
.



INSTITUTO POLITÉCNICO NACIONAL
CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN



Recuperación de imágenes mediante rasgos descriptores globales y locales

TESIS

Que para obtener el grado de Doctor en Ciencias de la
Computación

Presenta

José Félix Serrano Talamantes

Directores

Dr. Juan Humberto Sossa Azuela

Dr. Carlos Avilés Cruz

México D.F., Enero de 2011



INSTITUTO POLITÉCNICO NACIONAL SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de México, D.F. siendo las 17:30 horas del día 13 del mes de Enero de 2011 se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación del:

Centro de Investigación en Computación

para examinar la tesis titulada:

“RECUPERACIÓN DE IMÁGENES MEDIANTE RASGOS DESCRIPTORES GLOBALES Y LOCALES”

Presentada por el alumno:

SERRANO

Apellido paterno

TALAMANTES

Apellido materno

JOSÉ FÉLIX

Nombre(s)

Con registro:


| | | | | | | |
|---|---|---|---|---|---|---|
| A | 0 | 7 | 0 | 2 | 5 | 4 |
|---|---|---|---|---|---|---|

aspirante de: **DOCTORADO EN CIENCIAS DE LA COMPUTACIÓN**


Después de intercambiar opiniones los miembros de la Comisión manifestaron **APROBAR LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISIÓN REVISORA

Directores de Tesis



Dr. Juan Humberto Sossa Azuela



Dr. Carlos Avilés Cruz



Dr. Edgardo Manuel Felipe Riverón

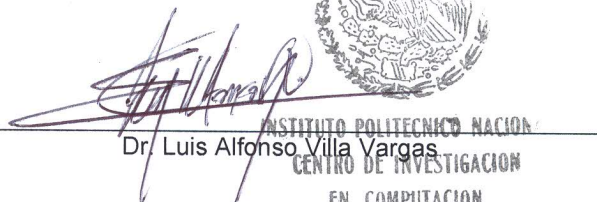


Dr. Olexsiy Pogrebnyak



Dr. Ricardo Barrón Fernández

PRESIDENTE DEL COLEGIO DE PROFESORES



Dr. Luis Alfonso Villa Vargas



INSTITUTO POLITÉCNICO NACIONAL
CENTRO DE INVESTIGACIÓN
EN COMPUTACIÓN
DIRECCIÓN

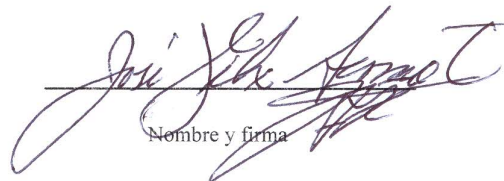


INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESIÓN DE DERECHOS

En la Ciudad de México D.F. el día 13 del mes de **Enero** del año 2011, el (la) que suscribe **Serrano Talamantes José Félix** alumno (a) del Programa de **Doctorado en Ciencias de la Computación** con número de registro **A07254**, adscrito al Centro de Investigación en Computación _____, manifiesta que es autor (a) intelectual del presente trabajo de Tesis bajo la dirección del **Dr. Juan Humberto Sossa Azuela** y el **Dr. Carlos Avilés Cruz** y cede los derechos del trabajo **titulado “Recuperación de imágenes mediante rasgos descriptores globales y locales”** _____, al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección **jfserrano@ipn.mx** _____. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.


Nombre y firma

RESUMEN

En este trabajo se propone una metodología para la recuperación de imágenes de escenarios naturales. Para esto se usa un conjunto de rasgos descriptivos obtenidos a partir de un conjunto de puntos aleatoriamente seleccionados más un clasificador, cuyo resultado es una *base de datos indexada*. Uno de los propósitos de esta investigación es verificar qué tan discriminantes son las características de la media, la desviación estándar y la homogeneidad proveniente de una *matriz de coocurrencia* para describir las diferentes clases de objetos presentes en una escena. Se propone el uso del *algoritmo de las K medias* con el fin de encontrar grupos (*clusters*) no correlacionados de objetos en imágenes de escenarios naturales. Mediante experimentación se ha demostrado que los grupos obtenidos mediante el algoritmo de las *K medias* no están asociados necesariamente con etiquetas generadas manualmente por un observador humano al analizar este tipo de imágenes. La metodología propuesta ha sido comparada con otros trabajos similares relacionados con la categorización de conceptos presentes en escenarios naturales.

Palabras clave: Recuperación de imágenes, puntos aleatorios, matriz de coocurrencia, algoritmo de *K -medias* y base de datos indexada.

ABSTRACT

In this work we describe a methodology for image retrieval of natural scenes. For this, a combination of features computed from a set of randomly selected points and a classifier is used. The output is an indexed database. One of the goals of this research is to test how the *mean, the standard deviation and the homogeneity* (from a cooccurrence matrix) allow differentiating among different object classes in a scene. We propose to use a *K means algorithm* to find uncorrelated groups or clusters of objects in images of natural scenes. Throughout experimentation we have shown that the groups obtained by means of the *K means* algorithm are not necessarily associated with the labels generated by a human user when analyzing images of natural scenes. The proposed methodology has been compared with other reported methodologies in the content of concept categorization in images of natural scenarios.

Key words: CBIR, Image Retrieval, random points, K means algorithm, and cooccurrence matrix.

AGRADECIMIENTOS

En primer lugar a **Dios**, por concederme la vida y la salud y así poder culminar y concluir este proyecto de vida.

A **Ludy, mi amada esposa**, que me sugirió la idea de hacer un estudio doctoral; gracias por haberme apoyado en todo, por su AMOR, aliento, esperanza y paciencia (MUCHA PACIENCIA) para poder realizar y terminar este trabajo. Siempre estuvo al pendiente de mí en todos los aspectos y siempre creyó en mí.

A mis **queridos enanos Keylita y Eduardito**, que siempre me preguntaban que estaba haciendo y que si me faltaba mucho por terminar, les agradezco por su paciencia y tolerancia cuando querían jugar conmigo mientras me dedicaba a mis estudios. Gracias queridos hijitos...

A mis **asesores el Dr. Carlos Avilés Cruz (UAM)** por haberme transmitido sus grandes enseñanzas, por su Gran Paciencia, por su amistad y que siempre creyó en mí.

y al **Dr. Humberto Sossa Azuela (CIC-IPN)** por sus enseñanzas, motivación y dedicación hacia mí en la realización de esta tesis.

A los miembros de comité revisor:

- **Dr. Edgardo Manuel Felipe Riverón** por las charlas e intercambios de opiniones en los corredores del CIC., y por las múltiples revisiones de la tesis.
- **Dr. Ricardo Barrón Fernández**, que siempre me preguntaba acerca del desarrollo de esta tesis y me dio ánimos, consejos y sugerencias para realizar esta tesis.
- **Dr. Oleksiy Pogrebnyak** por sus comentarios y sugerencias.

Al **Dr. Gustavo Olague Caballero (CICESE)**, que también estuvo al pendiente de mí con sus enseñanzas, convivencias y su gran hospitalidad en Ensenada Baja California durante mis estudios y que siempre creyó en mí.

También al grupo de **estudiantes de (Evo-visión)** por sus comentarios y sugerencias al respecto.

Al **doc Fernando Galindo Soria** por su amistad y sus sabios consejos.

Al **Dr. Eduardo de Jesús García García (UVM)** por su gran amistad en todos estos años, y porque siempre ha confiado en mí y me ha motivado para seguir adelante y por transmitirme sus enseñanzas y darme consejos y sugerencias.

Al **Ing. Francisco Apolinar Cruz Lázaro** (Director de la ESCOM-IPN) por su apoyo incondicional.

Al **Dr. Juan Villegas Cortés**, por su amistad y por su gran colaboración en el desarrollo de esta tesis.

A **Itztli** y **Jorge Solórzano** por su apoyo moral y su gran amistad.

A **Jesús Alfredo Martínez Nuño (NUÑIN)**, a **Jorge Servín** y **Roberto Antonio Vázquez Espinoza de los Monteros** por su amistad y por sus comentarios y sugerencias.

A los compañeros estudiantes y profesores del Laboratorio de Reconocimiento de Patrones e Inteligencia Artificial del (CIC-IPN).

Al **Consejo Nacional de Ciencia y Tecnología (CONACyT)** por la beca otorgada durante mis estudios doctorales.

Al COTEPABE por su gran apoyo por el permiso para poderme dedicar a tiempo completo al desarrollo de mi tesis.

A la DAAD por el apoyo económico para la estancia realizada en Jena, Alemania, bajo el proyecto PROALMEX J000.426/2009.

A los compañeros estudiantes del Laboratorio de Visión por Computadora, de la Facultad de Matemáticas de la Universidad Friedrich Schiller de Jena, Alemania.

Este documento se ha realizado con la ayuda financiera de la Unión Europea y el CONACYT, bajo el proyecto FONCICYT 93829.

El contenido de este documento es responsabilidad exclusiva del IPN y en modo alguno debe considerarse que refleja la posición de la Unión Europea.

A todas las personas que nunca creyeron en mí, por motivarme día a día a luchar en lograr esta meta que me propuse conquistar.

Finalmente a todas aquellas personas que de alguna forma u otra contribuyeron conmigo e hicieron posible la realización de este trabajo...

“...MIL GRACIAS A TODOS Y CADA UNO DE USTEDES.....”

Índice general

| | |
|--|-----------|
| Índice general | 11 |
| Lista de figuras | 15 |
| Lista de tablas | 21 |
| 1 INTRODUCCIÓN | 25 |
| 1.1. Planteamiento del problema | 28 |
| 1.2. Consideraciones generales de la metodología | 31 |
| 1.3. Justificación | 31 |
| 1.4. Identificación del problema | 32 |
| 1.4.1. Objetivo general | 32 |
| 1.4.2. Objetivos específicos | 33 |
| 1.5. Aportaciones | 33 |
| 1.6. Organización de la tesis | 34 |
| 2 ESTADO DEL ARTE | 35 |
| 2.1. Antecedentes | 35 |
| 2.2. Definición general del problema | 42 |
| 2.3. Eficiencia y carga computacional | 42 |
| 2.4. Tipos de consulta | 43 |

| | | |
|----------|---|-----------|
| 2.5. | Técnicas de recuperación | 44 |
| 2.5.1. | Emparejamiento directo | 45 |
| 2.5.2. | Estructuración del espacio de búsqueda | 48 |
| 2.6. | Sistemas en línea | 51 |
| 2.6.1. | CIRES | 51 |
| 2.6.2. | FIRE | 52 |
| 2.6.3. | IRMA | 53 |
| 2.6.4. | Análisis de los resultados | 53 |
| 3 | MARCO TEÓRICO | 55 |
| 3.1. | Reconocimiento de patrones | 56 |
| 3.2. | Atributos o descriptores | 60 |
| 3.2.1. | Textura | 60 |
| 3.2.2. | Matriz de coocurrencia | 61 |
| 3.2.2.1. | Cálculo de la matriz de coocurrencia | 62 |
| 3.3. | Agrupamiento no supervisado | 65 |
| 3.3.1. | Algoritmo de K-medias | 65 |
| 3.4. | Clasificadores | 68 |
| 3.4.1. | Clasificador de los K vecinos más próximos (K-NN) | 68 |
| 3.4.1.1. | Principio teórico | 68 |
| 3.4.1.2. | Cálculo de la distancia | 72 |
| 3.4.2. | Clasificador bayesiano | 72 |
| 3.4.2.1. | Patrones y clases de patrones | 72 |
| 3.4.2.2. | Métodos de decisión teórica | 73 |
| 3.4.2.3. | El clasificador de Bayes para clases de patrones gaussianos | 75 |
| 3.4.3. | Redes neuronales | 77 |
| 3.4.3.1. | La red de retropropagación (<i>Backpropagation</i>) | 79 |

| | | |
|----------|--|------------|
| 3.4.3.2. | Estructura y aprendizaje de la red de retropropagación . . . | 80 |
| 3.4.3.3. | Estructura de una sola neurona | 81 |
| 3.4.3.4. | Algoritmo de entrenamiento de la red de retropropagación | 82 |
| 3.4.3.5. | Diagrama de flujo de los pasos y procedimientos del al- goritmo de la red de retropropagación | 86 |
| 3.4.3.6. | Dimensiones de la red (número de neuronas ocultas) . . . | 87 |
| 4 | METODOLOGÍA PROPUESTA | 89 |
| 4.0.4. | Etapas de recuperación | 94 |
| 4.0.5. | Metodología basada en puntos aleatorios | 96 |
| 4.0.6. | Construcción de la base indexada | 98 |
| 5 | RESULTADOS EXPERIMENTALES | 101 |
| 5.1. | Recuperación de imágenes | 101 |
| 5.2. | Identificación del tipo de imagen | 117 |
| 5.3. | Resultados obtenidos con la metodología modificada (solamente con pun- tos aleatorios) | 121 |
| 6 | CONCLUSIONES Y TRABAJO FUTURO | 133 |
| 6.1. | Conclusiones | 133 |
| 6.2. | Trabajo futuro | 134 |
| 7 | APÉNDICES | 137 |
| 7.1. | Definición de imagen digital | 137 |
| 7.2. | Comprensión de las imágenes | 138 |
| 7.3. | Representación de las imágenes | 139 |
| 7.4. | Características de una imagen | 140 |
| 7.5. | Fundamentos del color | 143 |
| 7.5.1. | Modelos de color | 144 |

Referencias

147

Lista de figuras

| | |
|--|----|
| 1.1. Diagrama de bloques de un sistema de visión artificial. | 26 |
| 1.2. Idea general del problema de la recuperación de imágenes. | 28 |
| 1.3. Diagrama de bloques del modelo para recuperar imágenes. | 30 |
| 2.1. Ilustración con diversos tipos de imagen: a) laguna, b) mapa, c) persona, d) automóvil, e) avión, f) teléfono celular, g) puesta de sol, h) automóvil Ford T, i) automóvil Jeep | 36 |
| 2.2. Resultados inconsistentes en el buscador de Google cuando se quiere buscar la imagen de un león y se escribe en el buscador la etiqueta “león”. | 40 |
| 2.3. Taxonomía de las técnicas de recuperación. | 45 |
| 3.1. Diagrama general de un sistema de reconocimiento de objetos. | 58 |
| 3.2. Imagen con tres niveles de gris | 62 |
| 3.3. Matriz de coocurrencia para $d=1$ a 0^0 | 63 |
| 3.4. Matriz de coocurrencia para $d=1$ a 45^0 | 63 |
| 3.5. Matriz de coocurrencia para $d=1$ a 90^0 | 63 |
| 3.6. Diagrama de flujo del algoritmo de las k -medias. | 67 |
| 3.7. Clasificador K-NN | 68 |

| | |
|---|----|
| 3.8. Selección de los k -vecinos, donde el patrón 'x' está representado por el pequeño círculo blanco, el cual es clasificado con la clase \otimes dado que de sus $k(3)$ próximos vecinos, "uno" pertenece a la clase \star , y "dos" a la clase \otimes | 70 |
| 3.9. Selección de los k vecinos por "volumen". El círculo blanco es clasificado en la clase \otimes , dado que $K=3$ próximos vecinos están más próximos que los 3 elementos próximos de la clase de puntos \star | 72 |
| 3.10. Ejemplo de clasificación de tres clases. | 75 |
| 3.11. Funciones de densidad de probabilidades para dos clases de patrones unidimensionales. El punto x_0 es la frontera de decisión si las 2 clases son equiprobables. | 76 |
| 3.12. Modelo de la arquitectura de una red de retropropagación. | 81 |
| 3.13. Estructura de una célula. | 82 |
| 3.14. Diagrama de flujo de los pasos y procedimientos de la red de retropropagación. | 87 |
| 4.1. Diagrama de flujo para la etapa del entrenamiento. | 91 |
| 4.2. (a).-Para la descripción de las sub-imágenes, 300 pixeles de imagen son aleatoriamente seleccionadas uniformemente. (b).-Para lograr una segmentación automática de la imagen, alrededor de cada uno de los 300 pixeles se abre una ventana cuadrada de tamaño $M \times N$. En esta figura se muestran solamente 20 puntos como ejemplo. | 92 |
| 4.3. Escenas de costa, río/lago, bosque, montaña, pradera y cielo/nubes respectivamente. | 93 |
| 4.4. Diagrama de flujo para la etapa de la prueba. | 94 |
| 4.5. (a) Una imagen es uniformemente dividida en 100 subimágenes para obtener 100 regiones descriptivas de características. (b) Para cada una de las subimágenes, una ventana de tamaño 10×10 pixeles es seleccionada para calcular el correspondiente vector de características. | 95 |

| | | |
|------|---|-----|
| 4.6. | Diagrama de flujo para el entrenamiento de la metodología modificada. | 97 |
| 4.7. | Diagrama de flujo para la prueba modificada | 98 |
| 5.1. | <i>Clusters</i> formados en una imagen de una escena natural usando el algoritmo de K-Medias y puntos aleatorios. | 102 |
| 5.2. | (a) Imagen rotada 90 ⁰ . (b) Imagen rotada 180 ⁰ . (c) Imagen escalada al 50%. Obsérvese como el resultado presenta invarianza ante estas transformaciones. . | 102 |
| 5.3. | <i>Clusters</i> formados en una imagen de una escena natural usando el algoritmo de K-Medias y puntos aleatorios para imágenes de una imagen costera. | 103 |
| 5.4. | Imágenes recuperadas dada una escena consulta tipo costa. | 104 |
| 5.5. | Imágenes recuperadas dada una imagen consulta tipo bosque. | 105 |
| 5.6. | Imágenes recuperadas dada una imagen consulta tipo costa. | 105 |
| 5.7. | Eficiencia de nuestra propuesta comparada contra el método descrito en [51]. Mediante nuestra propuesta se obtiene 88.68% de eficiencia (gráfica azul), mientras que en [51] se obtiene 85.60% de eficiencia (gráfica en negro) cuando se aplica una consulta de una escena de una puesta de sol con cielo rojo. | 107 |
| 5.8. | Eficiencia de nuestra propuesta comparada contra el método descrito en [51]. Mediante nuestra propuesta se obtiene 81.58% de eficiencia (gráfica azul), mientras en [51] se obtiene 77.66% de eficiencia (gráfica negra) cuando se aplica una consulta de la escena de un bosque. | 108 |
| 5.9. | Eficiencia de nuestra propuesta al comparar contra el método descrito en [25]. Se obtiene una eficiencia del 81.7% (gráfica azul) mientras que en [25] se ob- tiene una eficiencia de 77.71% (gráficas en rojo y negro respectivamente). | 109 |

5.10. (a).-Para la descripción de las subimágenes, 300 píxeles de imagen son automáticamente seleccionados aleatoriamente. (b).-Para lograr una segmentación automática de la imagen, alrededor de cada uno de los 300 píxeles se abre una ventana cuadrada de tamaño $M \times N$. En esta figura solamente 20 puntos se muestran para dar un ejemplo. 109

5.11. (a) Una imagen es uniformemente dividida en 100 subimágenes para obtener 100 regiones descriptivas de características. (b) Para cada una de las subimágenes, una ventana de tamaño 10×10 píxeles es seleccionada para calcular el correspondiente vector de características. 110

5.12. Recuperación de imágenes de cielo completamente nublado cuando se aplica al sistema una imagen consulta tipo cielo nublado. 110

5.13. Eficiencia de nuestra propuesta comparada contra el método descrito en [14]. . 111

5.14. Las 6 clases de objetos presentes en las imágenes del entrenamiento. 112

5.15. Propuesta de la existencia de 4 clases adicionales a las 6 que ya están propuestas, las cuales les llamaremos “clases de borde, o de frontera”. 113

5.16. Imágenes recuperadas dada una imagen consulta tipo bosque. 114

5.17. Imágenes recuperadas dada una imagen consulta tipo costa. 114

5.18. Imágenes recuperadas dada una imagen consulta tipo lago. 114

5.19. Imágenes recuperadas dada una imagen consulta tipo montaña. 115

5.20. Imágenes recuperadas dada una imagen consulta tipo cielo/nubes. 115

5.21. Imágenes recuperadas dada una imagen consulta tipo pradera. 115

5.22. Eficiencia de nuestra propuesta, usando 10 clases y 700 imágenes de entrenamiento. 116

5.23. Eficiencia de nuestra propuesta de “Identificación de la imagen consulta”. . . . 120

5.24. Tipos de imagen de la base de imágenes COREL. 121

5.25. Tipos de imagen de la base de imágenes CALTEK. 122

5.26. Tipos de imagen de la base de imágenes TORRALBA. 122

| | |
|---|-----|
| 5.27. Precisión del clasificador 1-NN en la base de imágenes CALTEK. | 125 |
| 5.28. Precisión del clasificador Bayesiano en la base de imágenes CALTEK. | 126 |
| 5.29. Precisión de una red neuronal artificial (RNA) usando la combinación de neuronas del caso 1 en la base de imágenes CALTEK. | 126 |
| 5.30. Precisión de una red neurona artificial (RNA) usando la combinación de neuronas del caso 2 en la base de imágenes CALTEK. | 127 |
| 5.31. Precisión de una red neuronal artificial (RNA) usando la combinación de neuronas del caso 3 en la base de imágenes CALTEK. | 127 |
| 5.32. Precisión de un clasificador 1-NN en la base de imágenes TORRALBA. | 128 |
| 5.33. Precisión de un clasificador Bayesiano en la base de imágenes TORRALBA. | 128 |
| 5.34. Precisión de una red neuronal artificial (RNA) usando la combinación de neuronas del caso 1 en la base de imágenes TORRALBA. | 129 |
| 5.35. Precisión de una red neurona artificial (RNA) usando la combinación de neuronas del caso 2 en la base de imágenes TORRALBA. | 129 |
| 5.36. Precisión de una red neuronal artificial (RNA) usando la combinación de neuronas del caso 3 en la base de imágenes TORRALBA. | 130 |
| 5.37. Precisión de un clasificador 1-NN en la base de imágenes COREL. | 130 |
| 5.38. Precisión de un clasificador Bayesiano en la base de imágenes COREL. | 131 |
| 5.39. Precisión de una red neuronal artificial (RNA) usando la combinación de neuronas del caso 1 en la base de imágenes COREL. | 131 |
| 5.40. Precisión de una red neuronal artificial (RNA) usando la combinación de neuronas del caso 2 en la base de imágenes COREL. | 132 |
| 5.41. Precisión de una red neuronal artificial (RNA) usando la combinación de neuronas del caso 3 en la base de imágenes de COREL. | 132 |
| 7.1. Convenio de ejes utilizado para la representación de una imagen digital. | 138 |

7.2. Cubo de color RGB. Los puntos a lo largo de la diagonal principal tienen valores de gris, desde el punto negro en el origen al punto blanco. 146

7.3. Triángulo de color en el modelo HSI. Color e Intensidad en el modelo HSI. . . 146

Lista de tablas

| | |
|---|-----|
| 2.1. Clases de imágenes del CIRES | 52 |
| 2.2. Características visuales de FIRE. | 54 |
| 2.3. Eficiencia del sistema IRMA. | 54 |
| 2.4. Comparación con los sistemas en línea. | 54 |
| 4.1. Distribución de los 210,000 características entre las 10 clases o componentes de imagen (CI) seleccionadas para el conjunto de las 700 imágenes de las imágenes de los escenarios naturales usadas para construir la base indexada de datos. | 93 |
| 4.2. Estructura de la base de datos indexada. | 96 |
| 4.3. Estructura de la base indexada usando solamente puntos aleatorios | 99 |
| 5.1. Base de datos indexada para 6 clases de objetos y 300 imágenes de entrenamiento. | 104 |
| 5.2. Promedio de eficiencia para la metodología descrita en [48] y [50]. | 117 |
| 5.3. Resultados obtenidos con nuestra propuestas (valores promedio obtenidos de la figura 5.22). | 117 |
| 5.4. Matriz de confusión para el grupo 1. | 119 |
| 5.5. Matriz de confusión para el grupo 2. | 120 |
| 5.6. Resultados obtenidos para “Identificando la escena consulta” (valores promedio obtenidos de la figura 5.23 en la página 120. | 120 |
| 5.7. Eficiencia de cada uno de los clasificadores para cada base de datos de imágenes. | 124 |

Siglas y acrónimos

- RGB.-Modelo en color del inglés (Red,Green,Blue), que significa, rojo verde y azul respectivamente.
- HSI.-Modelo en color del inglés (Hue, Saturation,Intenity), que significa Tono,Saturación e Intensidad, respectivamente.
- CBIR.- Del Inglés (Content Based Image Retrieval), que significa Recuperación de Imágenes en Base a su Contenido.
- SARO.- Sistema Automático de reconocimiento de Objetos.
- K-means.- Algoritmo de Agrupamiento no supervisado. K-medias o K-centros. Es un algoritmo de *clustering* más conocido y utilizado ya que es de muy simple aplicación y eficaz, creado por MacQueen.
- GLCM.- Del Inglés (Grey Level Coocurrence Matriz) que significa Matriz de coocurrencia en niveles de gris.
- K-NN.-Del Inglés (K-Nearest Neighbour) que significa los K próximos vecinos.
- RNA.- Red Neuronal Artificial.
- Bases de imágenes de entrenamiento:
 - COREL: 668 imágenes { 103 bosques, 255 costas, 179 montañas y 131 praderas }.

- TORRALBA:1471 imágenes{328 bosques, 359 costas; 379 montañas y 410 praderas}.
- CALTEK: 373 imágenes{128 bonsai, 60 Joshua, 85 girasoles, 64 lotos y 36 lirios}.

Capítulo 1

INTRODUCCIÓN

Uno de los principales problemas a los que se enfrenta la sociedad de la información en la actualidad, es la gestión óptima y productiva de la información disponible. En otras palabras, diariamente se generan grandes cantidades de datos, por lo que es imprescindible disponer de técnicas que nos ayuden a localizar, en el menor tiempo posible, la información que es relevante para nuestras necesidades. Uno de los paradigmas que en la última década ha experimentado un amplio desarrollo dentro de la visión artificial, es el estudio de técnicas de recuperación en grandes bases de datos y de imágenes a través de imágenes clave.

El tratar de dotar a los sistemas artificiales de capacidades de captación y procesamiento similares a las de los seres humanos, ha sido uno de los retos más interesantes para el ser humano [28].

Para que un sistema de visión artificial (ver figura 1.1) pueda interactuar eficientemente con el medio que lo rodea, como lo hace el ser humano, es necesario que cuente con las capacidades adecuadas de captura y análisis automático de la información visual que recibe [28].

De acuerdo a la figura 1.1, la visión por computadora es la capacidad de una máquina para ver el mundo que lo rodea de una forma tridimensional a partir de varias imágenes

1. INTRODUCCIÓN

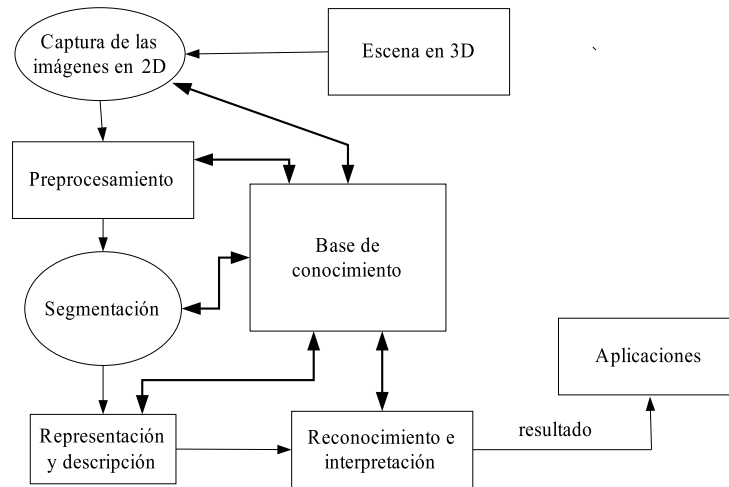


Figure 1.1: Diagrama de bloques de un sistema de visión artificial.

bidimensionales. El problema de la visión por computadora es un problema abierto para el cual no existe ningún algoritmo eficaz que reconozca todo tipo de objeto en cualquier ambiente de forma similar a como lo realiza nuestro sentido de la vista [20].

Cuando hablamos del reconocimiento automático de objetos (ver 1.1), se dice que el proceso se inicia con la captura en 2D de una imagen en tres dimensiones, la cual se procesa para mejorar la calidad y eliminar posibles imperfecciones; el siguiente paso es separar los objetos del fondo de la imagen mediante alguna técnica de segmentación, seguida de la extracción de las características que describen a los objetos (color, textura, geometría, etc); finalmente se comparan estas características con las de otros objetos que se tienen en la base de conocimiento y así determinar la clase a la que pertenecen [21].

En cada una de las etapas de la figura 1.1 se requiere de un conocimiento previo, como puede ser, el tamaño de los objetos, el número de objetos a reconocer en la imagen, etc, para aplicar determinadas técnicas y algoritmos adicionales para su reconocimiento. Dados los conceptos anteriores podemos hacer la siguiente pregunta: ¿Cómo es posible que una computadora pueda realizar millones de cálculos por segundo y no sea capaz de reconocer una simple imagen e interpretarla como un coche, un escenario natural, una persona, etc? el

enfoque dado a este trabajo consiste en utilizar técnicas y herramientas de la computación para que una computadora pueda reconocer patrones en la imagen y con ello realizar el proceso de su recuperación.

Actualmente no existen buscadores eficientes para buscar imágenes mediante imágenes, aunque hay muchos de ellos para texto como Google, Yahoo, Lycos, Altavista, etc. Cuando el usuario busca imágenes de la manera descrita, los buscadores le devuelven muchas imágenes, de las que no está buscando, ya que el texto asociado a la imagen no está acorde con el contenido de la misma, haciéndole perder tiempo y recursos; por ejemplo, si el usuario teclea la palabra *playas*, los buscadores responderán con imágenes que tengan asociada la etiqueta *playas* que algún usuario mal o bien les asignó; entonces habrá una imagen o varias que su contenido sea efectivamente de playas y otras no precisamente serán imágenes cuyo contenido sea de *playas*, porque la búsqueda se realiza por la etiqueta asignada y no por su contenido.

Con esta tesis se logró evitar este tipo de problemas, al darle un enfoque más específico al reconocer los objetos locales de las imágenes descargando solo aquellas realmente asociadas a las de la imagen consulta.

Mientras que para el ser humano no presenta dificultad el reconocer y recuperar datos multimedia [5], los sistemas de información actuales presentan varios problemas debido a que en lo fundamental están diseñados para procesar información alfanumérica, aunque algunas veces son expandidos con herramientas de desarrollo gráfico y con simples técnicas de diagramación y dibujo [7]. Por lo tanto, hasta ahora, no hay muchos sistemas que hayan sido diseñados enfocados hacia las tareas de reconocimiento y recuperación de imágenes de forma eficiente.

En este trabajo se desarrolló un sistema de información visual que utiliza paradigmas orientados particularmente al procesamiento de la información visual en imágenes con escenas de tipo natural (objetos no hechos por el hombre dentro del conjunto de imágenes de paisajes), específicamente mediante la organización y recuperación de las imágenes

1. INTRODUCCIÓN

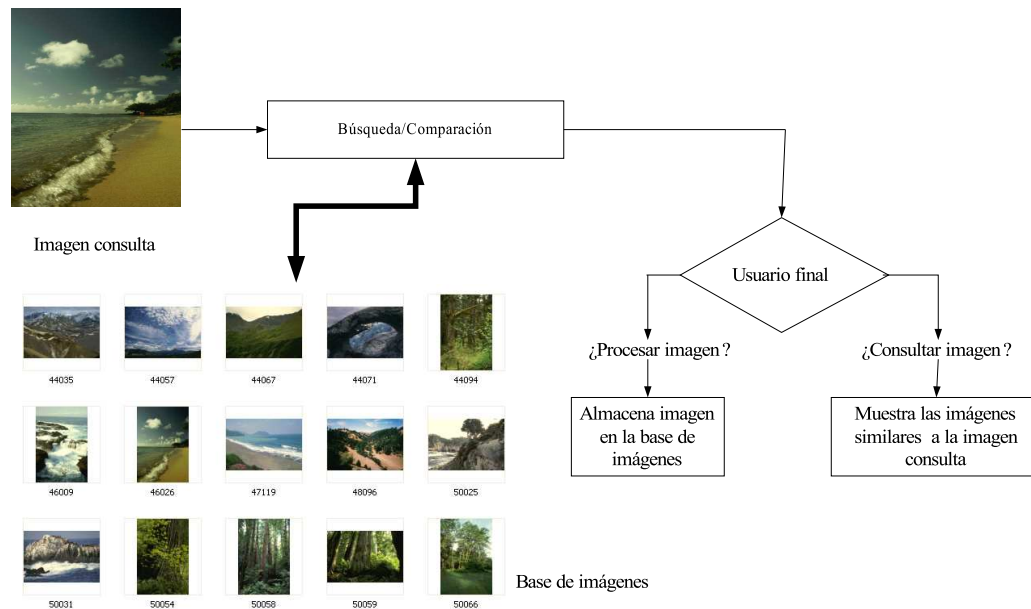


Figure 1.2: Idea general del problema de la recuperación de imágenes.

sobre la base de su contenido, en términos del color y la textura así como la clasificación mediante métodos estadísticos, bayesianos y de redes neuronales.

1.1. Planteamiento del problema

En general, el problema de la recuperación de imágenes consiste en que dada una imagen consulta I_c , extraer de un banco de imágenes aquellas “más parecidas” a I_c sobre la base de su contenido (ver figura 1.2). Esto se logra al comparar la imagen consulta con cada una de las imágenes del banco de imágenes.

Dicha comparación se puede hacer píxel a píxel [23]. Sin embargo, en este proyecto de tesis la comparación se realiza al transformar la imagen consulta en un conjunto de vectores descriptores de n rasgos cada uno. En el presente trabajo se puede decir que dos imágenes son “similares o parecidas” si sus respectivos conjunto de vectores descriptores son parecidos con respecto a una métrica dada.

En los sistemas de recuperación de imágenes, debe existir la capacidad de comparar

eficientemente dos imágenes para determinar si tienen contenido similar con respecto a las características extraídas. Dichas características representan la información discriminante.

Desde este punto de vista, el problema de la recuperación de imágenes se puede plantear de la siguiente manera: una imagen digital constituye una distribución bidimensional. Para las imágenes en blanco y negro, $f(x, y)$ denota el valor de la intensidad del píxel (x, y) en la escala de grises. Para las imágenes en color, $f(x, y)$ denota el valor del color compuesto del píxel (x, y) en sus componentes RGB o HSI [23]. Si la información en color se representa en términos de los tres colores primarios RGB (rojo, verde y azul) [23], la función imagen se describe como se muestra en la ecuación (1.1).

$$f(x, y) = \{f_R(x, y), f_G(x, y), f_B(x, y)\} \quad (1.1)$$

Si la información de color se descompone en términos de los tres canales para la caracterización del color HSI, la función imagen se describe como la ecuación (1.2).

$$f(x, y) = \{f_H(x, y), f_S(x, y), f_I(x, y)\} \quad (1.2)$$

(H).-representa al tono, el cual está relacionado con la longitud de onda dominante en un conjunto de ondas luminosas. Describe un color puro (amarillo puro, naranja puro, etc.).

(S).-representa a la saturación, la cual está relacionada con la pureza relativa o cantidad de luz blanca relacionada con un tono en particular. Proporciona una medida de grado en que un color puro está diluido en luz blanca.

(I).- representa al brillo, el cual se usa para describir qué tan claro u oscuro es un color, y se refiere a la cantidad de luz percibida. Representa lo claro u oscuro que es un color respecto de su color patrón.

Sea F un mapeo desde el espacio imagen hacia un espacio n -dimensional, $X = \{x_1, x_2, x_3, \dots, x_n\}$ como se observa en la ecuación (1.3):

$$F : f \rightarrow X \quad (1.3)$$

Donde n es el número de características que se utilizan para representar a una imagen. La diferencia vectorial entre 2 imágenes f_1 , y f_2 se puede expresar como una distancia d ,

1. INTRODUCCIÓN

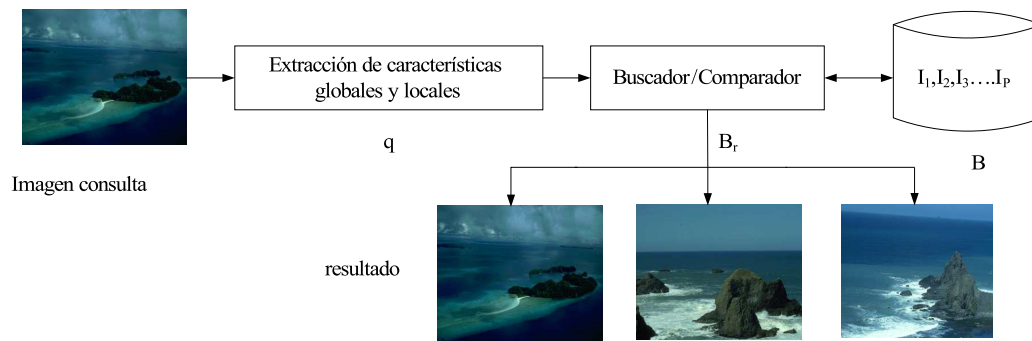


Figure 1.3: Diagrama de bloques del modelo para recuperar imágenes.

entre los respectivos vectores de características x_1 y x_2 .

Dadas las ecuaciones anteriores, el problema de la recuperación de imágenes con el criterio de la distancia mínima se puede proponer de la siguiente manera:

Dada una imagen consulta q , para recuperar una imagen f_r desde una base de datos de imágenes B , es necesario que se cumpla la ecuación (1.4) con respecto a la distancia mínima entre la imagen consulta q y la imagen recuperada f_r .

$$d(F(q), F(f_r)) \leq d(F(q), F(f)) \quad (1.4)$$

para todo $f \in B$, $f \neq f_r$

O expresado en otros términos:

Dadas p imágenes, $I_1, I_2, I_3, \dots, I_p$ con $p \gg 0$ pertenecientes a un conjunto heterogéneo B , extraer de B un subconjunto limitado de imágenes B_r , dada una consulta q formulada en términos de un grupo de rasgos globales y locales. Ver figura 1.3.

Como se puede apreciar en la figura se presenta una imagen consulta de una escena costera, la cual se pretende recuperar. Como resultado se obtienen las tres imágenes más parecidas con respecto a la imagen consulta, con ésta en primer lugar.

1.2. Consideraciones generales de la metodología

Para el desarrollo de esta implementación, se requiere de un conocimiento previo, como pueden ser:

- Los rasgos que se deben considerar para describir el contenido de una imagen.
- Tomar una parte de una imagen en base a estos rasgos, para que mediante éstos se pueda descomponer en sus partes.
- Diseñar y estructurar un diccionario indexado para organizar las imágenes descritas.
- Demilitar los criterios que deben tener las imágenes consulta para extraer las imágenes del diccionario indexado.
- Acotar y demilitar el tipo de imágenes usadas para realizar el entrenamiento.

Cada una de estas consideraciones se describen co más detalle en los siguientes capítulos de este trabajo.

1.3. Justificación

Aproximadamente el 73% de la información circulante en Internet está en forma de imágenes [4]. Esta información, en general, no se encuentra bien organizada ni bien estructurada. En Internet podemos encontrar imágenes de todo tipo: gentes, flores, animales, automóviles, paisajes, etc. Por este motivo, día a día, aumentan las colecciones de imágenes digitales. Esta información hay que organizarla, ordenarla y clasificarla de una manera automática.

Si se habla de una metodología capaz de diferenciar entre 10,000 clases de objetos diferentes, por decir algo, entonces hablamos de un problema de investigación aún abierto.

1. INTRODUCCIÓN

Los sistemas de recuperación de imágenes se han venido desarrollando de manera amplia como un campo activo de investigación; se han implementado sistemas de recuperación por contenido mediante varias técnicas y enfoques.

1.4. Identificación del problema

El concepto *indexar* se refiere a la acción de ordenar un conjunto de datos para elaborar un índice [1]. La manera clásica de indexar imágenes está basado en un proceso de “etiquetado”, el cual consiste en realizar anotaciones manuales que describen el contenido de cada imagen. Esta es una tarea tediosa, imprecisa, costosa, subjetiva y en muchos casos no está completamente disponible. Para recuperar imágenes sobre la base de su contenido, es necesario aplicar técnicas de procesamiento de imágenes y extraer aquellas características que permitan identificar la información contenida en cada imagen de acuerdo al contexto de cada aplicación [7]. El recuperar imágenes de escenas naturales desde una base de datos indexada requiere de la aplicación de técnicas computacionales para organizar e indexar los registros automáticamente de acuerdo a su significado. Dado un conjunto grande de imágenes, se desea implementar una metodología para recuperar imágenes que permita a los usuarios recuperar solo aquellas imágenes similares o iguales a partir de una imagen consulta el sistema debe mostrar el subconjunto de imágenes provenientes de la base de datos indexada más parecidas sobre la base de los rasgos globales y locales de las mismas.

1.4.1. Objetivo general

Diseñar y poner en operación una metodología para indexar imágenes digitales mediante descriptores globales y locales para recuperar imágenes similares desde una base de datos de gran tamaño.

1.4.2. Objetivos específicos

1. Identificar y extraer características visuales de una imagen digital que proporcione suficiente información para diferenciarla de otras imágenes similares.
2. Organizar las características de una imagen de tal forma que se pueda procesar el contenido visual de la misma.
3. Presentar una imagen consulta al sistema que recupere el subconjunto de imágenes similares de acuerdo al contenido de la imagen consulta.

1.5. Aportaciones

La metodología propuesta hace uso de puntos aleatorios para hacer la extracción de características y así poder describir las partes a los objetos presentes en una imagen. Las características o rasgos descriptores son: media, la desviación estándar [19] y la homogeneidad, proveniente esta última de la matriz de coocurrencia [42] y [39]. Estos rasgos son extraídos de una subimagen a partir de los canales H, S e I [21] , [22] y [23]. Se propone el uso del algoritmo *K-medias* [16], [19], [17] el clasificador K-NN (el vecino más cercano, donde $K=1$), [30], [11], [19] y [17], el clasificador Bayesiano [19], y una red neuronal cuyo tipo de aprendizaje es la retropropagación [54]. Todos estos elementos se usan para construir una base de datos indexada de imágenes [49], [48] y [50] logrando las siguientes aportaciones:

1. Se realiza la recuperación de imágenes sin necesidad de etiquetar el contenido de las escenas consulta.
2. Se realiza la recuperación de imágenes en forma automática desde una base de datos.
3. Se construye una base de datos indexada de forma automática usando toda la base de imágenes del entrenamiento.

1. INTRODUCCIÓN

4. Al realizar una consulta, además de hacer la recuperación de imágenes, se puede determinar de forma automática el tipo de escena que conforma a la imagen consulta.

1.6. Organización de la tesis

Este documento está organizado de la siguiente manera:

- Capítulo 1.-Introducción.
- Capítulo 2.- Estado del arte.
- Capítulo 3.-Marco teórico de referencia de las herramientas que se utilizaron para resolver el problema planteado.
- Capítulo 4.-Metodología propuesta.
- Capítulo 5.-Resultados experimentales y su discusión.
- Capítulo 6.-Conclusiones y trabajo futuro.
- Apéndices.
- Referencias.

Capítulo 2

ESTADO DEL ARTE

2.1. Antecedentes

Aproximadamente el 73% de información en la Internet se encuentra en forma de imágenes [4]. Esta información, en general, no se encuentra bien organizada ni estructurada. En la Internet podemos encontrar imágenes con diverso contenido: gente, flores, animales, automóviles, paisajes, etc. razón por la cual día a día aumentan las colecciones de imágenes digitales. Esta información hay que organizarla, ordenarla y clasificarla de una manera automática.

Si se habla de una metodología capaz de diferenciar entre 10,000 clases de objetos diferentes, entonces se habla de un problema de investigación todavía abierto [28].

Los sistemas de recuperación de imágenes se han venido desarrollando de manera amplia como un campo activo de investigación y se han implementado sistemas de recuperación por contenido utilizando varias técnicas y enfoques [57].

La selección y extracción de características es una decisión importante en el proceso de investigación que exige un mejor entendimiento de las imágenes para desarrollar una metodología enfocada a la organización y búsqueda de un cierto tipo de imágenes con buenos puntajes de precisión [7]. Por eso, el desarrollo de esta metodología debe estar

2. ESTADO DEL ARTE

orientado por el conocimiento específico en el área de aplicación.

El indexado de las imágenes para manejar grandes volúmenes de información es otra de las consideraciones técnicas que se deben tener en cuenta para integrar los módulos de extracción de características, almacenamiento físico de las imágenes, cálculos de similitud, procedimientos de consulta, interfaz del usuario y arquitectura del sistema [57].

Hay algunos problemas que continúan sin resolver y que hacen más interesante el trabajo de investigación, como la definición de una medida de similitud entre imágenes para calcular la equivalencia aproximada de contenidos entre ellas. Estas medidas de similitud se aplican sobre las características de las imágenes que –dependiendo de la aplicación– pueden estar en términos estadísticos, matriciales, histogramas, vectores, etc.

Otras aplicaciones [43] realizan consultas a partir de regiones identificadas por una serie de puntos que aproximan zonas de interés particular en imágenes con diversos tipos de contenidos (ver figura 2.1) o en algunos documentos de aplicaciones médicas.



Figura 2.1: Ilustración con diversos tipos de imagen: a) laguna, b) mapa, c) persona, d) automóvil, e) avión, f) teléfono celular, g) puesta de sol, h) automóvil Ford T, i) automóvil Jeep .

- QBIC [57] y[43] .- Hace consultas por el contenido de imagen. Se trata del primer

sistema comercial basado en un sistema CBIR. Soporta hacer consultas mediante imágenes de ejemplo, dibujos, patrones de textura. Sus características de textura es una versión mejorada de la representación de textura de Tamura [24].

- VIRAGE [43].- Es un sistema similar al QBIC basado en el contenido de la imagen. Hace consultas visuales basadas en el color, composición del color y textura.
- WebSEEK [43] y [31].- Es un buscador de www orientado a la búsqueda de texto e imágenes. Sus características visuales son el color y la transformada wavelet basadas en las características de textura.
- MARS [43] y [31].-Es un sistema que difiere de los sistemas anteriores puesto que fue diseñado para la investigación, involucra a la comunidad científica de visión por computadora, involucra el manejo de bases de datos y la recuperación de información. Para describir la forma de las imágenes hace uso de los descriptores de Fourier, para describir la textura hace uso de la transformada de Fourier y los wavelets mientras que para la descripción del color hace uso de los histogramas.
- IMAGE ROVER [31].-Permite el uso de varias imágenes en la consulta. Para describir la textura hace uso de histogramas para describir los contornos y el tipo de consulta es textual basado en imágenes.
- DIOGENES [31].-Su principal característica es que solo fue diseñado para la identificación del rostro de celebridades. Tiene rastreadores para enlazarse a Google y Altavista. Su tipo de consulta es textual.
- ATLAS WISE [31].-Hace análisis de la textura; hace uso de los histogramas en los contornos; para el análisis del color hace uso de histogramas.
- Gonzalez-Garcia A.C. et al en [7].-Propone en su trabajo hacer la recuperación de imágenes, con la ayuda de imágenes. Mediante la transformada Wavelet Daubechies

2. ESTADO DEL ARTE

4 que son 4 coeficientes que tienden a representar la semántica de la imagen, es decir, la variación local del color de los objetos y el fondo. Se extraen las tres bandas (RGB) (ver figura 7.2 en la página 146 en la parte del apéndice) del color de una imagen porque es el más comúnmente usado. Usan histogramas para la extracción del color. Debido a que el histograma no aporta suficiente información acerca de la posición de los píxeles, hacen uso de la multiresolución. Para hacer la clasificación de las características hacen uso de un perceptrón multicapa. La recuperación de las imágenes se basa en el color.

- C. Schmid en [44].-Su trabajo está enfocado a la recuperación de imágenes. Hace cálculos de descriptores genéricos invariantes a rotaciones y aplicados a cada píxel. Sus imágenes están en niveles de gris. Ordena sus descriptores haciendo clusterización (agrupaciones) mediante el algoritmo de **K**-medias. Hace uso de la distancia euclideana para hacer la comparación entre los descriptores. Usa una gaussiana para calcular la probabilidad de un descriptor.
- **Nota :** definiendo a un descriptor.- Un descriptor consiste en realizar un cálculo en cierta área visual de una imagen para extraer propiedades o atributos tales como forma, color, textura, etc., para su uso en las aplicaciones [23].
- Julia Vogel en [48] y [50].-Su trabajo está orientado a la recuperación de imágenes, pero recupera las imágenes con base en las anotaciones o etiquetas que éstas tienen asociadas. Usa una máquina de soporte vectorial para entrenar un clasificador de características donde obtiene un 71.7% de entrenamiento, y un clasificador tipo K-NN. A cada región de la imagen se le extraen histogramas de HSI. [Ver figura (7.3 en la página 146) en la parte del apéndice].
- J.Li et al en [35].- Describe en su trabajo que el aumento de la información representada en imágenes digitales ha complicado el manejo y la administración

de las mismas, por lo que se ha intentado administrarlas mediante un etiquetado automático en tiempo real. Este trabajo describe la implementación de un sistema de etiquetado, pero con una mayor cantidad de restricciones al momento de asignar un nombre, utilizando etiquetas de la semántica del tema de la imagen.

- En el año 2000 se publicó un artículo donde hacen mención que el etiquetado de imágenes tiene inconsistencias (ver figura 2.2) al momento de realizar una búsqueda de imágenes [9] y propone una recuperación/búsqueda de imágenes mediante el contenido de la misma.
- El modelo de campo aleatorio en [53] fue propuesto para la integración de la información. El modelo trata de identificar los rasgos de una imagen de manera global y local, dándole una etiqueta a cada uno de los rasgos en particular. Esto ayuda a clasificar la imagen prediciendo su escena. La incorporación de información global ayuda a resolver ambigüedades. La información local hace referencia a las características de la imagen que son extraídas y usadas por segmento, mientras que la información global describe a la imagen en su totalidad y se utiliza como predicción de la misma [52]
- Como resultado de una investigación encaminada al diseño y evaluación de búsquedas por contenido, surge un prototipo de un sistema para la recuperación de imágenes de histología [10]. Este utiliza una semántica que convierte características de bajo nivel extraídas de la imagen, en su concepto asociado de la histología (estudio de las lesiones celulares, órganos y tejidos en el organismo humano).
- Con respecto a la clasificación de imágenes [8] se tienen algunas técnicas de aproximación tales como el modelado de bajo nivel y el modelado semántico. El modelado de bajo nivel utiliza los rasgos de bajo nivel (color, textura) además de la información proveniente del histograma para determinar directamente el tipo de

2. ESTADO DEL ARTE

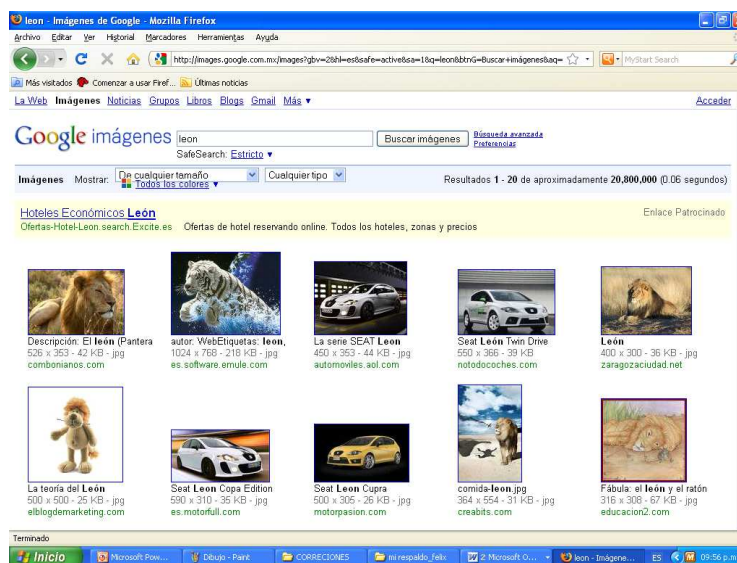


Figura 2.2: Resultados inconsistentes en el buscador de Google cuando se quiere buscar la imagen de un león y se escribe en el buscador la etiqueta “león”.

escenario a consultar. Sin embargo, esto resulta poco efectivo ya que aunque hace una distinción correcta de dos tipos de imagen (ciudad, paisaje, interiores, exteriores, etc.), no hace una clasificación más específica (selva, bosque, pradera, etc.). La problemática de la clasificación es resuelta mediante la teoría de decisión de Bayes [8]. Cada imagen es representada por un vector característico extraído de la misma imagen. Los modelos probabilísticos requeridos para la aproximación del Bayesiano son calculados durante la etapa del entrenamiento.

Las posibilidades de producción de imágenes crece más rápido que las metodologías para administrar y procesar esa información visual, imponiéndose un nuevo reto para su eficiente recuperación, utilizando técnicas apropiadas para el almacenamiento y recuperación.

Las bases de datos de imágenes han sido estudiadas desde hace varios años. Las primeras aproximaciones para indexar grandes volúmenes de imágenes se realizaron utilizando palabras clave, pero la construcción del índice se convertiría en una tarea costosa y subjetiva. El Photobook [57] y [43], el QBIC [57] y [43] entre otros han sido algunos de los primeros trabajos para bases de datos de imágenes en donde los autores se preocuparon

por las propiedades visuales de la imagen y sus características de forma, área y textura, implementando un sistema de recuperación de imágenes que utiliza operadores visuales.

La exploración de grandes cantidades de imágenes es una tarea donde los sistemas de información pueden contribuir a organizar y clasificar los registros automáticamente. Una base de datos de imágenes debe permitir al usuario recuperar una imagen del archivo a partir de sus propiedades visuales, como la forma o composición estructural. Los usuarios necesitan encontrar imágenes con ciertas características, sin tener que explorar demasiados registros, obteniendo aquellas que tengan un alto grado de importancia de acuerdo a los criterios definidos.

Varios trabajos se han realizado para representar el significado de la imagen a partir de sus características visuales [57] y [43] obteniendo resultados apropiados principalmente para especialidades artísticas o gráficas. Por otro lado los trabajos de visión artificial orientados hacia otras aplicaciones de tiempo real y control, tienen requerimientos más específicos para su desarrollo, como la fuerte identificación de objetos, el seguimiento de los mismos en secuencias de imágenes y el reconocimiento de patrones. Muchos de estos problemas continúan aún sin resolver, principalmente por la dependencia que existe del completo entendimiento de las imágenes en donde los investigadores han identificado dos vacíos fundamentales: sensorial y semántico (Ver apéndices).

La creciente necesidad de implementar sistemas que permitan acceder a imágenes a partir de su contenido visual, ha llevado a impulsar uno de los campos de investigación más activos de hoy en día: Recuperación de Imágenes sobre la Base de su Contenido (CBIR). **CBIR** viene de las siglas en inglés. Content Based Image Retrieval [36],[40],[55] y [2]. La investigación en esta área comparte conceptos y resultados provenientes de trabajos de visión artificial, reconocimiento de rostros, biometría, exploración de extensos álbumes de fotografías, recuperación y clasificación de secuencias de video, entre otras.

2.2. Definición general del problema

El objetivo de los sistemas de recuperación de imágenes sobre la base de su contenido (Content Based Image Retrieval (CBIR)) [40] consiste en administrar grandes cantidades de datos multimediales en aplicaciones concretas. En muchos campos de trabajo de hoy en día, se tiene una creciente tasa de almacenamiento de imágenes, haciéndose necesario ordenar, organizar, clasificar y automatizar esta información para facilitar el acceso y aprovechar la colección de imágenes en la toma de decisiones diaria. Clásicamente, las imágenes eran indexadas utilizando palabras clave, pero cuando se trata de un elevado número de imágenes, las anotaciones se convierten en un esfuerzo costoso e impreciso y la subjetividad se hace sentir por la imposibilidad de expresar algunos conceptos visuales en términos del lenguaje natural, terminando todo esto en una recuperación no muy satisfactoria para el usuario.

2.3. Eficiencia y carga computacional

Los sistemas CBIR deben trabajar eficientemente con una gran cantidad de imágenes. Para el caso de la metodología de evaluación propuesta en [44] se habla sobre la necesidad de contar con sistemas interactivos, definidos como sistemas que sean capaces de responder a una consulta en un tiempo menor a un segundo. Aunque parezca obvio, muchos trabajos han sido desarrollados utilizando técnicas cuyo tiempo promedio de ejecución es superior a este requerimiento. Sin embargo, estos resultados se deben principalmente a la complejidad en el manejo de las estructuras de representación, haciendo necesario balancear el compromiso entre precisión y rendimiento [44]. No solamente es importante la evaluación de un sistema en términos del tiempo de ejecución, sino también con respecto a la precisión (ver ecuación 5.1 en la página 107) para recuperar registros correctamente clasificados. En [49] y [44] se propone un esquema de evaluación

para los sistemas de recuperación de imágenes. También señala la forma en la que los resultados de la experimentación deben ser presentados para facilitar la comparación de técnicas y la evaluación de resultados.

2.4. Tipos de consulta

Existen varios paradigmas de consulta en los sistemas CBIR:

1. **Consultas por palabras clave.** Las palabras clave sirven para recuperar imágenes que tengan asociado el concepto gramatical expresado por el usuario. Sin embargo, las anotaciones (etiquetas) no son confiables y en pocos registros se encuentran completamente disponibles. En [48] se realizan anotaciones automáticas a partir de las propiedades visuales de la imagen, pero requiere de un conocimiento específico del área de aplicación.
2. **Consultas por controles visuales.** En este caso, se utilizan controles que permiten al usuario seleccionar propiedades visuales deseadas en las imágenes resultantes. Los controles incluyen selección de color, textura, contrastes, brillo, combinaciones de éstos y otros mas [57] y [46]. Aunque los resultados corresponden a las selecciones del usuario, se deben tener conocimientos artísticos para combinar adecuadamente los criterios sin tener resultados frustantes, haciendo de estas interfases un sistema complejo para usuarios en otras áreas.
3. **Consultas mediante imágenes ejemplo.** Los sistemas basados en este tipo de consulta permiten seleccionar una imagen que tenga características deseadas en los ejemplos. El sistema toma la imagen de ejemplo, la analiza y luego busca en la base de datos los patrones más sobresalientes que fueron encontrados. Uno de los problemas que pueden presentarse en las imágenes ejemplo es que pueden contener

2. ESTADO DEL ARTE

detalles irrelevantes para la consulta, cuando el usuario desea concentrarse solamente en alguna de sus propiedades.

4. **Consultas por trazos.** Este tipo de consulta permite al usuario realizar trazos de la formas que considera más importantes en los resultados. Puede combinar los trazos con operadores visuales como color o textura. Las formas pueden construirse con ayuda del sistema. Los problemas de este tipo de consulta llegan cuando se requiere de habilidades artísticas para expresar la forma aproximada que se desea y puede fallar cuando se requieren formas a cierto nivel de detalle en los resultados.

Los problemas de consulta aparecen por la imposibilidad de los seres humanos de expresar algunas de sus propiedades o significados de las imágenes, que no pueden representarse en lenguaje natural. Las investigaciones en sistemas de interacción hombre-máquina pueden aportar conceptos que faciliten a los usuarios expresar sus necesidades de consulta, de una manera simple e intuitiva para incrementar la satisfacción de utilización. Las dificultades y problemas inherentes al tratamiento de imágenes hacen mas interesantes las investigaciones en el área de la Recuperación por Contenido. Las contribuciones se realizan activamente atacando uno o varios problemas a la vez.

2.5. Técnicas de recuperación

Dado un patrón de búsqueda de acuerdo a los paradigmas de consulta, las técnicas de recuperación deben seleccionar de la base de datos aquellas imágenes cuyas representaciones emparejen satisfactoriamente con el patrón de consulta seleccionado. La mayoría de las técnicas de recuperación involucran medidas de similitud o de distancia definidas en el dominio de la representación de las imágenes. Estas medidas tratan de identificar, con respecto a las características que conforman la representación, qué tan parecida es una imagen a otra.

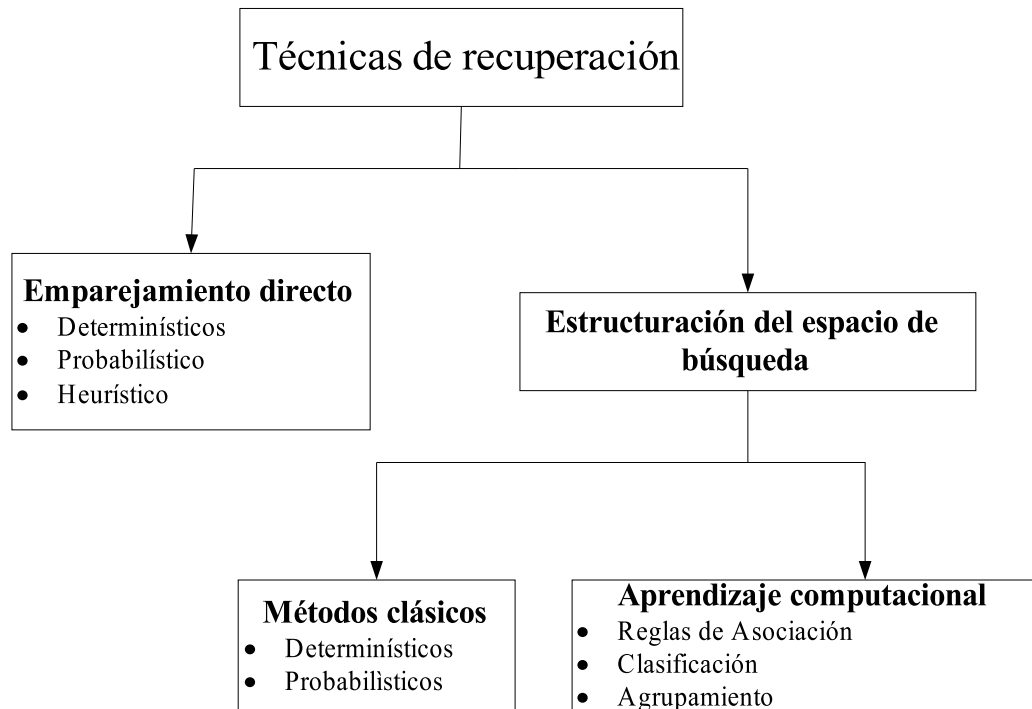


Figura 2.3: Taxonomía de las técnicas de recuperación.

A continuación se define que es la recuperación de imágenes \implies Ciencia de la búsqueda de información (imágenes). Búsqueda de un metadato que describa imágenes. O también la búsqueda en bases de datos de imágenes ya sea en Internet o intranet de forma pertinente o relevante.

Para encontrar imágenes relevantes en una base de datos, el usuario debe estar interesado en dos tipos de resultados: ubicar una imagen objetivo o navegar por categorías de imágenes similares. Las técnicas de recuperación (ver figura 2.3) pueden responder a los tipos de requerimientos. De acuerdo al tipo y propósito de las técnicas podemos clasificar los trabajos realizados de la siguiente manera:

2.5.1. Emparejamiento directo

Este tipo de técnica se enfoca en recuperar imágenes objetivo según los criterios de búsqueda del usuario. Aunque las características de una imagen en la base de datos

2. ESTADO DEL ARTE

hagan que ésta sea la única, las técnicas de emparejamiento deben de encontrar registros que tengan un alto grado de similitud con respecto al patrón de búsqueda proporcionado [3]. El emparejamiento puede ser visto como un proceso de optimización en donde debe minimizarse la distancia entre el patrón de búsqueda y los resultados presentados.

Generalmente se utiliza un índice para orientar la búsqueda, que contiene la representación de cada imagen para ser evaluada. Este tipo de procedimiento se ha inspirado en los trabajos de reconocimiento de patrones en el área de la visión artificial. Después de proporcionar los patrones de búsqueda, el sistema debe aplicar la técnica interactivamente para representar los resultados. Para esto se utiliza un número de imágenes que cumplen con los criterios dados. El número de imágenes puede controlarse mediante la cantidad de resultados o mediante la definición de un umbral de emparejamiento.

Las técnicas de emparejamiento se dividen en tres grupos:

1. Emparejamientos determinísticos.
2. Emparejamientos probabilísticos.
3. Emparejamientos heurísticos.

a) Emparejamientos Determinísticos.- Utilizan el índice de la base de datos y una función de comparación para determinar la similitud. Los principales métodos utilizados en este tipo de aplicaciones son:

- **K-NN:** A partir de un patrón de búsqueda, se localizan los K vecinos más cercanos en el conjunto de los datos. Es utilizado para clasificar las imágenes a manera de función de aproximación según la distribución de los datos. También es utilizada cuando no se tiene un conocimiento explícito y manejable de la distribución de los datos, sino que se prefiere realizar una comparación de los registros para obtener aquellos

mas similares. Algunos ejemplos de aplicaciones que utilizan esta técnica son [19], [11], [30], etc.

- **Entropía:** Es utilizada como medida de similitud para dirigir el emparejamiento. Ubica las regiones más interesantes de una imagen aplicando técnicas para medir la cantidad de información en estas regiones. A partir de ellas se comparan las características de las imágenes, evitando la introducción de regiones que no contienen datos relevantes. La entropía es medida a nivel global o a nivel local. Algunos trabajos que utilizan la entropía para emparejar las imágenes se pueden estudiar en [15].

b) **Emparejamientos probabilísticos.-** Los métodos probabilísticos de emparejamiento miden la similitud de dos imágenes de acuerdo a funciones de probabilidad de cada componente en la representación. Para optimizar la comparación de las imágenes, se llevan a cabo procesos aleatorios que determinan si una imagen puede llegar a ser relevante o no, tras la selección de ciertas características. Estas medidas de similitud suelen ser más rápidas en sus tiempos promedios de ejecución [3].

- **Bayes:** Diferentes Modelos Bayesianos han sido planteados para el emparejamiento de imágenes, principalmente en trabajos de visión artificial.
- **Markov:** Los Modelos de Markov no solo han sido utilizados para emparejamiento de imágenes, también han sido propuestos como elementos de clasificación.

c) **Emparejamiento heurístico:** Dirige la búsqueda de acuerdo al conocimiento previo en el dominio específico de las imágenes almacenadas. Definir una buena heurística es un aspecto importante para obtener tiempos de respuesta adecuados y soluciones óptimas. Las heurísticas permiten estimar una

2. ESTADO DEL ARTE

solución adecuada según los criterios dados. También sirven para medir la similitud entre 2 grafos mediante alguna métrica de distancia y un algoritmo para localizar las operaciones de menor costo que pueden obtener un grafo a partir del otro [3].

2.5.2. Estructuración del espacio de búsqueda

En las aplicaciones CBIR se considera el espacio de búsqueda como la totalidad de los registros en la base de datos de las imágenes. Las técnicas de emparejamiento por sí solas pueden llegar a ser ineficientes cuando el espacio de búsqueda se presenta completo sin ninguna guía adicional, mientras que una organización adecuada de los registros puede contribuir a la reducción de la complejidad. La estructuración del espacio de búsqueda no solamente es útil para facilitar el trabajo de las técnicas de emparejamiento, sino que también son utilizadas para proporcionar facilidades de los sistemas CBIR que permiten al usuario localizar imágenes por grupos y categorías. La idea de esta tipo de técnicas consiste en construir índices multinivel que permitan asociar las imágenes a una determinada categoría, en donde las características del resto del grupo siguen un patrón de similitud. Este concepto permite imponer una relación de orden en la base de datos de imágenes agilizando los procesos de recuperación. Las categorías, clases y grupos emergen naturalmente en los grandes bancos de datos, convirtiéndose en información que vale la pena aprovechar para dirigir con mayor eficiencia y precisión las tareas de recuperación. Los procedimientos de estructuración por lo general son aplicados en etapas de preprocesamiento o preparación de la base de datos. No es utilizado como técnica de recuperación porque la complejidad es alta y el resultado se convierte en grupos de imágenes a través de toda la base de datos, proceso que puede llevarse a cabo sin la presencia de criterios de búsqueda. La estructuración del espacio de búsqueda está dividido en (ver figura 2.3):

1. Métodos clásicos.

■ **Métodos clásicos:** Los métodos clásicos de estructuración comprenden aquellos métodos que construyen un mapa para guiar los procesos de búsqueda. Estas técnicas han sido inspiradas en estructuras de indexado para bases de datos relacionales, espaciales o geográficas. También existen tendencias de indexado provenientes desde técnicas de recuperación de textos tanto determinísticos como no determinísticos. Otros métodos organizan los registros por la probabilidad de que un usuario requiera justamente esos registros.

a) **Determinísticos.**- Existen varias técnicas en esta categoría, con diferentes propuestas que explotan algún tipo de información particular en las estructuras de representación. Algunas de estas informaciones son: ecuaciones diferenciales para particionar el espacio de búsqueda, grafos que expresan las relaciones entre los grupos de imágenes, valores propios de las matrices de adyacencia de los grafos [45].

b) **Probabilísticos.**- Los Métodos probabilísticos utilizan información estadística del conjunto de datos para dirigir la búsqueda. La información estadística se extrae de las características de cada imagen, obteniendo datos que pueden facilitar la localización de las imágenes más relevantes. Varios modelos de han propuesto, incluso para trabajos de recuperación de imágenes de la web. En [27] se utiliza un modelo Bayesiano para construir una tabla Hash Geométrica basada principalmente en las propiedades visuales de la forma.

2. Aprendizaje computacional.

■ **Aprendizaje computacional:** Estas técnicas [3] tratan de encontrar patrones

2. ESTADO DEL ARTE

ocultos o distribuciones frecuentes en el conjunto de los datos, y a partir de ellos se construyen los índices que guiarán las consultas. La información encontrada se representa como un resumen del conocimiento subyacente al conjunto de datos, que en algunos casos puede ser explícito (como las reglas de asociación) o puede permanecer codificado en las estructuras de aprendizaje (como en las redes neuronales). Cada que se presenta un patrón de búsqueda, existe un algoritmo que puede determinar la categoría en la que debe efectuarse una búsqueda detallada, ahorrando extensas operaciones de emparejamiento sobre muchos registros irrelevantes.

- a) **Reglas de asociación.**-Éstas pueden ser extraídas mediante los operadores visuales junto con algunos conceptos del dominio de las imágenes. Suelen extraerse del conjunto de datos sobre las características visuales de los registros. En [49] se aplican diversas técnicas de minería de datos para encontrar una relación entre los descriptores de bajo nivel de una imagen y su significado semántico.
- b) **Clasificación.**- Los trabajos de clasificación reciben como entrada un conjunto de datos correctamente clasificados y un conjunto de datos de entrenamiento. El sistema debe encontrar las características más importantes y utilizar este resumen como base para clasificar nuevos registros [11].
- c) **Agrupamiento.**- Las técnicas de agrupamiento no reciben como entrada las categorías sino que a partir de la distribución de los datos, los grupos se identifican . Existen técnicas adaptables, las cuales conforme se incrementan los registros de imágenes, se reorganizan los grupos. Estos grupos son utilizados en sistemas de exploración y navegación [19].

- d) **Otros.-** En [48] se presenta un modelo de aprendizaje computacional, para construir un índice de orientación en la base de datos de imágenes. En [44] se aplica un algoritmo de cuantificación vectorial para agrupar imágenes con puntos de interés.

2.6. Sistemas en línea

A continuación hablaremos de sistemas en línea diseñados para la recuperación de imágenes.

Los sistemas diseñados para la recuperación de imágenes deben proporcionar funcionalidades suficientes para recuperar una imagen coherente con los criterios de búsqueda del usuario. También deben proporcionar una arquitectura que permita extender sus características y funcionalidades en otras direcciones. Uno de los principales retos de los sistemas CBIR es la representación de las imágenes y la recuperación óptima de resultados relevantes para el usuario.

2.6.1. CIRES

CIRES [26] es un sistema en línea de recuperación de imágenes basado en su contenido que combina los principios de las características de alto nivel y las de bajo nivel. En el análisis de alto nivel utiliza organización perceptiva, y principios de agrupamiento para extraer información semántica que describa la estructura del contenido de una imagen. En el análisis de bajo nivel describe la textura de la imagen y utiliza histogramas de color para mapear todos los colores en una paleta de colores fija. El sistema está disponible para realizar consultas de imágenes que contienen objetos naturales como vegetación, árboles, cielos, etc, además de objetos hechos por el hombre tales como construcciones, torres, puentes, etc. La base de datos que utiliza 6 tipos de imágenes diferentes tal y como se muestra en la tabla 2.1.

2. ESTADO DEL ARTE

| Clase | número de imágenes |
|---------------------|---------------------------|
| hecho por el hombre | 1980 |
| aves | 811 |
| insectos | 1134 |
| mamíferos | 2496 |
| flores | 1161 |
| paisajes | 2711 |
| Total | 10,221 |

Tabla 2.1: Clases de imágenes del CIRES

El CIRES en general, el análisis de color y textura no siempre alcanza el nivel adecuado de ejecución en las consultas y satisfacción del usuario, particularmente en imágenes que contienen objetos hechos por el hombre. Logra un porcentaje de precisión del 77.4% en la recuperación.

2.6.2. FIRE

FIRE [47] es un sistema en línea de recuperación basado en su contenido que contiene 7 características visuales disponibles para representar a una imagen y diversas medidas para expresar la distancia. En este trabajo se tiene la posibilidad de elegir una imagen aleatoria de su base de datos, o bien de cargar una imagen desde cualquier ubicación de la computadora. Una vez que se ha seleccionado la imagen deseada, el sistema busca las que son similares, dando la opción al usuario de marcar las imágenes como relevantes, irrelevantes o indiferentes. Una de las bases de datos de imágenes que se utilizó fue la de fotografías históricas de San Andrés (España). En general FIRE obtiene mejores resultados en las consultas automáticas cuando se usan características visuales; se obtiene una eficiencia de 39.4% al utilizar solamente características visuales y 58.7 al combinar

información en texto.

En la tabla 2.2 se presenta una lista de características visuales y sus medidas de distancia asociadas.

2.6.3. IRMA

IRMA [33] es un sistema de recuperación de imágenes de radiografías. Su objetivo es recuperar imágenes de las etapas de terapia del mismo paciente o también recuperar imágenes con diagnóstico similar en bases de datos de imágenes muy grandes; métodos de reconocimiento de patrones y análisis estructurado son utilizados para describir el contenido de una imagen en una firma característica. Usa una base de datos de 1617 imágenes de radiografías donde se presentan 110 imágenes de abdomen, 706 imágenes de manos, 103 imágenes de senos, 110 imágenes de cráneos, 410 imágenes de torsos, y 178 imágenes de columnas vertebrales. En general logra una eficiencia de precisión de 87.5 %.

En la tabla (2.3) se presenta la eficiencia de cuatro medidas de distancia usando dos métodos de clasificación. Se observa que el mejor resultado se obtiene con la distancia tagente distorsionada y la densidad de núcleo.

2.6.4. Análisis de los resultados

A continuación en la tabla 2.4 se presenta un estudio comparativo de los sistemas que hacen recuperación de imágenes mediante imágenes con el propuesto en el presente trabajo. Se observa la diversidad de metodologías utilizadas que tratan de reconocer la forma, el color, y la textura. Obviamente, estos no son los únicos trabajos desarrollados, pero si son los que más se relacionan con la metodología propuesta en esta tesis.

2. ESTADO DEL ARTE

| Número | Característica visual | Medida de distancia |
|--------|----------------------------------|---------------------|
| 1 | Imagen escalada a 32 x 32 | Euclidea |
| 2 | Imagen escalada a 32 x X | IDM |
| 3 | Descriptor global de textura | Euclidea |
| 4 | Histograma de textura Tamura | Divergencia Jeffrey |
| 5 | Histograma con núcleo monomial | Divergencia Jeffrey |
| 6 | Histograma con núcleo relacional | Divergencia Jeffrey |
| 7 | Color o tonos de gris | Selección Binaria |

Tabla 2.2: Características visuales de FIRE.

| Medida de la distancia | 1-NN | Densidad de núcleo |
|----------------------------------|--------|--------------------|
| Distancia Euclidea | 81.9 % | 83.8 % |
| Distancia Tangente | 84.7 % | 85.2 % |
| Modelo Distorsionado | 83.5 % | 85.3 % |
| Distancia Tangente Distorsionada | 85.8 % | 87.5 % |

Tabla 2.3: Eficiencia del sistema IRMA.

| Sistema en línea | Metodología | Número de imágenes | Eficiencia |
|---|--|--------------------|------------|
| Semantic Scene Modeling and Retrieval [48] (2004). | Recuperación de imágenes mediante etiquetas ligadas a las zonas de interés. | 700 | 77.1 % |
| CIRES[26] (2005). | Uso de histogramas de textura para describir el contenido de una imagen. | 10,221 | 77.4 % |
| FIRE[47] (2006). | Usa 7 características visuales para representar una imagen. | 10,000 | 58.7 % |
| IRMA[33] (2006). | Recuperación de la imagen basada en la apariencia. Cada píxel de la imagen es una característica. | 1617 | 87.5 % |
| Recuperación de imágenes en base a su contenido[7] (2007). | Ondeletas de histogramas de una ventana circular. Clasificación mediante redes neuronales. | 1068 | 96 % |
| TESIS.-Recuperación de Imágenes mediante rasgos descriptores globales y locales. | Uso de ventanas aleatorias de 10 x 10 para describir características. Uso del K-means, y varios tipos de clasificadores. | 2512 | 77.07 % |

Tabla 2.4: Comparación con los sistemas en línea.

Capítulo 3

MARCO TEÓRICO

En este capítulo se presenta el marco teórico sobre el que se fundamenta la operación de la metodología propuesta en esta tesis.

La recuperación de imágenes basada en su contenido posee la habilidad de recuperar información visual utilizando como llave de búsqueda una imagen. Se trata de buscar en una base de datos de imágenes aquellas n imágenes más parecidas a la imagen consulta.

El esquema de la generación de la firma utilizando un preprocesamiento de la imagen para obtener un vector de características como una representación numérica simplificada, sirve para almacenar su firma en una base de datos y así acelerar el proceso de la recuperación de las imágenes, ya que el preprocesamiento caracteriza de forma efectiva las propiedades locales de la imagen, tales como el color y la textura.

Las herramientas utilizadas en el diseño de nuestra metodología la conforman una combinación de puntos aleatorios, el uso del algoritmo de K-medias, tres diferentes clasificadores: (K-NN, Bayesiano y una red neuronal artificial) y la distancia euclidiana como criterio de comparación para ver qué tan similar es una imagen de otra cuando se hace una consulta.

Uno de los procesos fundamentales del análisis de una imagen es la extracción de características de las imágenes.. La característica de más bajo nivel es el punto [57].

3. MARCO TEÓRICO

Básicamente un píxel se puede describir por medio de dos coordenadas: $p = p(x,y)$ en 2D en donde $x,y \in \mathbb{Z}$ Los puntos son identificados dentro de una imagen digital en forma de un píxel distinto a sus vecinos. Los puntos son necesarios en tareas como el reconocimiento de objetos, reconstrucción 3D, por mencionar algunas de ellas. En otras palabras, los puntos son necesarios e indispensables, sin los cuales no se podrían desarrollar otras aplicaciones dentro del campo de la visión por computadora.

Veamos a continuación la definición de algunos conceptos.

3.1. Reconocimiento de patrones

El reconocimiento de patrones es útil para la identificación de formas, figuras, objetos, etc. Es un proceso fundamental que se encuentra en casi todas las acciones humanas. Un sistema automático de reconocimiento de objetos (SARO) permite a una máquina (reconocer y posicionar) objetos en el mundo real a partir de una o mas imágenes del mundo, usando modelos de los objetos conocidos a priori [28].

La cadena de pasos en un sistema de reconocimiento de patrones se ilustra en la figura (3.1)

1. **Reconocimiento del objeto.-** Es el ente u objeto a reconocer, que puede ser algún tipo de señal, una base de datos de imágenes, un cultivo, alguna enfermedad, etc.
2. **Preprocesamiento en el dominio del objeto.-** Se hace un tipo de preprocesamiento para eliminar información que no es útil, por ejemplo, ruido ambiental, cancelación de eco; se aplica algún tipo de filtro (pasa altas, pasa bajas, morfológico, etc), con el objetivo de mejorar la calidad de la imagen para futuros tratamientos.
3. **Preprocesamiento en el dominio de los atributos.-** Aplica operadores sobre una imagen permitiendo identificar la presencia de un objeto en una escena. Los

atributos (rasgos o características) utilizados por el sistema dependen del tipo de objetos a ser identificados o reconocidos.

Permite eliminar información redundante y reducir la dimensionalidad de trabajo . Si este módulo funciona bien se producen 2 cosas:

- Una alta clasificación.
- Una reducción de tiempo de cómputo.

4. **Clasificación.-** Sirve para calcular las similitudes entre los objetos que pertenecen a cierta clase.

a) **Objeto o forma.-** Es algo visible y cuantificable que será descrito por un conjunto de medidas. Estas medidas forman un conjunto descriptivo del objeto en R^n [11].

b) **Clase.-** Es el conjunto de objetos que tienen el mismo significado, es decir, que comparten características comunes. La noción de clase es subjetiva y depende del contexto y de la cultura.

5. **Evaluación de la clasificación.-** Mediante una matriz de confusión se evalúa el porcentaje para verificar que tan bueno o malo fue el reconocimiento de los objetos pertenecientes a determinada clase.

Aprendizaje.- Es el proceso de estimación de una relación desconocida (entrada, salida) o estructura de un sistema utilizando un número limitado de muestras.

En este trabajo de tesis las muestras son los vectores de atributos de entrenamiento. Esto equivale a estimar las propiedades de alguna distribución estadística a partir de las muestras del entrenamiento. De este modo, la información contenida en las muestras de entrenamiento, que corresponde a experiencias pasadas puede utilizarse para responder a cuestiones sobre

3. MARCO TEÓRICO

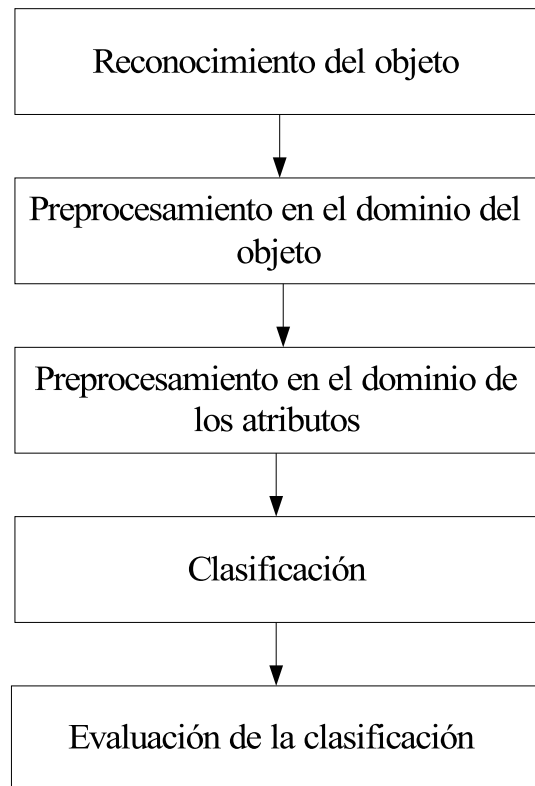


Figura 3.1: Diagrama general de un sistema de reconocimiento de objetos.

datos o muestras futuras. Por lo tanto, podemos distinguir dos estados en la operación de un sistema de aprendizaje:

1. Aprendizaje/estimación a partir de de las muestras del entrenamiento.
2. Operación/predicción, cuando las predicciones se hacen para muestras futuras o de prueba.

La minería de datos consiste en la extracción no trivial de información que reside de manera implícita en los datos. En otras palabras, la minería de datos prepara, sondea y explora los datos para sacar la información oculta en ellos. Para un experto, o para el responsable de un sistema, normalmente no son los datos en sí lo más relevante, sino la información que se encierra en sus relaciones y dependencias. Bajo el nombre de minería de

datos se engloba todo un conjunto de técnicas orientadas a la extracción del conocimiento procesable, implícito en las bases de datos.

Cuando se hace un análisis de los datos los algoritmos utilizados se clasifican en:

- **Supervisados**

- **No supervisados**

Un aprendizaje **supervisado** se utiliza para estimar una relación desconocida (entrada/salida) a partir de muestras conocidas (entrada/salida). El término supervisado corresponde con el hecho de que los valores de salida para las muestras del entrenamiento son conocidos y por tanto son proporcionados por un supervisor. Este tipo de aprendizaje se presenta en los siguientes casos o situaciones.

- Interviene el humano.

- Se conocen las clases de pertenencia.

- Se busca la convergencia de los parámetros.

- Se optimiza la convergencia:
 - a).-mejorando las velocidades de convergencia
 - b).-optimizando las funciones separatrices

- Se trabaja con algoritmos probados y establecidos.

- Cada clasificador tiene sus propios parámetros, por ejemplo el clasificador Bayesiano (hay que calcular la media y la matriz de varianza-covarianza)

- La separabilidad puede ser lineal, cuadrática o cúbica.

Un aprendizaje **no supervisado** consiste en que solamente se proporciona al sistema de aprendizaje las muestras de entrada y no existe noción alguna de la salida durante el

3. MARCO TEÓRICO

aprendizaje. El objetivo del aprendizaje no supervisado es de estimar la distribución de la probabilidad de las entradas o descubrir la estructura natural de los grupos de datos de entrada. En este tipo de aprendizaje se descubren patrones o tendencias entre ellos. En otras palabras, ni se conocen las clases de pertenencia ni su cantidad.

En este tipo de aprendizaje destaca el algoritmo de *K-means o K-medias*.

Las técnicas de la minería de datos provienen de la inteligencia artificial y de la estadística, dichas técnicas son algoritmos sofisticados que se aplican sobre un conjunto de datos para obtener un determinado resultado. Entre estas destacan las técnicas de:

- **Agrupamiento (formación de *clusters*).**- Es un procedimiento de agrupación de una serie de vectores según criterios (habitualmente) de distancia; se trata de disponer los vectores de entrada de tal forma que estén más cercanos aquéllos que tengan características comunes; entre ellos destaca el algoritmo de *K-means o K-medias*.

3.2. Atributos o descriptores

Primeramente damos una definición de los rasgos y para que sirven.

Rasgos.-Una manera de modelar un objeto es a través de una descripción del mismo en términos de un tuplo x de atributos usualmente denominados rasgos o características.

Rasgo o característica .-Es cualquier propiedad física de un objeto que puede ser usada para describir dicho objeto [28].

3.2.1. Textura

Los descriptores de textura se basan siempre en una vecindad, ya que la textura se define para regiones y no para píxeles individuales. Es difícil encontrar un solo descriptor de textura, ya que existen varios problemas asociados a ellos [20]. El detector perfecto debería ser insensible a rotaciones y a escalamientos.

La textura es una característica importante en la identificación de los objetos o regiones de interés en una imagen. Haralick [42] propuso 14 medidas de textura basadas en la dependencia espacial de los tonos de grises. En [29] se sugieren variables de textura basadas en estadísticas de primer orden (media, desviación estándar, varianza), estadísticas de segundo orden basadas en la matriz de coocurrencia entre las más usadas para medir textura. La suposición es que la información textural en una imagen está contenida en la relación espacial que los tonos de grises tienen entre ellos. Esa relaciones están especificadas en la matriz de coocurrencia espacial (o de niveles de gris) que son calculadas en una dirección específica o bien para todas (0° , 45° , 90° , y 135°) entre los píxeles vecinos dentro de una ventana móvil dentro de la imagen.

La clasificación de un píxel puede variar cuando se le analiza aisladamente o cuando se consideran también sus vecinos; además, cuando se utilizan imágenes de muy alta resolución, donde cada píxel hace referencia a una parte de un objeto, el tratamiento basado en un píxel pierde validez [39].

El modelo matemático más común para medir la textura es la matriz de coocurrencia de niveles de grises (**GLCM**) (*Grey Level Cooccurrence Matrix*), basado en estadísticas de segundo orden. Es un histograma de los niveles de gris de dos dimensiones para un par de píxeles (píxel de referencia y vecino). Esta matriz aproxima la probabilidad de la distribución conjunta de un par de píxeles. Diversos estudios han corroborado que los datos texturales conjuntamente con los datos espectrales mejora la precisión de la clasificación [37].

3.2.2. Matriz de coocurrencia

En el análisis de texturas, la extracción de características se realiza a partir de la distribución estadística con la que se observan combinaciones de determinadas intensidades en posiciones relativas de la imagen. La matriz de coocurrencia es una matriz cuadrada

3. MARCO TEÓRICO

en la que el número de filas y columnas coincide con el número de niveles de gris en la imagen a analizar y donde cada elemento de la matriz $C(i, j)$ contiene la frecuencia relativa con la que dos píxeles de la imagen I , con intensidades i y j respectivamente, y separados por una distancia d y un ángulo θ , ocurren en una determinada vecindad. Dicho de otro modo, el elemento $C(i, j|d, \theta)$ contiene la probabilidad de que dos píxeles cualesquiera a una distancia d y un ángulo θ tengan respectivamente niveles de gris i y j .

La matriz de coocurrencia describe la frecuencia de un nivel de gris que aparece en una relación espacial específica con otro nivel de gris, dentro del área de una ventana determinada. La matriz de coocurrencia es un resumen de la forma en que los valores de los píxeles ocurren al lado de otro valor en una pequeña ventana.

3.2.2.1. Cálculo de la matriz de coocurrencia

Tomando como base la matriz de la figura 3.2, con una distancia de un píxel “ $d=1$ ” y dirección a 0^0 se procede ahora a determinar el cálculo de la matriz de coocurrencia.

Como esta matriz de ejemplo (tomado de la referencia [20]) únicamente contiene tres niveles de gris (0,1,y 2) se crea una matriz de 3×3 para cada rotación . El cálculo de la matriz de coocurrencia para cada dirección se muestra en las figuras 3.3, 3.4 y 3.5 respectivamente.

| | x | | | | |
|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 |
| 0 | 0 | 0 | 1 | 1 | 2 |
| 1 | 0 | 0 | 1 | 1 | 2 |
| 2 | 0 | 0 | 1 | 1 | 2 |
| 3 | 0 | 0 | 1 | 1 | 2 |
| 4 | 0 | 0 | 1 | 1 | 2 |

Figura 3.2: Imagen con tres niveles de gris

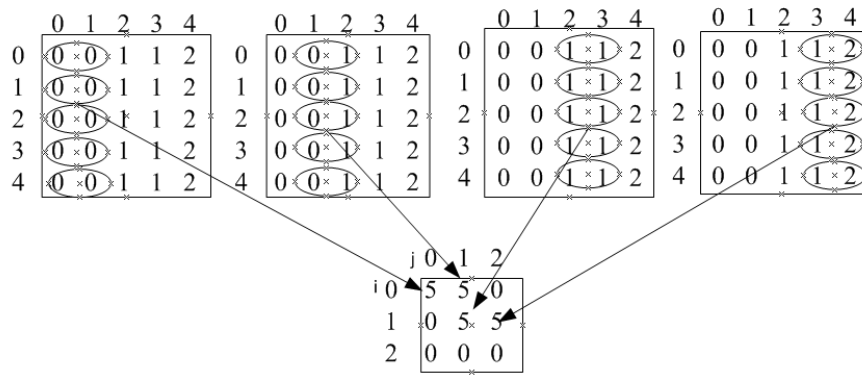


Figura 3.3: Matriz de coocurrencia para $d=1$ a 0°

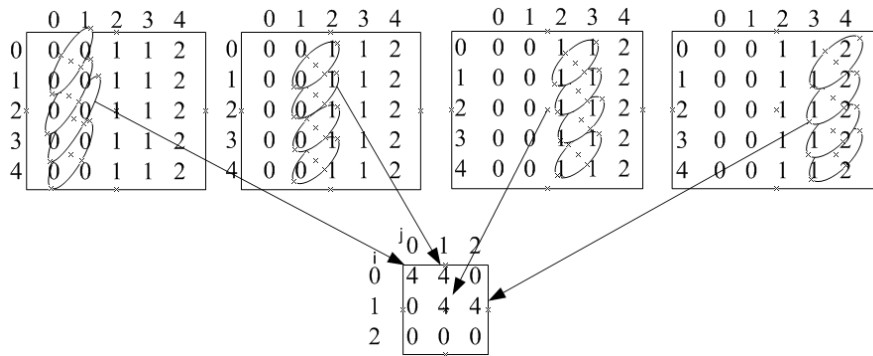


Figura 3.4: Matriz de coocurrencia para $d=1$ a 45°

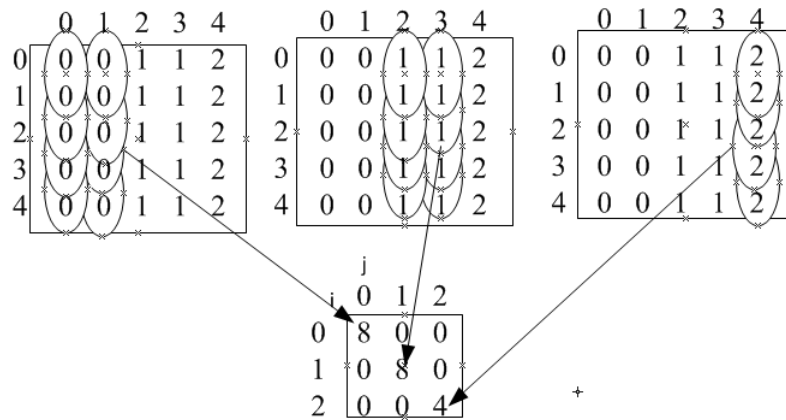


Figura 3.5: Matriz de coocurrencia para $d=1$ a 90°

3. MARCO TEÓRICO

Así, cuanto mayores sean los valores de la diagonal principal de la matriz de coocurrencia, más homogénea será la textura que representa, mientras que cuanto más repartidos estén los valores fuera de la diagonal principal más heterogénea será.

A continuación se presentan algunas medidas de detección de rasgos de textura que se calculan a partir de una matriz de coocurrencia.

- **Energía.**- El descriptor de energía de una textura extraído de la matriz de coocurrencia está dado por la siguiente ecuación

$$Energía = \sum_i \sum_j C_{ij}^2 \quad (3.1)$$

donde C_{ij} se refiere a cada elemento de la matriz de coocurrencia y varían desde 0 hasta el número total de niveles de gris.

De la ecuación 3.1 se puede decir que cuando todos los valores de C_{ij} son semejantes, (mayor dispersión en la diagonal principal de la matriz), el valor de la energía será menor; por el contrario, si ocurre que en la diagonal principal se dan mayores picos de intensidad, el descriptor será mayor. La propiedad de la energía da una idea de la suavidad de la textura, y esto se refleja en la ubicación de sus probabilidades en la matriz de coocurrencia.

- **Entropía.**- Es una medida de la aleatoriedad contenida en la matriz de coocurrencia. Se calcula mediante la siguiente ecuación

$$Entropía = - \sum_i \sum_j C_{ij} \log_2 |C_{ij}| \quad (3.2)$$

A medida que todos los elementos de la matriz son similares, este descriptor aumenta su valor, siendo máximo en el caso de que todos los elementos en la matriz fueran iguales.

- **Contraste.**- El contraste de una textura proporciona información acerca de las variaciones bruscas de color en la imagen y está dado por la siguiente ecuación.

$$Contraste = \sum_i \sum_j |i - j| C_{ij} \quad (3.3)$$

Como se observa en la ecuación 3.3, el valor del contraste aumentará si existen más elementos de la matriz de coocurrencia alejados de la diagonal principal. En una textura de características suaves y uniformes su contraste será bajo, mientras que si presenta un aspecto rugoso o irregular su contraste presentará un valor alto.

- **Homogeneidad local.**- Este descriptor proporciona información sobre la regularidad local de la textura. Se calcula mediante la siguiente ecuación:

$$Homogeneidad = \sum_i \sum_j \frac{C_{ij}}{1 + (i - j)^2} \quad (3.4)$$

Este descriptor aumentará cuando la distancia $i-j$ sea mínima, lo cual indica que mientras los elementos de la matriz de coocurrencia estén más próximos a la diagonal principal, mayor será el valor de la homogeneidad local.

3.3. Agrupamiento no supervisado

3.3.1. Algoritmo de K-medias

En este trabajo, se propuso probar con el algoritmo de K-medias creado por MacQueen. Es el algoritmo de *clustering* más conocido y utilizado ya que es de muy simple aplicación y eficaz. Sigue un procedimiento simple de clasificación de un conjunto de objetos en un determinado número \mathbf{K} de *clusters*. \mathbf{K} es determinado a priori. El nombre de K medias

3. MARCO TEÓRICO

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 es un significado gráfico y estadístico. Cada *cluster*, por lo tanto, se caracteriza por su centro o centroide, que se encuentra en el centro o en medio de los elementos que conforman al *cluster*.

Partimos de una base de datos X con N muestras y n atributos

donde:

- g_K = número de grupos;
- $Z_K(i)$ son los centros de masa de cada grupo o valor medio de las muestras contenidas en g_k
- N_K = número de muestras en cada grupo g_K

El algoritmo de las K-medias se realiza en cuatro etapas:

- Etapa 1:
 - elegir aleatoriamente K objetos que forman los K *clusters* iniciales.
- Etapa 2:
 - Se distribuyen cada una de las muestras X entre los K grupos según la siguiente regla: $X \in g_k(Z)$ si se cumple que la distancia entre la muestra x y el centroide $Z_K(1)$ es mínima.
- Etapa 3:
 - Recalcular los centros de los K *clusters*. Se buscan los centros $Z_k(2)$ de manera tal que el criterio de dispersión: $J_K = \sum_{X \in g_K(2)} D^2(x, Z_k(2))$ sea mínimo. Si el crite-

rio de similitud es la distancia euclideana, entonces: $J_K = \sum_{X \in g_K} (X - Z_K)^T (X - Z_K)$

para J_K siempre > 0 , luego $Z_K(2) = \frac{1}{N_K} \sum_{x \in g} X$

■ Etapa 4:

- Repetir los pasos 2 y 3 hasta que se consiga la convergencia, es decir, los centroides ya no se muevan.

En la figura 3.6 podemos ver el diagrama de flujo del algoritmo de las K medias.

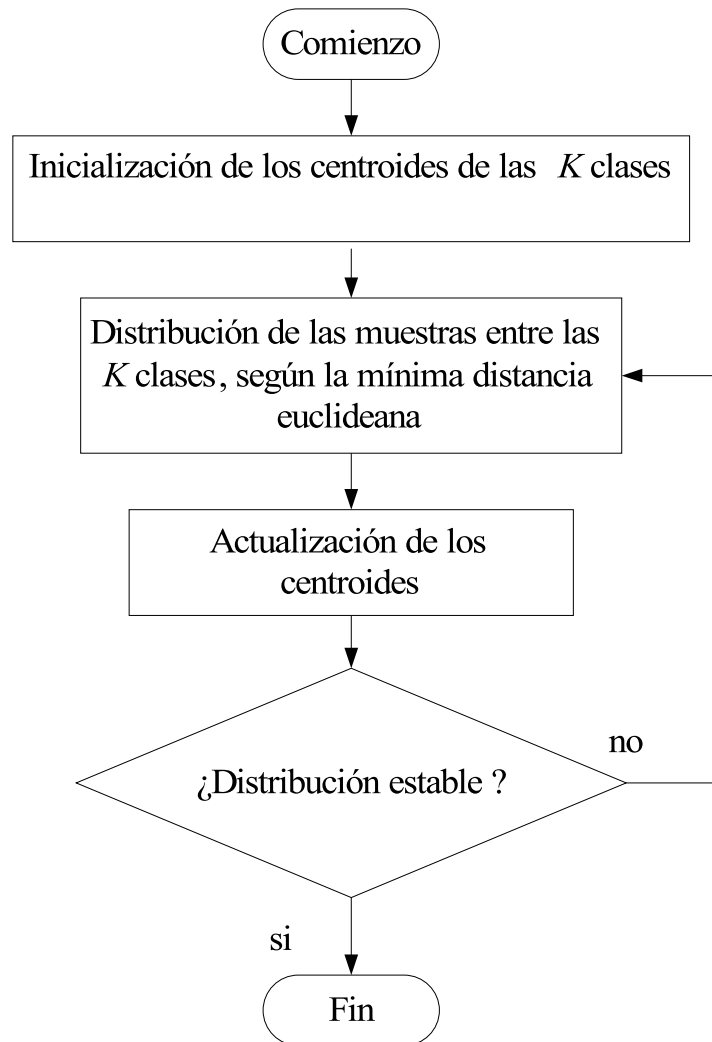


Figura 3.6: Diagrama de flujo del algoritmo de las k -medias.

3. MARCO TEÓRICO

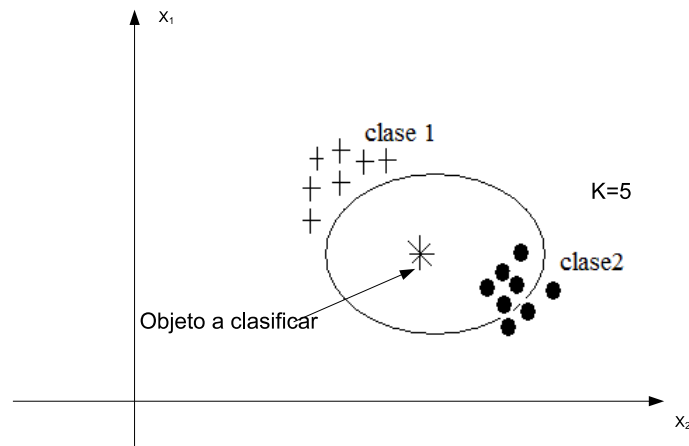


Figura 3.7: Clasificador K-NN

3.4. Clasificadores

3.4.1. Clasificador de los K vecinos más próximos (K-NN)

Este clasificador es ampliamente usado en el reconocimiento de formas; dado un vector a clasificar y un conjunto de diversas clases existentes (base del conocimiento) donde dicho vector pueda ser agrupado. La regla consiste en calcular la distancia del vector a clasificar a todos y cada uno de los vectores que conforman la base del conocimiento, después seleccionar los “**K**” vecinos más próximos y decidir por la clase más votada entre los mismos.

3.4.1.1. Principio teórico

Sea \vec{x} un vector de dimensión “ n ” a clasificar, sea M una base de datos de referencia construída a partir de N vectores de dimensión “ n ” y además se conoce la clase C_i a la cual pertenecen los vectores de la clase de referencia M . El clasificador de **K** próximos vecinos se basa en la estimación local de la densidad de probabilidad de la muestra \vec{x} a partir de los “ K -próximos vecinos” de la base de referencia [13] y [19]. Ver figura 3.7.

Sea $p(\vec{x}/C_i)$ la densidad de probabilidad. A partir de esta estimación, la regla de BAYES nos permite expresarlo en términos de la probabilidad a posteriori que la mues-

tra \vec{x} pertenezca a la clase C_i tal que:

$$p_r(C_i/\vec{x}) = \frac{p(\vec{x}/C_i) * p_r(C_i)}{p(\vec{x})} = \frac{p_r(\vec{x}/C_i) * p_r(C_i)}{\sum_{k=1}^c p(\vec{x}/C_k) * p_r(C_k)} \quad (3.5)$$

donde:

- $p_r(C_i)$ = probabilidad de aparición de la clase C_i
- $p_r(\vec{x})$ = probabilidad de que la muestra \vec{x} pertenezca a la clase C_i
- $p_r(C_i/\vec{x})$ = densidad de probabilidad condicional de la muestra \vec{x} conociendo la clase C_i

Partiendo de la base de referencia M (base del aprendizaje), se estiman las densidades de probabilidad $p(\vec{x}/C_i)$ para todas las clases C_i siguiendo 2 métodos diferentes, produciendo 2 reglas de decisión o afectación diferentes. El principio se basa sobre la búsqueda de los K -próximos vecinos de \vec{x} sin importar la clase (**método de reagrupamiento general**) o en una clase C_i (**método de reagrupamiento por clase**).

1. a) Método “de reagrupamiento general”:

- Sea “V” el volumen hipersférico definido por la distancia “D” entre la muestra \vec{x} y el K -ésimo vecino, la densidad de probabilidad conjunta $p(\vec{x}/C_i)$ es definida como $\frac{K_i}{(N*V)}$ siendo K_i el número de muestras que pertenecen a la clase C_i entre los K vecinos, normalizando con respecto al número total de muestras y dividido por el volumen que engloban los K -vecinos.
- Si se hace la hipótesis que las probabilidades de aparición de cada clase son equiprobables, es decir, $\forall i, j p_r(C_i) = p_r(C_j)$, entonces la ecuación 3.5 se transforma en:

$$p_r(C_i/\vec{x}) = \frac{\frac{K_i}{N*V}}{\sum_{j=1}^c \frac{K_j}{N*V}} = \frac{K_i}{\sum_{j=1}^c K_j} = \frac{K_i}{K} \quad (3.6)$$

3. MARCO TEÓRICO

donde: C = número total de clases y K =Número de los “ k ” vecinos buscados.

La clase a la cual pertenece la muestra \vec{x} es determinada al considerar el número más grande de prototipos pertenecientes a la clase C_i (entre los k prototipos). Es decir, que \vec{x} es asociado a la clase mayoritariamente representada de entre los K próximos vecinos. Generalmente el valor de K debe ser impar para evitar ambigüedades de clases que tienen el mismo número K de vecinos. En el caso a 2 clases C_1 y C_2 si $k_1/k_2 > 1$ entonces la clase ganadora será C_1 , en el caso contrario, la muestra \vec{x} será asignada a la clase C_2 . En la figura 3.8) se muestra un ejemplo, la cual, fue tomada de la referencia [30].

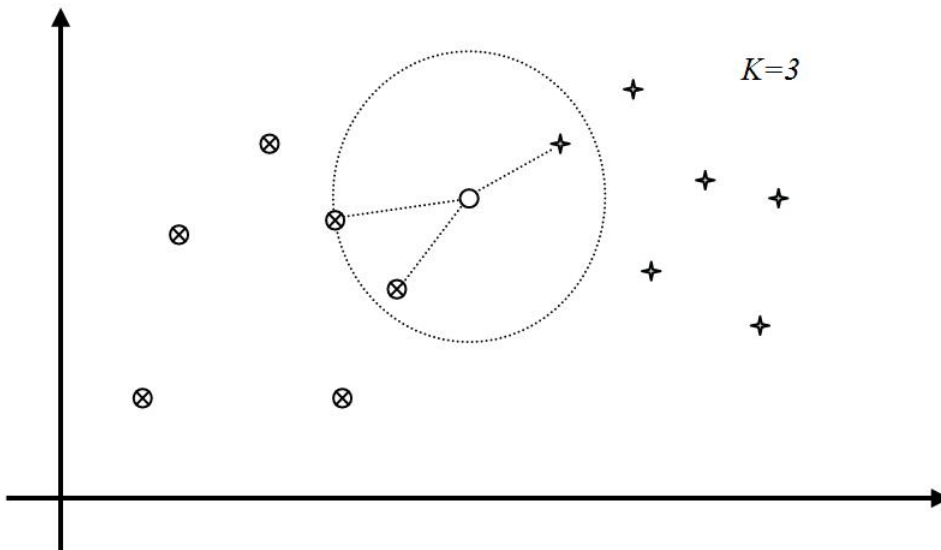


Figura 3.8: Selección de los k -vecinos, donde el patrón 'x' está representado por el pequeño círculo blanco, el cual es clasificado con la clase \otimes dado que de sus $k(3)$ próximos vécinos, “uno” pertenece a la clase $+$, y “dos” a la clase \otimes .

2. a) Método “ de reagrupamiento por clase

- La densidad de probabilidad conjunta $p(\vec{x}/C_i)$ se define ahora como $\frac{K}{N * V_i}$.
El número de K prototipos pertenecientes a la clase C_i son normalizados con

respecto al total de los prototipos (N) y divididos por el volumen V_i generado a partir de la k -ésima distancia. Si se realiza la misma hipótesis que las probabilidades a priori de cada clase C_i son equiprobables, es decir, $\forall i, j p_r(C_i) = p_r(C_j)$, entonces la ecuación 3.5 se transforma en la ecuación 3.7.

$$p_r(C_i/\vec{x}) = \frac{\frac{K_i}{N*V_i}}{\sum_{j=1}^c \frac{K_j}{N*V_j}} = \frac{\frac{1}{V_i}}{\sum_{j=1}^c \frac{1}{V_j}} \quad (3.7)$$

- Para determinar la clase a la que pertenece la muestra \vec{x} se definen tantos volúmenes como clases existentes. El volumen de la clase C_i es determinado por sus k representantes, los más próximos de la muestra \vec{x} . La clase ganadora C_i es la que posee el volumen más pequeño, es decir, la distancia más pequeña entre la muestra \vec{x} y los prototipos de la clase C_i

$$volumen = \frac{4\pi r^n}{3} = \frac{\pi d^n}{6} \quad (3.8)$$

- En el caso de dos clases C_1 y C_2 si $V_2/V_1 > 1$ ó en distancias $D_2 > D_1$, entonces la clase ganadora será C_1 y en el caso contrario, la muestra \vec{x} será asignada a la case C_2 . Como se puede observar en la figura 3.9, tomada de la referencia [30] V_1 ó clase \otimes es $<$ que V_2 o clase $\star \implies$ la clase ganadora es C_1

3. MARCO TEÓRICO

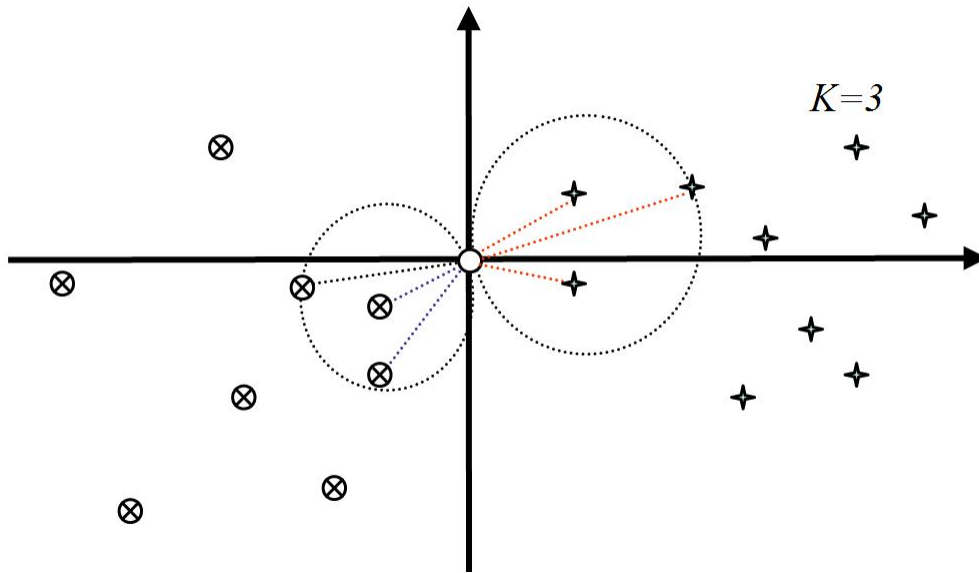


Figura 3.9: Selección de los k vecinos por “volumen”. El círculo blanco es clasificado en la clase \otimes , dado que $K=3$ próximos vecinos están más próximos que los 3 elementos próximos de la clase de puntos \star .

3.4.1.2. Cálculo de la distancia

Para calcular la distancia entre la muestra \vec{x} y los puntos de la base de datos M existen diferentes formas de medirla, pero en este caso solo usamos la distancia Euclídeana, la cual se define como:

- Distancia Euclídeana:

$$D(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3.9)$$

3.4.2. Clasificador bayesiano

3.4.2.1. Patrones y clases de patrones

En el análisis de una imagen es fundamental realizar, en algún nivel, un reconocimiento de patrones [30]. Un patrón es una descripción estructural o cuantitativa de un objeto o de alguna otra entidad de interés en una imagen. En general, un patrón está formado por

uno o más descriptores. En otras palabras, un patrón es una disposición de descriptores o características. Una *clase de patrones* es una familia de patrones que comparten algunas propiedades comunes. Las clases de patrones se representan por $\omega_1, \omega_2, \omega_3, \dots, \omega_M$, donde M es el número de clases. El reconocimiento de patrones mediante una máquina supone la utilización de técnicas que permitan asignar los patrones a sus respectivas clases, automáticamente y con la menor intervención humana posible.

Las tres representaciones de patrones principalmente utilizadas en la práctica son los vectores (para descripciones cuantitativas), cadenas y los árboles (para descripciones estructurales). Los patrones vectoriales se representan utilizando letras minúsculas en negritas como \mathbf{x}, \mathbf{y} , y \mathbf{z} y tienen la forma :

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad (3.10)$$

donde cada componente x_i representa el *i-ésimo* descriptor y n es el número de descriptores. Los patrones vectoriales se representan como columnas (esto es, como matrices $n \times 1$). De aquí que un patrón vectorial se puede expresar en la forma mostrada por la ecuación 3.10 o en la forma equivalente $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$, donde T indica la transposición.

La naturaleza de los componentes de un patrón vectorial \mathbf{x} depende de la técnica de medida utilizada para describir el propio patrón físico.

3.4.2.2. Métodos de decisión teórica

Los métodos de decisión teórica para el reconocimiento se basan en la utilización de *funciones de decisión* (o *discriminantes*). Con lo anterior, podemos hacer las siguientes

3. MARCO TEÓRICO

consideraciones. Supongamos que $x = |x_1, x_2, x_3 \dots x_n|^T$ representa un patrón vectorial de dimensión n . Para M clases de patrones podemos hacer $\omega_1, \omega_2, \dots, \omega_M$, el problema básico en el reconocimiento de patrones utilizando reglas de decisión teórica consiste en encontrar M funciones de decisión $d_1(x), d_2(x), d_3(x) \dots d_M(x)$, que tengan la propiedad de que, si un patrón pertenece a la clase ω_i , entonces

$$d_i(x) > d_j(x) \quad j = 1, 2, 3 \dots M \quad j \neq i \quad (3.11)$$

En otras palabras, un patrón desconocido x pertenece a la i -ésima clase de patrones si, al sustituir la x en todas las funciones de decisión, $d_i(x)$ toma el mayor valor numérico. Si se presenta un caso de empate, los conflictos se resuelven arbitrariamente.

La *frontera de decisión* que separa la clase ω_i de la clase ω_j viene dada por valores de x para los que $d_i(x) - d_j(x) = 0$. De esta forma $d_i(x) > 0$, para los patrones de la clase ω_i y $d_j(x) < 0$ para los patrones de la clase ω_j . En la figura (3.10) se muestra gráficamente para tres clases, tomada de la referencia [30].

La probabilidad a priori de que un patrón concreto, x , pertenezca a una clase ω_i se representa por $p(\omega_i/x)$. Si el clasificador de patrones decide que x proviene de ω_j cuando realmente proviene de ω_i , incurre en un error de clasificación, de notación L_{ij} . Como el patrón x puede pertenecer a cualquiera de las M clases consideradas, el error medio que se produce al asignar x a la clase ω_j es:

$$r_j(x) = \sum_{k=1}^M L_{kj} p(\omega_k/x) \quad (3.12)$$

La ecuación (3.12) se le suele llamar riesgo medio condicional o pérdidas esperadas según la terminología de la teoría de decisiones.

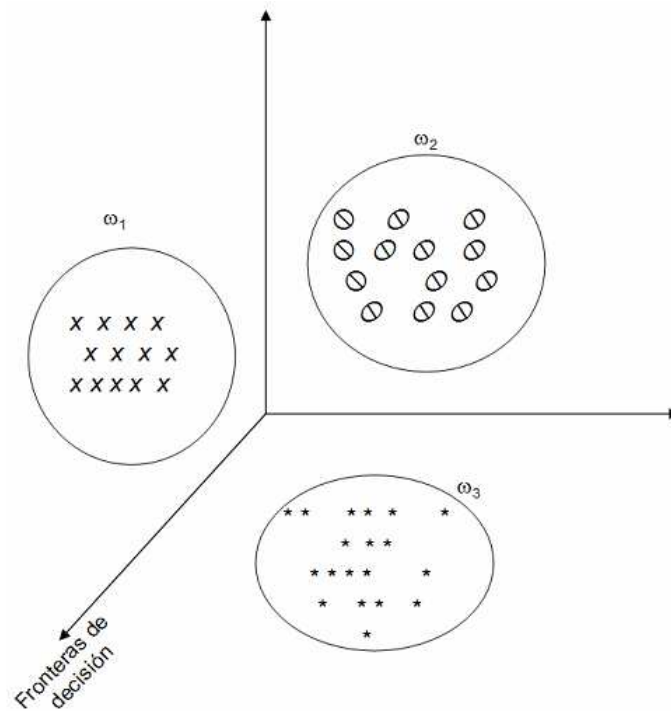


Figura 3.10: Ejemplo de clasificación de tres clases.

3.4.2.3. El clasificador de Bayes para clases de patrones gaussianos

Para empezar, consideremos un problema unidimensional, de tal manera que $n = 1$) que incluye dos clases de patrones ($M = 2$) gobernadas por intensidades gaussianas, con medidas m_1 y m_2 y desviaciones estándar σ_1 y σ_2 , respectivamente. Las funciones de decisión de Bayes tienen la siguiente forma:

$$d_j(x) = p(x/\omega_j) p(\omega_j) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left[-\frac{(x-m_j)^2}{2\sigma_j^2}\right] p(\omega_j) \quad j = 1, 2 \dots \quad (3.13)$$

donde los patrones son ahora escalares, de notación x . La figura (3.11) (la cual fue tomada de la referencia ([30]) muestra una representación gráfica de las funciones de densidad de probabilidad para las dos clases, la frontera entre las 2 clases es un punto único, denominado x_0 , tal que $d_1(x) = d_2(x)$. Si las 2 clases fueran equiprobales, $p(\omega_1) = p(\omega_2) = \frac{1}{2}$, y la frontera de decisión es el valor x_0 , para el cual $p(x_0/\omega_1) = p(x_0/\omega_2)$. Este punto es

3. MARCO TEÓRICO

la intersección de las 2 funciones de densidad de probabilidad, tal como se observa en la figura 3.11. Cualquier patrón (punto) situado a la derecha de x_0 se clasifica como miembro de la clase ω_1 , de manera similar, cualquier punto situado a la izquierda de x_0 se clasifica como miembro de la clase ω_2 . Si las 2 clases no fuera equiprobables, x_0 , se mueve a izquierda si existe más probabilidad de ocurrencia de la clase ω_1 o, a la derecha si tiene mas probabilidad de ocurrir la clase ω_2 . Este resultado es de esperarse, ya que el clasificador trata de minimizar el error de clasificación. En el caso de que la clase ω_2 nunca ocurriera, nunca se equivocaría el clasificador al asignar todos los patrones en la clase ω_1 . Esto es, el resultado de que x_0 se movería hacia menos infinito.

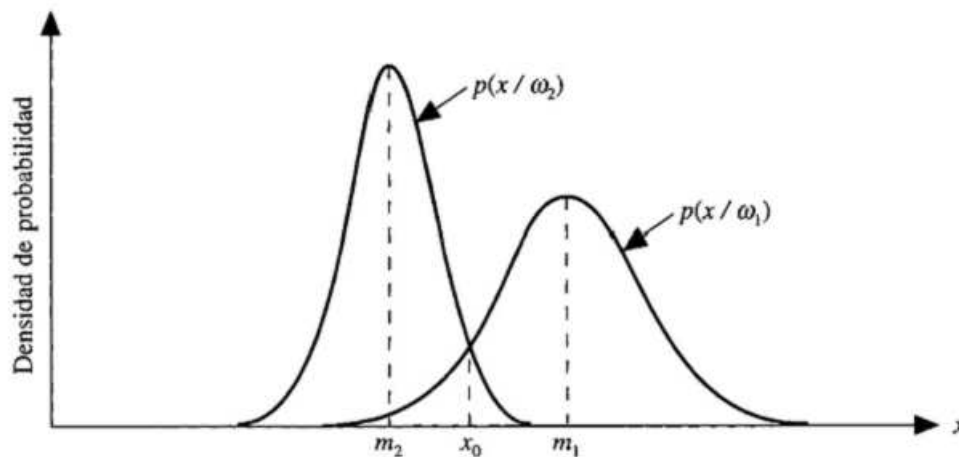


Figura 3.11: Funciones de densidad de probabilidades para dos clases de patrones unidimensionales. El punto x_0 es la frontera de decisión si las 2 clases son equiprobables.

En el caso mas generalizado de n , la densidad gaussiana de los vectores de la j -ésima clase de patrones tiene la forma:

$$p(x/\omega_j) = \frac{1}{(2\pi)^{n/2} \|C_j\|^{1/2}} \exp\left[-\frac{1}{2}(x - m_j)^T C_j^{-1} (x - m_j)\right] \quad (3.14)$$

donde cada densidad queda especificada completamente por su vector de la media m_j y la matriz de varianza-covarianza C_j , los cuales se definen a contiución:

$$m_j = E_j[x] \quad (3.15)$$

y

$$C_j = E_j[(x - m_j)(x - m_j)] \quad (3.16)$$

donde $E_j[\cdot]$ representa el valor esperado del argumento para los patrones de la clase ω_j . En la ecuación 3.14, n es la dimensión de los patrones vectoriales, y $|C_j|$ es el determinante de la matriz C_j . Al aproximar el valor esperado $E_j[\cdot]$ por el valor medio de las catidades en cuestión, se obtiene una estimación del vector de la media y la matriz de la varianza-covarianza, los cuales se muestran a contiución:

$$m_j = \frac{1}{N} \sum_{x \in C_j} x_j \quad (3.17)$$

y

$$C_j = \frac{1}{N_j} \sum_{x \in C_j} (x - m)(x - m)^T \quad (3.18)$$

donde N_j es el número de patrones vectoriales de la clase ω_j y la sumatoria se calcula para estos vectores.

La matriz de varianza-covarianza C_j es simétrica y no negativa.

3.4.3. Redes neuronales

Existen numerosas formas de definir el concepto de las redes neuronales, por lo que mencionaremos algunas de ellas a continuación:

- Una nueva forma de computación, inspirada en modelos biológicos.
- Un modelo matemático compuesto por un gran número de elementos procesales organizados en niveles.

3. MARCO TEÓRICO

- Un sistema de cómputo hecho por un gran número de elementos simples, elementos de proceso muy interconectados, los cuales procesan información por medio de su estado dinámico como respuesta a entradas externas [41]
- Son redes interconectadas masivamente en paralelo de elementos simples y con organización jerárquica, las cuales intentan interactuar con los objetos del mundo real del mismo modo que lo hace el sistema nervioso biológico [32].

En las redes neuronales **biológicas**, las neuronas corresponden a los elementos de procesos anteriores. Las interconexiones se realizan por medio de las ramas de salida(axones) que producen un número variable de conexiones(sinapsis) con otras neuronas (o quizá con otras partes, como músculos o glándulas). En las redes neuronales **artificiales** presentan un gran número de características semejantes a las de cerebro. Por ejemplo, son capaces de aprender de la experiencia, de generalizar de casos anteriores a casos nuevos, de abstraer o considerar aisladamente características esenciales a partir de entradas que representan información irrelevante, etc. Esto hace que ofrezcan numerosas ventajas y que este tipo de tecnología se esté aplicando en numerosas áreas. De acuerdo a [6] éstas ventajas pueden ser :

- **Aprendizaje adaptativo.** Capacidad de aprender para realizar tareas basadas en un entrenamiento o una experiencia inicial. Las redes neuronales pueden aprender a diferenciar patrones mediante ejemplos y entrenamientos, no es necesario que se elaboren modelos a priori ni necesitamos especificar funciones de distribución de probabilidad. Las redes neuronales son adaptables debido a su capacidad de autoajustarse de los elementos procesales(neuronas) que componen el sistema. Son dinámicos, porque constantemente están cambiando para adaptarse a las nuevas condiciones.
- **Autoorganización.** Una red neuronal puede crear su propia organización o representación de la información que recibe mediante una etapa de aprendizaje. Mientras

que el aprendizaje es la modificación de cada elemento procesal, la autoorganización consiste en la modificación de la red neuronal completa para llevar a cabo un objetivo específico. Cuando las redes neuronales se usan para reconocer ciertas clases de patrones, éstas se autoorganizan con la información usada. Por ejemplo, la red llamada “ **backpropagation**” (propagación hacia atrás) crea su propia representación característica, mediante la cual puede reconocer ciertos patrones. Esta autoorganización provoca la generalización: que es la facultad que tienen las redes neuronales de responder adecuadamente cuando se les presentan datos o situaciones a los que no habían sido expuestos anteriormente. El sistema puede generalizar la entrada para obtener una respuesta. Permite que el sistema proporcione una solución cuando la información de entrada es poco clara o incompleta.

- **Tolerancia a fallos.** La destrucción parcial de una red conduce a una degradación de su estructura; sin embargo, algunas capacidades de la red se pueden retener, incluso sufriendo algún daño en su estructura.
- **Operación en tiempo real.** Las redes neuronales se diseñan para trabajar en paralelo, puesto que la necesidad de cambio en los pesos de las conexiones o entrenamiento, es mínima. Una de las mayores áreas de aplicación es la necesidad de procesar grandes cantidades de información de forma muy rápida. Por lo tanto las redes neuronales son la mejor alternativa para el reconocimiento y clasificación de patrones en tiempo real.

3.4.3.1. La red de retropropagación (*Backpropagation*)

El algoritmo de la retropropagación o propagación hacia atrás, es una regla de aprendizaje que se puede aplicar en modelos de redes con más de 2 capas de células. Una característica importante de este algoritmo es la representación interna del conocimiento que es capaz de organizar en la capa intermedia de las células para conseguir cualquier correspon-

3. MARCO TEÓRICO

dencia entre la entrada y salida de la red. El funcionamiento de una red de retropropagación consiste en un aprendizaje de un conjunto predefinido de pares de entradas-salidas dados como ejemplo, empleando un ciclo de propagación-adaptación de dos fases: primero se aplica un patrón de entrada como estímulo para la primera capa de neuronas de la red, se va propagando a través de todas las capas interiores hasta generar una salida, se compara el resultado obtenido en las neuronas de salida que se desea obtener y se calcula un valor de error para cada neurona de salida. Después estos errores se transmiten hacia atrás, partiendo de la capa de salida, hacia todas las neuronas de la capa intermedia que contribuyan directamente a la salida recibiendo el porcentaje de error aproximado a la participación de la neurona intermedia en la salida original. Este proceso se repite por cada capa, hasta que todas las neuronas de la red hayan recibido un error que describa su aportación relativa al error total. Basándose en el valor de error recibido, se reajustan los pesos de conexión de cada neurona, de manera que cada vez que se presente el mismo patrón, la salida esté más cerca a la deseada, disminuyendo así el error.

La importancia de la red de retropropagación consiste en que la red debe encontrar una representación interna que le permita generar las salidas deseadas cuando se le dan las entradas de entrenamiento y que además las entradas no presentadas en la etapa de aprendizaje pueda clasificarlas según las características que comparte con los ejemplos de entrenamiento.

3.4.3.2. Estructura y aprendizaje de la red de retropropagación

En una red de retropropagación existe una capa de entrada con n neuronas y una capa de salida con m neuronas y al menos una capa oculta de neuronas internas. Cada neurona de una capa (excepto las de entrada) recibe entradas de todas las neuronas de la capa anterior y envía su salida a todas las neuronas de la capa posterior (excepto las de salida). No hay conexiones hacia atrás (retroalimentación) ni laterales entre neuronas de la misma capa (ver figura 3.12).

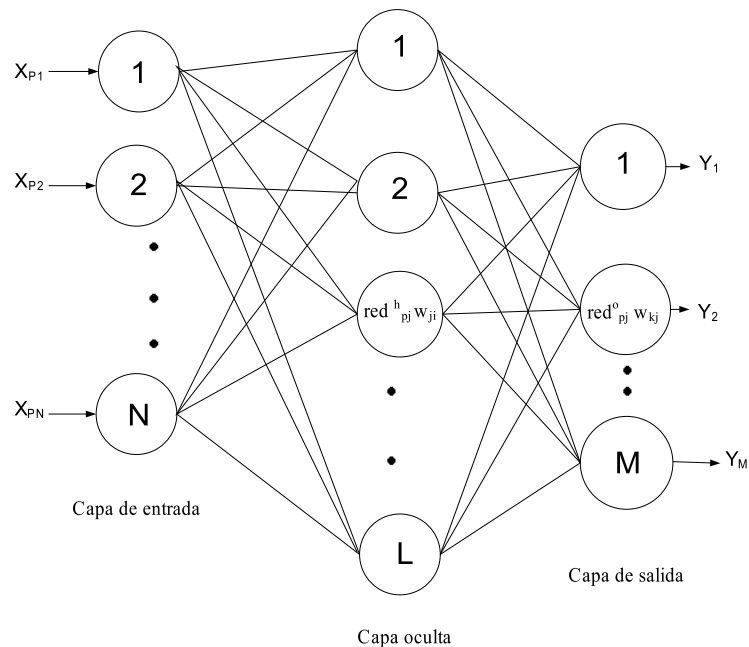


Figura 3.12: Modelo de la arquitectura de una red de retropropagación.

La estructura de la retropropagación tiene 2 etapas: una hacia adelante y otra hacia atrás. Durante la primera etapa el patrón de entrada es presentado a la red propagado a través de las capas hasta llegar a la capa de salida. Obtenidos los valores de salida de la red, se inicia la segunda etapa, al compararse estos valores con la salida esperada para obtener el error. Se ajustan los pesos de la última capa proporcionalmente al error. Se pasa a la capa anterior con una retropropagación del error, ajustando convenientemente los pesos y continuando con este proceso hasta llegar a la primera capa. De esta manera se han modificado los pesos de las conexiones de la red para cada ejemplo o patrón de aprendizaje.

3.4.3.3. Estructura de una sola neurona

Dada la siguiente figura (3.13) sea la neurona j : Cada célula j de la red cambia su salida y_j según la ecuación:

$$y_j = f(\text{Red}_j) \quad (3.19)$$

3. MARCO TEÓRICO

donde

$$Red_j = \sum_i w_{ji} \cdot y_i \quad (3.20)$$

siendo w_{ji} el peso de la conexión de la célula i en la célula j e y_i la salida que se produce en la célula i

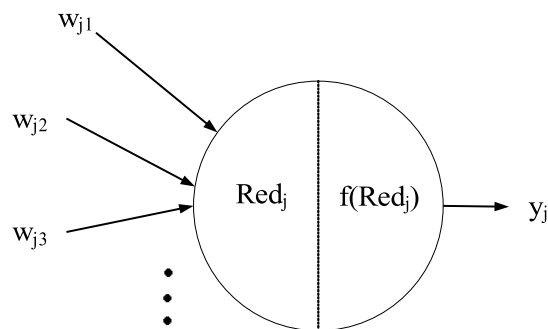


Figura 3.13: Estructura de una célula.

3.4.3.4. Algoritmo de entrenamiento de la red de retropropagación

A continuación se mencionan los pasos y procedimientos del algoritmo de entrenamiento de una red de retropropagación.

1. Iniciar los pesos de la red con valores aleatorios.
2. Presentar un patrón de entrada $X_P : X_{p1}, X_{p2}, \dots, X_{pN}$ y especificar la salida deseada que debe generar la red: d_1, d_2, \dots, d_M . En este caso la red se utilizará como un **clasificador**, por lo tanto, todas las salidas deseadas serán cero, excepto una, que será la de la clase a la que pertenece el patrón de entrada.
3. Calcular la salida actual de la red, y para lograr esto, se presentan las entradas de la red y se va calculando la salida que presenta cada capa hasta llegar a la capa de salida, ésta será la salida de la red: y_1, y_2, \dots, y_M . Los pasos son los siguientes:

- Se calculan las entradas netas para las neuronas ocultas procedentes de las neuronas de entrada.

- Para cada neurona oculta j :

$$red_{pj}^h = \sum_{i=1}^N w_{ji}^h x_{pi} \quad (3.21)$$

- donde el índice h se refiere a magnitudes de la capa oculta; el subíndice p , al p -ésimo vector de entrenamiento, y la j a la j -ésima neurona oculta.

- Se calculan las salidas de las neuronas ocultas:

$$y_{pj} = f_j^h(red_{pj}^h) \quad (3.22)$$

- Se realizan los mismos cálculos para obtener las salidas de las neuronas de salida:

$$red_{pk}^o = \sum_{j=1}^L w_{kj}^o y_{pj} \quad (3.23)$$

y

$$y_{pk} = f_k^o(red_{pk}^o) \quad (3.24)$$

4. Calcular los términos de error para todas las neuronas. La función o superficie de error asociada a la red busca el estado estable de mínima energía o de mínimo error a través del camino descendiente de la superficie de error. Por ello, retroalimenta el error del sistema para realizar la modificación de los pesos en un valor proporcional al gradiente decreciente de dicha función de error. En las redes multinivel, a diferencia de las redes sin neuronas ocultas, en un principio no se puede conocer la salida deseada de las neuronas de las capas ocultas para poder determinar los pesos en la función del error cometido. Sin embargo, inicialmente si podemos conocer la salida deseada de las neuronas de salida. entonces, de acuerdo a lo anterior, si la neurona k es una neurona de la capa de salida, el valor de la delta es:

$$\delta_{pk}^o = (d_{pk} - y_{pk}) f_k^{\prime o}(red_{pk}^o) \quad (3.25)$$

3. MARCO TEÓRICO

- donde d_{pj} es la salida deseada en la neurona k para el patrón p y red_{pk} es la entrada neta que recibe la neurona k para el patrón p . La ecuación 3.25 representa la modificación que hay que realizar en la entrada en la entrada de la neurona k . En el caso de que dicha neurona no sea de salida, el error que se produce estará en función del error que se cometa en las neuronas que reciban como entrada la salida de dicha neurona. **Esto es lo que se denomina procedimiento de propagación del error hacia atrás.**

- La función f , debe cumplir el requisito de ser derivable, lo que implica la imposibilidad de utilizar una función escalón. En general, se disponen de dos formas de función de salida:

$$f_k(red_{jk}) = red_{jk} \quad (3.26)$$

- y la función sigmoideal definida por la siguiente ecuación:

$$f_k(red_{jk}) = \frac{1}{1 + e^{-red_{jk}}} \quad (3.27)$$

- La selección de la función de salida depende de la forma en que se decida representar los datos de salida: Si se desea que las neuronas de salida sean binarias, se utiliza una función sigmoideal, en otros casos se puede aplicar cualquier otra función.
- Los términos de error para las neuronas de salida quedan, para una salida lineal:

$$\delta_{pk}^o = (d_{pk} - y_{pk}) \quad (3.28)$$

- y para una salida sigmoideal:

$$\delta_{pk}^o = (d_{pk} - y_{pk})y_{pk}(1 - y_{pk}) \quad (3.29)$$

- Si la neurona j no es de salida, entonces la derivada parcial del error no puede ser evaluada directamente. Por lo tanto, se obtiene el desarrollo a

partir de valores que son conocidos y otros que pueden ser evaluados. La expresión obtenida en este caso es:

$$\delta_{pj}^h = f_j^{h'}(\text{red}_{pj}^h) \sum_k \delta_{pk}^o w_{kj}^o \quad (3.30)$$

- Se observa que el error en las capas ocultas depende de todos los términos de error de la capa de salida. De aquí surge el término de *retropropagación hacia atrás*.
- Y para la función sigmoidea:

$$\delta_{pk}^h = x_{pi}(1 - x_{pi}) \sum_k \delta_{pk}^o w_{kj}^o \quad (3.31)$$

- donde k se refiere a todas las neuronas de la capa superior a la de la neurona j . Así el error que se produce en una neurona oculta es proporcional a la suma de los errores conocidos que se producen en las neuronas a las que está conectada la salida de ésta, multiplicando cada uno de ellos por el peso de la conexión.

5. Actualización de los pesos. Para lograr esto, utilizamos el algoritmo de forma recursiva, comenzando por las neuronas de salida y trabajando hacia atrás hasta llegar a la capa de entrada ajustando los pesos de la forma siguiente:

- Para los pesos de las neuronas de la capa de salida:

$$w_{kj}^o(t+1) = w_{kj}^o(t) + \Delta w_{kj}^o(t+1) \quad (3.32)$$

y:

$$\Delta w_{kj}^o(t+1) = \alpha \delta_{pk}^o y_{pj} \quad (3.33)$$

- donde Δ . es un incremento de paso que influye en la velocidad con la que converge el algoritmo. Esta velocidad se controla a través de la constante de proporcionalidad o tasa de aprendizaje α . Esta constante α normalmente debe

3. MARCO TEÓRICO

ser un número pequeño (del orden de 0.05 a 0.25), para asegurar que la red llegue a asentarse en una solución. Un valor pequeño de α significa que la red tendrá que hacer un gran número de iteraciones. Si esta constante es muy grande, los cambios de peso son muy grandes, avanzando rápidamente por la superficie de error, con el riesgo de saltar el mínimo y estar oscilando alrededor de él, pero sin poder alcanzarlo. Lo habitual es aumentar el valor de α a medida que disminuye el error de la red durante la fase de convergencia. Así aceleramos la convergencia aunque sin llegar nunca a valores de α demasiado grandes, que hicieran que la red oscile alejándose demasiado del valor mínimo.

6. El proceso se repite hasta que el término de error

$$E_p = \frac{1}{2} \sum_{k=1}^M \delta_{pk}^2 \quad (3.34)$$

- resulte aceptablemente pequeño para cada uno de los patrones aprendidos.

3.4.3.5. Diagrama de flujo de los pasos y procedimientos del algoritmo de la red de retropropagación

A continuación se muestra en la figura (3.14) el diagrama de flujo de los pasos y procedimientos que se han estado mencionando en párrafos anteriores.

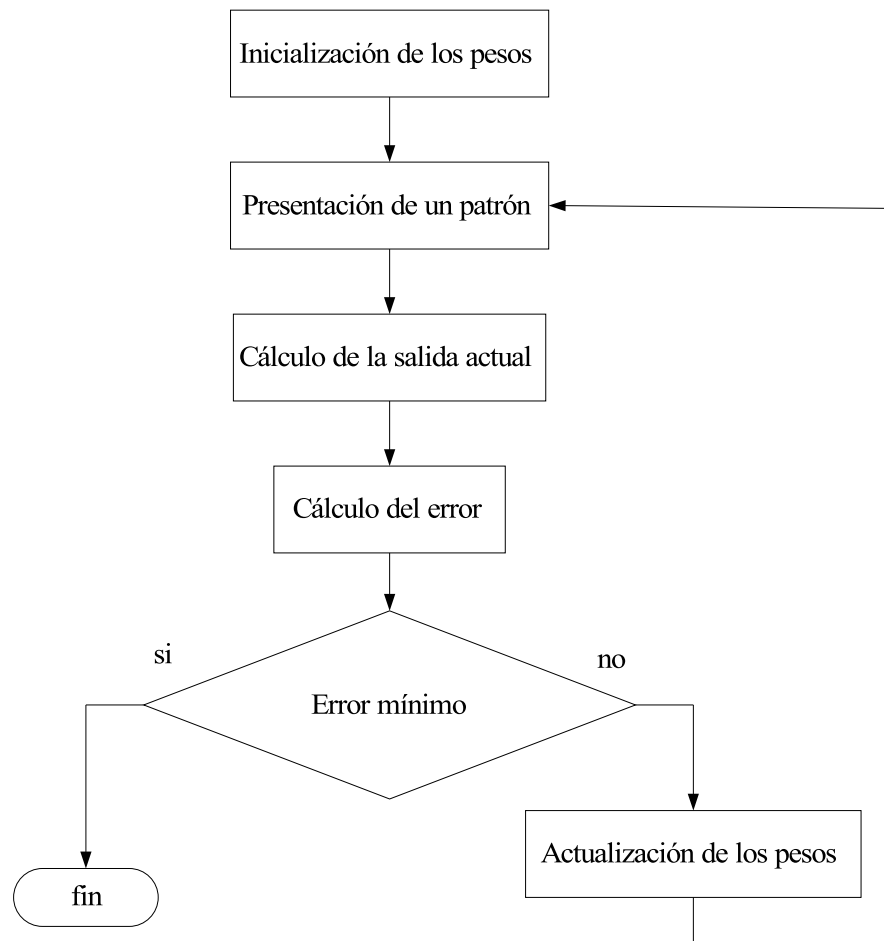


Figura 3.14: Diagrama de flujo de los pasos y procedimientos de la red de retropropagación.

3.4.3.6. Dimensiones de la red (número de neuronas ocultas)

No se pueden dar reglas concretas para determinar el número de neuronas ocultas o el número de capas en una red para resolver, en este caso, un problema de clasificación aplicando una red de retropropagación. Lo mismo ocurre al seleccionar al seleccionar el conjunto de vectores de entrenamiento. Respecto al número de capas de la red, en general tres capas son suficientes(entrada-oculta-salida). Sin embargo, hay veces en que un problema es más fácil de resolver (la red aprende mas rápidamente) con más de una capa oculta.

3. MARCO TEÓRICO

El tamaño de las capas, tanto de entrada, como el de salida, se determina por la naturaleza de la aplicación. En cambio, decidir cuántas neuronas debe tener la capa oculta no suele ser tan evidente. El número de neuronas ocultas interviene en la eficacia de aprendizaje y de generalización de la red. No hay ninguna regla que indique el número óptimo. En cada tipo de problema se debe estar probando con distintos números de neuronas para organizar la representación interna y escoger el mejor. Si la red no converge, es posible que sea necesario aumentar el número de neuronas. Por otro lado, examinando los valores de los pesos de las neuronas ocultas periódicamente en la etapa del aprendizaje, se pueden detectar a aquellas cuyos pesos cambian muy poco durante el aprendizaje respecto a sus valores iniciales, y así reducir el número de neuronas que apenas participan en el proceso de aprendizaje.

Capítulo 4

METODOLOGÍA PROPUESTA

En este capítulo se describe en detalle cada uno de los pasos involucrados en la aplicación de la metodología propuesta.

La extracción de características es un problema clave en lo referente a la recuperación de imágenes sobre la base de su contenido (CBIR). La metodología que se propone en esta tesis ha sido diseñada para la extracción y clasificación de características aplicada a la recuperación de imágenes. Esta metodología combina conjuntos puntos aleatorios y fijos para la extracción de características. Los rasgos descriptores que se proponen usar son: la media, la desviación estándar [19] y la homogeneidad; este rasgo descriptor proviene de la matriz de coocurrencia [39]. Estos tres rasgos se aplican a una subimagen bajo el dominio de los canales del formato (HSI) de una imagen. Se propone también el uso de un algoritmo de K -Medias [19] y algunos tipos de clasificadores como son:

- Clasificador 1-NN.
- Clasificador bayesiano
- Red Neuronal Artificial (RNA) con aprendizaje por retropropagación.

Se ha decidido combinar el algoritmo de K -Medias y algún tipo de clasificador para construir una base de datos indexada de 700 imágenes (por el momento). Una de las

4. METODOLOGÍA PROPUESTA

ventajas principales de la metodología que se propone es que no necesita hacer un etiquetado manual para la recuperación de las imágenes.

La Metodología propuesta involucra dos etapas principales:

- Etapa de entrenamiento.
- Etapa de prueba.

Esta etapa se divide en dos fases principales como se muestra en la figura 4.1. Durante la primera etapa (Parte A), un conjunto de 700 imágenes en formato RGB (720 x 480) o (480 x 720) es primeramente leído desde una base de imágenes de escenarios naturales. Posteriormente cada una de las imágenes es convertida al formato HSI. A cada imagen se le seleccionan automáticamente 300 píxeles aleatorios uniformemente distribuidos. Tomando cada uno de estos 300 puntos como centros, se abre una ventana cuadrada de tamaño 10 x 10 alrededor de cada uno de ellos. La figura 4.2 (b) muestra varios ejemplos. A cada una de las 300 ventanas se le extraen las siguientes características: (promedio del nivel de gris), desviación estándar y la homogeneidad obtenida desde una matriz de coocurrencia.

Nota: En la figura 4.1 parte B, la idea de hacer una selección de 100 ventanas fijas de (tamaño 10 x 10) se toma de las referencias [49] y [50].

Todo este proceso es aplicado a cada ventana en cada uno de sus tres canales de tono (H), saturación (S) e intensidad (I) de una imagen. El correspondiente vector de características para cada ventana de cada una de las imágenes tiene 9 componentes, tres para el canal H, tres para el canal S y tres para el canal I. Por lo que se obtiene una base de datos compuesta de 210,000 vectores descriptores (300 por cada una de las 700 imágenes), posteriormente se aplica el algoritmo de *K*-Medias de tal manera de obtener cuantos de estos 210,000 vectores se reparten entre 10 clases de objetos que se supone conforman a las imágenes: agua, roca, cielo, vegetación, pastos, y nubes más cuatro clases adicionales frontera o de borde que se forman entre cielo y agua, cielo y pastos, cielo y nubes, y cielo y

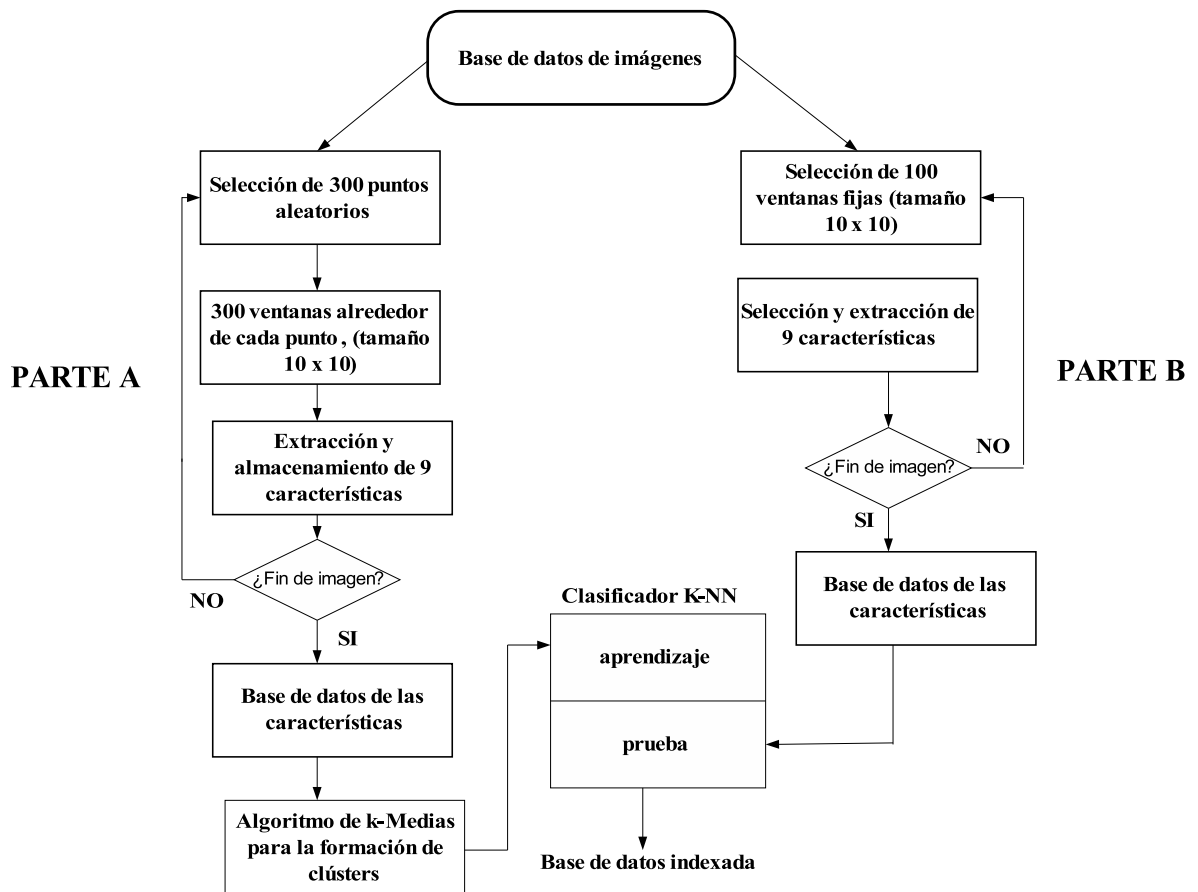


Figura 4.1: Diagrama de flujo para la etapa del entrenamiento.

rocas, dando como resultado, un total de 10 clases presentes en las imágenes de escenarios naturales.

Para las 700 imágenes seleccionadas durante la etapa de entrenamiento, la tabla 4.1 muestra cómo se reparten los 210,000 vectores entre las 10 posibles clases que se forman después de aplicar el algoritmo de *K*-Medias, es decir, cuántos vectores caen dentro de la clase 1, cuántos vectores caen en la clase 2, y así sucesivamente hasta la clase 10. Esto, de alguna manera, proporciona la probabilidad de que dada una clase, ésta pertenezca a alguna de las 700 imágenes. Se usaron 700 imágenes de escenarios naturales provenientes de la base de datos de COREL [48] y [50] durante la etapa del entrenamiento, las cuales están divididas en 6 diferentes tipos de imágenes: costas, ríos, lagos/lagunas, bosques, montañas,

4. METODOLOGÍA PROPUESTA



Figura 4.2: (a).-Para la descripción de las sub-imágenes, 300 píxeles de imagen son aleatoriamente seleccionados uniformemente. (b).-Para lograr una segmentación automática de la imagen, alrededor de cada uno de los 300 píxeles se abre una ventana cuadrada de tamaño $M \times N$. En esta figura se muestran solamente 20 puntos como ejemplo.

praderas y cielos/nubes respectivamente (ver figura 4.3).

Durante la segunda fase (Figura 4.1, parte B), para el mismo conjunto de entrenamiento de las 700 imágenes, una partición automática es fabricada como se muestra en la figura 4.5 (a). Cada imagen es dividida en 100 regiones de 10×10 de 72×48 píxeles cada una. Por cada una de estas 100 subimágenes [50], se toma una ventana de tamaño 10×10 píxeles como se muestra en la figura 4.5(b). A estas 100 ventanas fijas, se les extraen las mismas características: promedio del nivel de gris, desviación estándar y la homogeneidad calculadas en los mismos tres canales.

Cada ventana es descrita en forma de un vector de nueve componentes. De esta forma se tienen 70,000 vectores (100 por cada uno de las 700 imágenes). Para crear la base de datos indexada de las 700 imágenes que conforman el entrenamiento, se procede como sigue. Se toman los 210,000 vectores descriptivos (300 regiones por imagen y 700 imágenes) los cuales fueron obtenidos en la primera etapa del entrenamiento (figura 4.1 parte A). Éstos van a la entrada del aprendizaje de un clasificador 1-NN. Los 70,000 vectores descriptivos (obtenidos en la figura 4.1 parte B) entran a la parte de prueba de un clasificador 1-NN.



Figura 4.3: Escenas de costa, río/lago, bosque, montaña, pradera y cielo/nubes respectivamente.

A la salida del clasificador 1-NN se obtiene la base indexada compuesta de 700 vectores descriptivos, los cuales representan la información de cada una de las imágenes de escenarios naturales que conforman el proceso de entrenamiento.

| Número de componente de imagen (CI) | Número de características por cada componente de imagen |
|-------------------------------------|---|
| 1 | 22086 |
| 2 | 23267 |
| 3 | 23899 |
| 4 | 16127 |
| 5 | 23926 |
| 6 | 24506 |
| 7 | 30262 |
| 8 | 10708 |
| 9 | 10957 |
| 10 | 24252 |
| Total: 210,000 | |

Tabla 4.1: Distribución de los 210,000 características entre las 10 clases o componentes de imagen (CI) seleccionadas para el conjunto de las 700 imágenes de las imágenes de los escenarios naturales usadas para construir la base indexada de datos.

4. METODOLOGÍA PROPUESTA

4.0.4. Etapa de recuperación

Esta etapa fue diseñada como se muestra en la figura 4.4. Como se puede ver, solamente consta de una etapa. El procedimiento es como sigue: una imagen consulta es presentada al sistema. A esta imagen se le extraen las mismas características que se usaron en la etapa del entrenamiento (ver figura 4.1), de tal manera que se obtienen 100 vectores descriptores. Estos 100 vectores son inyectados directamente a un clasificados 1-NN previamente entrenado, el cual tiene una base de datos de referencia de 210,000 vectores aleatorios.

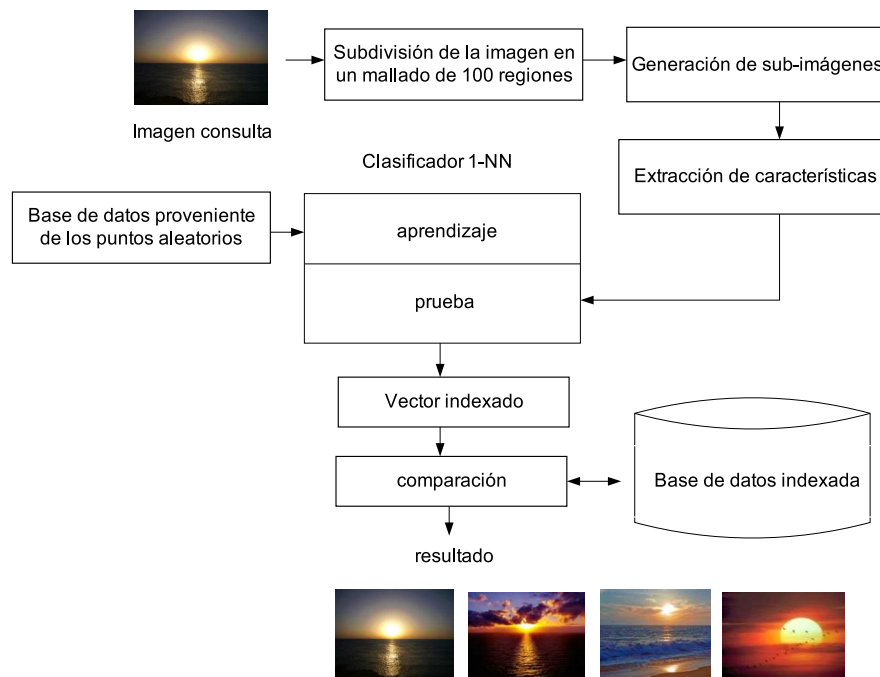


Figura 4.4: Diagrama de flujo para la etapa de la prueba.

A la salida de este proceso, solamente se obtiene un solo vector. Este vector contiene la probabilidad de cada una de las 10 clases, $C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, C_9$ y C_{10} esté contenido en la imagen consulta. Este vector es comparado con los 700 vectores almacenados en la base de datos indexada. Para reducir el tiempo de cálculo y obtener mejores resultados en la recuperación se toman las 4 clases con el mayor índice de votos de las 10 clases

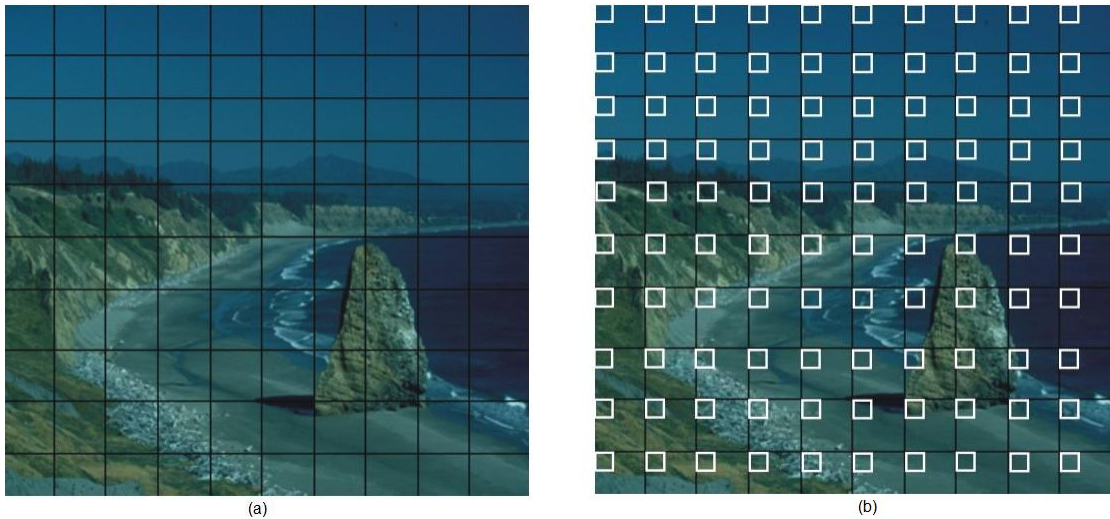


Figura 4.5: (a) Una imagen es uniformemente dividida en 100 subimágenes para obtener 100 regiones descriptivas de características. (b) Para cada una de las subimágenes, una ventana de tamaño 10 x 10 píxeles es seleccionada para calcular el correspondiente vector de características.

posibles. (Por ejemplo, en el último renglón de la tabla 4.2, las clases con más votos son $C_4, C_7, C_8,$ y C_{10}). Como medida de distancia, se usa la distancia euclideana. Para propósitos de recuperación, se escogieron manualmente 6 diferentes tipos de imágenes como se muestra en la figura 4.3.

Nota. Para probar nuestra propuesta, se han seleccionado 700 imágenes de escenarios naturales provenientes de la base de datos de imágenes COREL, cuya resolución es de 720 x 480 o de 480 x 720. Esta base de datos de imágenes fue proporcionada por J. Vogel [48], [49], [51] y [50]. Las 700 imágenes fueron utilizadas para formar el conjunto de entrenamiento, las cuales fueron divididas en 6 diferentes tipos de imágenes como sigue: 179 imágenes de montañas, 111 imágenes de ríos/lagos, 144 imágenes de costas, 103 imágenes de bosques, 131 imágenes de praderas y 32 imágenes de cielo/nubes.

En la tabla 4.2 se muestra cómo queda armada la base de datos indexada con las 700 imágenes de escenarios naturales pertenecientes al conjunto del entrenamiento.

4. METODOLOGÍA PROPUESTA

| C_1 | C_2 | C_3 | C_4 | C_5 | C_6 | C_7 | C_8 | C_9 | C_{10} | \rightarrow | Nombre de la imagen |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|---------------|---------------------|
| 3 | 3 | 2 | 10 | 60 | 6 | 0 | 0 | 16 | 0 | \rightarrow | imagen1.jpg |
| \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots |
| 1 | 22 | 0 | 21 | 50 | 2 | 0 | 0 | 4 | 0 | \rightarrow | imagen k.jpg |
| \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots |
| 6 | 3 | 0 | 25 | 0 | 3 | 22 | 28 | 4 | 9 | \rightarrow | imagen 700.jpg |

Tabla 4.2: Estructura de la base de datos indexada.

4.0.5. Metodología basada en puntos aleatorios

Después de una ardua experimentación se propone nuevamente hacerle una modificación a esta metodología y en vez de usar una combinación de puntos fijos y aleatorios como se muestra en la figura 4.1 simplemente nos quedamos con la parte aleatoria, entonces, la metodología modificada queda de la siguiente manera como se observa en la figura 4.6.

Para cada una de las n imágenes, 300 píxeles son seleccionados uniformemente (ver figura 4.2) aleatoriamente. Para cada uno de estos 300 puntos se abre una ventana cuadrada de tamaño 10×10 píxeles como se observa en la figura 4.2.

A cada una de estas 300 ventanas se le extraen las siguientes características: (promedio del nivel de gris), desviación estándar y la homogeneidad obtenida desde una matriz de co-ocurrencia.

Todo este proceso es aplicado a cada ventana en cada uno de sus 3 canales tono (H), saturación (S) e intensidad (I) de una imagen. El correspondiente vector de características para cada ventana de cada una de las imágenes tiene 9 componentes, tres para el canal H, tres para el canal S y tres para el canal I.

La parte de la prueba de la metodología modificad usando solamente puntos aleatorios se muestra en la figura (4.7)

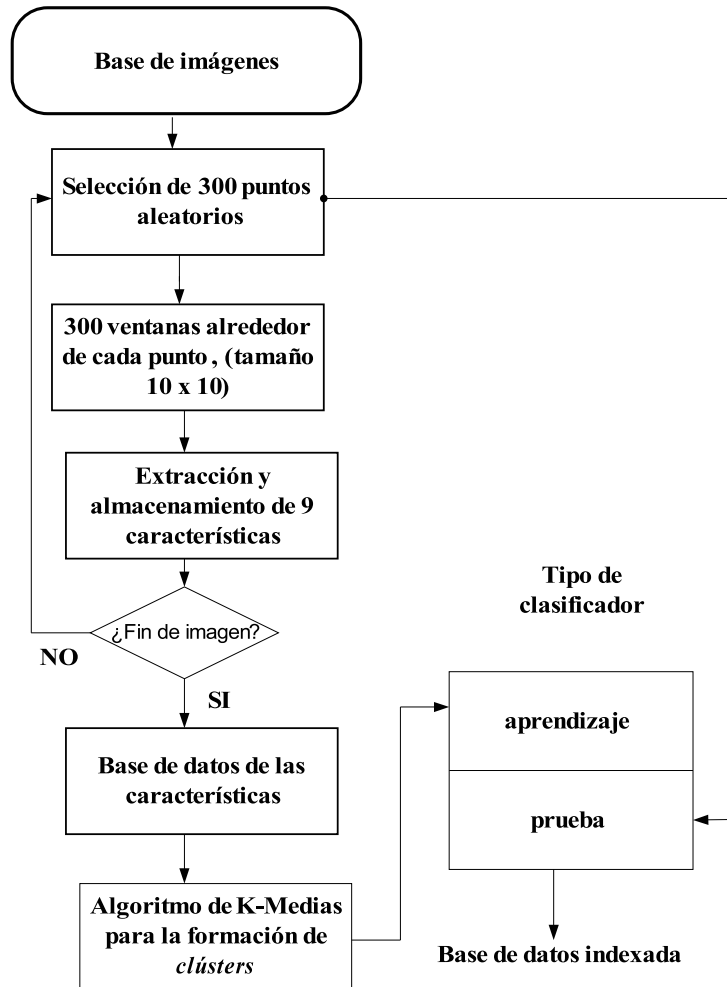


Figura 4.6: Diagrama de flujo para el entrenamiento de la metodología modificada.

Tomando los $n \times 300$ vectores de características (300 por cada una de las n imágenes que conformen el entrenamiento) y después de aplicar el algoritmo de las K -medias, se observa cuantos de los $n \times 300$ vectores de características se dividen en 10 grupos (*clústers* o componentes de imagen). Una componente de imagen (CI) es una parte representativa de la imagen. En el caso de un escenario natural, una CI puede ser una región de cielo o una región de arena. Por ejemplo, tomando una base de entrenamiento ($n=700$ imágenes). La tabla 4.1 muestra cuantos de estos 210,000 (700×300) caen en el CI uno, cuantos caen en el CI dos, y así hasta cuantos CI número diez. Esto nos dice que de alguna manera la probabilidad de que un determinado (CI) pertenece a las 700 imágenes. De la tabla 4.1,

4. METODOLOGÍA PROPUESTA

podemos ver que el número 7 del CI es la más representativa, mientras que las CI 8 y 9 son las menos representativas.

4.0.6. Construcción de la base indexada

Para construir la base de datos indexada, procedemos de la siguiente manera. Para cada una de las n imágenes de la base seleccionada para el entrenamiento, se toman sus 300 vectores de características (éstos fueron obtenidos durante la etapa del entrenamiento) y éstos son presentados a la entrada del tipo de clasificador seleccionado (el cual ya está entrenado). A la salida del clasificador, se obtienen los componentes de imagen (CI) para verificar a cual de éstas componentes de imagen (CI) pertenece.

Cada vez que un vector de características es clasificado como perteneciente a una clase o componente de imagen se acumula un voto dentro de 10 clases o elementos de imagen propuestas cuando se implementó el algoritmo del K-medias (ver primera columna de la figura 4.6) y más a detalle en la tabla 4.3.

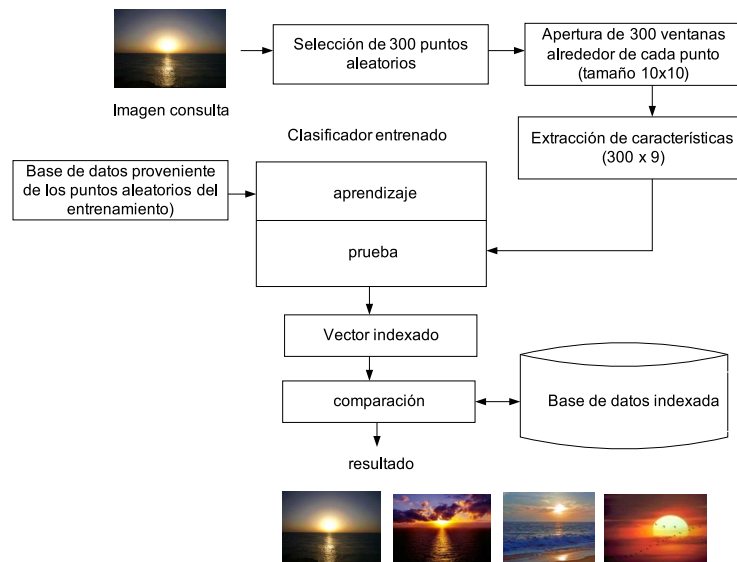


Figura 4.7: Diagrama de flujo para la prueba modificada

| IC1 | IC2 | IC3 | IC4 | IC5 | IC6 | IC7 | IC8 | IC9 | IC10 | → | nombre de la imagen |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|---|---------------------|
| 9 | 9 | 6 | 30 | 180 | 18 | 0 | 0 | 48 | 0 | → | imagen1.jpg |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | | ⋮ |
| 3 | 66 | 0 | 63 | 150 | 6 | 0 | 0 | 12 | 0 | | imagenk.jpg |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | | ⋮ |
| 18 | 9 | 0 | 75 | 0 | 19 | 66 | 74 | 12 | 27 | → | imagen n.jpg |

Tabla 4.3: Estructura de la base indexada usando solamente puntos aleatorios

La diferencia entre la tabla 4.2 y la tabla 4.3 respectivamente es que en la primera la metodología se contempla para su diseño y construcción puntos fijos y aleatorios y la suma en cada uno de sus componentes vectoriales es igual a 100 que son los puntos fijos, que indican como esos 100 puntos fijos se distribuyen entre las 10 clases o componentes de imagen (CI) (subimagen) cuando aplicamos el tipo de clasificador seleccionado, mientras que en la segunda tabla, su diseño y construcción contempla el uso de los 300 vectores de características provenientes de los 300 puntos aleatorios que se obtienen en la fase del entrenamiento y por lo tanto la suma de los componentes vectoriales de la tabla indexada es igual a 300, que indican como esos 300 puntos aleatorios se distribuyen entre las 10 clases o componentes de imagen (CI) (subimagen) cuando aplicamos el tipo de clasificador seleccionado.

En la etapa de prueba, cuando se lee la imagen consulta se obtiene un vector consulta con 300 componentes aleatorios divididos entre las 10 componentes de imagen, el cual es comparado los n vectores de la base indexada. Para reducir el tiempo de cómputo, se toman las 4 componentes más significativas o con el mayor número de votos. La distancia de comparación entre el vector consulta y la base de datos indexada es la distancia Euclideana.

Cabe señalar que para esta metodología de los puntos aleatorios se probó con 3 tipos de bases de datos que son:

4. METODOLOGÍA PROPUESTA

- COREL con 668 imágenes
- TORRALBA con 1471 imágenes
- CALTEK con 373 imágenes

También se probó con 3 tipos de clasificadores diferentes que son:

- clasificador 1-NN (o de distancia mínima)
- clasificador Bayesiano
- red neuronal artificial cuyo tipo de aprendizaje es de una red de retropropagación.

Capítulo 5

RESULTADOS EXPERIMENTALES

En este capítulo se describe el conjunto de experimentos realizados para probar el desempeño de la metodología desarrollada en esta investigación. Primeramente se muestran resultados relativos a la capacidad de la metodología para recuperar imágenes. Enseguida se muestran resultados relativos a como se aplica la metodología en el proceso de identificación de una imagen a través de sus imágenes o escenas consulta.

5.1. Recuperación de imágenes

Cuando se aplica el algoritmo de K -Medias, (en este caso $K=10$), se puede observar como se forman los *clusters* en una imagen de una escena natural usando puntos aleatorios. En la figura 5.1 se puede observar que la imagen está conformada por los *clusters* (grupos) más representativos de cada clase los cuales son los puntos negros, los cuales representan a la clase pasto, los puntos grises representan la clase cielo y los puntos blancos representan la clase foliage.

5. RESULTADOS EXPERIMENTALES

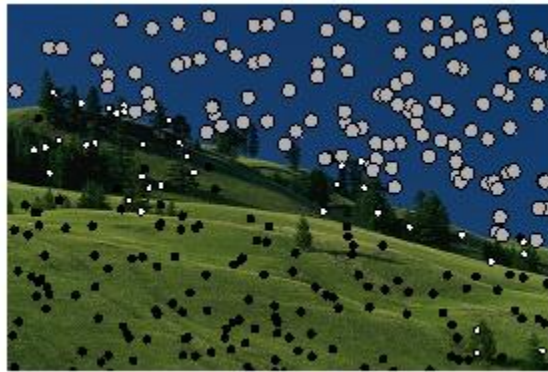


Figura 5.1: *Clusters* formados en una imagen de una escena natural usando el algoritmo de K-Medias y puntos aleatorios.

A continuación en las figuras 5.2(a), 5.2(b) y 5.2(c) se puede observar como ante transformaciones respectivamente las transformaciones de rotación y cambios de escala que puede sufrir la imagen, el resultado que se obtiene al aplicar el algoritmo de *K*-Medias que prácticamente es el mismo, es decir, el resultado es invariante ante transformaciones de imagen.

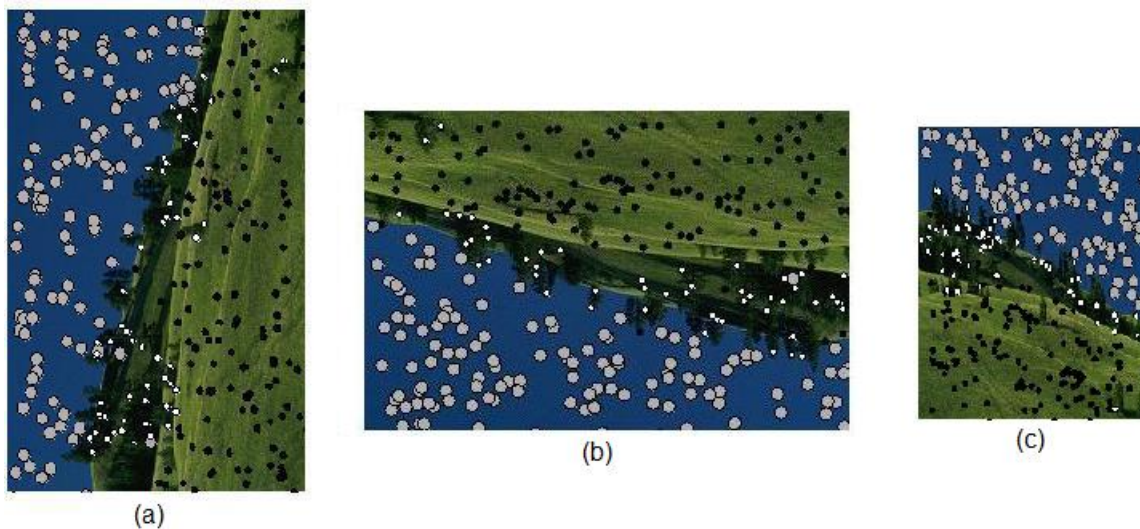


Figura 5.2: (a) Imagen rotada 90° . (b) Imagen rotada 180° . (c) Imagen escalada al 50%. Obsérvese como el resultado presenta invarianza ante estas transformaciones.



Figura 5.3: *Clusters* formados en una imagen de una escena natural usando el algoritmo de *K-Medias* y puntos aleatorios para imágenes de una imagen costera.

En la figura 5.3 se puede observar un ejemplo de como se conforman los escenarios del mismo tipo (en este caso, es un escenario de costas) cuando se aplica el algoritmo de *K-Medias* y los *clusters* mas representativos de los objetos presentes en el escenario de los cielos son los puntos negros, los cuales representan a la clase de los cielos, los puntos grises representan a la clase de las rocas y los puntos blancos representan a la clase agua. El tamaño de los puntos no tiene significado alguno, solo es para que no se pierda su visibilidad entre los puntos grises y los puntos blancos dentro de la figura.

Inicialmente se propuso usar 6 clases de objetos presentes en las imágenes de los escenarios naturales al aplicar el algoritmo de *K-Medias* probando con 300 imágenes de entrenamiento de la base de imágenes de COREL, las cuales fueron 54 imágenes de bosques, 54 imágenes de lagos/lagunas, 54 imágenes de costas, 54 imágenes de praderas, 54 imá-

5. RESULTADOS EXPERIMENTALES

| C_1 | C_2 | C_3 | C_4 | C_5 | C_6 | → | nombre de la imagen |
|-------|-------|-------|-------|-------|-------|---|---------------------|
| 40 | 16 | 23 | 20 | 1 | 0 | → | imagen 1.jpg |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 28 | 19 | 9 | 9 | 15 | 20 | → | imagen k.jpg |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 7 | 23 | 7 | 32 | 19 | 12 | → | imagen 300.jpg |

Tabla 5.1: Base de datos indexada para 6 clases de objetos y 300 imágenes de entrenamiento.

genes de montañas y 30 imágenes de cielos/nubes y las clases de objetos presentes en dichas imágenes de escenarios naturales son: agua, roca, pasto, cielo, vegetación y nubes, la base de datos indexada que se obtuvo se indica en la tabla 5.1.

Nota: Cuando se muestren los resultados experimentales, como se propuso acotar el problema de la recuperación de imágenes usando escenarios naturales y el conjunto de imágenes de entrenamiento está formado por imágenes de escenarios naturales.



Figura 5.4: Imágenes recuperadas dada una escena consulta tipo costa.



Figura 5.5: Imágenes recuperadas dada una imagen consulta tipo bosque.



Figura 5.6: Imágenes recuperadas dada una imagen consulta tipo costa.

En la figura 5.4 se observan los resultados al aplicar la metodología propuesta en la recuperación de imágenes dada una imagen consulta de tipo “puesta de sol”.

5. RESULTADOS EXPERIMENTALES

En la figura 5.5 se observan los resultados al aplicar la metodología propuesta en la recuperación de imágenes dada una imagen consulta de tipo “bosque”.

En la figura 5.6 se observan los resultados al aplicar la metodología propuesta en la recuperación de imágenes dada una imagen consulta de tipo “puesta de sol” completamente roja.

En esta sección se presentan los resultados experimentales obtenidos para validar nuestra propuesta. Para esto, se seleccionaron 221 imágenes desde Internet. Estas 221 imágenes de escenarios naturales no forman parte del conjunto del entrenamiento de la base de imágenes de COREL. Se le presentaron estas 221 imágenes al sistema de consulta y el sistema respondió desplegando en pantalla con las 10 imágenes más similares extraídas de la base indexada de imágenes. La figura 5.4 muestra un ejemplo de consulta. En dicha figura se puede observar que el sistema recupera correctamente 9 imágenes y solamente recupera incorrectamente 1 imagen (imagen 10). Esto nos arroja un resultado de 90% de eficiencia para esta recuperación. En la figura 5.5 se recuperan correctamente las 10 imágenes correspondientes a un bosque. Esto nos arroja un resultado de 100% de eficiencia para esta recuperación. En la figura 5.6, el sistema recupera correctamente 7 imágenes y solamente 3 son incorrectas (imagen 2, 4 y 8). Esto nos arroja un resultado de 70% de eficiencia para esta recuperación.

La prueba completa se puede observar en la figura 5.7 para una imagen consulta tipo “costas” usando una imagen con una puesta de sol completamente roja donde se prueba con la clase “cielo” dada en [49], [50], [51]; en la figura 5.8 se observa la prueba completa para imágenes tipo “bosque” donde se prueba con la clase “pasto” dada en [49], [50], [51]; y en la figura 5.9 se observa la prueba completa para imágenes tipo “costas” al comparar contra la Transformada Curvelet usada para las texturas en el trabajo [25] respectivamente.

Para probar la eficiencia de nuestra propuesta, se usaron las siguientes 2 mediciones, Precisión (P) y la recuperación (R):

$$P = \frac{\text{Número de imágenes relevantes recuperadas}}{\text{Número total de imágenes recuperadas}} \times 100\% \quad (5.1)$$

$$R = \frac{\text{Número de imágenes relevantes recuperadas}}{\text{Número total de imágenes relevantes en la base de imágenes}} \times 100\% \quad (5.2)$$

La ecuación (5.1) representa el número relevante de escenas recuperado con respecto al número total de escenas consultadas en el sistemas: La ecuación (5.2) representa el número relevante de escenas recuperadas con respecto al número total de escenas usadas durante el entrenamiento para una clase dada.

En la figura 5.7 se observa el resultado al comparar nuestra propuesta contra el método reportado en [51]. Como se puede apreciar, la eficiencia de nuestra propuesta es superior a la reportada en [51].

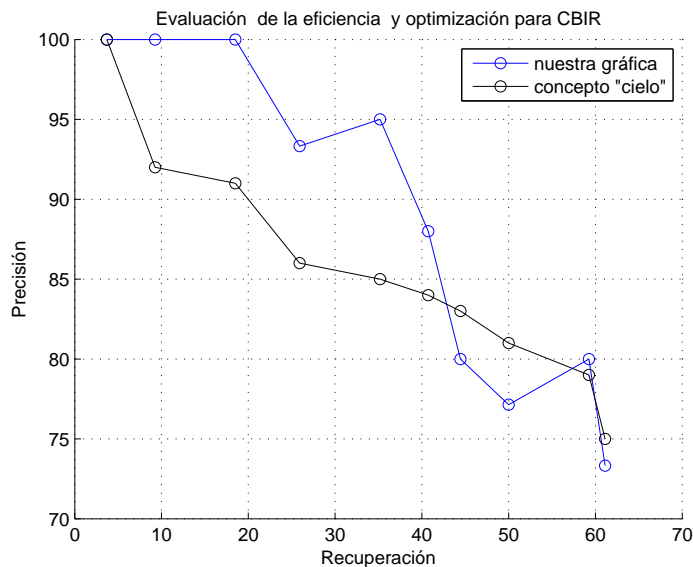


Figura 5.7: Eficiencia de nuestra propuesta comparada contra el método descrito en [51]. Mediante nuestra propuesta se obtiene 88.68% de eficiencia (gráfica azul), mientras que en [51] se obtiene 85.60% de eficiencia (gráfica en negro) cuando se aplica una consulta de una escena de una puesta de sol con cielo rojo.

5. RESULTADOS EXPERIMENTALES

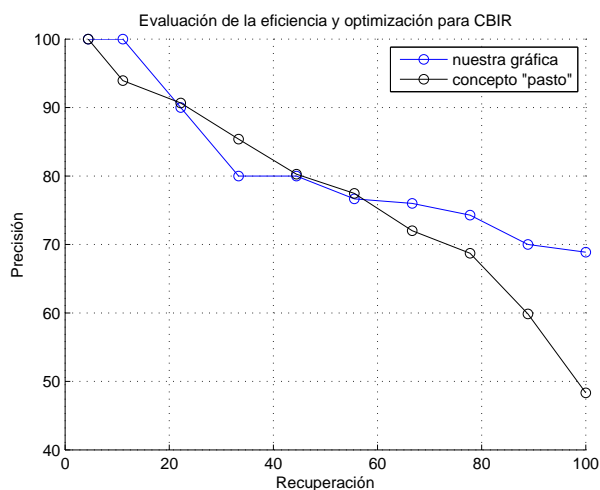


Figura 5.8: Eficiencia de nuestra propuesta comparada contra el método descrito en [51]. Mediante nuestra propuesta se obtiene 81.58 % de eficiencia (gráfica azul), mientras en [51] se obtiene 77.66 % de eficiencia (gráfica negra) cuando se aplica una consulta de la escena de un bosque.

En la figura 5.5 se observa el resultado obtenido al comparar nuestra propuesta contra el método reportado en [51]. Como se puede apreciar, la eficiencia de nuestra propuesta es nuevamente superior que la reportada en [51].

En la figura 5.9 se observa el resultado obtenido al comparar nuestra propuesta contra el método reportado en [25]. Como se puede apreciar, la eficiencia de nuestra propuesta es superior que la reportada en [25].

Siguiendo con el criterio de considerar solamente 6 clases, se procedió a probar nuestra metodología pero con una base de imágenes de la catedral de SACRE COEUR (París) cuya resolución de imagen es (768×1024). Estas imágenes fueron proporcionadas por Mauricio Díaz [14]. Se usaron 300 imágenes para el entrenamiento. La base de datos consiste en 3 tipos de imágenes principalmente: imágenes cuyo cielo es muy brillante, imágenes con cielo parcialmente nublado e imágenes con el cielo completamente nublado, por lo que la base de datos del entrenamiento tiene 100 imágenes de cada uno de estos tipos de cielos.

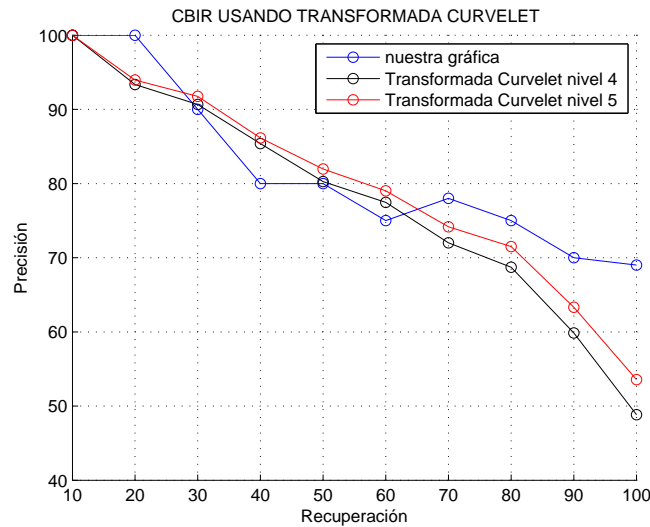


Figura 5.9: Eficiencia de nuestra propuesta al comparar contra el método descrito en [25]. Se obtiene una eficiencia del 81.7% (gráfica azul) mientras que en [25] se obtiene una eficiencia de 77.71% (gráficas en rojo y negro respectivamente).

Al usar nuestra metodología y al aplicarla ahora a las imágenes donde está presente la clase cielo se obtiene la figura 5.10 y la figura 5.11 respectivamente.

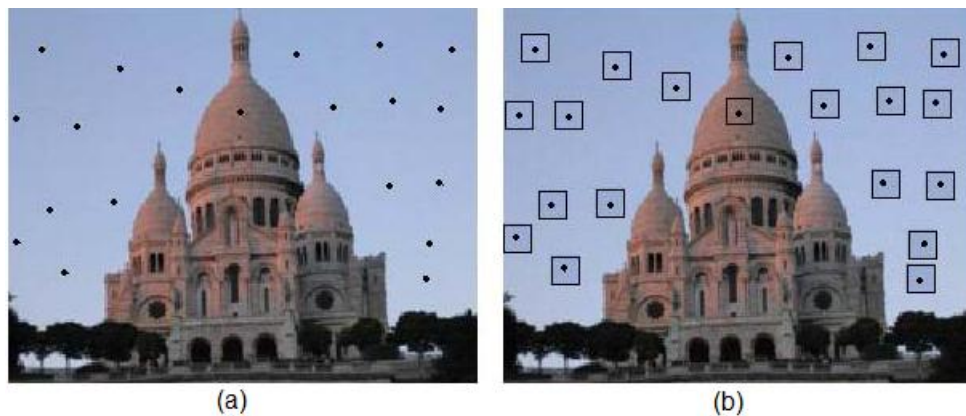


Figura 5.10: (a).-Para la descripción de las subimágenes, 300 píxeles de imagen son automáticamente seleccionados aleatoriamente. (b).-Para lograr una segmentación automática de la imagen, alrededor de cada uno de los 300 píxeles se abre una ventana cuadrada de tamaño $M \times N$. En esta figura solamente 20 puntos se muestran para dar un ejemplo.

5. RESULTADOS EXPERIMENTALES

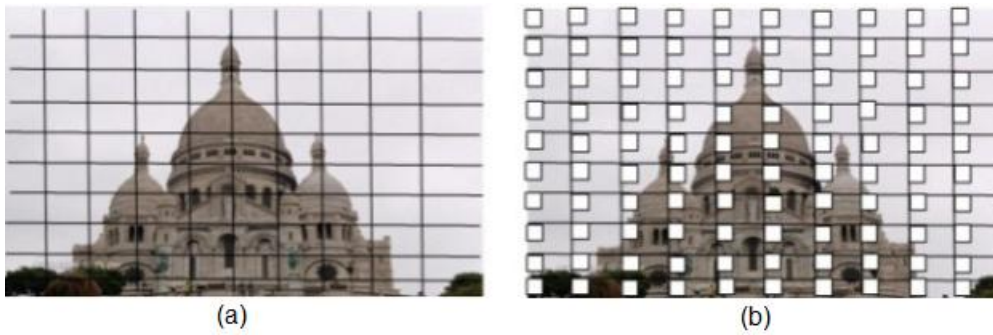


Figura 5.11: (a) Una imagen es uniformemente dividida en 100 subimágenes para obtener 100 regiones descriptivas de características. (b) Para cada una de las subimágenes, una ventana de tamaño 10 x 10 píxeles es seleccionada para calcular el correspondiente vector de características.

En la figura 5.12 se muestra un ejemplo del resultado obtenido de la recuperación de las escenas de diferentes tipos de cielos. Como se puede observar, se obtuvieron como resultado imágenes de cielos completamente nublados cuando se aplica una consulta de una imagen (que no forma parte del conjunto del entrenamiento de la base de imágenes “SACRE COEUR (París)”) de cielo completamente nublado.

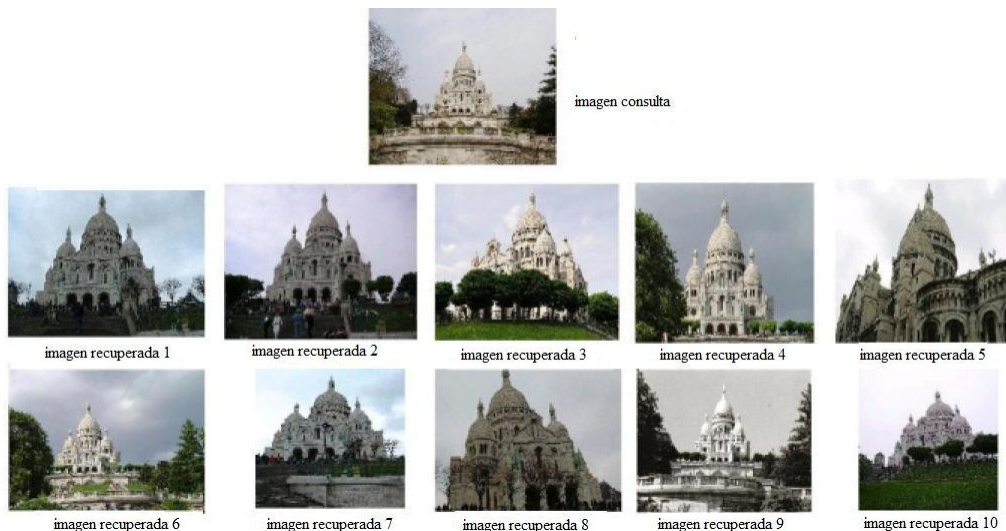


Figura 5.12: Recuperación de imágenes de cielo completamente nublado cuando se aplica al sistema una imagen consulta tipo cielo nublado.

Nota: En la figura (5.11) parte a) y b), la idea de hacer una selección de 100 ventanas fijas de (tamaño 10 x 10) se toma de la referencia [49] y [50].

Para probar la eficiencia de nuestra propuesta, en este caso, se usaron nuevamente las ecuaciones (5.1) y (5.2).

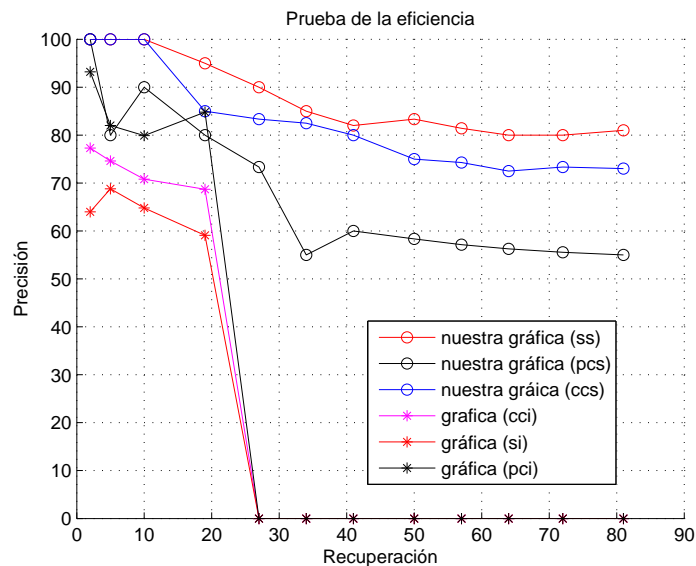


Figura 5.13: Eficiencia de nuestra propuesta comparada contra el método descrito en [14].

La figura 5.13 muestra que con nuestra propuesta (ver puntos (ss)) se obtiene 88.14% de eficiencia, mientras que en [14] los autores obtienen 64.17% de eficiencia (ver puntos (si)) cuando se hace la consulta de un cielo completamente brillante.

La figura 5.13 muestra que con nuestra propuesta (ver puntos (pcs)) se obtiene 63.75% de eficiencia, mientras que en [14] los autores obtienen 84.97% de eficiencia (ver puntos (pci)) cuando se hace la consulta de una escena de un cielo parcialmente nublado.

La figura 5.13 muestra que con nuestra propuesta (ver puntos (ccs)) se obtiene 83.24% de eficiencia, mientras que en [14] los autores obtienen 72.85% de eficiencia (ver puntos (cci)) cuando se hace la consulta de una escena de un cielo completamente nublado (ccs).

Nota: Las mediciones en [14] están solamente disponibles en un intervalo de 1 a 4 imágenes (por lo que se ponen asteriscos en negro sobre el eje x para representar la no

5. RESULTADOS EXPERIMENTALES

disponibilidad de sus mediciones) para lograr la recuperación de imágenes mientras que en nuestra propuesta está implementada con 12 medidas en un intervalo de 2 a 100 imágenes para lograr la recuperación de imágenes. Las dos bases de imágenes anteriores (la de COREL [49] y [50] y la de la catedral de de SACRE COEUR (París) [14] se probaron con 6 clases de objetos presentes en las imágenes (ver figura 5.14) y 300 imágenes de entrenamiento. Para mejorar aún los resultados de la recuperación de imágenes, se hizo la consideración de que en las imágenes pueden existir más clases de objetos de las que están consideradas en la figura 5.14. Por lo tanto, se considera que existen cuatro clases adicionales a las que llamaremos “clases de borde o de frontera”, las cuales se muestran en la figura 5.15. También se probaron estas cuatro clases adicionales, pero ahora probando con las 700 imágenes de toda la base de imágenes de COREL.

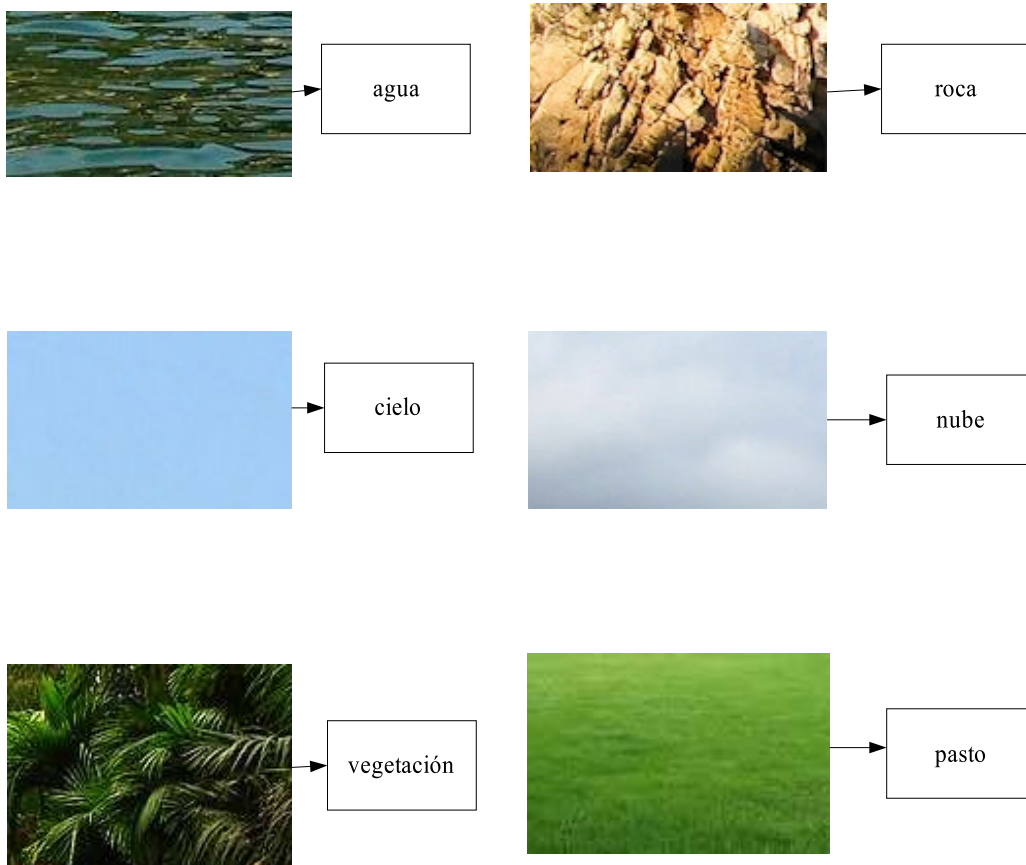


Figura 5.14: Las 6 clases de objetos presentes en las imágenes del entrenamiento.

En la tabla (4.1 en la página 93) se muestran como se distribuyen las 210,000 características (700 imágenes con 300 puntos aleatorios por cada imagen) entre las 10 clases seleccionadas.



Figura 5.15: Propuesta de la existencia de 4 clases adicionales a las 6 que ya están propuestas, las cuales les llamaremos “clases de borde, o de frontera”.

En la tabla (4.2 en la página 96) se muestran los resultados de como se formó la base de datos indexados tomando las 700 imágenes del COREL. Para probar nuestra propuesta, usamos las 700 imágenes de escenarios naturales de la base de imágenes de COREL (720 x 480) o (480 x 720). Las 700 imágenes de los escenarios naturales usados para el entrenamiento están agrupados en 6 diferentes tipos de imagen de la siguiente manera: 179 imágenes de montañas, 111 imágenes de ríos/lagos, 144 imágenes de costas; 103 imágenes de costas, 131 imágenes de praderas y 32 imágenes de cielo/nubes.

Algunos resultados de la recuperación de las imágenes se muestran a continuación en las figuras (5.16), (5.17), (5.18), (5.19), (5.20) y (5.21) respectivamente.

5. RESULTADOS EXPERIMENTALES

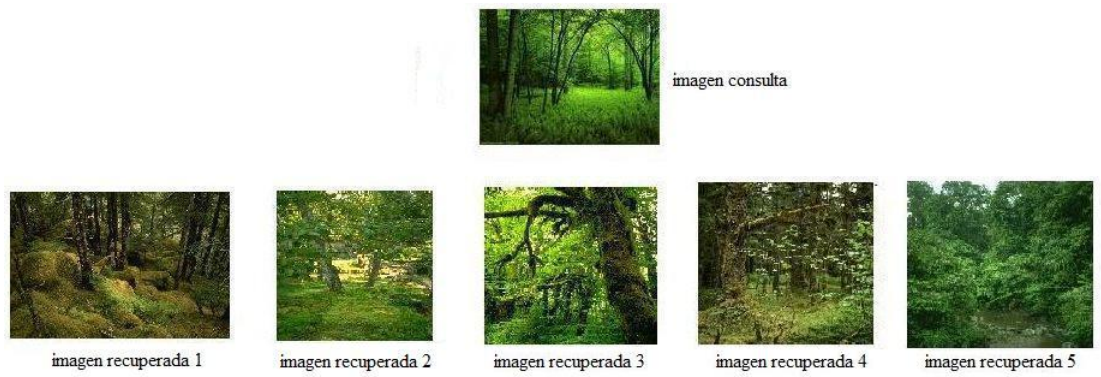


Figura 5.16: Imágenes recuperadas dada una imagen consulta tipo bosque.



Figura 5.17: Imágenes recuperadas dada una imagen consulta tipo costa.

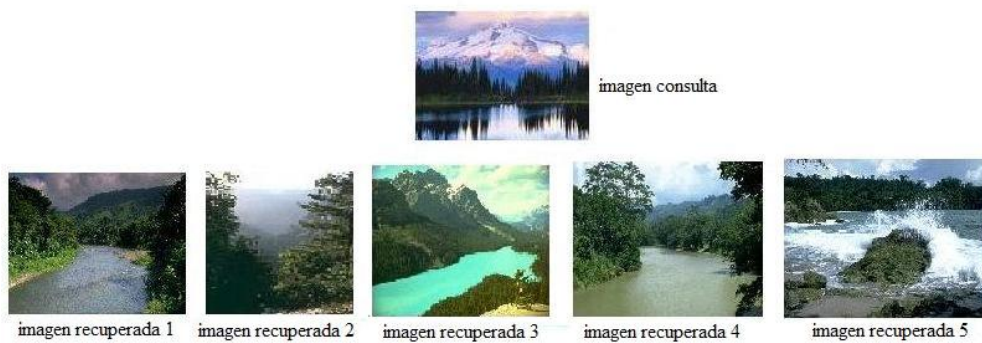


Figura 5.18: Imágenes recuperadas dada una imagen consulta tipo lago.



Figura 5.19: Imágenes recuperadas dada una imagen consulta tipo montaña.

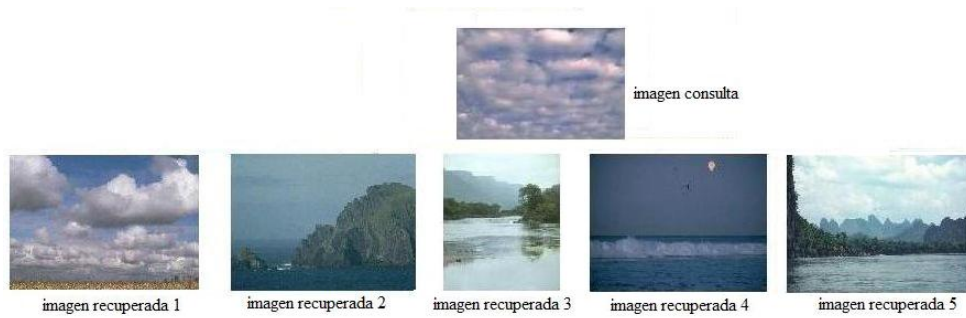


Figura 5.20: Imágenes recuperadas dada una imagen consulta tipo cielo/nubes.



Figura 5.21: Imágenes recuperadas dada una imagen consulta tipo pradera.

Para probar la eficiencia de nuestra propuesta en este caso usamos nuevamente las ecuaciones (5.1) y (5.2).

5. RESULTADOS EXPERIMENTALES

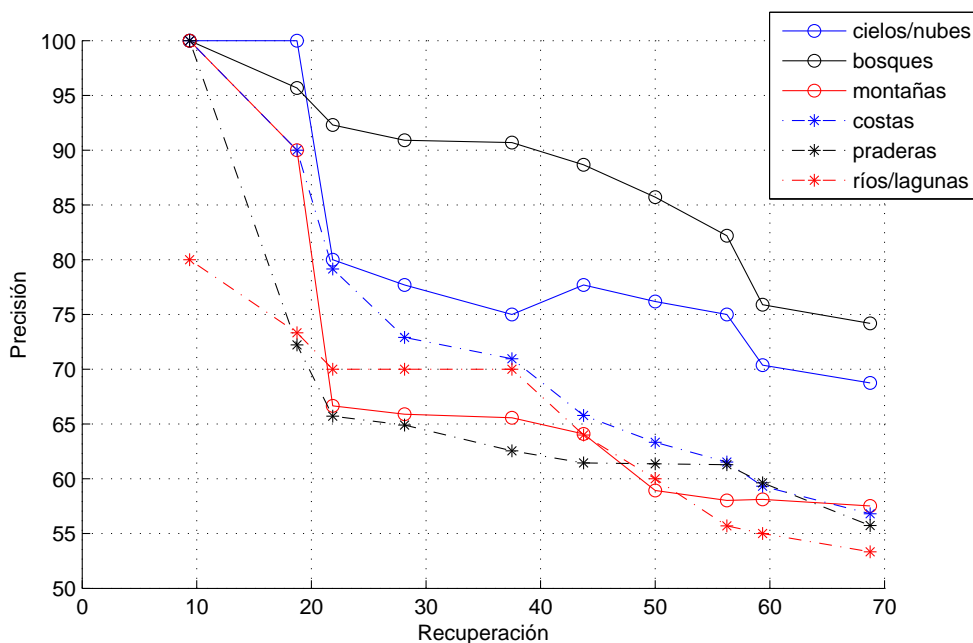


Figura 5.22: Eficiencia de nuestra propuesta, usando 10 clases y 700 imágenes de entrenamiento.

En la figura (5.22) muestra que nosotros obtenemos **70.29 %** de eficiencia (ver tabla 5.3) mientras que en la propuesta de [48] y [50] obtienen 58 % cuando se aplica recuperación de imágenes a imágenes tipo costas (ver tabla 5.2).

En la figura (5.22) muestra que nosotros obtenemos **63.71 %** de eficiencia (ver tabla 5.3) mientras que en la propuesta de [48] y [50] obtienen 40 % cuando se aplica recuperación de imágenes a imágenes tipo ríos/lagos.(ver tabla 5.2).

En la figura (5.22) muestra que nosotros obtenemos **86.18 %** de eficiencia (ver tabla 5.3) mientras que en la propuesta de [48] y [50] obtienen 83 % cuando se aplica recuperación de imágenes a imágenes tipo bosques (ver tabla 5.2).

En la figura (5.22) muestra que nosotros obtenemos **65.29 %** de eficiencia (ver tabla 5.3) mientras que en la propuesta de [48] y [50] obtienen 30 % cuando se aplica recuperación de imágenes a imágenes tipo praderas (ver tabla 5.2).

En la figura (5.22) muestra que nosotros obtenemos 66.43 % de eficiencia (ver tabla 5.3) mientras que en la propuesta de [48] y [50] obtienen **70 %** cuando se aplica recuperación de imágenes a imágenes tipo montañas (ver tabla 5.2).

En la figura (5.22) muestra que nosotros obtenemos 80.77 % de eficiencia (ver tabla 5.3) mientras que en la propuesta de [48] y [50] obtienen **87 %** cuando se aplica recuperación de imágenes a imágenes tipo cielos/nubes (ver tabla 5.2).

| | costas | ríos/lagos | bosques | praderas | montañas | cielos/nubes |
|-----------------|--------|------------|---------|----------|----------|--------------|
| tipos de imagen | 58 % | 40 % | 83 % | 30 % | 70 % | 87 % |

Tabla 5.2: Promedio de eficiencia para la metodología descrita en [48] y [50].

| | costas | ríos/lagos | bosques | praderas | montañas | cielos/nubes |
|-----------|---------|------------|---------|----------|----------|--------------|
| precisión | 70.29 % | 63.71 % | 86.18 % | 65.29 % | 66.43 % | 80.77 % |

Tabla 5.3: Resultados obtenidos con nuestra propuestas (valores promedio obtenidos de la figura 5.22).

Como se puede ver, mediante nuestra propuesta, en general, para todas las clases, la eficiencia es más alta; solamente para el caso de imágenes tipo cielos/nubes y tipo montañas nuestra propuesta es un poco menor.

5.2. Identificación del tipo de imagen

A continuación se muestran los resultados experimentales, los cuales llamaremos: “Identificación de la imagen consulta” A los diferentes tipos de imágenes con escenarios naturales se dividen en 2 grupos: El primer grupo contiene imágenes tipo bosques, cielos/nubes y montañas. El segundo grupo está conformado por imágenes tipo praderas, costas y ríos/lagos.

5. RESULTADOS EXPERIMENTALES

El procedimiento es como sigue: Se tomaron manualmente desde Internet 15 imágenes pertenecientes a cada tipo de imagen de (ver figura 4.3 en la página 93), luego se aplica nuestra metodología para la recuperación de las imágenes dando un total de las 90 imágenes, (es decir, 15 imágenes para cada uno de los 6 tipos de imágenes); una vez que se realiza la recuperación de imágenes, se obtiene un vector indexado para cada una de estas 90 imágenes consulta y se seleccionan las 2 clases (de las 10 posibles que se forman cuando se aplica el algoritmo de K-Medias) con el mayor número de votos de ocurrencia y el resultado queda como sigue:

- Grupo 1: formado por imágenes tipo bosques, cielo/nubes y montañas
 - *imágenes tipo bosques:*
 - están conformadas por las clases 8 y 9. Es decir, las clases 8 y 9 tienen los más altos índices de votación del resto de las 10 posibles clases.
 - *imágenes tipo cielo/nubes:*
 - están conformadas por las clases 2 y 5. Es decir, las clases 2 y 5 tienen los más altos índices de votación del resto de las 10 posibles clases.
 - *imágenes tipo montañas:*
 - están conformadas por las clases 1 y 5. Es decir, las clases 1 y 5 tienen los más altos índices de votación del resto de las 10 posibles clases.
- Grupo 2: formado por imágenes tipo praderas, costas y ríos/lagos
 - *imágenes tipo praderas:*
 - están conformadas por las clases 4 y 7. Es decir, las clases 4 y 7 tienen los más altos índices de votación del resto de las 10 posibles clases.
 - *imágenes tipo costas:*

- están conformadas por la clase 4 y 5. Es decir, las clases 4 y 5 tienen los más altos índices de votación del resto de las 10 posibles clases.

- **imágenes tipo ríos/lagos:**

- éstas están conformadas por la clase 1 y 9. Es decir, las clases 1 y 9 tienen los más altos índices de votación del resto de las 10 posibles clases.

Para probar la eficiencia de la propuesta respecto a la identificación del tipo de imagen, se obtiene una matriz de confusión para cada grupo y así poder validar y clasificar el tipo de imagen (ver tabla (5.4) para el grupo 1), la cual muestra una eficiencia de 81.10% (valor promedio) y (ver tabla (5.5) para el grupo 2), la cual muestra una eficiencia de 82.22% (valor promedio), respectivamente.

Para probar la eficiencia de nuestros resultados de lo que llamamos “Identificación de la escena consulta”, usamos las siguientes 2 medidas P =precisión e I =Identificación representadas por las ecuaciones (5.3) y (5.4) respectivamente . La prueba completa se puede observar en la figura 5.23.

$$P = \frac{\text{Número de imágenes relevantes identificadas}}{\text{Número total de imágenes identificadas}} \times 100\% \quad (5.3)$$

$$I = \frac{\text{Número de imágenes relevantes identificadas}}{\text{Número total de imágenes relevantes en la base de datos}} \times 100\% \quad (5.4)$$

| | bosques | cielos/nubes | montañas |
|--------------|----------------|----------------|-------------|
| bosques | 76.66 % | 6.66 % | 16.66 % |
| cielos/nubes | 0 % | 86.66 % | 13.33 % |
| montañas | 0 % | 20 % | 80 % |

Tabla 5.4: Matriz de confusión para el grupo 1.

5. RESULTADOS EXPERIMENTALES

| | praderas | costas | ríos/lagunas |
|--------------|----------------|-------------|--------------|
| praderas | 86.66 % | 6.66 % | 6.66 % |
| costas | 13.13 % | 80 % | 6.66 % |
| ríos/lagunas | 3.33 % | 16.66 % | 80 % |

Tabla 5.5: Matriz de confusión para el grupo 2.

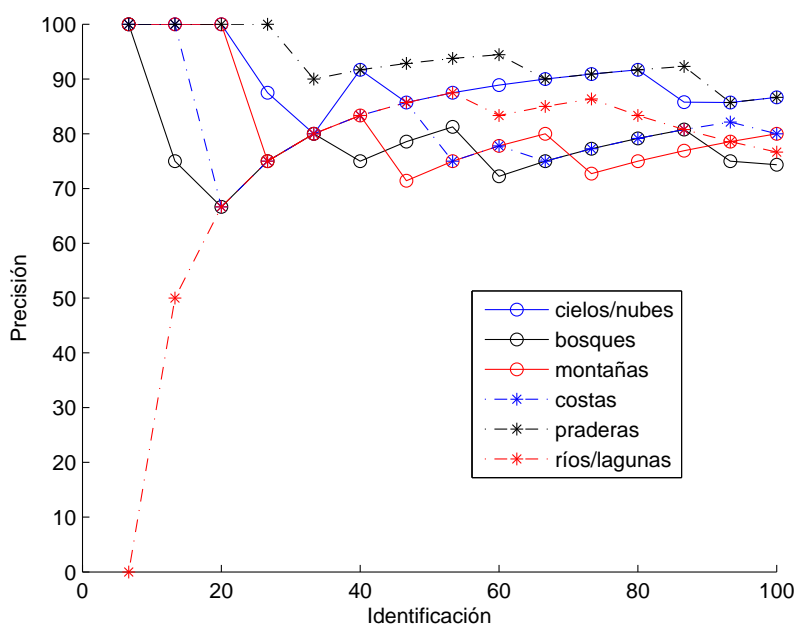


Figura 5.23: Eficiencia de nuestra propuesta de “Identificación de la imagen consulta”.

De la figura 5.23 y usando las ecuaciones 5.3 y 5.4 se obtiene el resultado mostrado en la tabla 5.6.

| | costas | ríos/lagunas | bosques | praderas | montañas | cielos/nubes |
|-----------|---------|--------------|---------|----------|----------|--------------|
| precisión | 81.18 % | 73.48 % | 77.68 % | 93.32 % | 81.71 % | 80.77 % |

Tabla 5.6: Resultados obtenidos para “Identificando la escena consulta” (valores promedio obtenidos de la figura 5.23).

5.3. Resultados obtenidos con la metodología modificada (solamente con puntos aleatorios)

La metodología modificada usa solamente puntos aleatorios (la cual se describe en el capítulo 4), la cual se probó usando tres diferentes clasificadores que son:

- Clasificador 1-NN
- Clasificador Bayesiano
- Red Neuronal Artificial (RNA) y su tipo de aprendizaje es del tipo de “retropropagación”.

La metodología también fue probada usando tres bases de datos validadas por la Comunidad Científica de Visión por Computadora:

- COREL (668 imágenes) con 4 tipos de imágenes diferentes (ver figura(5.24))
- CALTEK (373 imágenes) con 5 tipos de imágenes diferentes (ver figura (5.25))
- TORRALBA (1473 imágenes) con 4 tipos de imágenes diferentes (ver figura(5.26))

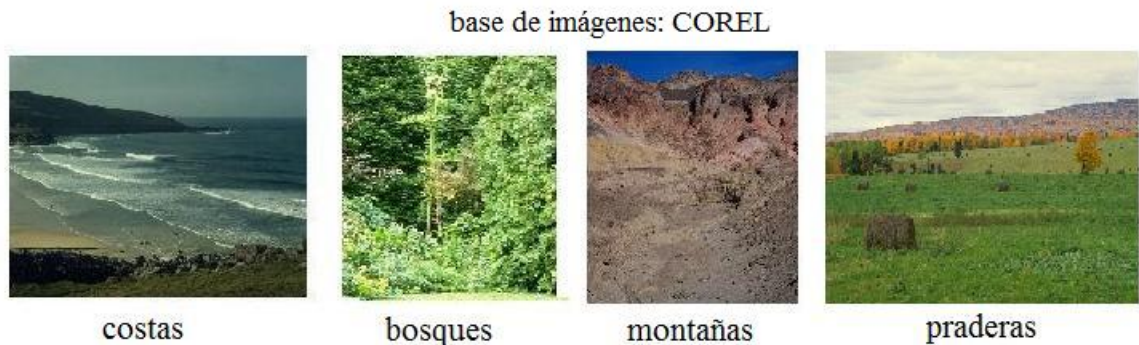


Figura 5.24: Tipos de imagen de la base de imágenes COREL.

5. RESULTADOS EXPERIMENTALES

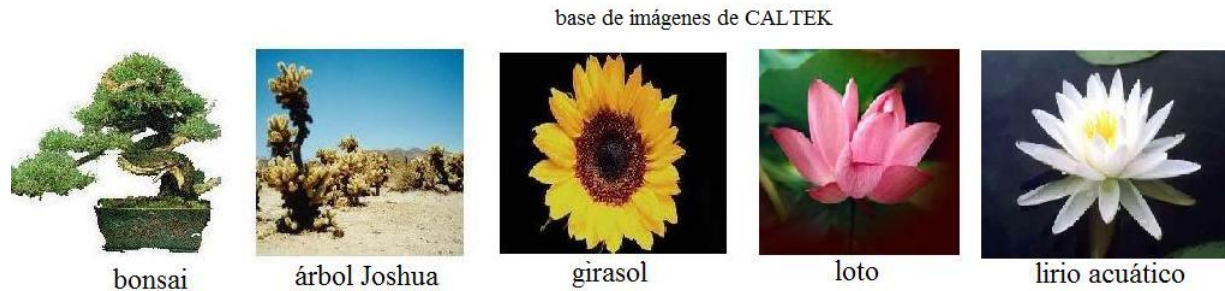


Figura 5.25: Tipos de imagen de la base de imágenes CALTEK.



Figura 5.26: Tipos de imagen de la base de imágenes TORRALBA.

En la tabla (5.7) se muestra a comparación de la eficiencia de cada uno de los diferentes tipos de clasificadores. En dicha tabla se observa que las redes neuronales presentan mejores resultados, salvo en algunos casos aislados, el clasificador 1-NN y el Bayesiano presentan mejores resultados que las redes neuronales.

Para probar la eficiencia de cada tipo de clasificador, se procedió de la siguiente manera: Para la base de imágenes CALTEK se tomaron 5 imágenes aleatorias de cada tipo o de cada clase que conforman a esta base de datos, es decir, 5 bonsais, 5 de árboles Joshua, 5 girasoles, 5 lotos y 5 lirios acuáticos. Para la base de datos COREL se tomaron 12 imágenes

5.3. Resultados obtenidos con la metodología modificada (solamente con puntos aleatorios)

de forma aleatoria de la base de imágenes TORRALBA de cada una de las 4 clases que la conforman, es decir, 12 imágenes de bosque, 12 imágenes de costas, 12 imágenes de montañas, y 12 imágenes de praderas. Finalmente para la base de imágenes de TORRALBA se tomaron 12 imágenes de cada tipo provenientes de la base de datos de COREL, es decir, 12 imágenes de bosque, 12 imágenes de costas, 12 imágenes de montañas y 12 imágenes de praderas.

La configuración de la Red Neuronal Artificial (RNA) cuya regla de aprendizaje es la de retropropagación, quedó definida de la siguiente manera:

Se realizaron muchos experimentos para ver donde se obtenían los mejores resultados, aquí se presentan las tres mejores combinaciones entre las capas de entrada, oculta y salida

- caso 1:

- Capa de entrada: nueve neuronas por tener una dimensión del vector de características de nueve.
- Como capa oculta se tiene una sola capa con diez neuronas, es decir, un arreglo [1, 10]
- Capa de salida diez neuronas por tener diez componentes de imagen (CI) o (subclases) donde se deben clasificar los patrones de entrada.

- caso 2:

- Capa de entrada: nueve neuronas por tener una dimensión del vector de características de nueve.
- Como capa oculta se tiene un arreglo bidimensional de diez neuronas con diez neuronas cada una, es decir, un arreglo [10, 10]
- Capa de salida diez neuronas por tener diez componentes de imagen (CI) o (subclases) donde se deben clasificar los patrones de entrada.

5. RESULTADOS EXPERIMENTALES

| Base de datos/Clasificador | 1-NN | BAYESIANO | RNA (caso 1) | RNA (caso 2) | RNA (caso 3) |
|--------------------------------|---------|-----------|--------------|--------------|--------------|
| CALTEK (373 imágenes) | | | | | |
| bonsai (128) | 94.44 % | 90.97 % | 88.88 % | 86.80 % | 95.13 % |
| árbol Joshua(60) | 60.69 % | 67.91 % | 70.83 % | 66.67 % | 75.20 % |
| girasoles(85) | 81.80 % | 74.30 % | 74.16 % | 75.28 % | 81.25 % |
| lotos(64) | 73.88 % | 78.47 % | 68.94 % | 69.44 % | 66.25 % |
| lirios acuáticos(36) | 76.38 % | 82.50 % | 75.00 % | 75.69 % | 75.00 % |
| promedio | 74.44 | 78.83 % | 75.56 % | 74.77 % | 78.02 % |
| TORRALBA(1471 imágenes) | | | | | |
| bosques(328) | 81.12 % | 83.48 % | 83.50 % | 91.36 % | 86.62 % |
| costas(359) | 68.56 % | 64.32 % | 63.61 % | 74.07 % | 67.51 % |
| montañas(374) | 64.97 % | 67.33 % | 58.33 % | 72.13 % | 73.50 % |
| praderas(410) | 39.64 % | 76.36 % | 59.20 % | 68.02 % | 61.91 % |
| promedio | 63.57 % | 72.86 % | 66.16 % | 76.39 % | 72.38 % |
| COREL(668 imágenes) | | | | | |
| bosques(103) | 74.93 % | 64.30 % | 75.56 % | 82.73 % | 82.51 % |
| costas(255) | 60.70 % | 70.33 % | 81.41 % | 78.73 % | 74.83 % |
| montañas(179) | 68.34 % | 76.69 % | 62.24 % | 69.05 % | 83.24 % |
| praderas(131) | 51.47 % | 67.07 % | 66.32 % | 53.13 % | 63.46 % |
| promedio | 63.86 % | 69.60 % | 71.38 % | 70.91 % | 76.01 % |

Tabla 5.7: Eficiencia de cada uno de los clasificadores para cada base de datos de imágenes.

■ caso 3:

- Capa de entrada: nueve neuronas por tener una dimensión del vector de características de nueve.
- Como capa oculta se tiene un arreglo bidimensional de quince neuronas con quince neuronas cada una, es decir, un arreglo [15, 15]
- Capa de salida diez neuronas por tener diez componentes de imagen (CI) o (subclases) donde se deben clasificar los patrones de entrada.

5.3. Resultados obtenidos con la metodología modificada (solamente con puntos aleatorios)

Para medir la eficiencia a esta modificación a la metodología usamos las medidas 5.1 y 5.2 respectivamente

En las siguientes imágenes: como se puede observar en la base de datos de CALTEK fueron usados 40 imágenes para consulta debido a que en esta base de datos no contiene muchas imágenes para consultar ni desplegar en pantalla. También se puede observar como el número de imágenes comienza a expandirse en las demás bases de datos, y la eficiencia de nuestra propuesta tiende a atenuarse.

A continuación se muestran las gráficas de precisión en la base de imágenes CALTEK. (Ver figuras 5.27 , 5.28, 5.29, 5.30 , y 5.31 respectivamente).

En las figuras (5.32, 5.33, 5.34, 5.35, 5.36, respectivamente) se muestran las gráficas de precisión en la base de imágenes TORRALBA.

En las figuras (5.37, 5.38, 5.39, 5.40, 5.41 respectivamente.) se muestran las gráficas de precisión en la base de imágenes COREL.

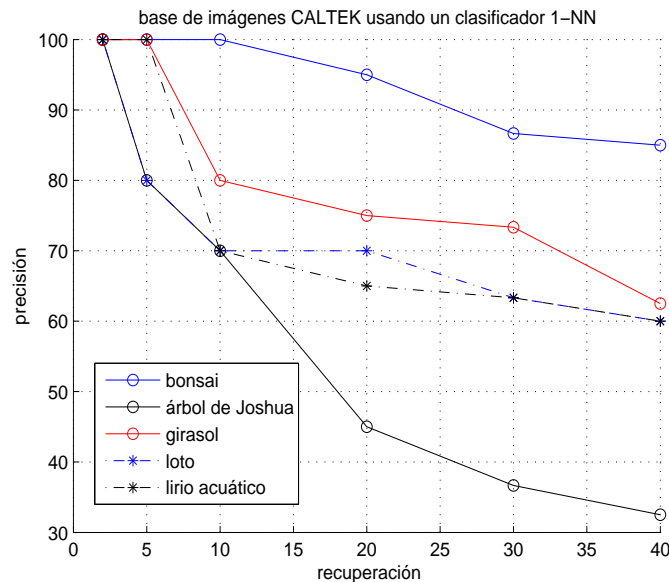


Figura 5.27: Precisión del clasificador 1-NN en la base de imágenes CALTEK.

5. RESULTADOS EXPERIMENTALES

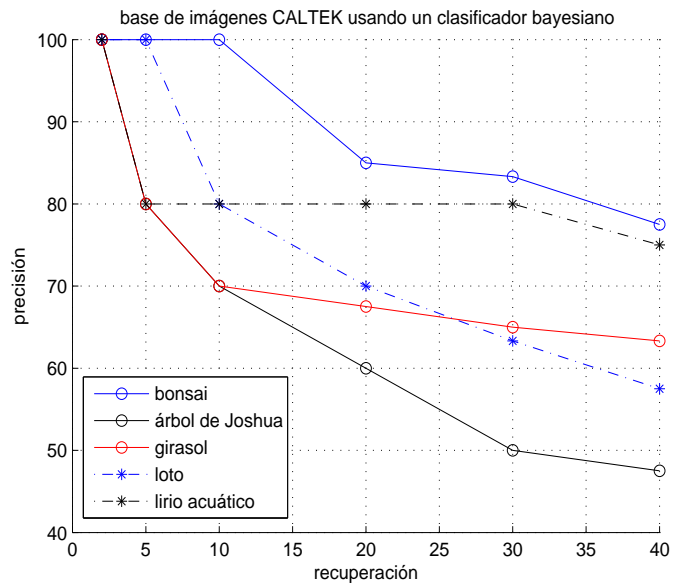


Figura 5.28: Precisión del clasificador Bayesiano en la base de imágenes CALTEK.

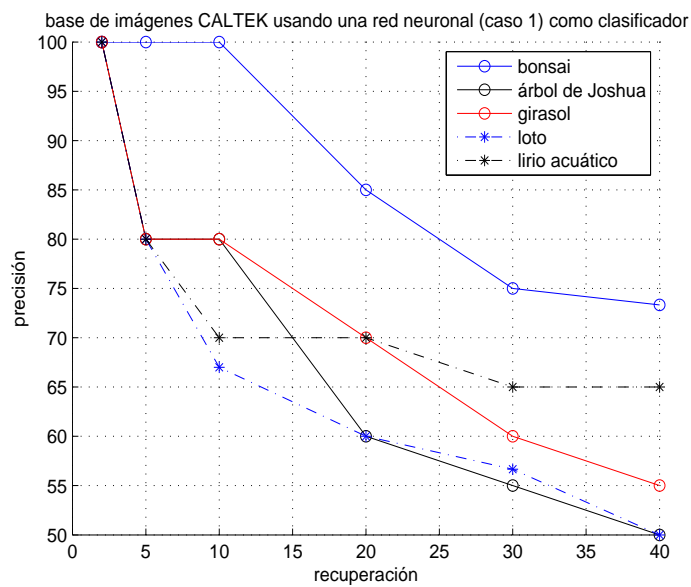


Figura 5.29: Precisión de una red neuronal artificial (RNA) usando la combinación de neuronas del caso 1 en la base de imágenes CALTEK.

5.3. Resultados obtenidos con la metodología modificada (solamente con puntos aleatorios)

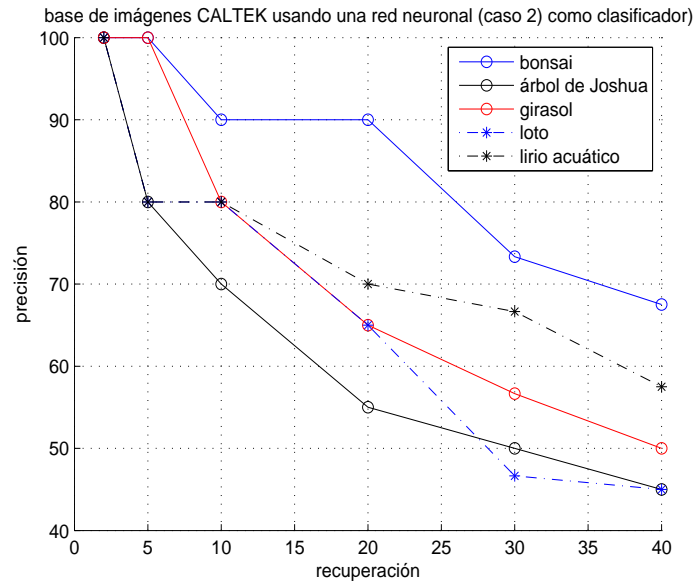


Figura 5.30: Precisión de una red neuronal artificial (RNA) usando la combinación de neuronas del caso 2 en la base de imágenes CALTEK.

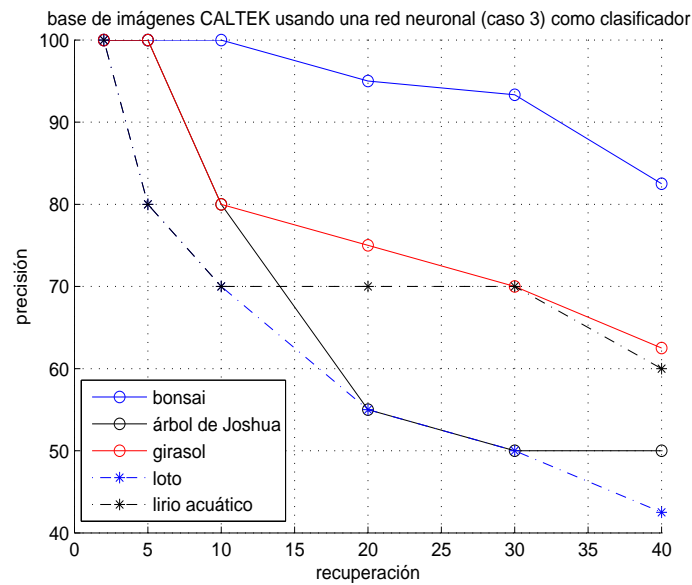


Figura 5.31: Precisión de una red neuronal artificial (RNA) usando la combinación de neuronas del caso 3 en la base de imágenes CALTEK.

5. RESULTADOS EXPERIMENTALES

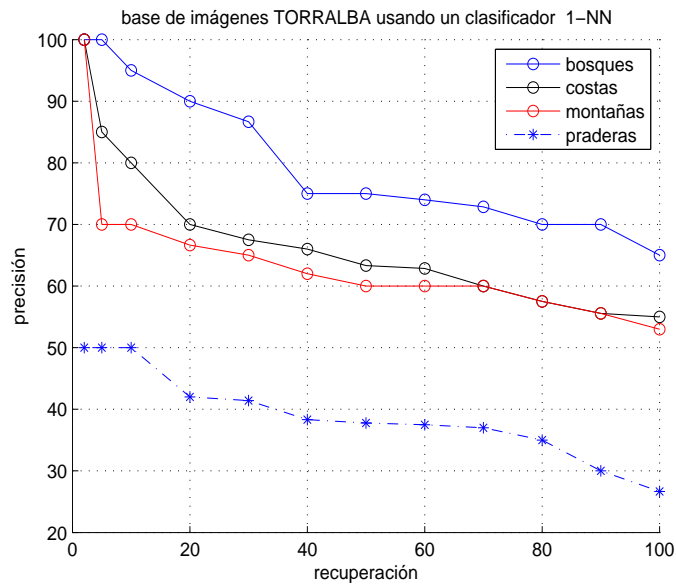


Figura 5.32: Precisión de un clasificador 1-NN en la base de imágenes TORRALBA.

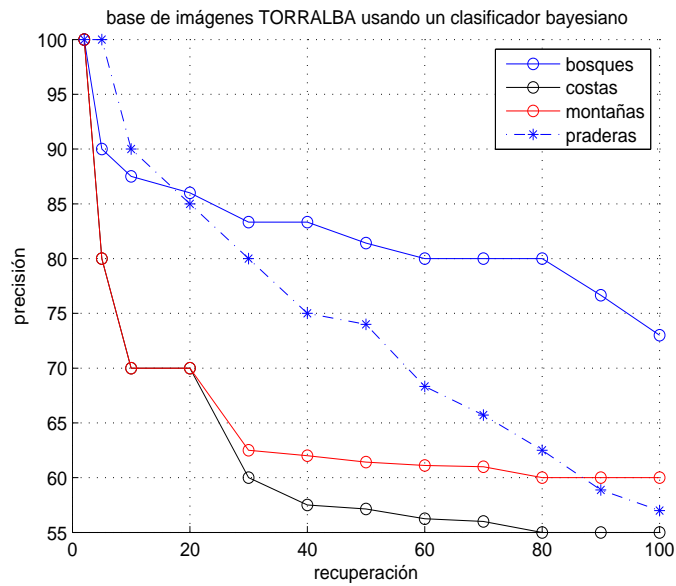


Figura 5.33: Precisión de un clasificador Bayesiano en la base de imágenes TORRALBA.

5.3. Resultados obtenidos con la metodología modificada (solamente con puntos aleatorios)

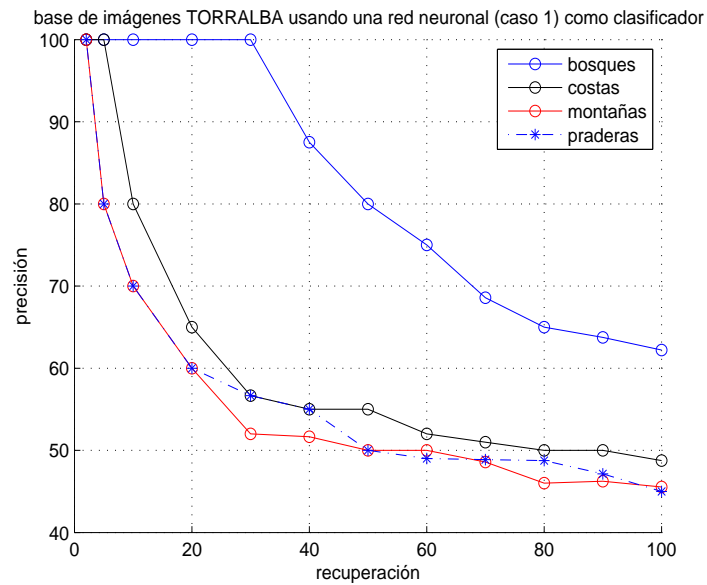


Figura 5.34: Precisión de una red neuronal artificial (RNA) usando la combinación de neuronas del caso 1 en la base de imágenes TORRALBA.

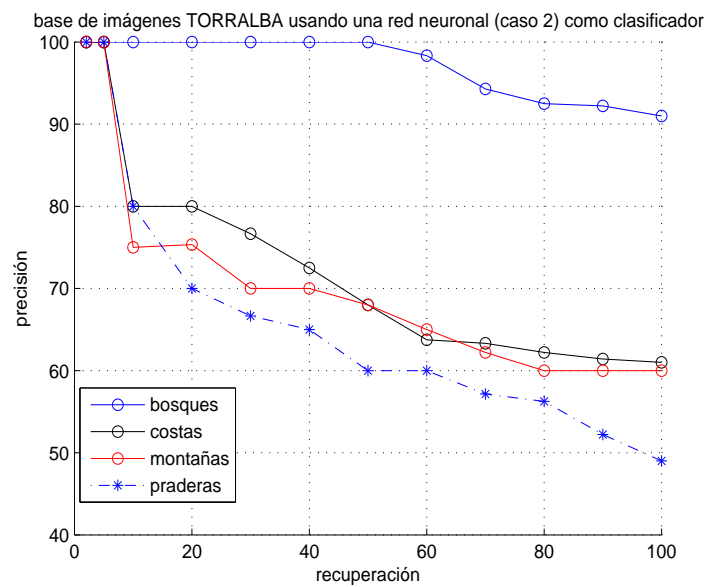


Figura 5.35: Precisión de una red neuronal artificial (RNA) usando la combinación de neuronas del caso 2 en la base de imágenes TORRALBA.

5. RESULTADOS EXPERIMENTALES

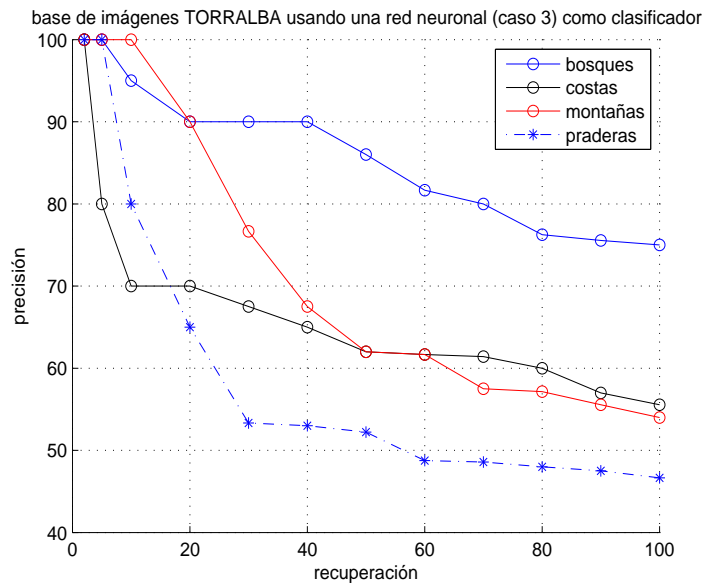


Figura 5.36: Precisión de una red neuronal artificial (RNA) usando la combinación de neuronas del caso 3 en la base de imágenes TORRALBA.

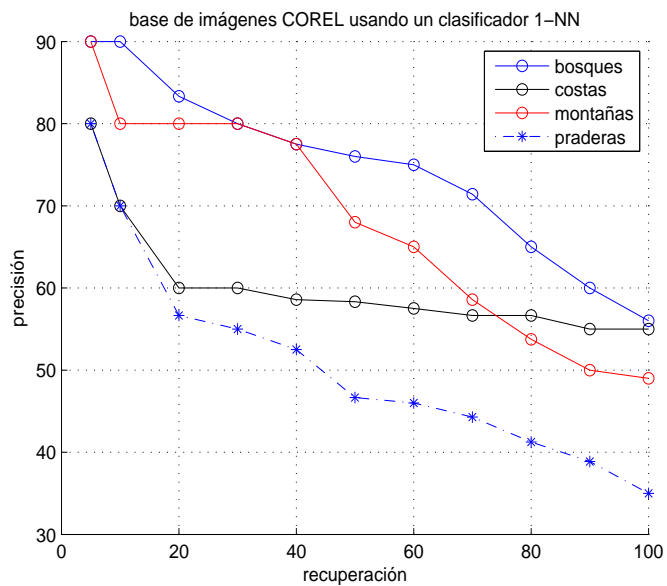


Figura 5.37: Precisión de un clasificador 1-NN en la base de imágenes COREL.

5.3. Resultados obtenidos con la metodología modificada (solamente con puntos aleatorios)

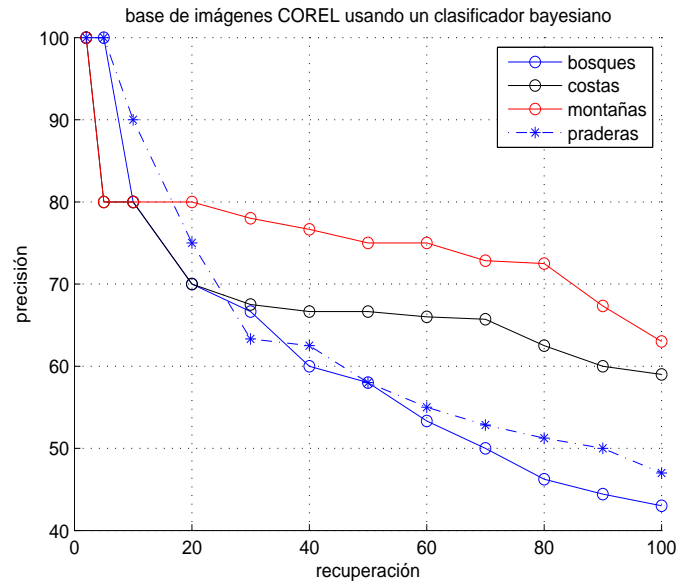


Figura 5.38: Precisión de un clasificador Bayesiano en la base de imágenes COREL.

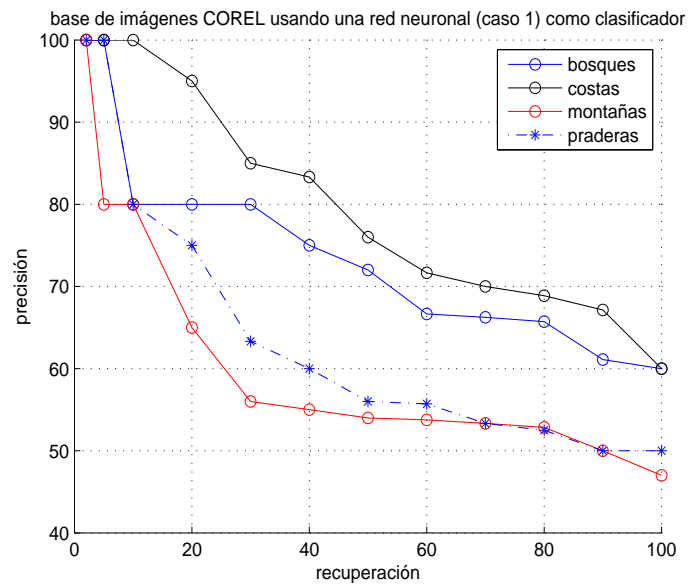


Figura 5.39: Precisión de una red neuronal artificial (RNA) usando la combinación de neuronas del caso 1 en la base de imágenes COREL.

5. RESULTADOS EXPERIMENTALES

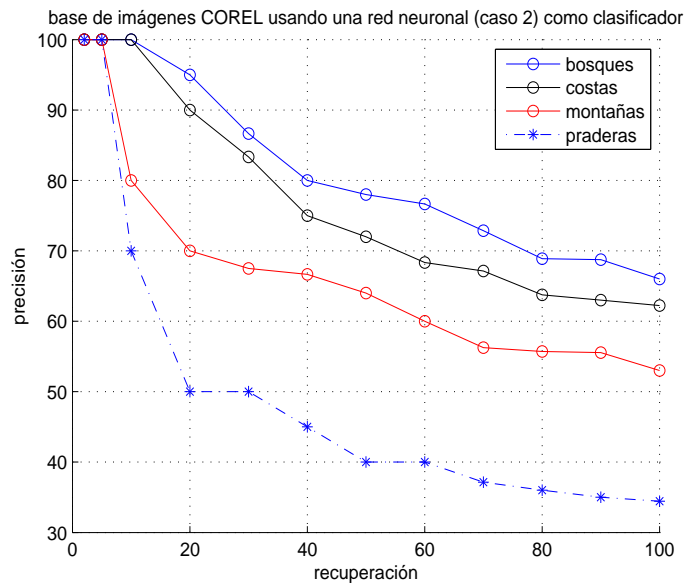


Figura 5.40: Precisión de una red neuronal artificial (RNA) usando la combinación de neuronas del caso 2 en la base de imágenes COREL.

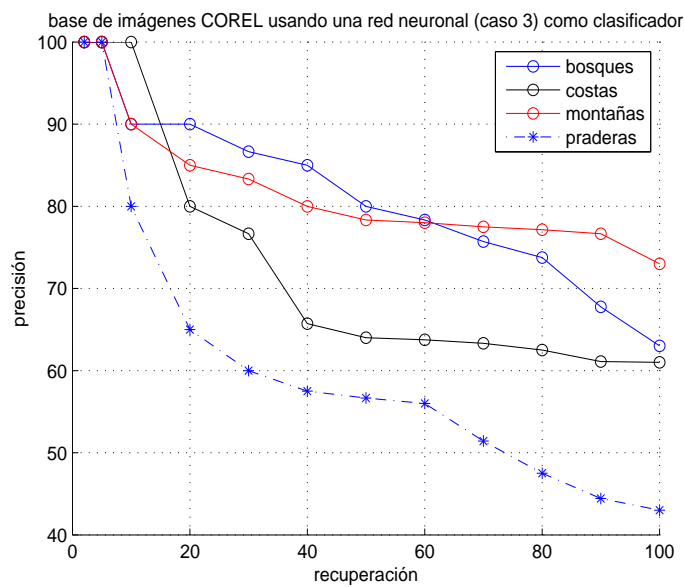


Figura 5.41: Precisión de una red neuronal artificial (RNA) usando la combinación de neuronas del caso 3 en la base de imágenes de COREL.

Capítulo 6

CONCLUSIONES Y TRABAJO FUTURO

En este capítulo se da por un lado, el conjunto de conclusiones a las que se ha llegado después de esta investigación. Por otro lado, se mencionan las acciones a seguir para continuar con futuras investigaciones derivadas del desarrollo de esta tesis.

6.1. Conclusiones

En este trabajo se describe una metodología que permite recuperar automáticamente imágenes de escenarios naturales desde una base de datos de imágenes. Como un resultado adicional, nuestra propuesta también permite la identificación de la imagen consulta.

Durante la etapa del aprendizaje, nuestra propuesta toma como entrada un conjunto de imágenes de escenarios naturales, los cuales están divididos en 6 tipos de imágenes: costas, ríos/lagos, montañas, bosques, praderas y cielos/nubes. Nuestra propuesta extrae desde cada tipo de imagen vectores descriptores usando puntos aleatorios los cuales son seleccionados automáticamente. Se usa el algoritmo de K -Medias para formar diez *clústers*. Se usan tres tipos de clasificadores : el de 1-NN, un bayesiano y uno de una red neuronal artificial para construir una base indexada donde se obtiene un vector descriptivo con información de cada una de las imágenes que conforman el conjunto del entrenamiento.

6. CONCLUSIONES Y TRABAJO FUTURO

Durante la etapa de recuperación, alguno de los clasificadores ya mencionados anteriormente (los cuales ya están entrenados) se usaron para recuperar desde la base de datos indexada las imágenes mas similares dada una imagen consulta. Para poder validar los resultados obtenidos hacemos uso de de las mediciones de precisión (ecuación) 5.1 en la página 107 y recuperación (ecuación) 5.2 en la página 107. y para la identificación de las imágenes consulta hacemos uso de las mediciones de precisión (ecuación) 5.3 en la página 119 e identificación (ecuación) 5.4 en la página 119.

También se probó nuestra metodología con la base de imágenes de la Catedral de SACRE COEUR (París) proporcionada por [14] que permite recuperar automáticamente imágenes de la base de datos indexada de imágenes bajo condiciones similares de iluminación. En la etapa del aprendizaje, nuestra propuesta tomó como entrada un conjunto de imágenes a imágenes bajo condiciones similares de iluminación aplicadas a regiones de cielos brillantes, parcialmente nublados y completamente nublados.

- Una de las ventajas de nuestra propuesta es que no se necesita hacer un etiquetado de las escenas de consulta para poder recuperar escenas desde una base de datos indexada.
- Nuestra metodología es insensible ante las transformaciones que puede sufrir una escena, tales como: rotaciones y cambios de escala.
- Nuestra Metodología puede ser aplicada a la recuperación de imágenes y para la identificación y clasificación de diferentes tipos de imagen.

6.2. Trabajo futuro

Actualmente estamos probando nuestra propuesta con más imágenes y con mas clases de objetos representativos de las imágenes con la idea de formar mas regiones de *clusters*.

También se pretende buscar otro tipo de descriptor y probar con otro tipos de clasificadores para así mejorar aún los índices de recuperación e identificación de las imágenes.

De la misma manera la idea es ir incrementando la base de datos indexada en tiempo real sin tener la necesidad de estar haciendo de nuevo todo el proceso de aprendizaje y además poder hacer búsquedas de imágenes desde el internet y así poder aplicar la recuperación e identificación de la escena.

Capítulo 7

APÉNDICES

7.1. Definición de imagen digital

El término *imagen* se refiere a una función bidimensional de intensidad de luz $f(i,j)$ donde i e j representan las coordenadas espaciales y el valor de f en un punto cualquiera (i,j) es proporcional al brillo (o nivel de gris) de la imagen en ese punto. La figura 7.1 ilustra el convenio de ejes que se utiliza en la representación de una imagen digital. Una imagen digital es una imagen $f(i,j)$ que se ha discretizado tanto en las coordenadas espaciales como en el brillo. Una imagen digital puede considerarse como una matriz cuyos índices de fila y columna identifican un punto de la imagen y el valor del correspondiente elemento de la matriz indica el nivel de gris en ese punto. Los elementos de una distribución digital de este tipo se denominan *elementos de la imagen*, o más comúnmente *pixels*, *píxeles* o *pels*, abreviaturas de su denominación inglesa “*picture elements*” [21].

7. APÉNDICES



Figura 7.1: Convenio de ejes utilizado para la representación de una imagen digital.

7.2. Comprensión de las imágenes

En los trabajos de investigación artificial y de los sistemas CBIR, se han identificado dos vacíos fundamentales cuando se pretende entender o trabajar con una imagen digital:

- **Vacío sensorial:** Se refiere al vacío existente entre el objeto del mundo real y la información capturada por alguno de los métodos de almacenamiento físico [57] y [49]
- **Vacío semántico:** Tiene que ver con la falta de coincidencia entre la información que se puede extraer de los datos visuales y la interpretación que hace el usuario de esos mismos datos [57] y [49]

Esto significa que la información que contiene una imagen digital tiene una gran diferencia con respecto a la forma en la que la misma imagen es visualizada por los humanos en el

mundo real, por la profundidad, iluminación y contraste. El primero de estos problemas es atacado por los investigadores que trabajan para incrementar el poder de los dispositivos de registro digital de imágenes, mejorando la resolución visual o desarrollando nuevos dispositivos de captura tridimensional y en rangos adicionales al espectro visual. El segundo es un problema que afecta más directamente a las aplicaciones CBIR. Para atacarlo se han realizado diferentes aproximaciones que van desde aquellas que clasifican las imágenes argumentando que no es necesario conocer su significado [57], hasta otras que tratan de completar el vacío al utilizar retroalimentación del usuario y minería de datos [49] y [51]

Las investigaciones para lograr un mejor entendimiento sobre las imágenes, son útiles en el caso de los sistemas CBIR para poder representar con mayor precisión el contenido de una imagen. De la representación seleccionada, depende gran parte el trabajo posterior en un sistema CBIR, y en esa representación quedarán encapsuladas las propiedades a las que un usuario tendrá acceso mediante las consultas.

7.3. Representación de las imágenes

Expresar el contenido de una imagen en una forma en la que las computadoras puedan entenderla, de la misma manera que lo hacen los seres humanos es todavía un problema de investigación abierto. Sería necesario algo equivalente a un gran sistema de inteligencia artificial [48] que trabajara como la mente humana con la habilidad de manipular ideas abstractas automáticamente para procesarlas. Esto, por supuesto, no está todavía al alcance de las capacidades de los sistemas actuales [48]. En los sistemas CBIR se utilizan se utilizan las características visuales de la imagen, para representarla y manipularla. La extracción de características de una imagen es un proceso típico en el momento del registro y la consulta. También debe construirse la estructura de representación formada por esas características que depende de la aplicación concreta.

7. APÉNDICES

- **Inteligencia Artificial.** Es una disciplina que se encarga de construir y diseñar procesos que al ser ejecutados sobre una arquitectura física producen acciones o resultados que maximizan una medida de rendimiento basándose en la secuencia de entradas percibidas y en el conocimiento almacenado en tal arquitectura [23].

7.4. Características de una imagen

Las características visuales de una imagen pueden clasificarse de acuerdo a su ámbito como **globales y locales** y su nivel de abstracción como **lógicas o físicas** [49].

1. **Físicas.**-Son aquellas que pueden expresarse cuantitativamente, y se extraen mediante la aplicación de técnicas de tratamiento digital de imágenes. También son llamadas características de bajo nivel [57].

- **Color.**-Se utiliza para describir las distribuciones cromáticas de la imagen, constituyendo el histograma de frecuencias. También son aplicados a imágenes en escalas de grises. Se han propuesto diversos espacios de colores, para poder describirlos de la misma forma como lo percibe el ojo humano.
 - **RGB (Red-Green-Blue).**- Contiene la codificación de los tres colores, de acuerdo a su intensidad en tres componentes. En una misma escena el mismo color puede cambiar en este espacio debido a problemas de iluminación y contrastes.
 - **HSI.**- Intensidad, Saturación y cromaticidad.- Es el más cercano a la percepción humana, pero no es un modelo perfecto.
- **Textura.**- Se encuentra en la categoría de las características locales. Esta puede definirse, en general, como una propiedad de homogeneidad en las regiones de la imagen [24]. Las técnicas para el análisis de textura incluyen: energía, entropía, homogeneidad, contraste, correlación, y otras más [12] y [56]

2. **Lógicas.**-Las características lógicas son también llamadas características de alto nivel. Por lo general contienen información de los objetos en la imagen así como sus relaciones espaciales.

- **Curvatura.**- Puede ser medida tanto en contornos identificados como en una región local de la imagen, utilizando la razón de cambio en la dirección de la línea tangente al contorno o curva.
- **Forma.**- Para la identificación de formas en una imagen existen diferentes aproximaciones o técnicas. Los descriptores elípticos de Fourier son utilizados para describir contornos cerrados en los objetos [38]. También existe segmentación por textura y otras series de técnicas que componen un amplio campo de investigación.
- **Puntos de interés.**-Dado que la identificación de las formas y objetos en una imagen es un problema abierto, se han realizado trabajos para representar la imagen a través de sus puntos de interés [34], reduciendo la complejidad de la imagen y enfocándose en las regiones con mayor interés visual.
- **Posición de las regiones.**- Basado en la identificación previa de las formas, la posición de las mismas es una característica interesante para algunas aplicaciones, la cual puede ser medida como posición absoluta (por cuadrantes) o posición relativa (con respecto a los otros objetos). Al medir la posición debe tenerse en cuenta la robustez frente a las rotaciones de la figura por errores de captura. La posición también incluye relaciones de contención, intersección y solapamiento.

3. **Locales.**- Las características están basadas en las características físicas o de bajo nivel. Estas características pueden medirse respecto a:

- Cada píxel .

7. APÉNDICES

- Una ventana de tamaño fijo.
- Una región previamente identificada.

4. **Globales.**-Son una combinación de características físicas, lógicas y locales. Este tipo de características proporcionan información sobre la totalidad de la imagen, como su tamaño, composición de colores, número de objetos, etc. La extracción de características es el primer paso en todo proceso de recuperación de imágenes. Con la información obtenida en este proceso se construirá la representación de cada imagen que servirá para crear índices, clasificaciones y realizar operaciones de similitud. En general, la estructura de representación contiene la información resumida de la imagen original, pero además de eso contiene datos de mayor interés que simplemente los píxeles ubicados en una matriz. Los criterios de diseño para la estructura dependen del dominio del problema y de la información requerida por el algoritmo de clasificación o recuperación.

Las estructuras utilizadas pueden clasificarse en los siguientes grupos:

- **Vectores n-dimensionales.**- En este tipo de representación cada imagen tiene asociado un vector de n características principalmente visuales. Esta es una de las representaciones más utilizadas principalmente por su simplicidad. Permite combinar varios tipos de características, sin dar mayor preferencia a ninguna de ellas. Las operaciones de similitud o distancia son simples y de baja complejidad. Ejemplos de cómo disponer de diferentes tipos de características en un arreglo n-dimensional pueden encontrarse en [5], etc. Existen otros trabajos como en [49] en donde el vector de características no contiene propiedades visuales sino conceptos semánticos, que han sido deducidos a partir de características físicas y lógicas.
- **Grafos de relaciones con atributos.**- Es una estructura compuesta por arcos y nodos [18]. Los nodos representan objetos en la imagen mientras que los arcos representan

relaciones entre los objetos. Tanto nodos como arcos contienen atributos o etiquetas que corresponden a las propiedades de los objetos o relaciones respectivamente. Es una estructura poderosa, porque permite no solamente las características de la imagen sino también la forma en la que están relacionadas las características. Contiene una mayor representación semántica y representa el contenido a un nivel de abstracción menos visual pero mas conceptual. Esta estructura no se usa mucho en aplicaciones CBIR porque requiere el resultado de una segmentación conceptual para representar objetos en la imagen, lo cual es todavía un problema de investigación abierto

- **Otras representaciones.**-Existen otros métodos de representación que permiten comparar el contenido de las imágenes, aunque algunos de ellos no son frecuentemente utilizados por su complejidad computacional o por la falta de información que representa para ciertas técnicas.

7.5. Fundamentos del color

El uso del color en el procesamiento de imágenes está motivado por 2 factores fundamentales. En primer lugar, el color es un potente descriptor que a menudo simplifica la identificación y extracción de objetos de una escena. En segundo lugar, el ojo humano puede distinguir una amplia gama de colores comparado con una imagen en niveles de gris [23].

En el tratamiento de imágenes el procesamiento de las imágenes en color se divide en dos áreas fundamentales: *color* propiamente dicho y *pseudocolor*. En la primer categoría se procesan las imágenes obtenidas con un sensor de color o multiespectral, mientras que en la segunda las imágenes monocromáticas son coloreadas por asignación de un color a un determinado nivel de intensidad [23].

7. APÉNDICES

7.5.1. Modelos de color

El propósito de un modelo de color es facilitar la especificación de los colores de alguna forma estándar. Los modelos de color en las imágenes están orientados hacia el hardware (monitores en color e impresoras) y una amplia gama de video cámaras como lo es el modelo RGB(Rojo, Verde, Azul); para impresoras a color se aplica el modelo CMY (Cyan, Magenta y Amarillo). Y en procesamiento de imágenes se usan principalmente el modelo RGB y el HSI (Matiz, Saturación e Intensidad).

- **Modelo RGB.** Del Inglés (Red,Green,Blue), que significa rojo,verde,azul.- Este modelo hace referencia a la composición de un color en términos de la intensidad de los colores primarios con que se forma el rojo, verde y azul. Es posible representar un color mediante la mezcla por adición de los tres colores primarios. No define por sí mismo lo que significa exactamente rojo, verde o azul, por lo que los mismos valores RGB pueden mostrar colores notablemente diferentes en diferentes dispositivos que usen este modelo de color. Cada color primario se codifica con un código de un byte (8 bits), por lo que la intensidad de cada una de las componentes se mide en una escala que va desde el cero al 255. Por lo que el rojo se obtiene con (255,0,0), el verde con (0,255,0) y el azul con (0,0, 255) obteniendo en cada caso un color monocromático, por lo que el color negro se obtiene con (0,0,0) y el blanco con (255,255,255). La combinación de dos colores a nivel 255 con un tercero en nivel 0 da lugar a 3 colores intermedios, de esta manera surgen los colores: amarillo(255,255,0), Cyan (0,255,255) y el Magenta(255,0,255). El conjunto de todos los colores se puede representar en forma de cubo como se observa en la figura (7.2) [23].
- **Modelo HSI.** Del Inglés (Hue, Saturation,Intensity) que significa Matiz,Saturación e Intensidad respectivamente.

- El matiz es un atributo del color que describe su pureza (amarillo puro, rojo, naranja) mientras la saturación proporciona una medida del grado en que el color puro es diluído con luz blanca. El modelo de color HSI debe su utilidad a dos hechos fundamentales: Primero la componente de Intensidad I, se puede separar de la información del color en la imagen. Segundo las componentes de matiz y saturación están íntimamente relacionadas con el modo en que los seres humanos perciben el color. Estas características hacen del modelo HSI una herramienta ideal para desarrollar algoritmos de procesamiento de imágenes basados en algunas de las sensaciones de color del sistema visual humano [23].
- Algunos ejemplos de la utilidad de este modelo varían desde el diseño de sistemas para verificar el grado de madurez de las frutas o la inspección del acabado de color de determinados productos. La idea consiste en utilizar las propiedades del color del mismo modo que lo haría un ser humano encargado de esta misión.
- Las componentes de color del modelo HSI (matiz y saturación) están definidas con respecto al triángulo mostrado en la figura 7.3(a), dicha figura es un triángulo en el que no existe perspectiva alguna y todos los puntos en él son coplanarios, el punto W es un punto de intersección de las medianas del mismo.
- Combinando matiz, saturación e intensidad en un espacio de color 3-D se obtiene la representación piramidal de la figura 7.3(b). Cualquier punto en la superficie de la estructura piramidal representará un color saturado puro, ya que los puntos de los lados de los triángulos representan colores desaturados. El matiz de dicho color se determina por su ángulo con respecto al eje rojo y su intensidad por su distancia perpendicular desde el punto correspondiente al negro (esto es cuanto mayor es la distancia desde el negro mayor es la intensidad del color)

7. APÉNDICES

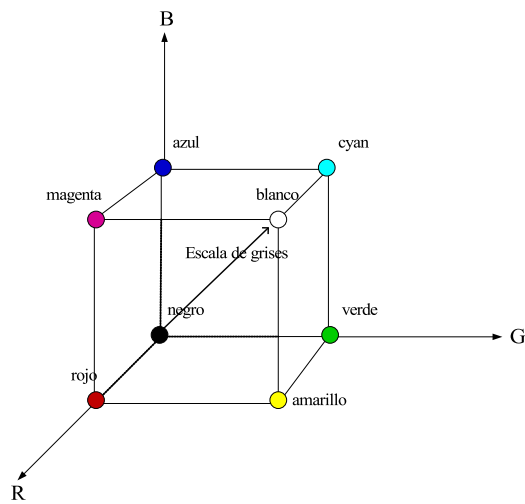


Figura 7.2: Cubo de color RGB. Los puntos a lo largo de la diagonal principal tienen valores de gris, desde el punto negro en el origen al punto blanco.

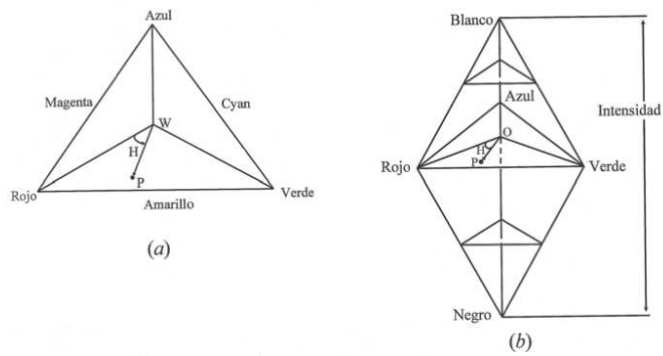


Figura 7.3: Triángulo de color en el modelo HSI. Color e Intensidad en el modelo HSI.

Referencias

- [1] *El Pequeño Larousse ilustrado*. Ediciones Larousse, S.A de C.V. Dinamarca num 81. México D.F., 2006.
- [2] Acton.Scott-T, Soliz-Peter, Russell.Stephen, and Pattichis.Marios-S. Content based image retrieval: The foundation for future case-based and evidence-based ophthalmology. *Digital Object Identifier 10.1109/ICME.2008.4607491*, pages 541–544, April 2008.
- [3] Acton.S.T., Rossi.A., and C.L.Brown. Matching and retrieval of tattoo images: Active contour cbir and local image features. *Image Analysis and Interpretation, 2008. SSIAI 2008. IEEE Southwest Symposium on 24-26 March 2008 Page(s):21 - 24*, pages 21–24, 2008.
- [4] A.Del-Bimbo. A perspective view on visual information retrieval systems, content based access of image and video libraries. *IEEE, Workshop on volume 21 IEEE 1998.:108–109*, 1998.
- [5] A.J.M. Content based image retrieval using approximate shape of objects. *17 th IEEE Symposium on Computer-based medical Systems (CBMS)*, 2004.
- [6] A.J.Maren, H.Ritter, and K.SChulten. *Handbook of Neuronal Computing Applications*. 1990.

- [7] Alain.C.Gonzalez-Garcia. *Image Retrieval Based on the Contents*. PhD thesis, Center for Research in Computing (CIC)-IPN, Mexico DF, September 2007.
- [8] Anna.Bosch, Xavier.Muñoz, and Robert.Marti. Which is the best way to organize/classify images by content? Technical report, Department of Electronics Informatics and Automaticas, University of Girona, Campus Montilivi, 2006.
- [9] A.W.Smeulders, M.Worring, S.Santini, A.Gupta, and R.Jain. Content based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22 No. 12:1349–1380, 2000.
- [10] Barb and A.Shyu. Semantics modeling in diagnostic medical image databases using customized fuzzy membership functions. *FUZZY SYSTEMS. THE 12 th IEEE INTERNATIONAL CONFERENCE*, 2, 2003.
- [11] C.Avilés-Cruz. *Analyse de Texture par Statistiques D ' Ordre Supérieur: Caractérisation et Performances*. PhD thesis, Instituto Nacional Politécnico de Grenoble-Francia, 1997.
- [12] C.C-Gotlieb and H.E.Kreyszig. Texture descriptors based on matrices. *Computer Vision, Graphics, and Image Processing*, 51, 1990.
- [13] Darío.Maravall.Gómez-Allende. *Reconocimiento de formas y visión artificial*. Addison -Wesley Iberoamericana, 1997.
- [14] Mauricio Díaz and Peter Sturm. Finding images with similar lighting conditions in large photo collections. *CIARP 2009. LNCS 5586, Springer*, pages 53–60, 2009.
- [15] D.B.Russakof. Image similarity using mutual information. *ECCV*, 3, Prague Czech republic, 2004.
- [16] P.R. Devijver and J. Kittler. *Pattern Recognition, A Statistical Approach*. Prentice Hall, New York, 1982.

-
- [17] Richar O. Duda and P.E. Hart. *Pattern Classification ans scene Analysis*. Wiley, New York., 2000.
- [18] E.G.Petrakis. Similarity searching in image databases. *IEEE Transactions on knowl- edge and Data Engineering*, vol 14, pp 1187-1201, (2005), 14:1187–1201, 2005.
- [19] K. Fukunaga. *Introduction to statistical Pattern Recognition*. Academic Press New York, 1990.
- [20] Elías García-Santilán. Detección y clasificación de objetos dentro de un salón de clases empleando técnicas de procesamiento digital de imágenes. Master's thesis, Universidad Autónoma Metropolitana, Mayo de 2008.
- [21] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1992.
- [22] Woods Eddins Gonzalez. *Digital Image Processing using Matlab*. Prentice Hall, Boston, MA, USA, 2004.
- [23] Gonzalo.Pajares.M. and Jesús.M.de.la-Cruz-García. *Visión por Computadora (Imá- genes Digitales y aplicaciones)*. Alfaomega Grupo Editor, S.A. de C.V., 2008.
- [24] H.Tamura, S. Mori, and T.Yamawaki. Texture features corresponding to visual per- ception,. *IEEE Trans.on Sys. Man and Cyb . SMC-8(6)*, 1978.
- [25] I.J.Sumana, Md.M.Islam, D.Zhang, and G.Lu. Content based image retrieval using curvelet transform. *IEEE 10th Workshop on Multimedia Signal Processing*, Pp. 11- 16. 8-10 Oct 2008, pages 11–16, 2008.
- [26] Qasim Iqbal and J.K.Aggarwal. Cires. a system for content-based retrieval in digi- tal image libraries. In *Seventh International Conference on Control, Automatization Robotics and Vision (ICARCV) Singapore pp 205-210*, 2002.

- [27] I.Rigoutsos and R.Hummel. A bayesian approach to model matching with geometric hashing. *Computer Vision and Image Understanding*, 62, 1995.
- [28] J.H.Sossa-Azuela. *Rasgos descriptores para el reconocimiento de objetos*. Instituto Politécnico Nacional, 2006.
- [29] Jensen J.R. *Introductory Digital Image Processing (Second Edition)*. Prentice Hall, 1996.
- [30] J.Villegas-Cortez. Identificación de tipos de letra. Master's thesis, Universidad Autónoma Metropolitana (Unidad Azcapotzalco), Junio 2005.
- [31] M.L. kherfi and D.Ziou. Image retrieval from the world wide web: Issues, techniques, and systems. *ACM Computing Surveys*, 36, Num 1:35–67, March 2004.
- [32] T. kohonen. An introduction to neuronal computing. *Neuronal Networks*, 1:3–16, 1988.
- [33] Lehmann.TM, Güld. MO, Thies.C, Fischer.B, Spitzer.K, Keysers.D, Ne.H, Kohnen.M, and Schubert Hand Wein.BB. Content-based image retrieva in medical applications. *Methods of Information in medicine*, 43(4):354–361, 2004.
- [34] Leonardo-Trujillo and Gustavo-Olague. Synthesis of interest point detectors throught genetic programming. *Genetic and Evolutionary Computation Conference (GECCO) Seattle EUA*,, 1:887–894., Julio 8-12 (2006).
- [35] J. Li and J. Z. Wang. Real-time computerized annotation of pictures. *Proceedings of the 14th annual ACM international conference on Multimedia Pp. 911-920, 2006.*, pages 911–920, 2006.
- [36] F. Long, H.J.Zhang, and D.Feng. Fundamentals of content image retrieval, in multimedia information retrieval and management. D Feng Eds, Springer 2003., 2003.

-
- [37] Marceau.D.J., Howarth., P.J.Dubois J.M., and Gratton.D.J. Evaluation of grey level co-occurrence matrix method for land classification using spot imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 28,Num 4:513–519, 1990.
- [38] M.Carlin. measuring the performance of shape similarity retrieval methods. *Computer vision and Image understanding, Vol 84, 2001*, 84, 2001.
- [39] Presutti.M. *Co-currency Matrix in Multispectral Classification: Tutorial for Educators textural measures. The 4th day Educacao em Sensoriamento Remoto Ambito not do Mercosul - 11 to August 13, 2004 - Sao Leopoldo RS. Brazil.*
- [40] P.S.Hiremath and J.Pujari. Content based image retrieval using color, texture and shape features. . *15th International Conference on Advanced Computing and Communications pp. 780-784, 2007.*, pages 780–784, 2007.
- [41] R.Hecht-Nielsen. Neurocomputing: Picking the human brain. *IEEE SPECTRUM Artificial Neural Networks: Theoretical Concepts"(V.Vemuri ed)*, 25:36–41, March 1988.
- [42] R.M.Haralick, K Shanmugan, and I.Dinstein. Texture features for image classification. *IEEE Trans. on Sys Man. and Cyb. SMC-3(6) 1973.*, 1973.
- [43] Y. Rui, Th.S.Huang, and Sh F.Chang. Image retrieval: Currente techniques, promissing directions, and open issues. *Journal of Visual Communication and Image Representation 10, 39-62, 1999.*, 10:39–62, 1999.
- [44] C. Schmid. Weakly supervised learning of visual models and its application to content-based image retrieval. *International Journal of Computer Vision*, 56, no. 12:7–16, 2004.

- [45] Shokoufandeh.A., Macrini.D., Dickinson.S., Siddiqi.K., and Zucker.S.W. Indexing hierarchical structures using graph spectra. *IEEE Trans Pattern Anal Mach Intell*, 27:1125–1140, 2005.
- [46] M. Stricker and M.Orengo. Similarity of color images, storage and retrieval for image and video databases. 1995.
- [47] Thomas.Desealaers, Daniel.Keysers, and Hermann.Ney. Fire. flexible image image retrieval engine. *Image CLEF Evaluation, C: Peters et al (Eds.) CLEF 2004 LNCS 3491, Springer-Verlag Berlin Heidelberg*, pages 688–698, 2005.
- [48] J. Vogel. *Semantic Scene Modeling and Retrieval PhD Thesis*. PhD thesis, Swiss Federal Institute of technology Zurich. Zurich Germany, 2004.
- [49] J. Vogel, A.Schwaninger, C. Wallraven, and H. H.Bülthoff. Categorization of natural scenes: Local vs. global information. *Proceedings of the Symposium on Applied Perception in Graphics and Visualization (APGV06), 33-40. ACM Press, New York, NY, USA (07 2006)*., pages 33–40, 2006.
- [50] J. Vogel and B. Schiele. Semantic modeling of natural scenes for content-based image retrieval. *International Journal of Computer Vision*, 72(2), pages 133–157, 2007.
- [51] J. Vogel and B. Schiele. Performance evaluation and optimization for content-based image retrieval. *Pattern Recognition*, 39(5):897–909, May 2006.
- [52] X.He, R.S.Zemel, and M.A.Carreira-Perpin. Multiscale conditional random fields for image labeling. *Proc. IEEE CS Conference Computer Vision and Pattern Recognition*, 2:695–702, 2004.
- [53] X.He, R.S.Zemel, and D Ray. Learning and incorporating top down cues in image segmentation. *Proc. IEEE CS Conference Computer Vision*, 1:338–351, 2006.

-
- [54] José Ramón Hilera y Víctor Martínez. *Redes Neuronales Artificiales. Fundamentos, modelos y aplicaciones*. Alfaomega Grupo Editor, S.A. de C.V. Pitágoras Num 1139, Col. del Valle. C.P.03100 MÉXICO D.F., 2000.
- [55] Yan.Gao, Kap.Luk-Chan, and Wei-Yun-Yau. Learning in content based image retrieval - a brief review. *10-13 Dec. 2007 Page(s):1 - 5 Digital Object Identifier 10.1109/ICICS.2007.4449869*, pages 1–5, 2007.
- [56] Y.Deng and B.S.Manjunath. Unsupervised segmentation of color-texture regions in images and video. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI '01)*, 23(8):800–810, Aug 2001.
- [57] Y.Liu, D.Zhang, and et al. A survey of content-based image retrieval with high-level semantics. *Pattern Recognition 40:262-282, 2007.*, 40:262–282, 2007.