadores débiles se pueden mejorar los resultados por las siguientes razones. Este es un algoritmo adaptativo que busca minimizar el error a casi cero mediante adaptación a los datos.

Como trabajo futuro se planea utilizar otras técnicas de Machine Learning sobre las tres bases de datos para realizar una comparativa entre los diversos resultados obtenidos de los mismos.

Parte VIII

CONCLUSIONES Y RECOMENDACIONES

CONCLUSIONES

1. Los resultados que fueron obtenidos a partir de la combinación de los histogramas de color RGB, HSV y el estándar de color MPEG-7 Color Layout Descriptor fueron los que se detallan en estas líneas. En primer lugar, el experimento con los esquemas de votación Borda aplicados a la base de datos Wang obtuvo una precisión mucho mayor que con la aplicación del histograma de color HSV, debido a que la distribución de color de las imágenes del corpus es muy similar en cada categoría, por ello Color Layout descriptor pudo clasificar mejor las imágenes. En segundo lugar, los experimentos realizados con las bases de datos Wang y ZuBud no pudieron superar los resultados obtenidos con la aplicación del histograma de color HSV, debido a que la precisión del estándar de color MPEG-7 no fue buena y la combinación de los histogramas con Color Layout Descriptor no pudo superar a los resultados presentados en [19].
2. Dados los resultados de los histogramas RGB, HSV y Color Layout descriptor se puede llegar a la conclusión de que la precisión de un descriptor o histograma está dado por la estructura que lo define. Ello quiere decir que un descriptor o histograma puede ser bueno en clasificar algunos tipos de imágenes, mientras que otro descriptor/histograma puede ser bueno en clasificar otro tipo de imágenes, lo que nos da una mayor precisión global.
3. Dado que el estándar MPEG-7 define un conjunto completo de descriptores visuales y de audio, es importante analizar si es posible combinar otros descriptores visuales tales como histograma de bordes, dominant color descriptor, etc. De la misma forma, es importante buscar otros métodos como Adaboost, árboles de decisión, etc.

RECOMENDACIONES

1. Se recomienda experimentar utilizando otros métodos de ensemble tales como: Marqués de Condorcet, Adaboost y Support Vector Machines, con la finalidad de comparar resultados con los esquemas de votación Borda. Debido a que los experimentos que fueron realizados en las bases de datos UW y ZuBud tuvieron malos resultados al momento de utilizar el conjunto de votaciones del Color Layout Descriptor, se puede intentar mejorar los resultados al utilizar este sistema de votación.
2. Se recomienda usar la técnica descrita en [87] con la finalidad de mejorar los resultados de precisión del proceso de combinación de features. Debido a que el procedimiento realizado en esta tesis solamente obtuvo buenos resultados en el corpus Wang, se puede utilizar este procedimiento con la finalidad de mejorar los resultados en los corpus UW y ZuBud.

Parte IX
GLOSARIO

GLOSARIO

CBIR: Content-Based Image Retrieval, *en español retrono de imágenes basado en contenido.*

KNN: K-Nearest Neighbours, *en español K-Vecinos más cercanos..*

DCT: Discrete Cosine Transform, *en español transformada discreta del coseno.*

DFT: Discrete Fourier Transform, *en español transformada discreta de Fourier.*

RGB: Red, green and blue, *en español rojo, verde y azul.*

HSV: Hue, Saturation and Value, *en español transformada discreta de Fourier.*

CLD: Color Layout Descriptor, *en español descriptor de la distribución de color.*

MPEG-7: Multimedia content description interface, *en español interface de descripción del contenido multimedia.*

GOF: Group of frames, *en español grupo de cuadros.*

GOP: Group of pictures, *en español grupo de imágenes.*

SIFT: Scale-invariant feature transform, *en español transformada de características a escala invariante.*

ASIFT: Affine-SIFT.

SURF: Speeded Up Robust Features.

BIBLIOGRAFÍA

- [1] K. Abou-Moustafa, M. Shah, F. De la Torre, and F. Ferrie. Relaxed Exponential Kernels for Unsupervised Learning. In *Symposium of the German Association for Patter Recognition (DAGM)*, 2011. (Cited on pages xvii and 57.)
- [2] K. Alsabti, S. Ranka, and V. Singh. An efficient k-means clustering algorithm. 1997. (Cited on page 59.)
- [3] J. R. Anderson, R. S. Michalski, R. S. Michalski, J. G. Carbonell, and T. M. Mitchell. *Machine learning: An artificial intelligence approach*, volume 2. Morgan Kaufmann, 1986. (Cited on page 56.)
- [4] R. Balasubramani and D. V. Kannan. Efficient use of MPEG-7 color layout and edge histogram descriptors in cbir systems. *Global Journal of Computer Science and Technology*, 9(4), 2009. (Cited on pages 46 and 47.)
- [5] J. D. Ballesteros and V. P. Noriega. Descripción, indexación, búsqueda y adquisición de secuencias de vídeo mediante descriptores MPEG-7. 2006. (Cited on pages 45, 46, 48, 49, and 50.)
- [6] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. *Computer Vision–ECCV 2006*, pages 404 – 417, 2006. (Cited on page 50.)
- [7] Dienekes Anthropology Blog. Genetic Classification of Populations Using Supervised Learning . 2011. URL <http://dienekes.blogspot.com/2011/06/genetic-classification-of-populations.html>. (Cited on pages xvii and 57.)
- [8] O. Boullosa. Estudio comparativo de descriptores visuales para la detección de escenas cuasi-duplicadas. Master’s thesis, 2011. URL <http://arantxa.ii.uam.es/~jms/pfcsteleco/lecturas/201103180scarBoullosa.pdf>. (Cited on pages 41, 47, 75, and 97.)
- [9] G. Bradski and A. Kaehler. *Learning OpenCV: Computer vision with the OpenCV library*. O’Reilly Media, 2008. (Cited on pages xvi, 17, 18, 19, 20, 22, 23, 39, 41, 47, 56, 58, 59, and 60.)
- [10] V. Robles Bykvaev. ESQUEMAS DE VOTACIÓN BORDA APLICADOS AL ETIQUETADO DE ROLES SEMÁNTICOS. Master’s thesis, 2010. URL <http://users.dsic.upv.es/~proso/resources/RoblesMSc.pdf>. (Cited on pages 63 and 64.)
- [11] F. Calderon. home. 2012. URL <http://opencvjaveriana.wikispaces.com/>. (Cited on page 78.)
- [12] C. García Cambronerero and I. Gómez Moreno. ALGORITMOS DE APRENDIZAJE: KNN & KMEANS. 2006. URL <http://www.it.uc3m.es/jvillena/irc/practicas/08-09/06.pdf>. (Cited on page 59.)

- [13] L. Cioni. A few notes on the Borda and Condorcet methods. Technical report, Technical Report TR-10-16, Computer Science Department, 2010. (Cited on pages 64 and 66.)
- [14] R. Dawkins. *The selfish gene*. Oxford University Press, USA, 2006. (Cited on pages 11 and 12.)
- [15] R. Dawkins. *The greatest show on earth: The evidence for evolution*. Transworld Digital, 2009. (Cited on page 11.)
- [16] R. Dawkins. *The greatest show on earth: The evidence for evolution*. Transworld Digital, 2009. (Cited on pages 9, 10, 12, and 13.)
- [17] A. de la Escalera. *Visión por Computador: Fundamentos y métodos*. Prentice Hall, 2010. (Cited on page 25.)
- [18] Universidad del País Vasco. TOMA DE DECISIÓN CON CERTIDUMBRE . 2011. URL <http://www.ehu.es/Degypi/Metodologia/montecarloyarboles.htm>. (Cited on pages xvii and 59.)
- [19] T. Deselaers, D. Keysers, and H. Ney. Features for image retrieval: An experimental comparison. *Information Retrieval*, 11(2):77 – 107, 2008. (Cited on pages xvi, xviii, 33, 71, 98, and 107.)
- [20] M. H. Diaz. Análisis Comparativo de Algoritmos para Reducción de Ruido en Señales Utilizando Wavelets . Master’s thesis, 2003. URL http://catarina.udlap.mx/u_dl_a/tales/documentos/lem/hernandez_d_m/capitulo2.pdf. (Cited on page 32.)
- [21] H. Ding, W. Pan, and Y. Guan. Image Acquisition, Storage and Retrieval. (Cited on page 34.)
- [22] Etienne Dupuis. Optimizing YUV-RGB Color Space Conversion Using Intels SIMD Technology. 2003. (Cited on page 43.)
- [23] J. E. Duque, Ó. H. Moscoso, and A. D. Cubillos. El lóbulo insular. *Acta Neurol Colomb*, 20(2), 2004. (Cited on page 15.)
- [24] et al E. Gardner. *Anatomia*. Salvat Editores, 1971. (Cited on pages xvi and 14.)
- [25] J. Eakins, M. Graham, and I. S. C. T. A. P. Joint. Content-based image retrieval, 1999. (Cited on pages 4 and 35.)
- [26] E.Izquierdo. Visual Information Retrieval from Vision to human computer interaction. 2012. (Cited on page 34.)
- [27] REAL ACADEMIA ESPAÑOLA. *DICCIONARIO DE LA LENGUA ESPAÑOLA - Vigésima segunda edición* . Anchor, 2012. URL <http://lema.rae.es/drae/>. (Cited on pages 12 and 15.)
- [28] A. Estévez-González, C. García-Sánchez, and L. Barraquer-Bordas. Los lóbulos frontales: El cerebro ejecutivo. *Revista de Neurología*, 31(6):566 – 577, 2000. (Cited on pages xvi, 14, and 15.)
- [29] J. P. Figueroa and V. R. Bykbaev. Image retrieval based on the combination of RGB and HSVs histograms and Colour Layout Descriptor. 2012. (Cited on pages xvi, xvii, 42, 46, 47, 48, 76, and 77.)

- [30] A. Ford and A. Roberts. Colour space conversions. *Westminster University, London*, 1998. (Cited on page 42.)
- [31] R. Garc a. *Peque o Larousse Ilustrado*. 1964. (Cited on pages 9 and 13.)
- [32] D. Gefner. EL CEREBRO ORGANIZACION Y FUNCION. 2012. URL <http://svneurologia.org>. (Cited on page 15.)
- [33] S. J. Gould. *El pulgar del panda*, volume 3. Editorial Critica, 2006. (Cited on pages 11 and 13.)
- [34] A. Guyton. *Fisiologia y Fisiopatologia basicas*. Interamericana, 1972. (Cited on pages xvi, 16, 17, 18, 20, 21, 22, 23, 24, 25, and 26.)
- [35] P. Harrington. *Machine Learning in Action*. Manning, 2012. (Cited on pages xvii, 55, 56, 58, 60, 62, and 63.)
- [36] Black Ice. HSV Color Space - Color Space Conversion, 2012. URL <http://www.blackice.com/colorspaceHSV.htm>. (Cited on pages xvi and 43.)
- [37] National Eye Institute. NEI catalog number NEA09. 2009. URL <http://www.nei.nih.gov/health/eyediagram/eyeimages3.asp>. (Cited on pages xvi and 25.)
- [38] Intersil. Color Spaces. URL http://www.compression.ru/download/articles/color_space/ch03.pdf. (Cited on page 75.)
- [39] J.Al-Khalili. Science and Islam: The empire of reason. 2010. (Cited on page 3.)
- [40] J.Branch and G.Olague. LA VISI N POR COMPUTADOR. UNA APROXIMACI N AL ESTADO DEL ARTE. 2001. URL <http://www2.unalmed.edu.co/dyna2005/133/vision.pdf>. (Cited on page 3.)
- [41] S. Jeong, C. S. Won, and R. M. Gray. Image retrieval using color histograms generated by Gauss mixture vector quantization. *Computer Vision and Image Understanding*, 94(1):44 – 66, 2004. (Cited on pages xvi, 42, and 97.)
- [42] A. Jos, I. Moreno, A. G. Gonz lez, G. Repetto, and A. M. Came n. Differentiation of sparkling wines (cava and champagne) according to their mineral content. *Talanta*, 63(2):377 – 382, 2004. (Cited on page 56.)
- [43] L. Juan and O. Gwun. A comparison of sift, pca-sift and surf. *International Journal of Image Processing*, 3(4):143 – 152, 2009. (Cited on page 50.)
- [44] H. Kekre, M. D. Mishra, and M. A. Kariwala. A survey of cbir techniques and semantics. *International Journal of Engineering Science*, 3, 2011. (Cited on pages xvi, 34, and 35.)
- [45] C. Kingsford and S. L. Salzberg. What are decision trees? *Nat. Biotechnol.*, 26(9):1011 – 1013, 2008. (Cited on page 58.)
- [46] P. Lammertsma. K-nearest-neighbor algorithm. URL paul.luminos.nl/download/document/knn.pd. (Cited on page 61.)

- [47] M. F. Land and R. D. Fernald. The evolution of eyes. *Annual review of neuroscience*, 15(1):1 – 29, 1992. (Cited on page 13.)
- [48] B. Leon and A. Flores Vega. *Estudio de las técnicas de detección de plagio textual y análisis de sinonimia en ensayos y desarrollo de un sistema prototipo*. Universidad Politecnica Salesiana, 2012. (Cited on page 97.)
- [49] C. S. Lima, J. H. Correia, J. Ramos, and D. Barbosa. Texture classification of images from Endoscopic Capsule by using MLP and SVM-A comparative approach. In *World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany*, pages 271 – 274. Springer, 2009. (Cited on page 75.)
- [50] D. Lippman. Voting Theory. URL <http://dlippman.imathas.com/mathinsociety/VotingTheory1.4.pdf>. (Cited on page 66.)
- [51] B. S. Manjunath, J. R. Ohm, V. V. Vasudevan, and A. Yamada. Color and texture descriptors. *Circuits and Systems for Video Technology, IEEE Transactions on*, 11(6):703 – 715, 2001. (Cited on page 86.)
- [52] J. C. Martin. *Introduction to Languages and the Theory of Computation*, volume 2. McGraw-Hill, 2003. (Cited on page 58.)
- [53] I. S. Martínez. *El estado social: Antecedentes, origen, desarrollo y declive*. Trotta, 2010. (Cited on page 13.)
- [54] S. MediaLab. MPEG-7 White Paper. *Outubro*, 2003. (Cited on pages xvi, 43, 44, and 45.)
- [55] K. Mekaldji, S. Boucherka, and C.S. Color quantization and its impact on color histogram based image retrieval. *Proceedings of the Second Conference International sur l'Informatique et ses Applications (CIIA'09)*, May 3-4 2009. (Cited on pages 41, 42, 75, and 84.)
- [56] G. Miller. *A Review of Sexual Selection and Human Evolution: How Mate Choice shaped Human Nature*. 1998. URL http://www.unm.edu/~psych/faculty/lg_gmiller.html. (Cited on page 11.)
- [57] G. Miller. *Sexual Selection*. 2007. URL http://www.unm.edu/~psych/faculty/lg_gmiller.html. (Cited on pages 10 and 11.)
- [58] G. Miller. *The mating mind: How sexual choice shaped the evolution of human nature*. Anchor, 2011. (Cited on pages 12 and 13.)
- [59] T. M. Mitchell. The discipline of machine learning. *Machine Learning Department technical report CMU-ML-06-108*, Carnegie Mellon University, 2006. (Cited on page 55.)
- [60] J. W. Moore. The Crisis of Feudalism. *Organization & environment*, 15(3):301 – 322, 2002. (Cited on page 13.)
- [61] J. M. Morel and G. Yu. ASIFT: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, 2(2):438 – 469, 2009. (Cited on page 51.)
- [62] OpenCV. OpenCV Wiki: Welcome. 2012. URL <http://opencv.willowgarage.com/wiki/>. (Cited on page 78.)

- [63] A. Plaza and J. L. Zambrano. *Estudio y selección de las técnicas SIFT, SURF y ASIFT de reconocimiento de imágenes para el diseño de un prototipo en dispositivos móviles*. Universidad Politecnica Salesiana, 2012. URL <http://dspace.ups.edu.ec/handle/123456789/1259?mode=full>. (Cited on pages 50 and 51.)
- [64] P. Urgiles and A. Paguay. RECUPERACIÓN DE IMÁGENES MEDIANTE EXTRACCIÓN DE BLOBS APLICANDO EL OPERADOR LAPLACIANO DE GAUSS Y EL KERNEL GAUSSIANO Y DESARROLLO DE UN PROTOTIPO. Master's thesis, 2012. URL <http://dspace.ups.edu.ec/handle/123456789/1710?mode=full>. (Cited on pages 3, 4, 78, and 93.)
- [65] rex20028. Lentes Concavas y Convexas. 2009. URL <http://www.slideshare.net/rex20028/las-lentes-cncavas-estn-curvadas-hacia-dentro>. (Cited on page 16.)
- [66] M. J. Roberts. *Signals and Systems: Analysis of Signals Through Linear Systems*. McGraw-Hill Science/Engineering/Math, 2003. (Cited on pages 46 and 47.)
- [67] H. Rouviere. *Compendio de Anatomia y Diseccion*. Salvat Editores, 1978. (Cited on pages xvi and 16.)
- [68] S. J. Russell and P. Norvig. *Artificial intelligence: A modern approach*. Prentice hall, 2010. (Cited on pages 3, 31, and 56.)
- [69] W. Schudy. When Voters Strategize, Approval Voting Elects Condorcet Winners but Condorcet Methods can Elect Condorcet Losers. 2007. (Cited on page 63.)
- [70] N. S. Sharma, P. S. Rawat, and J. S. Singh. Efficient CBIR using color histogram processing. *Signal & Image Processing*, 2(1), 2011. (Cited on page 41.)
- [71] T. Sikora. The MPEG-7 Visual standard for content description-an overview. *Circuits and Systems for Video Technology, IEEE Transactions on*, 11(6):696 – 702, 2001. (Cited on page 86.)
- [72] R. G. B. C. Space. Color Spaces. (Cited on pages 42 and 46.)
- [73] S. Sural, G. Qian, and S. Pramanik. Segmentation and histogram generation using the HSV color space for image retrieval. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 2. Ieee, 2002. (Cited on pages 42 and 86.)
- [74] R. Szeliski. *Computer vision: Algorithms and applications*. Springer-Verlag New York Inc, 2010. (Cited on pages xvi, 4, 18, 19, 31, 32, and 33.)
- [75] C. Theoharatos, V. K. Pothos, N. A. Laskaris, G. Economou, and S. Fotopoulos. Multivariate image similarity in the compressed domain using statistical graph matching. *Pattern recognition*, 39(10):1892 – 1904, 2006. (Cited on pages xvii and 48.)
- [76] Unai Mujika Torrontegi. Reconstrucción densa de modelos tridimensionales utilizando Visión Artificial. Master's thesis, 2010. URL http://www.ccia-kzaa.ehu.es/s0140-con/es/contenidos/informacion/tesis_master/es_t_master/adjuntos/10umujika.pdf. (Cited on pages 20 and 22.)

- [77] UniversityOfCaliforniaBerkley. Understanding Evolution Natural Selection. 2012. URL http://evolution.berkeley.edu/evolibrary/article/0_0_0/evo_25. (Cited on pages [xvi](#), [10](#), and [11](#).)
- [78] UniversityOfCaliforniaBerkley. Understanding Evolution: Glosary Definition, Natural Selection. 2012. URL http://evolution.berkeley.edu/evolibrary/glossary/glossary_popup.php?word=natural+selection. (Cited on page [9](#).)
- [79] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. Evaluating Color Descriptors for Object and Scene Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9): 1582 – 1596, 2010. URL <http://www.science.uva.nl/research/publications/2010/vandeSandeTPAMI2010>. (Cited on page [50](#).)
- [80] A. Vedaldi. An implementation of SIFT detector and descriptor. *University of California at Los Angeles*, 2006. (Cited on page [50](#).)
- [81] L. R. Winston. *Human instinct*. Bantham Books, 2002. (Cited on pages [11](#), [12](#), and [13](#).)
- [82] R. Winston. *Bad Ideas?: An Arresting History of Our Inventions*. Bantam, 2011. (Cited on page [13](#).)
- [83] K. B. Wolf and G. Krotzsch. El problema de las tres lentes. *Revista Mexicana de Física*, 47(3):291 – 298, 2001. (Cited on page [16](#).)
- [84] P. Farreras y Ciril Rozman. *Medicina Interna*, volume II. Editorial Marin, 1978. (Cited on page [14](#).)
- [85] R. Garc a-Pelayo y Gross. *evolucion*. Larousse, 1964. (Cited on page [55](#).)
- [86] M. Yamasaki, M. I. Tenaillon, I. V. Bi, S. G. Schroeder, H. Sanchez-Villeda, J. F. Doebley, B. S. Gaut, and M. D. McMullen. A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. *The Plant Cell Online*, 17(11):2859 – 2872, 2005. (Cited on page [11](#).)
- [87] Q. Zhang and E. Izquierdo. A multi-feature optimization approach to object-based image classification. *Image and Video Retrieval*, pages 310 – 319, 2006. (Cited on page [109](#).)

Parte X

ANEXOS

- Poveda, J. y Robles, V. (2012). "Image retrieval based on the combination of RGB and HSV's histograms and Colour Layout Descriptor". INGENIUS. N7, (enero/junio). pp 3-10. ISSN: 1390-650X
- Poveda, J. y Robles, V. (2012). "Combination of HSV and RGB histograms and MPEG-7 Descriptor: Preliminar Results and Future Work". INGENIUS. N8 , (julio/diciembre), pp. 3-10, ISSN: 1390-650X.
- Conferensista del "Segundo Congreso de Tecnologías" realizado en la Universidad Politécnica Salesiana el día 28 de noviembre del 2012.

IMAGE RETRIEVAL BASED ON THE COMBINATION OF RGB
AND HSV'S HISTOGRAMS AND COLOUR LAYOUT
DESCRIPTORJavier Poveda Figueroa^{1,*}, Vladímir Robles Bykbaev²**Abstract**

In this paper we present the first stage of a new approach to improve the precision and recall of the content-based image retrieval task. To do this, we aim to combine three colour features, RGB and HSV histograms, and MPEG-7 Colour Layout Descriptor. To perform the combination, we propose to use an approximation based on Borda Voting-Schemes. Under that the Borda Voting-Schemes needs at least three votes to perform the combination, we intend to use the K-Nearest Neighbors methods to select the candidate images, given a query image. In the second stage, we'll implement our approach using at least three image databases.

Keywords: HSV histogram, RGB histogram, Colour layout descriptor, Borda Voting Schemes, KNN.

Resumen

En este artículo se presenta la primera fase de experimentación de una nueva propuesta para mejorar la precisión y cobertura de la recuperación de imágenes basada en contenidos. Para realizar esta tarea, se propone combinar tres características de color, como son los histogramas RGB y HSV y el Descriptor de Distribución de Color del estándar MPEG-7. A fin de llevar a cabo la combinación, se plantea emplear una aproximación basada en Esquemas de Votación Borda. En virtud de que dichos esquemas requieren al menos tres votos, se propone usar la técnica de los K - Vecinos más cercanos, con el objetivo de seleccionar las imágenes candidatas a partir de imagen de consulta. En la segunda etapa se implementará nuestra propuesta empleando al menos tres bases de datos de imágenes.

Palabras clave: Histograma HSV, Histograma RGB, Colour Layout Descriptor, Borda Voting Schemes, KNN.

^{1,*} Colaborador del Laboratorio de Investigación en Sistemas Informáticos e Inteligencia Artificial, Carrera de Ingeniería de Sistemas, Universidad Politécnica Salesiana, sede Cuenca. Autor para correspondencia: jpoveda@est.ups.edu.ec

² Máster en Inteligencia Artificial, Reconocimiento de Formas e Imagen Digital, Ingeniero en Sistemas, Estudiante de Doctorado en Informática, Universidad Politécnica de Valencia, Encargado del Laboratorio de Investigación en Sistemas Informáticos e Inteligencia Artificial, Universidad Politécnica Salesiana, sede Cuenca

Recibido: 06 - Marzo - 2012; Aprobado tras revisión: 15 - Mayo - 2012

Forma sugerida de citación: Poveda, J. y Robles, V. (2012). "Image retrieval based on the combination of RGB and HSV's histograms and Colour Layout Descriptor". *INGENIUS*. N°7, (enero/junio). pp 3-10. ISSN: 1390-650X

1. Introduction

Nowadays, the Content-Based Image Retrieval (CBIR) is a technique that has several areas of application. For example, in the medical area, the CBIR applications help to detect non-melanoma skin lesions [1], breast cancer [2], while in forensics area it is possible to detect altered images [3]. Another application that successfully uses CBIR is a device that tries to help in navigation, and object location for blind people [4]. These successfully developed applications are a real example that CBIR techniques have a broad scope of applications.

The constant growth of the visual information, like pictures and videos, makes necessary to have efficient methods to retrieve this kind of information in a proper way [5].

The MPEG-7 standard defines a set of visual descriptors to extract several image features. An example of these features is: image edges and color distribution [6], the shape of objects in the image [7], the texture characteristics [8], etc.

With this approach, we try to improve the level of precision or recall in the images retrieval task, using the combination of three colour features: RGB and HSV's histograms, and MPEG-7 Colour Layout Descriptor. To conduct the combination, we use a new alternative, the Borda Voting-Schemes.

The rest of the paper is organized as follows. In the Section 2 we review the fundamentals of the colour features like RGB and HSV histograms and MPEG-7 Colour Layout Descriptor (CLD). The K - Nearest Neighbors method is described in Section 3. In Section 4 we review the Borda Voting-Schemes. In Section 5 we describe our approach to combine

the RGB, HSV and CLD features to perform the image retrieval task. Conclusions and future work are reviewed in the Section 6.

2. Colour features fundamentals

2.1 RGB histogram

RGB is the acronym for Red, Blue and Green. This histogram is the most used histogram in computer graphics and it uses its red, green and blue components to create a new color [5]. This color histogram combines red, green and blue colors to form a new color. To form a new color it is necessary to increase the values of one or more of the components of the RGB components [9].

The RGB color histogram can be represented on the Cartesian coordinate system, which is illustrated in Figure 1. As it shows a new color can be formed by changing the values of the RGB component. For example, if the magenta color is needed, it is necessary to combine green and blue to create it [9].

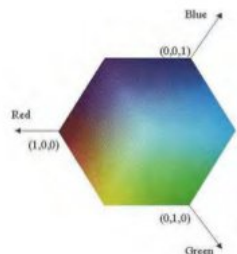


Figure 1. RGB coordinate system where the "x" axis represents red colors, "y" axis represents green colors and "z" represents blue colors [10]. As this figure shows changing the values of RGB component can form a new color.

2.2 HSV histogram

HSV color histogram is the acronym of Hue, Saturation and Value; this color histogram is

the closest perception of the human eye. As [11] explains, the human eye perceives colors by the excitation of two cells of the eye, which are rods and cones. "Excitation of the cone cells leads to perception of color, while rod cells help in perception of various shades of gray. [...] The HSV color histogram separates the luminance component (Intensity) of a pixel color from its chrominance components (Hue and saturation)." This representation works as the human eye because it works like the separation of the rods and cones.

On the representation of the components for this histogram Hue represents the chromatic component, saturation represents the predominance value of a hue a color and Value represents the intensity of the color [10].

The generation of colors of the HSV color space is represented in the next way. According to [12], Hue defines the color by changing its angle; hue is defined as an angle in the range $[0, 2\pi]$. Saturation is the depth or purity of the color and is measured as a radial distance from the central axis with value between 0 at the center to 1 at the outer space. Finally, value is represented by the vertical central axis.

In Figure 2 is illustrated how we represent an HSV color histogram.

2.3 Colour layout descriptor

This MPEG-7 visual descriptor represents the spatial layout of color images in a very compact form [6]. To extract the CLD descriptor, we assume the original image is represented in RGB color space, and the necessary steps to perform this task are the next:

- *CLD block calculation.* In this step the image is divided in 64 blocks of pixels. The width and height of each block are calculated

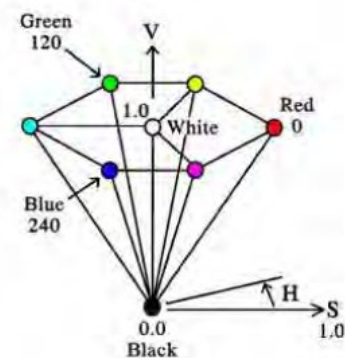


Figure 2. Representation of a HSV color descriptor. The value V is represented by the main axis orthogonal of the plane. The angle represents the Hue value, while radius represents the level of Saturation (purity of colour) [13].

by dividing the width and height of the image for 8, respectively. This process is performed for the three components of RGB colour space and we obtain 3 matrices of 64 blocks.

- *Representative colour selection.* Each block must have only one colour. To select the representative colour, the values of all pixels in the block are averaged (Figure 3). After that, each 8x8 matrix is transformed to YCbCr colour space ¹

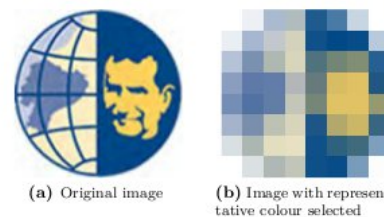


Figure 3. Result to applying representative colour selection in image in RGB colour space.

¹The YCbCr is a family of color spaces used in digital video and image transmission. The component Y represents the luminance and Cb and Cr are the blue-difference and red-difference chrominance components.

– *DCT (Discrete Cosine Transform)*. In this step the three colour components of image (Y, Cb and Cr) are transformed by 8x8 DCT to obtain three 8x8 DCT matrices of coefficients [6]. To calculate the DCT in a 2D matrix, the following formula is used:

$$B_{pq} = \alpha_p \alpha_q \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A_{mn} \cos \frac{\pi(2m+1)p}{2M} \cos \frac{\pi(2n+1)q}{2N} \quad (1)$$

Where:

- B_{pq} is the coefficient for p,q position of the DCT matrix.
- M is the number of rows of the matrix.
- N is the number of columns of the matrix.
- α_p and α_q are constants which value are calculated as follows:

$$\alpha_p \begin{cases} \frac{1}{\sqrt{M}}, & p = 0 \\ \frac{1}{\sqrt{2M}}, & 1 \leq p \leq M-1 \end{cases}$$

$$\alpha_q \begin{cases} \frac{1}{\sqrt{N}}, & p = 0 \\ \frac{1}{\sqrt{2N}}, & 1 \leq p \leq N-1 \end{cases}$$

– *Zigzag reordering*. The final step consists on zigzag reading of the DCT coefficients (Figure 4). This should be done in this way because the nonzero coefficients always appear in the top-left corner of the coefficients matrix. Usually, the CLD employs the first 6 coefficients of the Y colour component, and 3 coefficients of Cb and Cr colour components.

3. K - Nearest Neighbors

K-Nearest Neighbors, also known as KNN, is a supervised machine learning method, which

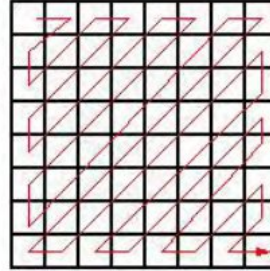


Figure 4. Zigzag scanning applied to Y, Cb and Cr components of the representative colour image.

classifies the data of a corpus. With this classifier a set of data can be classified in order to discover which elements are from the set of data.

To solve the classification problem the algorithm uses a series of clusters of data, which each one represent a category. When a new kind of data appears, the KNN algorithm tries to recognize which category the data belongs.

Figure 5 is illustrated how the KNN algorithm works [14].

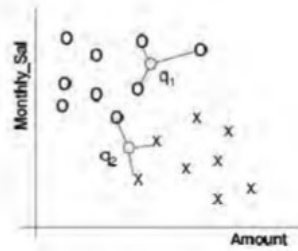


Figure 5. Representation of a KNN algorithm which main goal is to discover if the data, which we are seeking, is from Monthly-Sal or Amount.

To find to which set of elements the new data belongs to, a distance formula is used, which permits the calculation of the distance

of one element to another, to calculate we use equation 2 [14].

$$D_{a-b} = \sqrt{\sum_{i=1}^N (a_i - b_i)^2} \quad (2)$$

Equation 2. Euclidean distance to calculate the distance between two elements, where b_i represents the position of the first element and a_i represents the position of the second element [15].

4. Borda voting scheme

Borda voting - scheme is a technique, which can help to classify the data of a corpus; this method has its origins in France it the 18th century. It was proposed by Charles Borda because the voters did bad decisions in the process of voting; for this reason Borda proposed a method by which the voters must vote for the candidates by its merits [16].

Following the Borda scheme this method consists in a group of candidates, which vote for the elements of a set of data by its relevance. For this method we have a set of candidates $X = \{x_1, x_2, x_3, \dots, x_n\}$ where $n \geq 3^2$ is and m voters, where $m \geq 3$ [13]. To calculate the voting results it is necessary to follow the next steps:

Firstly, we choose a photograph and we have three categories, which gives us 3 options per each candidate. In the Table 1 shows the options and the candidates of each element.

Table 1. Voters with its different options

	Candidate 1	Candidate 2	Candidate 3
Option 1	2	2	3
Option 2	3	3	1
Option 3	1	3	2

Secondly, we represent the data taken by the candidates to form a square matrix, which

is formed by the number of options we have; in this case we have a 3 x 3 matrix. We fill the matrix with 1 in each row that the candidates have preference and 0 in the rest of positions [17].

$$\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$

Finally, we sum all the elements of each row and we get an array of data, which shows us the order of classification of the image [18].

$$\begin{bmatrix} 2 \\ 4 \\ 3 \end{bmatrix} \begin{bmatrix} c1 \\ c2 \\ c3 \end{bmatrix}$$

5. Experiment desing

The process for this experiment is divided in two parts. The first part of this experiment is the extraction of features of the images, quantization of the images features, shuffle of the images features and getting the train and test sets for experimentation. In the second part we use the KNN algorithm and the Borda voting - schemes to get the image candidate for classification.

The first part goes as follows:

- *Features Extraction:* It helps to extract the bins of all images, apply a conversion method to transform images to three descriptors, in this case HSV and CLD, and RGB descriptors, so that we have three descriptors to start our work.

- *Quantization:* The quantization method reduces the dimensionality of the image features. This method is used to improve the calculation process of the distance between two features set.

– *Shuffle*: The objective of this process is to mix the image set of features as a necessary stage for the experiment, because the system must not learn the image features in the same order each time.

– *Test and Train for experimentation*: This process is used for dividing the image feature set in two subsets to do experimentation. This two set of images features are the training set and the testing set. The training set has 70 % of the features, which is used as the base to compare the features to find a new one, and the testing set which is going to be compared with the training set of images to find to which category it fits in. In Figure 6 is explained the first part of the process.

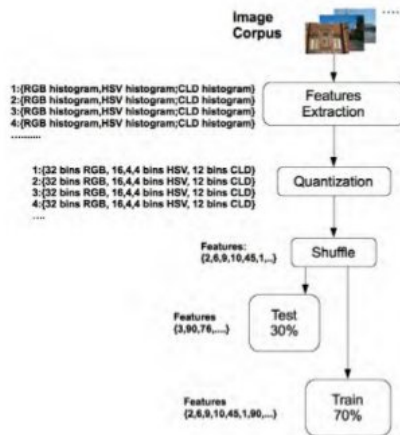


Figure 6. Representation of the extraction of features of the image corpus, quantization, shuffle of the images and get the test and train sets of images features.

The second part goes as follows:

– *KNN Algorithm*: It is used for performing the classification process based on each feature set. To do this process we take each image features of the test set and compare them to each image features of the training set. As

a result the system returns a set of votes and candidates for each descriptor.

– *Borda voting - schemes*: This ensemble method is used to classify the image, to do this step we use the votes we obtained on the earlier step. We use the candidates and votes that we obtained of the KNN process and generate a matrix with each result we had in the previous process, then we sum the matrixes and we get an array of data which results in the classification order of the candidate. As a result the system can classify the image in a category. Figure 7 explains the second part of the experimentation design.

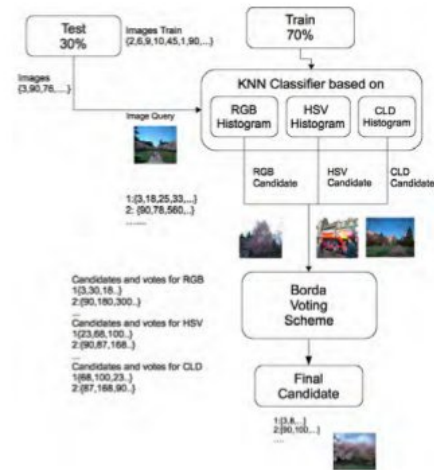


Figure 7. The second stage of the experiment process.

6. Conclusions and future work

The main goal of this project is to improve the precision and recall of the system, so that we can get best results at discovering and recognizing the patterns of the image.

Given that MPEG-7 standard defines a complete set of visual and audio descriptors, is

important to analyze if is possible to combine other visual descriptors like Edge Histogram, Dominant Color, etc. In the same way, it is important to study more combination methods like AdaBoost, Decision Trees, etc.

For future work we will start the experiments we designed in this paper with different image databases of different themes in order to improve the system.

Acknowledgment

This work was funded by the project "Using visual MPEG-7 descriptors for Content - Based Image Retrieval" of the Fourth Call for Research Funds of the Universidad Politécnic Salesiana.

References

- [1] L. Ballerini, X. Li, R. Fisher, and J. Rees, "A query-by-example content-based image retrieval system of non-melanoma skin lesions," *Medical Content-Based Retrieval for Clinical Decision Support*, pp. 31–38, 2010.
- [2] D. Tahmoush, "CBIR for mammograms using medical image similarity," *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 7628, p. 8, 2010.
- [3] T. Kijak, T. Furon, and L. Amsaleg, "Including image recognition in SIFT-based CBIR systems," in *Multimedia in Forensics, Security and Intelligence (MiFor)*, Florence, Italy, October 29 2010.
- [4] S. Caperna, C. Cheng, V. Fan, A. Luthra, B. O'Leary, J. Sheng, A. Sun, L. Stearns, R. Tessler, P. Wong, and J. Yeah, "A navigation and object location device for the blind," *University of Maryland*, 2009.
- [5] N. Singhai and S. Shandilya, "A survey on: Content based image retrieval systems," *International Journal of Computer Applications IJCA*, vol. 4, no. 2, pp. 22–26, 2010.
- [6] R. Balasubramani and V. Kannan, "Efficient use of MPEG-7 color layout and edge histogram descriptor in CBIR systems," *Global Journal of Computer Science and Technology*, vol. 9, no. 4, pp. 157–163, 2009.
- [7] S. Berretti, A. Del Bimbo, and P. Pala, "Retrieval by shape similarity with perceptual distance and effective indexing," *IEEE Transactions on Multimedia*, vol. 2, no. 4, pp. 225–239, 2000.
- [8] T. Deselaers, D. Keysers, and H. Ney, "Features for image retrieval: an experimental comparison," *Information Retrieval*, vol. 11, no. 2, pp. 77–107, 2008.
- [9] K. Mekaldji, S. Boucherka, and C. S., "Color quantization and its impact on color histogram based image retrieval," in *Proceedings of the Second Conference Internationale sur l'Informatique et ses Applications (CHIA'09)*, Saida, Algeria, May 3 - 4 2009.
- [10] J. Miralles, *Tutorial de GIMP*. [Online]. Available: <http://sites.google.com/site/tutorialdegimp/011---teoria-del-color-for%macion-y-mezcla-de-colores-rgb-y-cmyk>
- [11] S. Jeong, "Histogram-based color image retrieval," Stanford University, Palo Alto, CA, Psych221/EE362 Project Report, 2001.

- [12] A. Vadivel, S. Sural, and A. Majumdar, "Human color perception in the HSV space and its application in histogram generation for image retrieval," in *SPIE Proceedings meetings*, San José CA, United States of America, 2005.
- [13] S. Sural, G. Quian, and S. Pramatik, "Segmentation and histogram generation using the HSV color space for image retrieval," in *Proceedings International conference on Image Processing*, 2002.
- [14] University of Auckland. Department of Computer Science, *CBIR: Color Features*. [Online]. Available: <http://www.cs.auckland.ac.nz/compsci708s1c/lectures/Glect-html/topic3c708FSC.htm>
- [15] P. Cunningham and S. Delany, "K-nearest neighbors," Report of the UDC School of Computer Science and Informatics, Dublin, Ireland, Tech. Rep., 2007.
- [16] Carleton College, *Computer Science Comps Project. Netflix prize*. [Online]. Available: http://cs.carleton.edu/cs_comps/0910/netflixprize/final_results/knn/index.html
- [17] V. Robles, "Esquemas de votación borda aplicados al etiquetado de roles semánticos," Master's thesis, Universidad Politécnica de Valencia, Valencia, Spain, 2010.
- [18] —, "Borda based voting schemes for semantic role labeling," in *13th International Conference on Text, Speech, and Dialogue. Lecture Notes in Computer Science*, Brno, Czech Republic, September 2010.

Combination of HSV and RGB histograms and MPEG-7 Descriptor: Preliminar Results and Future Work.

Javer Poveda Figueroa, Vladímir Robles Bykbaev
Laboratory of Research in Computer Science and Artificial Intelligence - CIDII
Engineering Systems Career, Universidad Politécnica Salesiana,
Cuenca, Ecuador
jpoveda@est.ups.edu.ec, vrobles@ups.edu.ec

Abstract—In this paper we present the partial results obtained with a new approach to combine HSV and RGB histograms and MPEG-7 CLD descriptor. The combination was conducted using Borda Voting Schemes in three databases: Wang, ZuBud and UW. Despite the poor initial classification performed with CLD descriptor, our proposal has achieved good results for the Wang database (82.66%), outperforming the precision of HSV (72.33%) and RGB histograms (79.33%), and CLD descriptor (53%). In the other databases (ZuBud and UW) the combination approach was unable to perform a significant improvement.

Keywords—HSV histogram; RGB histogram; Colour Layout Descriptor; Borda Voting Schemes; KNN

INTRODUCTION

At present several descriptors exist to extract relevant features from images. One of the most used descriptors is the color histogram, because it is simple to implement and understand. However, the color histograms have limitations when they are used to classify images with a similar color distribution but different content. Other descriptors are more robust to classify image using textures, color distribution and shape descriptors.

Several approaches have been presented to perform image retrieval through the combination of different descriptors. Some proposals use the machine-learning approach (Support Vector Machines and Neuro-Fuzzy Networks) [1], while others directly combine the features set [2][3] to show the improvement achieved with combination.

In this paper we introduce the results of a new approach presented in [4], to combine HSV and RGB histograms and MPEG-7 CLD descriptor. In the previous paper we explained the color histograms RGB and HSV and the MPEG-7 color Layout descriptor. Then, we applied the supervised learning algorithm K-Nearest Neighbors to get the votes of each image, so that we can classify them and use the votes with the Borda Voting Scheme. Finally, we use the Borda Voting Scheme to combine the votes to classify the data we are working with.

The results achieved with our approach have obtained good results for the Wang database, improving the precision to 82.66

percent. With the other two databases that we used, the combination has not significantly improved the precision.

The rest of the paper is organized as follows. In section 2 we review the databases and tools used, and the process followed to extract the features. The technical details of the approach's implementation are reviewed in section 3. In section 4 we present the results obtained. Finally, discussion and future work are described in section 5.

DESIGN PHASE OF THE EXPERIMENT

A. Corpus selection

The process selected to choose the corpus images is explained in the next steps. UW, Wang and ZuBud corpuses were chosen because of their features. Each of these corpuses was made by a group of experts in the area of computer vision to apply machine-learning algorithms to them so that they can classify the classes of images that the image corpus has.

The University of Washington made the UW database and it contains 1109 images. The images were chosen from 18 vacation places. As [3] points out, the classification process is made with annotations, so that the image could be classified in a category based on its annotations. For example, given a query image of a category that is going to be compared with all the categories in order to determine to which category they belong to. A retrieval task could be looking for the image of the same building, or the image of any place of vacation. Figure 1 illustrates the UW database with its annotations.

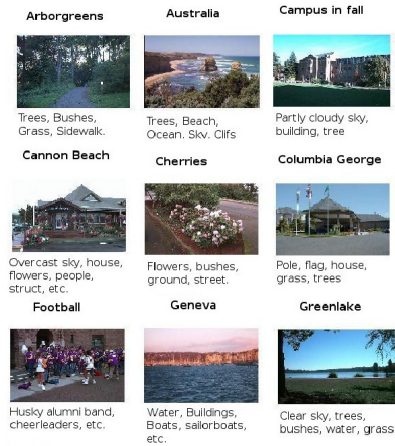


Figure 1: Description of the UW Database, which can be found at: <http://www.cs.washington.edu/research/imagedatabase/groundtruth>.

The Wang database is a collection of 1000 images from the Corel database, which are divided in 10 categories of 100 images each one. The Content-Based Image Retrieval process for this database searches from a query database comparing all the images of the database to seek where it belongs to so that it can be classified in any category. Figure 2 shows this database with its categories [3].

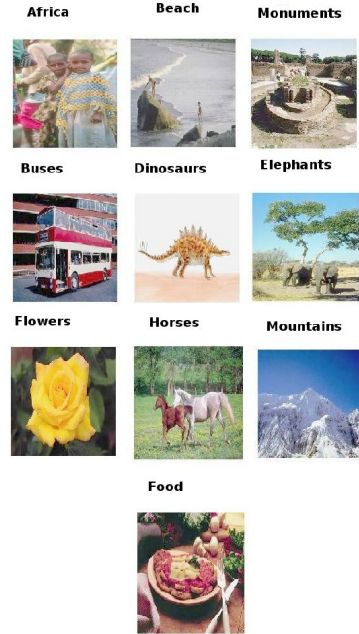


Figure 2: Images of the Wang database, which can be found at: wang.ist.psu.edu/docs/related

The ZuBud database was created at the Swiss Federal Institute of Technology in Zurich and it contains 1005 images of 201 buildings. Each building is represented in this database by 5 images; each one was taken by different viewpoints, under different conditions and with two different cameras. The Content Based Image retrieval process is the same process that was explained with the UW and Wang databases. The difference being that the query image, which is used must recover an image from the same building. Figure 3 shows this database with its categories [3].

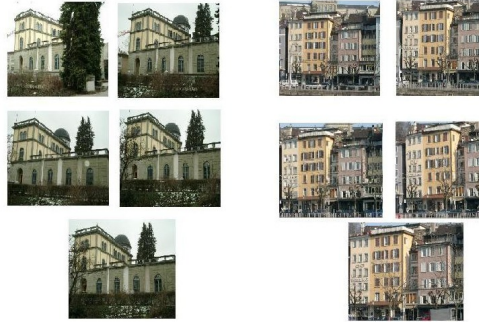


Figure 3: Images of ZuBud database, which can be found at:
<http://www.vision.ee.ethz.ch/showroom/zubud/>

B. Tools which were used to do the experiment

The tools that were used to do the experiment are: OpenCV 2.4.0 for UNIX, the C++ version 4.6 for UNIX, Visual Studio 2008 Express Edition, Java 1.6.1 and Net Beans 5.5. OpenCV2.4.0 was used to get the image features from the color spaces RGB, HSV and the MPEG-7 standard Color Layout Descriptor. Also, we used the OpenCV tool for transforming RGB to HSV color space and to the YCbCr Color space. After that, the GUI Visual Studio Standard Edition was used to program the K-Nearest Neighbors Algorithm. Then, the GUI NetBeans 5.5 with the programming language Java 1.6.1 to program the Borda Voting Schemes. Finally, the results of the experiment were saved on a Microsoft Word 2007 text document to analyze them and get the results.

C. Process followed to do the experiment

The process of the experiment was divided in two parts. The first part consists in the extraction of the characteristics of the images, the quantization and shuffle of them. The second part is the application of the machine-learning algorithm K-Nearest Neighbors and the Borda Voting Schemes. In this subsection we will explain in more detail the design phase of the experiment.

The process goes as follows:

- *Features Extraction:* This process generates the color RGB and HSV histograms and CLD descriptor. To obtain the bins of the HSV and the CLD it is necessary to convert the RGB color space of the images to the color spaces HSV and YCbCr. This process is done because the RGB histogram is the most basic histogram that can be converted to any color space [5].

To convert the RGB to the HSV color space we have to use the following formulas [5]:

$$H = \cos^{-1} \frac{0.5[(R - G) + (R - B)]}{\sqrt{(R - G)^2 + (R - G)(R - B)}}$$

Equation 1: Conversion from RGB to Hue.

$$S = 1 - \frac{3}{R + G + B} [\min(R, G, B)]$$

Equation 2: Conversion from RGB to Saturation.

$$V = \frac{1}{3}(R + G + B)$$

Equation 3: Conversion from RGB to Value.

To convert RGB to YCbCr we use the following formulas [6]:

$$Y = 0.299R + 0.587G + 0.1148B$$

Equation 4: Conversion RGB to Y.

$$Cb = -0.127R - 0.339G + 0.5211B + 128$$

Equation 5: Conversion of RGB to Cb.

$$Cr = 0.511R - 0.428G - 0.083B + 128$$

Equation 6: Conversion RGB to Cr.

- *Quantization:* The quantization process is used to reduce the dimensionality of the color features [4] and improve the performance of the system. The processing time needed to analyze them is limited because of the computer memory and the time to process them must be as shorter as possible. For this reason, the color data bins of each image are quantized to facilitate the elements that were explained. The color data bins are reduced the next form. RGB is reduced to 32 bins for each color channel, but for the experiment they were reduced to 256 bins for all the channels, instead of reducing them to 32 bins for each color channel; HSV is reduced to 16 bins in H, 4 bins in S, and 4 bins in V; finally, the CLD color descriptor produces 12 bins because its structure was defined previously.

- *Shuffle:* The objective of this process is to mix the image set of features as a necessary stage for the experiment, because the system must not learn the image features in the same order each time [4].

- *Test and Train for experimentation:* The set of color features is divided in two subsets, which are going to be used in the experiment: train and test. The train subset has 70% of the features that are going to be used as a base to compare the features to a new one, and the test subset of images that is going to be compared with the training set to determine which category the images belong to [7].

- *K-Nearest Neighbors Algorithm:* The K-Nearest Neighbor Algorithm was chosen to do the experiment because of its characteristics. The K-Nearest Neighbor Algorithm uses two sets of data, one as a standard and the other to compare each group of data with the standard set of data. Another important

characteristic of this algorithm is the structure, because it processes the group of data without any loss of it [7].

Because of the nature of the KNN algorithm, which is a machine-learning supervised algorithm, it uses two subsets of data that are the train and the test sets. As it was explained previously in the Test and Train subarea of experimentation we use the train subset as a base to compare features and the test subset to compare the images with the train subset so that we can determine which image belongs to each category of the set of images. In this case, the train and test set are used for comparing each color feature of the test set with each one of the training set, in this method the next equations [8] are applied, equation 7 represents the Euclidean distance which is used for measuring the distance on the RGB and HSV histogram, and equation 8 measures the distance of the CLD color descriptor:

$$Euclidean = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Equation 7: Euclidean distance formula which determines the distance between two points. In this case x_i represents the value of the test subset, while the

y_i represents the value of the train subset.

$$Dcd = \sqrt{\sum_i w_i^y (Y_i - Y_i')^2 + \sum_i w_i^{Cb} (Cb_i - Cb_i')^2 + \sum_i w_i^{Cr} (Cr_i - Cr_i')^2}$$

Equation 8: Formula which compares the distance between the color values of YCbCr.

Where:

- Y_i represents the value of Y of the test image and Y_i' represents the value of Y of the train image.
- Cb_i represents the value of Cb of the test image and Cb_i' represents the value of Cb of the train image.
- Cr_i represents the value of Cr of the test image and Cr_i' represents the value of Cr of the train image.
- w_i^y represents the weights of Cb, and the equation 9 represents the weight of Cb where j and k are the positions of the train and test images respectively :

$$w_i^y = \frac{0.4}{(j+k)^2}$$

Equation 9: Weight of the Cb value.

- w_i^{Cb} represents the weights of Y, and the equation 10 represents the weight of Y where j and k are the positions of the train and test images respectively :

$$w_i^{Cb} = \frac{0.2}{(j+k)^2}$$

Equation 10: Weight of the Y value.

- w_i^{Cr} represents the weights of Cr, and the equation 11 represents the weight of Cr where j and k are the positions of the train and test images respectively :

$$w_i^{Cr} = \frac{0.4}{(j+k)^2}$$

Equation 11: Weight of the Cr value.

-Borda Voting Schemes: The Borda Voting Schemes are an ensemble technique developed in France in the 18th century by the french Mathematician Charles Borda. This method was developed because the voters made very bad decisions in the voting process. This was because the voters chose the candidate not for its merits, but for his popularity. Because of this reason, Borda developed a method, which helped to choose the correct candidate by his merits [9].

The process of the Borda Voting Scheme goes as follows. The Borda Voting scheme consists of a group of candidates $X = \{1, 2, 3, \dots, X_n\}$ where $n \geq 3^2$ and m voters where $m \geq 3$ [7]. Defined in a formal ways as [9] explains: "The relationship P^k represent the preference of the voter k, $k = \{1, 2, 3, \dots, m\}$, over the collection of n alternatives of X". The relationship can be represented by the matrix [9]:

$$P^k = \begin{bmatrix} r^{k_{11}} & r^{k_{12}} & \dots & r^{k_{1n}} \\ r^{k_{21}} & r^{k_{22}} & \dots & r^{k_{2n}} \\ \dots & \dots & \dots & \dots \\ r^{k_{n1}} & r^{k_{n2}} & \dots & r^{k_{nm}} \end{bmatrix}$$

Where:

- $r_{ij}^k = \begin{cases} 1 & \text{(if } [x_i P x_j]) \\ 0 & \text{(in [other] \{case\})} \end{cases}$

It can be defined by the next equation [9]:

$$r_k(x_i) = \sum_{j=1}^n r_{ij}^k$$

Equation 12: The Borda Voting Schemes formula. If you want to know how it resolves a problem please go to [4].

APPROACH IMPLEMENTATION

This section will explain the process done to make the experiment. First, the processing of each corpus will be explained. Second, how the KNN and Borda Voting schemes algorithms were programmed.

distribution of color of them is not very similar. On figure 5 the correct and incorrect answers can be appreciated.



Figure 5: Images of the UW database where the Query Images are highlighted in red and the other four images are the top four results for this query image.

The Wang database had a precision of 82.66%. This is because the color distribution is very similar on each image of the categories. On figure 6 the correct and incorrect answers can be appreciated.

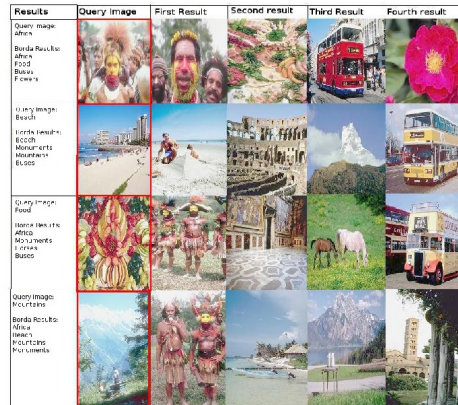


Figure 6: Images of the Wang database where the Query Images are highlighted in red and the other four images are the top four results for this query image.

The ZuBud database has a precision of 50%. This result is because the Color Layout Descriptor has more precision with images that have better color distribution and less border structure. On figure 7 the correct and incorrect answers can be appreciated.

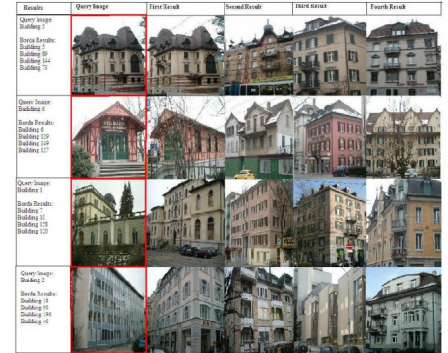


Figure 7: Images of the ZuBud database where the Query Images are highlighted in red and the other four images are the top four results for this query image.

The recall of the system was 100% because of the characteristics of the K-Nearest Neighbors.

G. Comparing Precision with other systems

1) *General Results of the system:* The results of the designed system were:

Table 1: Results of the experiment with RGB, HSV, CLD and the Borda Voting Schemes done with the experiments.

Corpus	RGB	HSV	CLD	Borda RGB+HSV+CLD
UW	87.31	94.72	31.11	50
Wang	79.33	72.33	53	82.66
ZuBud	77.18	82.08	38.39	66.42

CONCLUSIONS AND FUTURE WORK

The approach presented in this paper has outperformed the precision level of HSV and RGB Histograms, and MPEG-7 CLD Descriptor (Wang database). The precision of CLD descriptor was poor, and this fact significantly affected the overall process of classification.

In the same way, the classification process in the other databases presents low level of precision, due to CLD descriptor precision level.

Given the above, is important to count on descriptors with similar levels of precision or recall. In this way, one descriptor can be good to classify some types of images, and other descriptor can be good to classify other types, giving us better results in higher global precision.

For future work we will incorporate the Boosting-Based classification, using uncorrelated MPEG-7 descriptors to analyze if is possible to improve precision in CBIR tasks.

ACKNOWLEDGMENT

This work was funded by the project "Using visual MPEG-7 descriptors for Content – Based Image Retrieval" of the Fifth

Call for Research Funds of the Universidad Politécnica Salesiana.

REFERENCES

- [1] E. Spyrou, and H. Le Borgne, and T. Mailis, and E. Cooke, E. and Y. Avrithis, and N. O'Connor, "Fusing MPEG-7 visual descriptors for image classification", Lecture notes in computer science, pp. 847-852, 2005.
- [2] P.V.B. Reddy, and A.R.M. Reddy, and K.G.J. Devi, "HSV Color Histogram and Directional Binary Wavelet Patterns for Content Based Image Retrieval", vol. 4, International Journal on Computer Science and Engineering (IJCSE), 2012.
- [3] T. Deselaers, D. Keysers and H. Ney, "Features for image retrieval: an experimental comparison", Information Retrieval, vol. 11, no. 2, pp. 77-107, 2008.
- [4] J. Poveda, V. Robles, "Image retrieval based on the combination of RGB and HSV's and Color Layout Descriptor". INGENIUS No 7. (january/june). pp. 3-10. ISSN: 1390-650X.
- [5] K. Mekaldji, S. Boucherka, "Color quantization and its impact on color histogram base image retrieval". Proceedings of the second conference International sur l'Informatique et ses Applications (CIIA' 09), Saida, Algeria, May 3-4 2009.
- [6] Intersil. "YCbCr to RGB Considerations ". Available at: <http://www.intersil.com/content/dam/Intersil/documents/an97/an9717.pdf>.
- [7] P. Harrington, "Machine Learning in Action", Manning Publications, New York, 2012.
- [8] O. Boullosa, "Estudio Comparativo de descriptores visuales para la detección de escenas cuasi-duplicadas" Universidad Politécnica de Madrid, Madrid, Spain, 2011.
- [9] V. Robles, "Esquemas de Votación Borda aplicados al etiquetado de roles semánticos", Universidad Politécnica de Valencia, Valencia, Spain, 2010.
- [10] S. Jeong, "Histogram-Based Color Image Retrieval", Psych221/EE362, Project Report, Palo Alto, Stanford University, 2001.

**LA FEDERACION DE ESTUDIANTES DE LA UNIVERSIDAD POLITECNICA
SALESIANA SEDE CUENCA**

CERTIFICA:

Que, el Sr. **JAVIER RAMIRO POVEDA FIGUEROA**, ha colaborado en con nosotros como **EXPOSITOR**, durante el **SEGUNDO CONGRESO DE TECNOLOGIAS UPS 2012** demostrando durante su intervención responsabilidad, honestidad y dedicación al momento de realizar dicha disertación.

Se expide la presente a solicitud del interesado, para los fines que crea conveniente.

Cuenca, 02 de Diciembre del 2012



Sr. Christian Díaz
Representante Estudiantil de la FEUPS