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
M en C José Félix Serrano Talamantes

DIRECTORES:
Dr. Juan Humberto Sossa Azuela
Dr. Carlos Avilés Cruz.

México D.F. Enero de 2011

SIP-14 BIS

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	**TALAMANTES**	**JOSÉ FÉLIX**
Apellido paterno	Apellido materno	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. Oleksiy Pogrebnyak

Dr. Ricardo Barrón Fernández

PRESIDENTE DEL COLEGIO DE PROFESORES

Dr. Luis Alfonso Villa Vargas

INSTITUTO POLITECNICO NACIONAL
CENTRO DE INVESTIGACION
EN COMPUTACION
DIRECCION

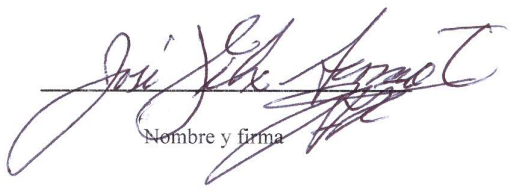
INSTITUTO POLITÉCNICO NACIONAL

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: La extracción de características es un problema clave cuando hablamos de la recuperación de las imágenes sobre la base de su contenido. Se han propuesto desde hace algunos años los descriptores de texturas. En este trabajo se propone una metodología para extraer y clasificar características aplicada a la recuperación de las escenas naturales. La propuesta consiste en usar puntos aleatorios como entrada de un clasificador 1-NN con el propósito de verificar que tan discriminantes son las características de la media, la desviación estándar y la homogeneidad proveniente de una matriz de co-ocurrencia para describir las diferentes clases de objetos presentes en una escena. También se propone el uso del algoritmo de las k-medias de forma no supervisada con el fin de encontrar grupos o clústeres que no estén correlacionados de tal manera que los objetos presentes en una escena no estén asociados con las etiquetas que un observador les inserta a las imágenes de escenarios naturales para describir su contenido.

Abstract: Feature extraction is a key issue in Content Based Image Retrieval (CBIR). In the past, a number of textures have been proposed in literature, including statistic methods. In this work is proposed an extraction and features classification methodology, applied to scenes retrieval of natural images. The proposed Methodology uses random points which are input to a 1-nn classifier with the purpose of testing how discriminating the mean, standard deviation are and homogeneity (from a co-occurrence matrix) features combination to describe different classes in a scene. It also proposes that the use of-K-means algorithm to find unsupervised groups or clusters (uncorrelated) that exist in a natural scene and the objects in scene are not associated with the labels normally a user makes an image to describe the contents.

Índice general

Índice de figuras

Índice de cuadros

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 y 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 acceso a 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 llamativos del ser humano [24].

Para que un sistema artificial pueda interactuar eficientemente con el medio que lo rodea, como lo hace el ser humano, es necesario que cuente con las capacidades adecuadas de adquisición y análisis automático de la información que recibe [24].

¿Cómo es posible que una computadora pueda realizar millones de cálculos por segundo y no sea capaz de reconocer una simple imagen e identificarla como un coche, un escenario natural, una persona, etc?. El enfoque que se pretende en esta tesis consiste en utilizar técnicas y herramientas de la computación para que una

computadora pueda reconocer patrones de la imagen y con ello realizar el proceso de la recuperación de imágenes.

Aún con la tecnología actual, no existen buscadores eficientes mediante imágenes, los hay y muchos de ellos para texto, como Google, Yahoo, Lycos, Altavista, Infoseek, etc. Cuando el usuario busca este tipo de información (imágenes) con la manera descrita, los buscadores le devuelven muchas imágenes, tal vez muchas de ellas de las que no esté buscando, inclusive no mostrándole mucha de a información, ya que el texto asociado a la imagen no va acorde al contenido de la misma, haciéndole perder una parte importante de tiempo y de recursos; por ejemplo, si el usuario teclea la palabra banco" los buscadores responderán con imágenes de banco de asiento, banco de animales o una institución bancaria, siendo ésta última la clase de imagen que realmente está buscando. Con este tesis, se tratará de evitar este tipo de problema, dándole un enfoque lo más específico posible, reconociendo los objetos locales de la imagen buscada y descargando imágenes asociadas a las de la imagen consulta.

El manejo de información involucra a menudo el reconocimiento, almacenamiento, tratamiento y recuperación de imágenes e información multimedia [5] y [36].

Aunque una gran cantidad de información multimedia se genera de forma continua para una variedad de aplicaciones, los sistemas de información actuales no son capaces de procesar la información multimedia de una forma eficiente, debido a que estos sistemas han sido diseñados para funcionar con datos simbólicos y estructurados.

La recuperación de imágenes se refiere a buscar y recuperar información visual en forma de imágenes, dentro de una colección de bases de datos de imágenes [3]y [36].

Los medios electrónicos actuales de almacenamiento, así como gran cantidad de imágenes que se almacenan en éstos, inducen al desarrollo de sistemas de información automatizados para la recuperación de imágenes.

Debido a lo anterior, se observa un incremento en el desarrollo de los sistemas de recuperación de imágenes, gracias a:

- El desarrollo de sistemas integrados multimedia con algoritmos de almacenamiento, compresión, procesamiento y recuperación de imágenes, así como de sistemas integrados de propósito general con funciones multimedia.

- Las mejoras en metodologías de desarrollo de programas y de estándares para el manejo efectivo de las imágenes.

- Los avances en comunicación digital, tales como la fibra óptica, el modo asíncrono de transferencia y otras tecnologías de redes de alta velocidad, que permiten anchos de banda mayores que hacen posible la transmisión y la entrega eficiente de imágenes.

Mientras que para el ser humano no presenta dificultad en reconocer y recuperar datos multimedia [4], los sistemas de información actuales presentan varios problemas, debido a que en lo fundamental están diseñados para procesar la información de tipo alfanumérica, y algunas veces son expandidos con herramientas de desarrollo gráfico y con simples técnicas de diagramación y dibujo[5]. Por lo tanto, hasta ahora, no hay muchos sistemas que se hayan sido diseñados enfocados hacia las tareas de reconocimiento y recuperación de imágenes de forma eficiente.

Mediante esta tesis, se pretende desarrollar un sistema de información visual que utilice paradigmas orientados particularmente al procesamiento de la información visual en imágenes con escenas de tipo natural (nada hecho por el hombre) específicamente mediante la organización y recuperación sobre la base de su contenido, en términos del color y la textura así como la clasificación mediante métodos bayesianos, métodos estadísticos, redes neuronales u otros.

1.1. Planteamiento del problema.

En general, el problema de la recuperación de imágenes consiste en: dada una imagen consulta I_c, extraer de un banco de imágenes aquellas "mas parecidas"a I_c sobre la base de su contenido. Esto se logra a comparar la imagen consulta con cada una de las imágenes del banco de imágenes..

Dicha comparación se puede hacer píxel a píxel. 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 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 función bidimensional de intensidad luminosa f(x,y) la cual se considera como una matriz de elementos cuyos índices de fila y columna identifican a un píxel de la imagen, (x, y) el cual representa las coordenadas espaciales, y el valor de f es un píxel cualquiera (x, y) es proporcional al brillo, ya sea en niveles de gris, o al color compuesto en sus componentes RGB ó HSI.

Sea f(x, y) donde $x, y = 1, 2 \cdots N$ es el arreglo en píxeles de una imagen en dos dimensiones. 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 ó HSI. Si la información en color se representa en términos de los tres colores primarios RGB

(rojo, verde y azul), la función imagen se describe como se muestra en la ecuación (1.1).

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

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

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

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

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

(I).- representa al brillo y está relacionado con la cromaticidad de la intensidad.

El modelo HSI es el modelo que más se asemeja al sistema visual humano (SVH), mientras que el modelo RGB se aplica más a los monitores a color y a cámaras de video[44].

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

$$F : f \to X \tag{1.3}$$

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

$$d(F(q), F(f_r)) \leq d(F(q), F(f)) \tag{1.4}$$

para todo f ϵ B, f$\neq f_r$

O expresado en otros términos:

Dadas p imágenes, $I_1, I_2, I_3 ... I_p$ con p>>0 pertenecientes a un conjunto heterogéneo B, extraer de B un subconjunto Br limitado de imágenes, dada una consulta q formulada en términos de un grupo de rasgos globales y locales. Ver figura 1.1.

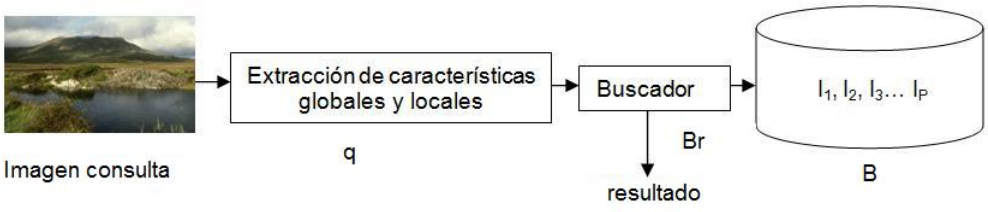


Figura 1.1: Diagrama a bloques del modelo para aplicar "recuperación de imágenes

Para el desarrollo de esta implementación se plantean las siguientes interrogantes:

- ¿Qué rasgos se deben considerar para describir el contenido de una imagen?
- ¿Cómo se convierte una parte de una imagen en rasgos para que mediante estos rasgos se pueda descomponer una imagen en sus partes?
- ¿Cómo se estructura o se diseña un diccionario indexado para organizar las imágenes descritas?

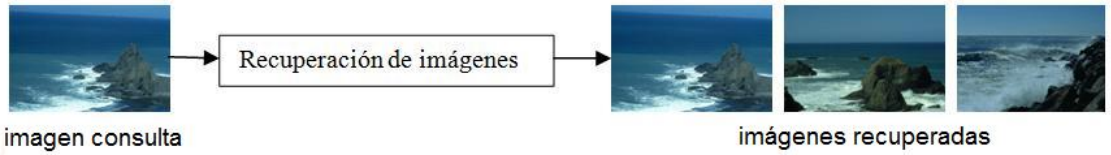


Figura 1.2: Imagen consulta, la cual es presentada al módulo de la recuperación de imágenes. A la salida de éste se muestran las imágenes más parecidas a la imagen consulta.

- ¿Qué criterios deben tener las imágenes consulta para extraer las imágenes del diccionario indexado?

Para ilustrar lo anterior con un ejemplo, en la Figura1.2 se presenta una imagen consulta de una escena costera, la cual se pretende recuperar. En dicha Figura se presentan las tres imágenes recuperadas; la primera de ellas es la escena consulta.

1.2. Justificación.

Aproximadamente el 73 % de información circulante en Internet se encuentra en forma de imágenes [3]. Esta información, en general, no se encuentra bien organizada ni estructurada. En Internet podemos encontrar imágenes de todo tipo: gente, 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, entonces hablamos de un problema de investigación abierto todavía.

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.

La selección de características constituye una decisión importante a tomar en

cuenta en el proceso de investigación, que exige un mejor entendimiento de las imágenes con el fin de desarrollar una metodología enfocada a la organización y búsqueda de un cierto tipo de imágenes con una buena medida de precisión, razón por la cual, esta metodología debe ser orientada hacia un conocimiento específico en el área de aplicación.

1.3. Identificación del problema.

La manera clásica de indexar imágenes 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 que representa cada imagen de acuerdo al contexto de cada aplicación [5]. 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 extenso de imágenes, se desea implementar una metodología para recuperar imágenes que permita a los usuarios encontrar imágenes similares o iguales a partir de una imagen "consulta", es decir, dada 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 la imagen de entrada.

1.3.1. Objetivo general.

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

1.3.2. Objetivo 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 permita procesar el contenido visual de la misma.

3. Implementar una interfaz de usuario que permita presentar una imagen consulta al sistema para poder recuperar un subconjunto de imágenes similares de acuerdo al contenido de la imagen consulta.

1.4. Aportaciones.

La metodología combina puntos aleatorios y puntos fijos 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: la media, la desviación estándar[18] y la homogeneidad, ésta proviene de la matriz de co-ocurrencia [37] y [33]. Estas rasgos son extraídos de una sub-imagen usando los canales H, S e I [19] , [20] y [21]. Se propone el uso del algoritmo K-medias [14],[18], [15] y el clasificador 1-NN (k-ésimo vecino más cercano), [26], [9],[18], y [15] En este caso K=1. Todos estos elementos se usan para construir una base de datos indexada de 700 imágenes [45],[44],[46] y [47] logrando las siguientes aportaciones:

1. Se realiza la recuperación de imágenes sin necesidad de describir o etiquetar el contenido de las escenas consulta.

2. Se realiza la recuperación de imágenes, mediante 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 (700 imágenes).

4. Al realizar una consulta, se puede identificar de forma paralela a la recuperación, la identificación de la escena de forma automática.

1.5. Organización de la tesis.

Este documento está organizado de la siguiente manera: El capítulo 1 se presenta el problema a resolver en esta tesis "la recuperación de imágenes".. En el capítulo 2 se da una descripción del estado del arte. En el capítulo 3 se presenta el marco teórico de referencia de las herramientas que se utilizaron para resolver el problema planteado, entre ellas el algoritmo de las K-Medias, El clasificador K-NN (el vecino más cercano). En el capítulo 4 se expone y se detalla la metodología propuesta. En el capítulo 5 se exponen y se presentan los resultados experimentales y su discusión. En el capítulo 6 se exponen las conclusiones y los trabajos futuros. Finalmente se presenta la relación de las referencias utilizadas en este documento.

Capítulo 2

ESTADO DEL ARTE

En este capítulo se presenta un breve, pero útil estado del arte de los trabajos mas importantes relacionados con la investigación presentada en esta tesis.

2.1. Estado del Arte

2.1.1. Antecedentes

Aproximadamente el 73 % de información en Internet se encuentra en forma de imágenes [3]. Esta información, en general, no se encuentra bien organizada ni estructurada. En Internet podemos encontrar imágenes de todo tipo: 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. 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.

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. Por eso, el desarrollo de esta metodología debe estar orientado por el conocimiento específico en el área de aplicación.

El indexado de las imágenes para manejar grandes volúmenes de información es otra de la 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 [53].

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 carácterísticas de las imágenes que –dependiendo de la aplicación- pueden estar en términos estadísticos, matriciales, histogramas, vectores, etc.

Otras aplicaciones [38] realizan consultas a partir de regiones identificadas por una serie de puntos que aproximan zonas interesantes en imágenes en tomografías computarizadas.

- QBIC [53] y[38] .- 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. Su características de textura es una versión mejorada de la representación de textura de Tamura [22].

- VIRAGE [38].- 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 [38] y [27].- 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 [38] y [27].-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 [27].-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 [27].-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 [27].-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 [5].-Propone en su trabajo hacer la recuperación de imágenes mediante imágenes. Mediante la transformada Wavelet Daubechies 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 3 bandas (RGB) del color de una imagen porque es el mas comúnmente usado.

Usan histogramas para la extracción del color. Debido a que el histograma no aporta suficiente información acerca de la posición de los pixeles, 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 [39].-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 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. Se usaron 4 clases de prueba.

- Julia Vogel en [44] y [47].-Su trabajo está orientado a la recuperación de imágenes, pero recupera las imágenes con base en las anotaciones 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 Cada región de la imagen le extrae histogramas de HSI.

- J.Li et al en [29].- 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. J. Li describe que la IEEE ha creado 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 de 2000 se publicó un artículo donde hacen mención que el etiquetado de imágenes tiene inconsistencias al momento de realizar una búsqueda de imágenes [7] y propone una recuperación/búsqueda de imágenes mediante el contenido de la misma.

- El modelo de campo aleatorio en [50] 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 [49]

- 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 [8]. 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). El asignador de la semántica se diseñó a partir de una máquina de aprendizaje para generar un espacio métrico-semántico en el que las distancias conceptuales entre las imágenes se pueden calcular.

- Con respecto a la clasificación de los escenarios [6] se tienen ciertas técnicas de aproximación: 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 escenario a consultar. Sin embargo, esto resulta poco efectivo ya que aunque hace una distinción correcta de 2 tipos de imagen (ciudad, paisaje, interiores, exteriores, etc), no hace una clasificación mas específica (selva, bosque, pradera, etc). La problemática de la clasificación es resuelta mediante la teoría de decisión de Bayes. Cada Imagen es representada por un vector característico extraído de la misma imágen. 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.

2.1.2. Introducción.

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 [53] y [38], el QBIC [53] y [38] 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 [53] y [38] 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 2 vacíos fundamentales: sensorial y semántico.

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 [30],[34],[51] y [1]. 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.

2.1.3. Definición del problema en general.

El objetivo de los sistemas de Recuperación de Imágenes por Contenido (CBIR) consiste en administrar grandes cantidades de datos multimediales en aplicaciones concretas [34]. En muchos campos de trabajo de hoy en día, se tiene una creciente tasa de almacenamiento de imágenes, haciéndose necesario ordenar, organizar, clasificar y sistematizar 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.

2.1.4. Entendimiento de imágenes

En los trabajos de investigación artificial y de los sistemas CBIR, se han identificado 2 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 [53] y [45]

- **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 [53] y [45]

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 [53], hasta otras que tratan de completar el vacío al utilizar retroalimentación del usuario y minería de datos [45] y [48]

Las investigaciones para lograr un mejor entendimiento sobre las imágenes, es útil 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.

2.1.5. 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 [39] 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. No solamente es importante la evaluación de un sistema en términos del tiempo de ejecución, sino también con respecto a la precisión para recuperar registros correctamente clasificados. En [45] y [39] 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.

2.1.6. 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 no son confiables y en pocos registros se encuentran completamente disponibles. En [44] se realizan anotaciones automáticas a par-

tir 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 [53] y [41]. 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 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 rep-

resentarse 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.

2.1.7. 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 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 [44]. 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.

2.1.7.1. 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 [45].**

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 [53].

- **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 3 colores, de acuerdo a su intensidad en 3 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 genera, como una propiedad de homogeneidad en las regiones de la imagen [22]. Las técnicas para el análisis de textura incluyen: energía, entropía, homogeneidad, contraste, correlación, y otrás más [10] y[52]

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 [32]. 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 [43], 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 contenencia, 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 .
   - 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 [4], etc. Existen otros trabajos como en [45]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 [16]. 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 re-

quiere 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.

2.1.8. 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 definida en el dominio de la representación de las imágenes. Estas medidas tratan de identificar, con respecto a las características que conforman la representación, que tan parecida es una imagen a otra.

Para encontrar imágenes relevantes en una base de datos, el usuario debe estar interesado en 2 tipos de resultados: ubicar una imagen objetivo o navegar por categorías de imágenes similares. Las técnicas de recuperación 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:

2.1.8.1. 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 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 [2]. 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 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 3 grupos:

- Emparejamientos determinísticos
- Emparejamientos probabilíticos
- Emparejamientos heurísticos

1. **Emparejamiento 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 [18], [9], [26], 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 [13].

2. **Emparejamientos probabilísticos.**- Los métodos probabilìsticos de emparejamiento, miden la similitud de 2 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. **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. Se utiliza una función heurística 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.

2.1.8.2. 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 den 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 prepocesamiento 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 búsquesa está dividido en:

- Métodos Clásicos.
- Aprendizaje Computacional.

1. **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 indexamiento para bases de datos relacionales, espaciales o geográficas. También existen tendencias de indexamiento traídas desde técnicas de recuperación de textos tanto determinísticos como no determinísticos. Otros métodos organizan los registros por la probabilidad de que un usuario requiera justo 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, eigenvalores de las matrices de adyacencia de los grafos [40].

b) **Probabilísticas.-** los Métodos probabilísticas utilizan información probabilística 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 [48] se utiliza un modelo Bayesiano para construir una tabla Geométrica basada principalmente en las propiedades visuales de la forma.

2. **Aprendizaje computacional.-** Estas técnicas tratan de encontrar patrones 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 [45] 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 [9].

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.

d) **Otros.-** En [44] se presenta un modelo de aprendizaje maquinal, para construir un índice de orientación en la base de datos de imágenes. En [39] se aplica un algoritmo de cuantificación vectorial para agrupar imágenes con puntos de interés.

En general las técnicas de emparejamiento no deberían ser utilizadas como único medio para localizar imágenes objetivo. Es muy deseable contar con un espacio de búsqueda estructurado, que permita reducir los cálculos aún si se desea realizar una comparación exhaustiva.

2.1.9. 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 sucientes 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 recuperación óptima de resultados relevantes para el usuario.

2.1.9.1. CIRES

CIRES[35] es un sistema en línea de recuperación de imágenes basado en su contenido que combina los pricipios 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.

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

Cuadro 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.

2.1.9.2. FIRE

FIRE [42] 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 cuaquier 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 fué 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.

2.1.9.3. IRMA

IRMA [28] es un sistema de recuperación de imágenes de radiografías. Su objetico es recuperar imágenes de las etapas del terapia del mismo paciente o también recuerar imágenes con dignóstico 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 imágenes de abdomen, mano, seno, cráneo, torso, y columna vertebral. En general logra una eficiencia de precisión de 87.5 %.

Capítulo 3

MARCO TEÓRICO

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

3.1. Marco Teórico.

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 pre-procesamiento 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 pre-procesamiento 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 con una combinación de puntos fijos, el uso del algoritmo de K-medias, un clasificador K-nn, y la distancia euclidiana como criterio

de comparación para ver que tan similar es una escena 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 [53]. Básicamente un píxel se puede describir por medio de dos coordenadas: $p = p(x, y)$ en 2D en donde $x, y \epsilon \mathbb{Z}$ Los puntos son identificados dentro de una imagen digital en forma de un pixel 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.

3.2. Reconocimiento de patrones

El reconocimiento de patrones, es útil para la identificación de formas, de figuras, objetos, etc. Es un proceso fundamental que se encuentra en casi todas las acciones humanas. Un sistema automático de reconocimiento de objetos (SARP) 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 [24].

La cadena de pasos en un sistema de recocimiento de patrones es:

1. **Ente a reconocer.**- Es el objetivo a reconocer, que puede ser: algún tipo de señal, una base de datos de imágenes, un cultivo, alguna enfermedad, etc.

2. **Pre-procesamiento en el dominio del ente.-** Se hace un tipo de pre-procesamiento 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), dicho en otras palabras, tiene como objetivo mejorar la calidad de la imagen para futuros tratamientos.

3. **Extracción de características.-**aplica operadores sobre una imagen permitiendo identificar la presencia de un objeto en una escena. Los rasgos utilizados por el sistema dependen del tipo de objetos a ser identificados o reconocidos.

   - **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 [24].

4. **Procesamiento en el dominio de las características.-** Permite eliminar información redundante y reducir la dimensionalidad de trabajo . Si este módulo funciona bien se producen 2 cosas:

   a) una alta clasificación.

   b) reducción de tiempo de cómputo.

5. **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 [9].

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

6. **Evaluación del desempeño.-** Mediante una matriz de confusión se evalúa el porcentaje de 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 entramiento. De este modo, la información contenida en las muestras de entrenamiento, que corresponde a experiencias pasadas puede utilizarse para responder a cuestiones sobre 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 depedencias. 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 aprendizaje. El objetivo del aprendizaje no supervisado es estimar la distribución de la probabilidad de las entradas o descubrir la estructura natural de los grupos o clústers en los 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 cuantas son.

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

Las técnicas de la minería de datos provienen de la Inteligenica 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 técnicas destacan las técnicas de:

- **Agrupamiento o de Clusterizado.-** Es un procedimiento de agrupación de una serie de vectores según criterios habitualmente de distancia; se trata de disponenr los vectores de entrada de tal forma que estén mas cercanos aquellos que tengan características comunes, entre ellos destaca el algoritmo de ***K*-means ó *K*-medias.**

3.3. Algoritmo de *K*-medias

En nuestro caso. decidimos probar con el de K-Medias. Es un algoritmo sencillo, y muy eficiente siempre que el número de cases se conozca a priori con exactitud.

El agrupamiento de las muestras se efectúa al minimizar un índice de dispersión. Para este algoritmo no hay un umbral por definir, sin embargo, hay que fijar a priori el número de grupos o clases a realizar, es decir, se fija los k grupos a encontrar.

El procedimiento es el siguiente:

- **Paso 1.-** Se establece previamente el número exacto de clases existentes, digamos k se escogen al azar entre los elementos a agrupar k vectores, de forma que van a constituir los centroides (al ser los únicos elementos) de las k clases, es decir:

$$C_1 : Z_1(1);\ C_2 : Z_2(1)...C_k(1) \tag{3.1}$$

en donde se ha introducido entre paréntesis el índice iterativo de este algoritmo

- **Paso 2.-**Como se trata de un proceso recursivo con un contador n, en la iteración genérica n se distribuyen todas las muestras $\{X\}1 \leq j \leq p$ entre las k clases, de acuerdo con la siguiente regla:

$$X \epsilon C_i(n) si ||X - Z_j(n)|| < ||X - Z_i(n)|| \forall i = 1, 2...K \; donde \; i \neq j \tag{3.2}$$

en donde se han indexado las clases (que son dinàmicas) y sus correspondientes centroides.

- **Paso 3.**- Una vez redistribuidos los elementos a agrupar entre las diferentes clases, es preciso recalcular o actualizar los centroides de las clases. El objetivo en el cálculo de los nuevos centroides es minimizar el índice de rendimiento siguiente:

$$J_i = \sum_{X \epsilon C_i(n)} si ||X - Z_i||^2; \; i = 1, 2....K \tag{3.3}$$

Este índice se minimiza mediante la media muestral o aritmètica de $C_i(n)$:

$$Z_i(n+1) = \frac{1}{N_i(n)} \sum_{X \epsilon C_i(n)} X; \; i = 1, 2...K \tag{3.4}$$

siendo $N_i(n)$ el nùmero de elementos de la clase C_i en la iteración n.

- **Paso 4.-** Se comprueba si el algoritmo ha alcanzado una posición estable, es decie si cumple:

$$Z_i(n+1) = Z_i(n) \; \forall i = 1, 2....K \tag{3.5}$$

Si se cumple, el algoritmo finaliza, de lo contrario regresa al paso 2.

En la figura 3.1 podemos ver el diagrama de flujo del algoritmo de las k-medias

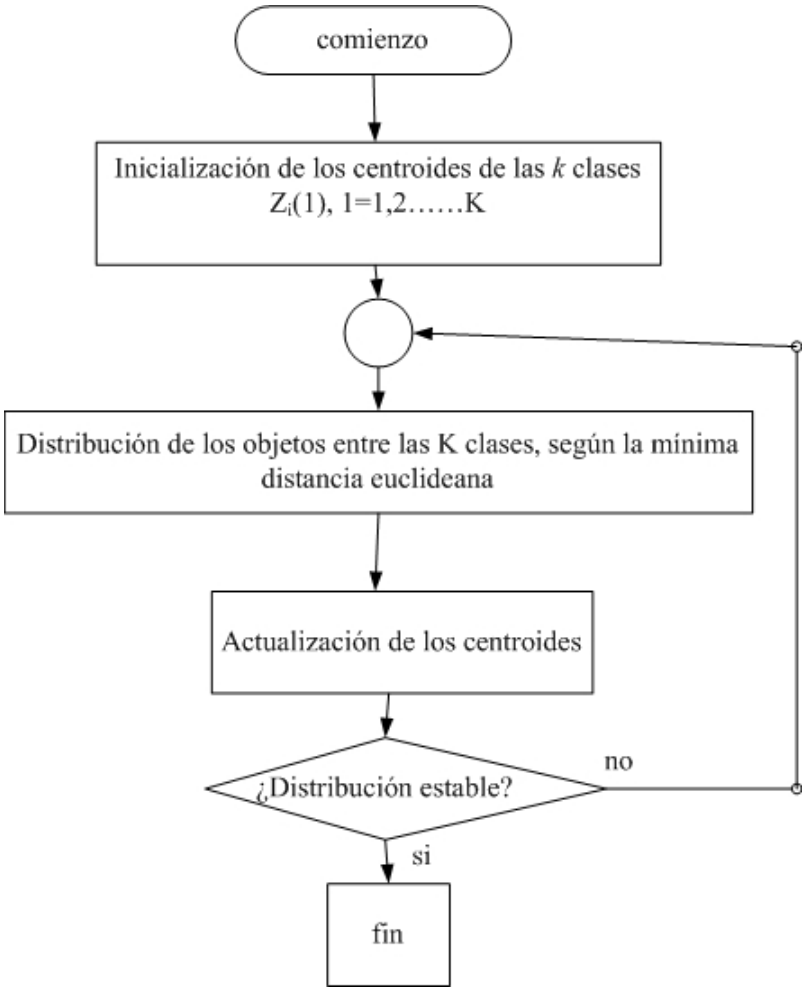


Figura 3.1: Diagrama de flujo de las k-medias

3.4. Matriz de co-ocurrencia.

3.4.1. 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 encontarar un solo descriptor de textura, ya que existen varios problemas asociados a ellos [17]. 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 [37] propuso 14 medidas de textura basadas en la dependencia espacial de los tonos de grises. En [25] sugiere variables de textura basadas en estadísticas de primer orden (media, desviación estándard, varianza), estadísticas de segundo orden basadas en la matriz de co-ocurrencia 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 co-ocurrencia espacial (o de niveles de gris) que son calculadas en una dirección específica o bien para todas ($0^0, 45^0, 90^0$,y 135^0) entre los píxeles vecinos dentro de una ventana móvil dentro de la imágen.

. 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 [33].

El modelo matemático más común para medir la textura es la matriz de co-ocurrencia de niveles de grises **(GLCM) (Grey Level Co-ocurrence Matrix),** basado en estadísticas de segundo orden. Es un histograma de los niveles de grises 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 corroborrado que los datos texturales conjuntamente con los datos espectrales se mejora la precisión de la clasificación [31].

3.4.2. Concepto de la matriz de co-ocurrencia.

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 co-ocurrencia es una matriz

cuadrada 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 imágen I(x,y), 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 co-ocurrencia 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 co-ocurrencia es un resumen de la forma en que los valores de los píxeles ocurren al lado de otro valor en una pequeña ventana.

3.4.2.1. Cálculo de la matriz de co-ocurrencia

Para ilustrar la manera en que se calcula la matriz de co-ocurrencia se presenta un ejemplo el cual se muestra en la figura 3.2. Tomando como base la matrix de la figura (1) con una distancia de un píxel d=1 y direcciones a 0^0determinemos la matriz de co-ocurrencia.

Como esta matriz únicamente contiene tres niveles de gris (0,1,y 2) se crea una matriz de 3x3 para cada rotación. El cálculo de la matriz de co-ocurrencia para cada dirección se muestra en las figuras 3.3, 3.4 y 3.5 respectivamente.

la cual se representa como una imagen de preuba donde los valores corresponden a niveles de gris. La imagen tiene 4 píxeles de lado y niveles de grises:0,1,2 y 3.

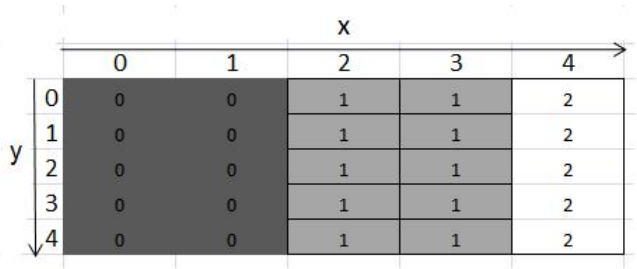


Figura 3.2: Imagen con tres niveles de gris

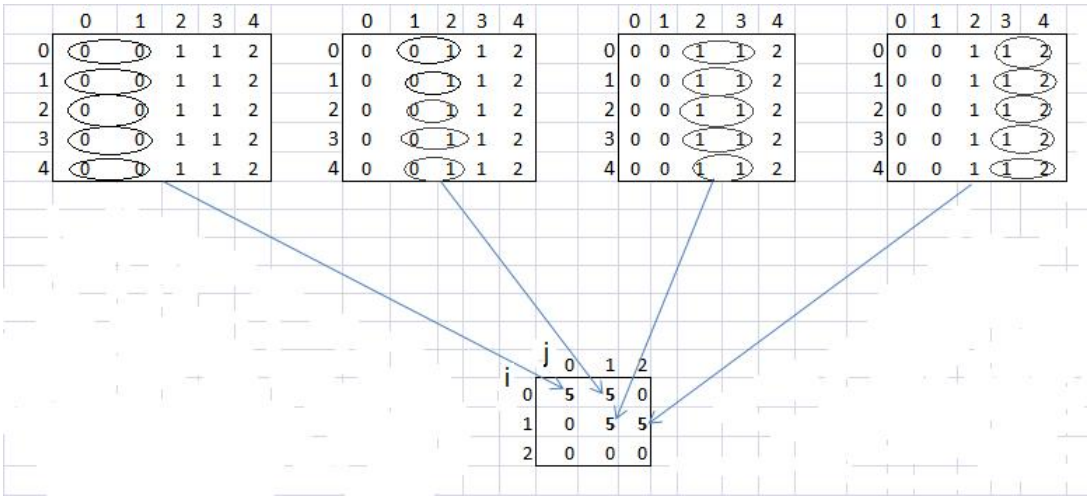


Figura 3.3: Matriz de co-ocurrencia para d=1 a 0^0

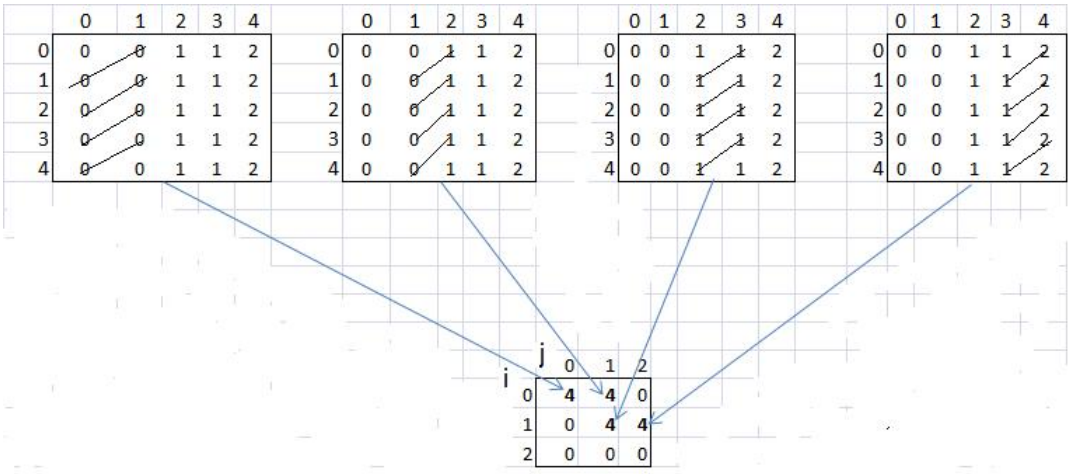


Figura 3.4: Matriz de co-ocurrencia para d=1 a 45^0

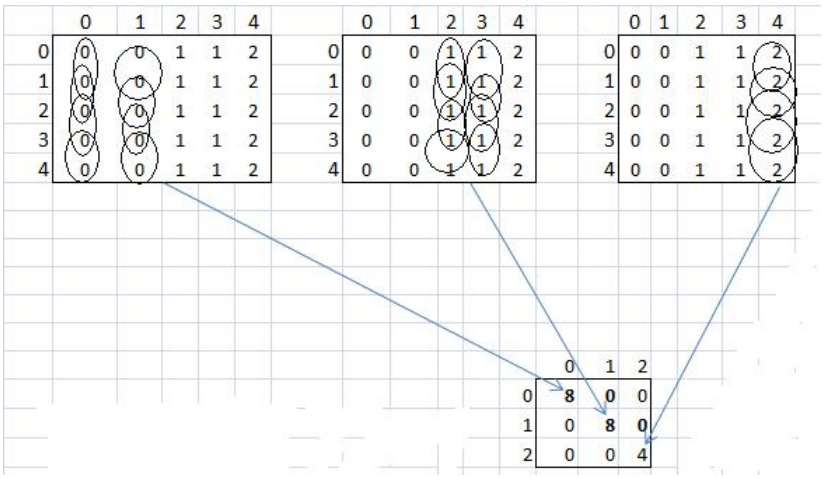


Figura 3.5: Matriz de co-ocurrencia para d=1 a 90^0

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

A continuación se presentan algunos métodos de detección de rasgos de textura que se calculan a partir de una matriz de co-ocurrencia.

Una vez obtenida la matriz de co-ocurrencia el siguiente paso es expresar esta matriz como una probabilidad. La definición más simple de probabilidad es: es número de veces que ocurre un evento, dividido por el número total de posibles eventos y la ecuación para su cálculo es:

$$P_{i,j} = \frac{V_{i,j}}{\sum_{i,j=0}^{N-1} V_{i,j}} \tag{3.6}$$

donde:

$\boldsymbol{i}$ es el número de filas y $\boldsymbol{j}$ es el número de columnas

$\boldsymbol{V}$ es el valor valor de la celda $\boldsymbol{i}$, $\boldsymbol{j}$ en la ventana

$P_{i,j}$es la probabilidad de la celda $\boldsymbol{i}$, $\boldsymbol{j}$

$\boldsymbol{N}$ es el número de las fias y columnas.

3.4.3. Estadísticas de primer orden.

Las medidas texturales de primer orden son calculadas a partir de los valores del nivel de gris originales de la imagen y su frecuencia, como la media, y la desviación estándard. En estas medidas no se considera la relación entre los píxeles.

- Media.-Es el cálculo de la media aritmética de los valores de grises de los píxeles de una ventana. Se calcula mediante las siguientes ecuaciones

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.7}$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{3.8}$$

- Desviación Estándar.-La varianza de un conjunto de mediciones $y_1, y_2, y_3 ... yn$ es la media de cuadrado de las desviaciones de las mediciones con respecto a su media. Simbólicamente la varianza de una muestra está dada por:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 \tag{3.9}$$

A mayor varianza de un conjunto de mediciones, corresponde una mayor variación dentro del conjunto. La varianza es útil en la comparación de una variación relativa de dos conjuntos de mediciones, pero solo aporta información con respecto a la variación en un solo conjunto cuando se interpreta en términos de la desviación estándard.

Las ecuaciones para el cálculo de la varianza se muestran en las ecuaciones 3.10 y 3.11 y dan el mismo resultado tanto para i como para j, porque la matriz es simétrica.

$$\sigma_i^2 = \sum_{i,j=0}^{N-1} P_{i,j}(i - \mu_i)^2 \tag{3.10}$$

$$\sigma_j^2 = \sum_{i,j=0}^{N-1} P_{i,j}(j - \mu_j)^2 \tag{3.11}$$

Mientras que las ecuaciones para el cálculo de la desviación estándard que a continuación se muestran en las ecuaciones 3.12 y 3.13 respectivamente

$$\sigma_i = \sqrt{\sigma_i^2} \tag{3.12}$$

$$\sigma_j = \sqrt{\sigma_j^2} \tag{3.13}$$

Las medidas texturales de primer orden son calculadas a partir de los niveles de gris originales de la imagen y su frecuencia como es la media, varianza y desviación estándar. En estas medidas no se considera la relación entre píxeles. Para este trabajo solamente se usó la media y la desviación estándar como rasgos descriptores.

3.4.4. Estadísticas de segundo orden.

Son las medidas que consideran la relación de co-ocurrencia entre grupos de dos píxeles de la imagen origina y una distancia dada.

- **Homogeneidad** .- Proporciona información sobre la regularidad local de la textura. Mide la cercanía o contigüidad de la distribución de elementos en la matriz de co-currencia con respecto a la diagonal principal, es decir, este descriptor aumentará cuando la distancia i-j sea mínima. Se calcula mediante la siguiente ecuación

$$\sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2} \tag{3.14}$$

siendo $P_{i,j}$ la probabilidad de co-ocurrencia de los valores de gris (i,j), para una distancia dada.

- **Contraste.-** Es una medida de la variación brusca local de color en una imagen. El valor del contraste aumentará si existen más elementos de la matriz de co-ocurrencia 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. Se calcula mediante la siguiente ecuación.

$$\sum_{i,j=0}^{N-1} P_{i,j}(1-j)^2 \tag{3.15}$$

- **Energía.-**Proporciona la suma de los elementos al cuadrado dentro una matriz de co-ocurrencia. También a ese rasgo se le conoce como momento angular se segundo orden. (Angular Second Moment ASM). Se calcula mediante la siguiente ecuación.

$$\sum_{i,j=0}^{N-1} P(i,j)^2 \tag{3.16}$$

Este rasgo proporciona valores altos cuando la matriz de co-ocurrencia tiene pocas entradas de gran magnitud y proporciona valores bajos cuando todas las entradas son similares. Dicho en otras palabras, se puede decir que cuando todos los elementos de la matriz de co-ocurrencia son semejantes (mayor dispersión en la diagonal principal) 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 energía da una idea de la suavidad de la textura y esto se refleja en la ubicación de sus probabilidades en la matriz de co-ocurrencia.

- **Correlación.**-Este rasgo mide la probabilidad de ocurrencia conjunta de los píxeles especificados. Se calcula mediante la siguiente ecuación.

$$\sum_{i,j=0}^{N-1} P_{i,j} \frac{(i-\mu_i)(j-\mu_j)}{\sigma_i \sigma_j} \tag{3.17}$$

Algunas propiedades de la correlación son:

- a).- Un objeto tiene más alta correlación dentro de él que entre objetos adyacentes.
- b).-Píxeles cercanos están más correlacionados entre sí que los objetos más distantes.

3.5. Clasificador de los *k*-próximos vecinos (*K*-NN)

Este clasificador es ampiamente usado en el reconocimiento de formas. Dado un vector a clasifcar (rasgos característicos del objeto a clasificar) y un conjunto de vectores prototipo asignados a las diversas clases existentes (base del conocimiento). La regla consiste en calcular la distancia del vector a clasificar 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.

3.5.1. Principio teórico.

Sea $\overrightarrow{x}$ un vector de dimensión *"n"* a clasificar, sea M una base de datos de referencia construiída 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 $\overrightarrow{x}$ a partir de los K-próximos vecinos de la base de referencia [11] y [18].

Sea $p(\overrightarrow{x}/C_i)$ 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 muestra $\overrightarrow{x}$ pertenezca a la clase C_i tal que:

$$p_r(C_i/\overrightarrow{x}) = \frac{p(\overrightarrow{x}/C_i) * p_r(C_i)}{p(\overrightarrow{x})} = \frac{p_r(\overrightarrow{x}/C_i) * p_r(C_i)}{\sum_{k=1}^{c} p(\overrightarrow{x}/C_k) * p_r(C_k)} \tag{3.18}$$

donde:

$p_r(C_i)$ = probabilidad de aparición de la clase C_i

$p_r(\overrightarrow{x})$ = probabilidad de que la muestra $\overrightarrow{x}$ pertenezca a la clase C_i

$p_r(C_i/\overrightarrow{x})$ = densidad de probabilidad condicional de la muestra $\overrightarrow{x}$ conociendo la clase C_i

Partiendo de la base de referencia M (base del aprendizaje), se estiman las densidades de probabilidad $p(\overrightarrow{x}/C_i)$ 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 $\overrightarrow{x}$" 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 hiperesférico definido por la distancia "D" entre la muestra $\overrightarrow{x}$ y el K-ésimo vecino, la densidad de probabilidad conjunta $p(\overrightarrow{x}/C_i)$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 3.18 se transforma en:

$$p_r(C_i/\overrightarrow{x}) = \frac{\frac{K_i}{N*V}}{\sum_{j=1}^{c} \frac{K_j}{N*V}} = \frac{K_i}{\sum_{j=1}^{c} K_j} = \frac{K_i}{K} \tag{3.19}$$

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

La clase a la cual pertenece la muestra $\overrightarrow{x}$ es determinada al considerar el número más grande de prototipos pertenecientes a la clase C_i(entre los k prototipos). Es decir, que $\overrightarrow{x}$ 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_1y C_2 si k_1/k_2>1 entonces la clase ganadora será C_1, en el caso contrario, la muestra $\overrightarrow{x}$ será asignada a la clase C_2. En la igura 3.6) se muestra un ejemplo, la cual, fue tomada de la referencia [26].

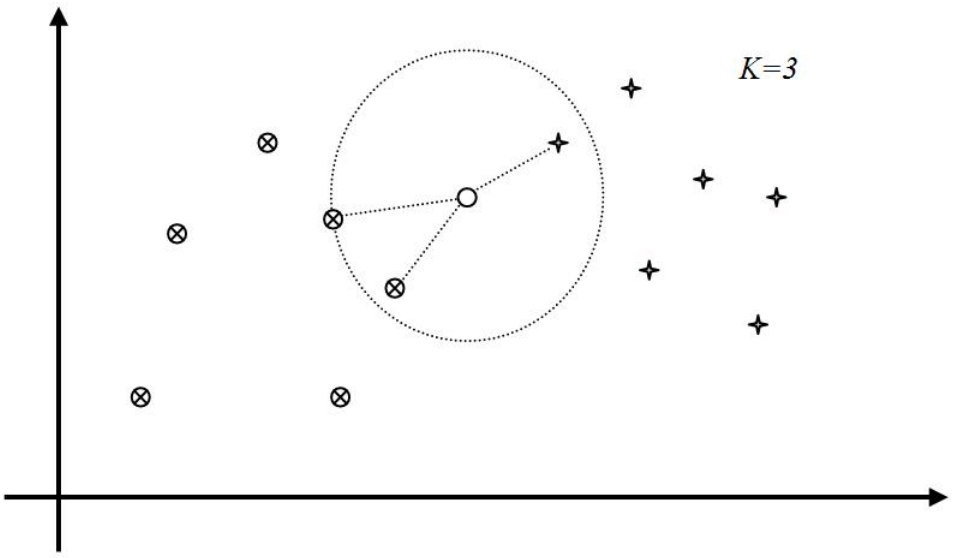


Figura 3.6: Selección de los k-vecinos, donde el patrón 'x' está representado por el pequeño círculo blanco, el cual es clasificado con la clase $\bigotimes$ dado que de sus k(3) próximos vécinos, "uno" pertenece a la clase $\star$, y "dos" a la clase $\bigotimes$

2. **a) Método " de reagrupamiento por clase**

   - La densidad de probabilidad conjunta $p(\overrightarrow{x}/C_i)$ 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 3.18 se transforma en la ecuación 3.20.

$$p_r(C_i/\overrightarrow{x}) = \frac{\frac{K_i}{N*V_i}}{\sum_{j=1}^{c} \frac{K_j}{N*V_i}} = \frac{\frac{1}{V_i}}{\sum_{j=1}^{c} \frac{1}{V_j}} \tag{3.20}$$

- Para determinar la clase a la que pertenece la muestra $\overrightarrow{x}$ 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 $\overrightarrow{x}$. La clase ganadora C_i es la que posee el volumen más pequeño, es decir, la distancia más pequeña entre la muestra $\overrightarrow{x}$ y los prototipos de la clase C_i

$$volumen = \frac{4\pi r^n}{3} = \frac{\pi d^n}{6} \tag{3.21}$$

- 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 $\overrightarrow{x}$ será asignada a la case C_2. Como se puede observar en la figura 3.7, tomada de la referencia [26] V_1 ó clase $\bigotimes$ es $<$ que V_2 o clase $\star \Longrightarrow$la clase ganaradora es C_1

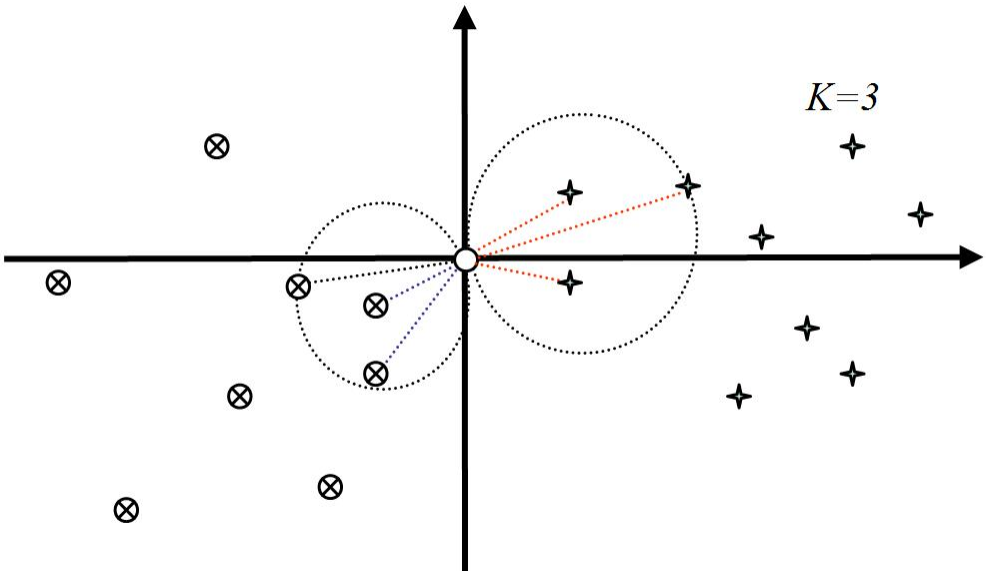


Figura 3.7: Selección de los k vecinos por "volumen". El círculo blanco es clasificado en la clase $\otimes$,dado que K=3 próximos vecinos están más próximos que los 3 elementos próximos de la clase de puntos $\star$

3.5.2. Cálculos de distancias

Para calcular la distancia entre la muestra $\overrightarrow{x}$ y los puntos de la base de datos M existen diferentes formas de medirla, por mencionar algunas:

- Distancia Euclidiana:

$$D(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)} \tag{3.22}$$

- Distancia Manhattan:

$$D(x,y) = \sum_{i=1}^{n}(x_i - y_i)^2 \tag{3.23}$$

- Distancia del Máximo

$$D(x,y) = max_i\,|x_i - y_i| \tag{3.24}$$

La distancia que se utiliza normalmente es la euclideana (3.22), pero la distancia Manhattan 3.23 y la del máximo 3.24 son más rápidas de calcular. El tipo de distancia

a utilizar depende de la aplicación, es decir, de factores como el tiempo de ejecución, el costo,, el desempeño, la precisión, etc..

Capítulo 4

METODOLOGÍA PROPUESTA

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

4.1. 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 [18] y la Homogeneidad, este rasgo descriptor proviene de la matriz de co-ocurrencia [33]. Estos 3 rasgos se aplican a una sub-imagen bajo el dominio de los canales del formato (HSI) de una imagen. Se propone también el uso de un algoritmo de K-Medias[18] y algunos tipos de clasificadores como son:

- Clasificador 1-NN.

- Clasificador del tipo: Red Neuronal Artificial (RNA).

Se ha decidido combinar el algoritmo de K-Medias y algún tipo de clasificador es para construir una base de datos indexada de 700 imágenes (por el momento). Una de las 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 2 etapas principales:

- Etapa de entrenamiento.
- Etapa de prueba.

Estas dos etapas principales se explican a detalle enseguida.

4.2. Etapa del entrenamiento

Esta etapa se divide en dos fases principales como se muestra en la figura 4.1. Durante la primer etapa (Parte A), un conjunto de 700 imágenes en formato RGB (720 x 480) ó (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 pixeles 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 co-ocurrencia.

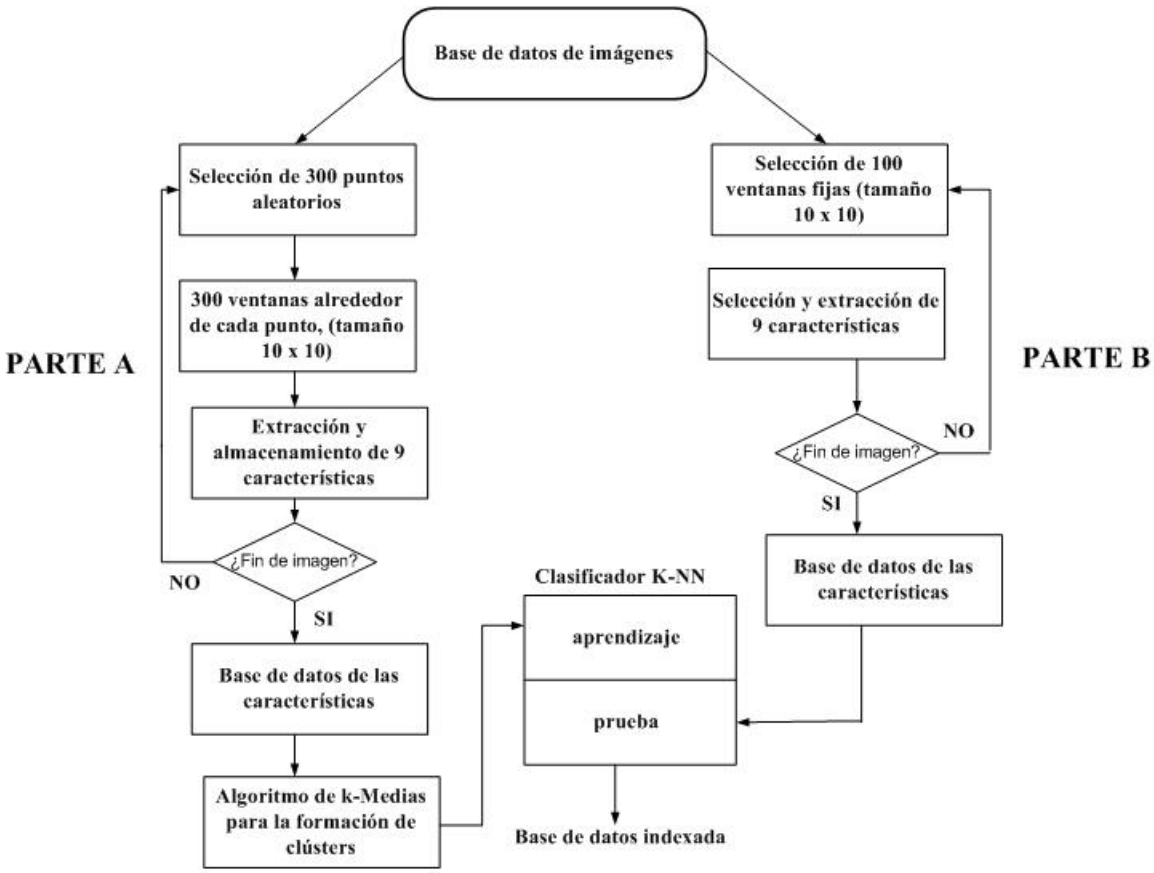


Figura 4.1: Diagrama de flujo para la etapa del entrenamiento

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. 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 escenas: agua, roca, cielo, vegetación, pastos, y nubes mas cuatro clases adicionales frontera o de borde que se forman entre cielo y agua, cielo y pastos, cielo y nubes, y cielo y rocas, dando como resultado, un total de 10

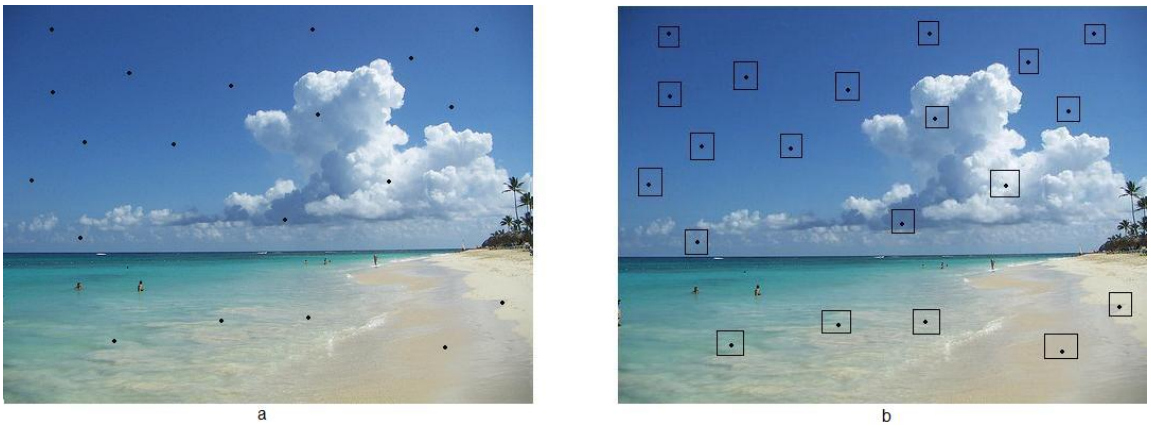


Figura 4.2: (a).-Para la descripción de las sub-imágenes, 300 pixeles de imagen son aleatoriamente seleccionadas uniformemente. (b).-Para lograr una segmentación automática de la imagen, alrededor de cada uno de los 300 pixeles se abre una ventana cuadrada de tamaño $M\ x\ N$. En esta figura se muestran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

clases presentes en las imágenes de escenarios naturales..

Para las 700 imágenes seleccionadas durante la etapa del entrenamiento, la tabla 4.1 muestra como se reparten los 210,000 vectores entre las 10 posibles clases que se forman después de aplicar el algoritmo de K-Medias, es decir, cuantos vectores caen dentro de la clase 1, cuantos 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 las 700 imágenes. Se usaron 700 imágenes de escenarios naturales provenientes de la base de datos de Corel [44], [46] y [47] durante la etapa del entrenamiento, las cuales están divididas en 6 diferentes tipos de escenas: costas, ríos,lagos/lagunas, bosques, montañas, 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 x 10 de 72 x 48 píxeles cada una. Por cada una de estas 100 sub-imágenes, se toma una ventana de

tamaño 10 x 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 3 canales.

Cada ventana es descrita en forma de un vector de 9 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 primer etapa del entrenamiento (figura 4.1 parte A) los cuales 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. 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.

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

Número de clase	Número de característica por clase
1	22086
2	23267
3	23899
4	16127
5	23926
6	24506
7	30262
8	10708
9	10957
10	24252
	Total: 210,000

Cuadro 4.1: Distribución de los 210,000 características entre las 10 clases seleccionadas para el conjunto de las 700 imágenes de los escenarios naturales usadas para construir la base indexada de datos.

4.3. Etapa de recuperación

Esta estapa fue diseñada como se muestra en la figura 4.4. Como se puede ver, solamente consta de una estapa. 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.

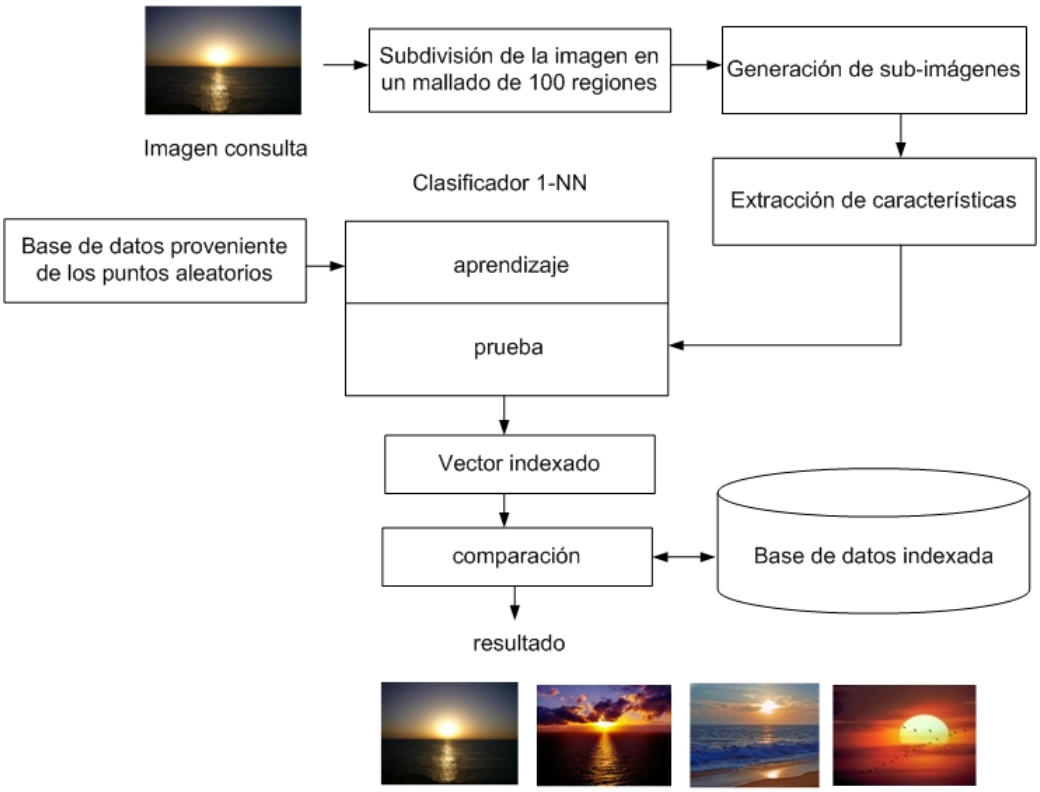


Figura 4.4: Diagrama de flujo para la etapa de la prueba.

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 más alto índice de probabilidad de las 10 clases posibles. 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 de Corel cuya resolución es de 720 x 480 ó de 480 x 720. Esta base de datos de imágenes fue proporcionada por J. Vogel [44], [46], [45],[48] y [47]. Las 700 imágenes fueron utilizadas para for-

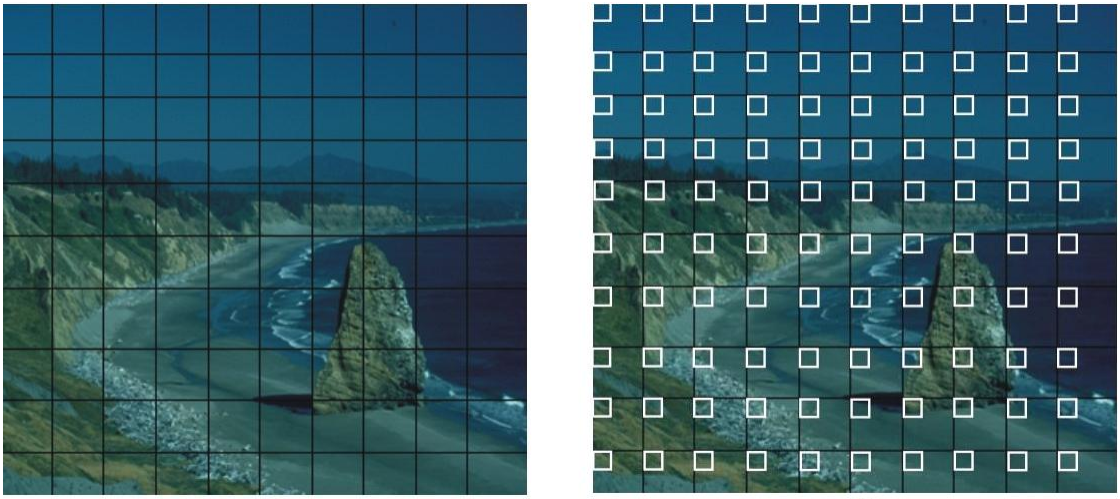


Figura 4.5: (a) Una imagen es uniformemente dividida en 100 sub-imágenes para obtener 100 regiones descriptivas de características. (b) Para cada una de las sub-imágenes, una ventana de tamaño 10 x 10 pixeles es selecionada para calcular el correspondiente vector de características.

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

Cuadro 4.2: Estructura de la base de datos indexada

mar 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 como queda armada la base de datos indexada con las 700 imágenes de escenarios naturales pertenecientes al conjunto del entrenamiento.

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 escena a través de sus imágenes o escenas consulta.

5.1. Recuperación de imágenes

Cuando se aplica el algoritmo de 10-Medias, se puede observar como se forman los clústeres en una escena natural usando puntos aleatorios. En la figura 5.1 se puede observar que la escena está conformada por los clústeres 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.

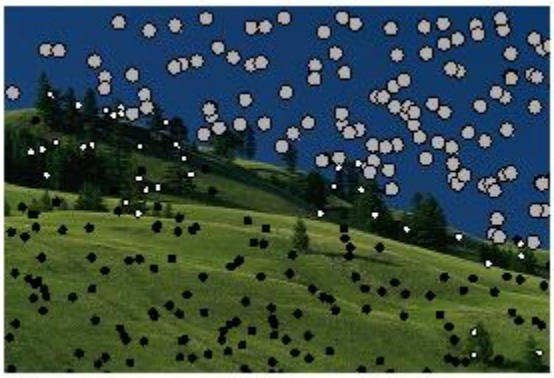


Figura 5.1: Clústeres formados en una escena natural usando el algoritmo de K-Medias y puntos aleatorios

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 .

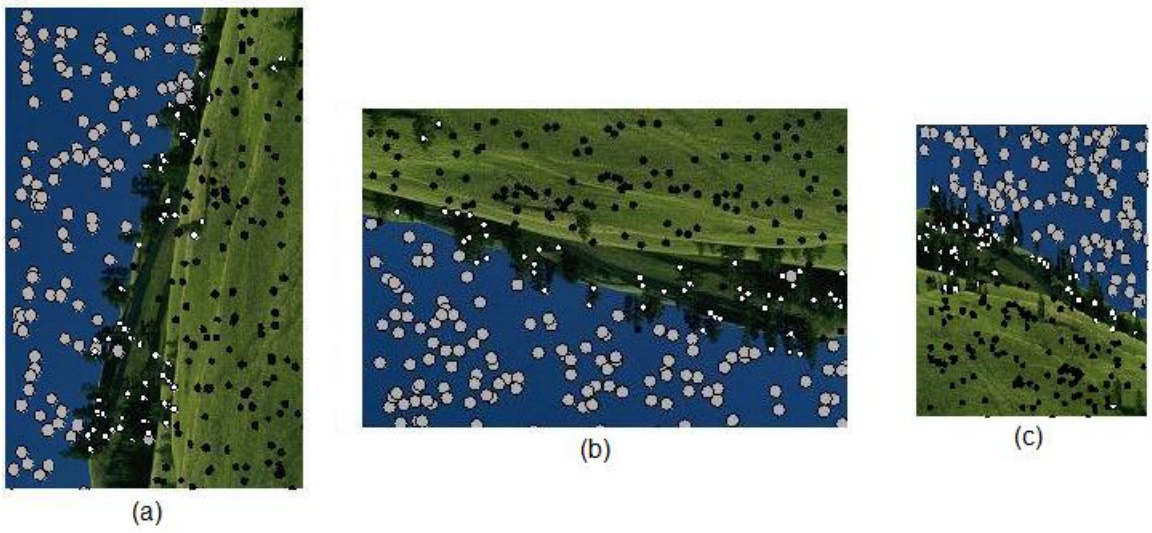


Figura 5.2: (a) Imagen rotada 90^0. (b) Imagen rotada 180^0. (c) Imagen escalada al 50 %. Obsérvese como el resultado presenta invarianza ante estas transformaciones.

Figura 5.3: Clústeres formados en una escena natural usando el algoritmo de K-Medias y puntos aleatorios para imágenes del mismo tipo de escenario.

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 clústeres 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 balancos representan a la clase agua.

Inicialmente se había propuesto usar 6 clases solamente al aplicar el algoritmo de K-Medias usando solamente 300 imágenes de entrenamiento, las cuales fueron 54 escenas de bosques, 54 escenas de lagos/lagunas, 54 escenas de costas, 54 escenas de praderas, 54 escenas de montañas y 30 escenas de cielos/nubes y las clases que contenían a dichas escenas son: agua, roca, pasto, cielo, vegetación y nubes, la base

de datos indexada que se obtuvo se indica en la tabla 5.1.

C_1	C_2	C_3	C_4	C_5	C_6	$\rightarrow$	nombre de la imagen
40	16	23	20	1	0	$\rightarrow$	imagen 1.jpg
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
28	19	9	9	15	20	$\rightarrow$	imagen k.jpg
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
7	23	7	32	19	12	$\rightarrow$	imagen 300.jpg

Cuadro 5.1: Base de datos indexada para 6 clases y 300 escenas de entrenamiento

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 escenarioss naturales, sustituiremos la palabra imagen por la de escena, por tratarse de un escenario natural.

Figura 5.4: Escenas recuperadas dada una escena consulta de una puesta de sol.

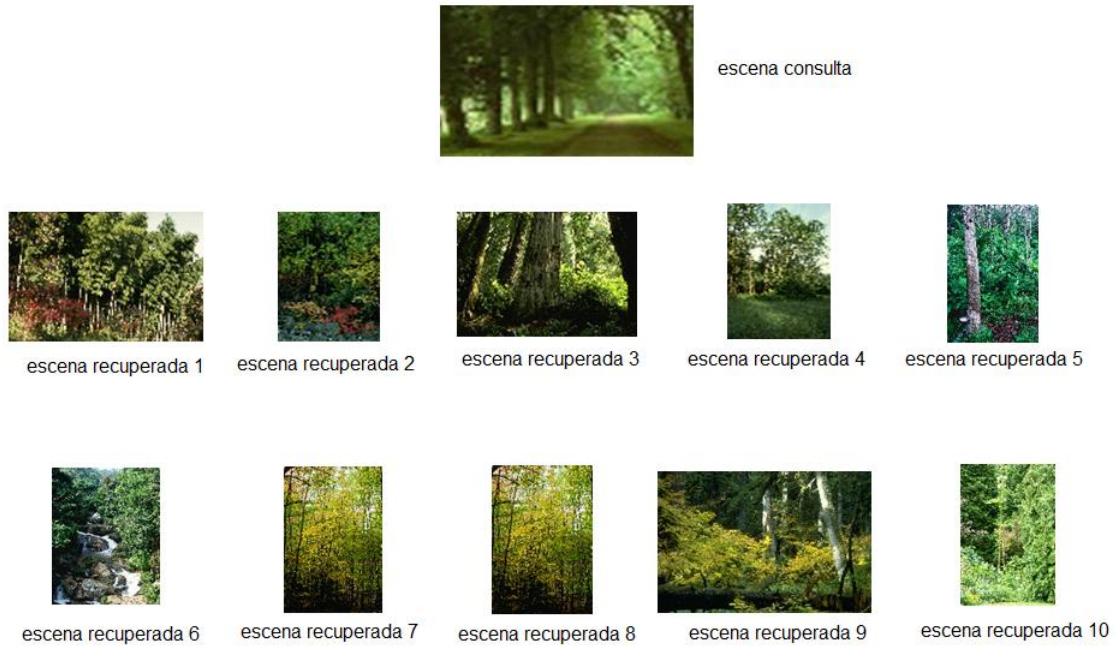


Figura 5.5: Escenas recuperadas dada una escena consulta de un bosque.

Figura 5.6: Escenas recuperadas dada una escena consulta de una puesta de sol roja.

En la figura 5.4 se observan los resultados al aplicar la metodología propuesta en

la recuperación de escenas dada una escena consulta de tipo “puesta de sol”.

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

Enla figura 5.6 se observan los resultados al aplicar la metodología propuesta en la recuperación de escenas dada una escena 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 escenas desde Internet. Estas 221 imágenes de escenarios naturales no forman parte del conjunto del entrenamiento. Se le presentaron estas 221 escenas al sistema de consulta y el sistema respondió desplegando en pantalla con las 10 escenas más similares extraídas de la base indexada de escenas. La figura 5.4 muestra un ejemplo de consulta. En la figura 5.4 se puede observar que el sistema recupera correctamente 9 escenas y solamente recupera incorrectamente 1 escena (escena 10). Esto nos arroja un resultado de 90 % de eficiencia para esta recuperación. La prueba completa se puede observar en la figura 5.7 para escenas de puestas de so completamente rojas, en la figura 5.8 para escenas de un bosque y en la figura 5.9 para escenas de costas al comparar contra la transformada curvelet. respectivamente. En la figura 5.5 se recuperan correctamente las 10 imágenes correspondientes a un bosque y en la figura 5.6, el sistema recupera corresctamente 7 escenas y solamente 3 incorrectas (escena 2, 4 y 8).

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

$$P = \frac{\text{Número de escenas relevantes recuperadas}}{\text{Número total de escenas recuperadas}} X 100\,\% \tag{5.1}$$

$$R = \frac{\text{Número de escenas relevantes recuperadas}}{\text{Número total de escenas relevantes en la base de datos}} X 100\,\% \tag{5.2}$$

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 [48] . Como se puede apreciar, la eficiencia de nuestra propuesta es superior a la reportada en [48].

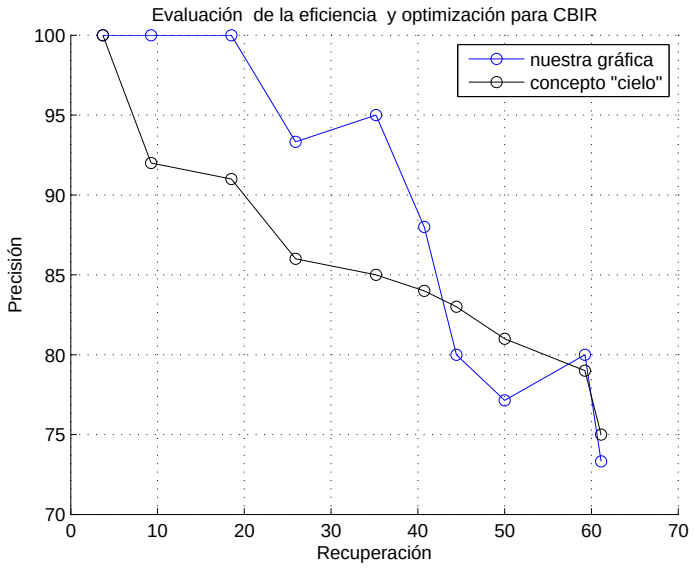


Figura 5.7: Eficiencia de nuestra propuesta comparada contra el método descrito en [48]. Mediante nuestra propuesta se obtiene 88.68 % de eficiencia (gráfica azul), mientras que en [48]se obtiene 85.60 % de eficiencia (gráfica en negro) cuando se aplica una consulta de una escena de una puesta de sol con cielo rojo.

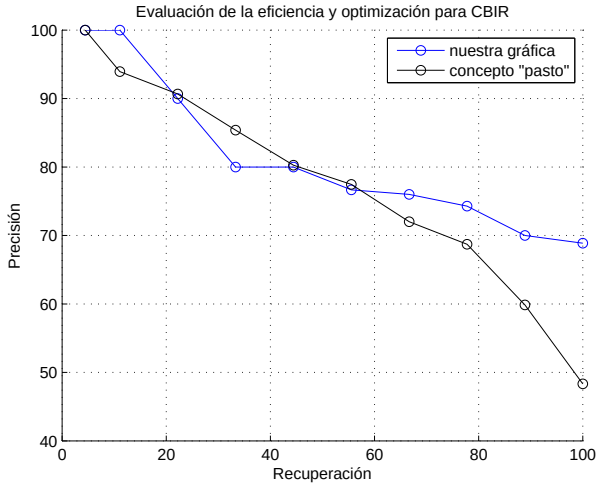


Figura 5.8: Eficiencia de nuestra propuesta comparada contra el método descrito en [48]. Mediante nuestra propuesta se obtiene 81.58 % de eficiencia (gráfica azul), mientras en [48] se obtiene 77.66 % de eficiencia (gráfica negra) cuando se aplica una consulta de la escena de un bosque.

En la figura 5.5 se observa el resultado obtenido al comparar nuestra propuesta contra el método reportado en [48]. Como se puede apreciar, la eficiencia de nuestra propuesta es nuevamente superior que la reportada en [48].

En la figura 5.9 se observa el resultado obtenido al comparar nuestra propuesta contra el método reportado en [23]. Como se puede apreciar, la eficiencia de nuestra propuesta es superior que la reportada en [23].

Siguiendo con el criterio de considerar solamente 6 clases, se procedió a probar nuestra metodología pero con una base de datos de la catedral de Sacre Coeur (París) cuya resolución de imagen es (768×1024). Estas imágenes fueron proporcionadas por Mauricio Díaz [12]. 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

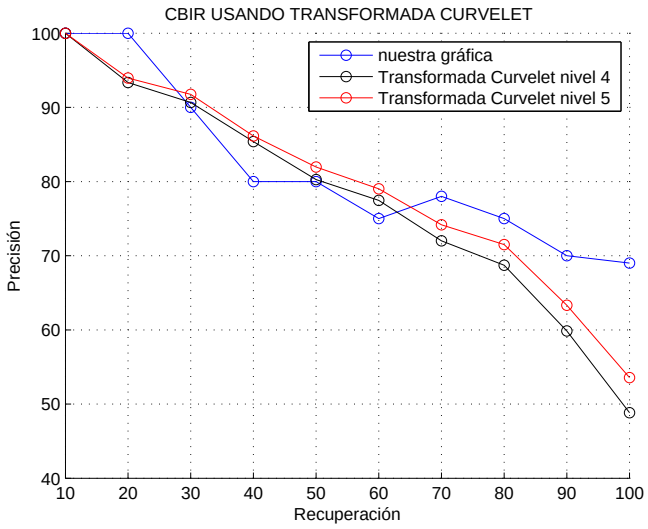


Figura 5.9: Eficiencia de nuestra propuesta al comparar contra el método descrito en [23]. Se obtiene una eficiencia del 81.7 % (gráfica azul) mientras que en [23] se obtiene una eficiencia de 77.71 % (gráficas en rojo y negro respectivamente).

estos tipos de cielos. Al usar nuestra metodología y al aplicarla ahora a las imágenes con escenarios de cielos se obtiene la figura 5.10 y la figura 5.11 respectivamente.

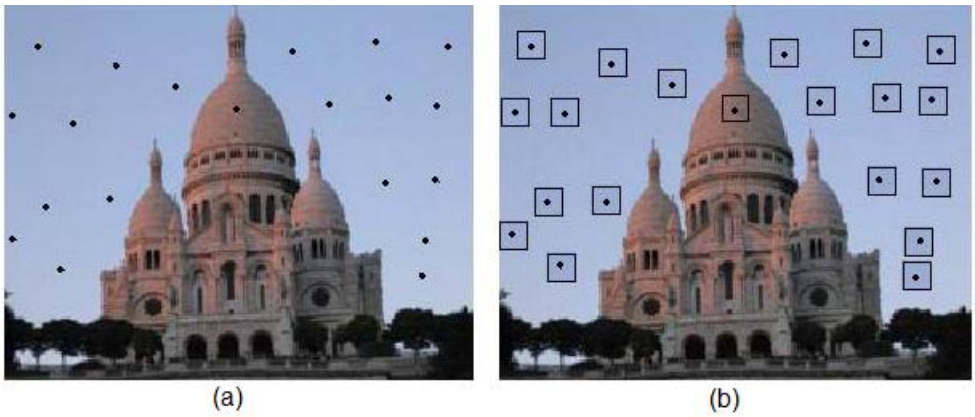


Figura 5.10: (a).-Para la descripción de las sub-imágenes, 300 pixeles de imagen son automática y uniformemente seleccionados aleatoriamente. (b).-Para lograr una segmentación automática de la imagen, alrededor de cada uno de los 300 pixeles se abre una ventana cuadrada de tamaño M x N. En esta figura solamente 20 puntos se muestran para dar un ejemplo.

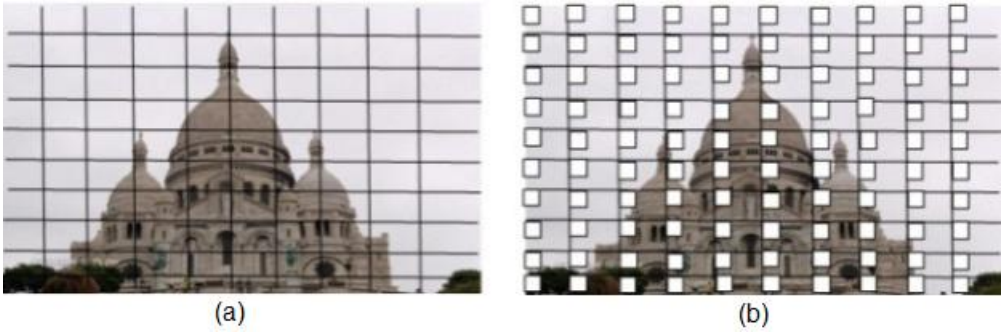


Figura 5.11: (a) Una imagen es uniformemente dividida en 100 sub-imágenes para obtener 100 regiones descriptivas de características. (b) Para cada una de las sub-imágenes, una ventana de tamaño 10 x 10 pixeles es selecionada para calcular el correspondiente vector de características.

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 escenas de cielos completamente nublados cuando se aplica una consulta de una escena (que no forma parte del conjunto del entrenamiento) de cielo completamente nublado.

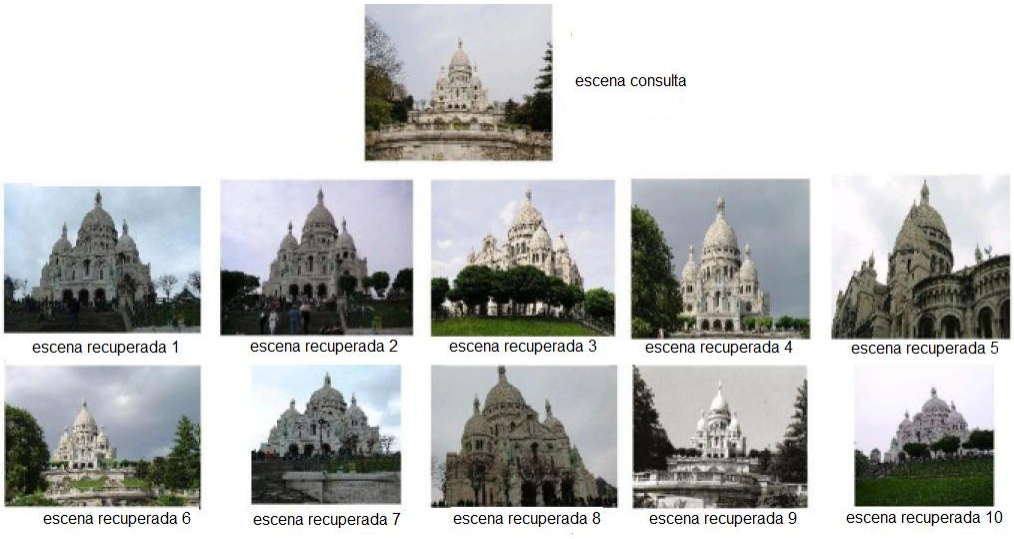


Figura 5.12: Recuperación de escenas de cielo completamente nublado cuando se aplica al sistema una escena consulta de un cielo nublado.

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

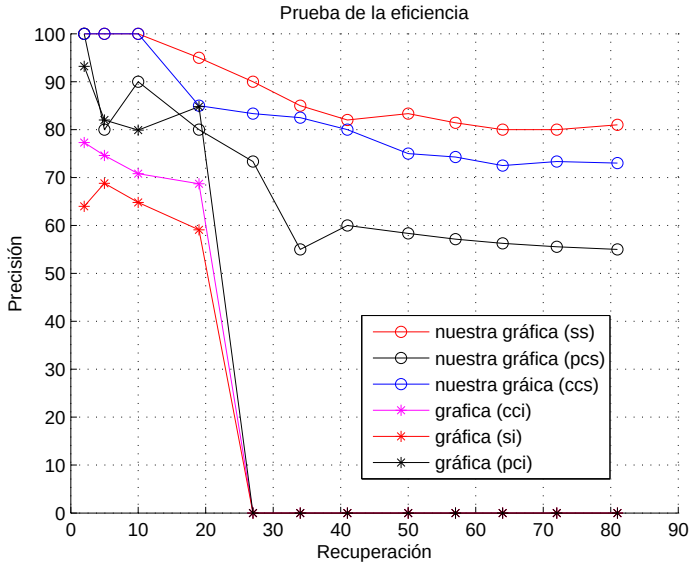


Figura 5.13: Eficiencia de nuestra propuesta comparada contra el método descrito en [12].

La figura 5.13 muestra que con nuestra propuesta se obtiene 88.14 % de eficiencia, mientras que en [12] los autores obtienen 64.17 % de eficiencia cuando se hace la consulta de una escena de un cielo completamente brillante (ss) en nuestra gráfica (gráfica con puntos rojos) y (si) es en la gráfica de [12] (gráfica con asteriscos rojos).

La figura 5.13 muestra que con nuestra propuesta se obtiene 63.75 % de eficiencia, mientras que en [12] los autores obtienen 84.97 % de eficiencia cuando se hace la consulta de una escena de un cielo parcialmente nublado (pcs) en nuestra gráfica (gráfica con puntos negros) y (pci) es en la gráfica de [12] (gráfica con asteriscos negros).

La figura 5.13 muestra que con nuestra propuesta se obtiene 83.24 % de eficiencia, mientras que en [12] los autores obtienen 72.85 % de eficiencia cuando se hace la consulta de una escena de un cielo completamente nublado (ccs) es en nuestra gráfica

(gráfica con puntos azules) y (cci) es en la gráfica de [12] (gráfica con asteriscos magenta).

Nota: Las mediciones en [12] 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 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 2 bases de datos anteriores (la de Corel [46], [45] y [47] y la de la catedral de de Sacre Coeur (París) [12] se probaron con 6 clases de objetos presentes en las imágenes (ver figura 5.14) y 300 imágenes de entrenamiento.

Figura 5.14: Las 6 clases de objetos presentes en las imágenes del 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. Se Considera que existen 4 clases adicionales que

llamaremos “clases de borde o de frontera” , las cuales se muestran en la figura 5.15. También se usaron todas las imágenes de la base de datos de Corel (700) imágenes.

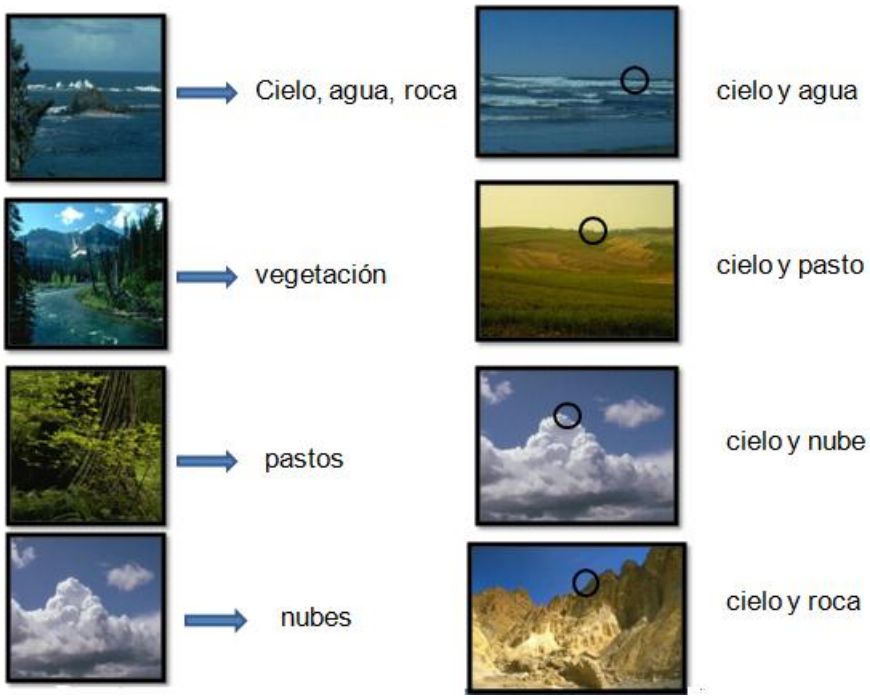


Figura 5.15: Propuesta de la existencia de 4 clases adicionales a las 6 que ya están propuestas, las cuales les llamaremos “clases de borde, o de frontera”

En la tabla (4.1 en la página 76) se muestran como se distribuyen las 210,000 características (700 imágenes con 300 puntos aleatorios por cada imagen) entre las 10 clases seleccionadas.

En la tabla (4.2 en la página 78) se muestran los resultados de como se formó la base de datos indexados tomando las 700 imágenes del Corel.

Para probar nuestra propuesta, usamos los 700 escenarios naturales de la base de datos de Corel (720 x 480) ó (480 x 720). Los 700 escenarios naturales usados para el entrenamiento están agrupados en 6 diferentes tipos de escenarios de la siguiente manera: 179 escenas de montañas, 111 escenas de ríos/lagos, 144 escenas de costas; 103 escenas de costas, 131 escenas de praderas y 32 escenas de cielo/nubes.

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

Figura 5.16: Escenas recuperadas dada una escena consulta de un bosque.

Figura 5.17: Escenas recuperadas dada una escena consulta de una costa.

Figura 5.18: Escenas recuperadas dada una escena consulta de un lago.

Figura 5.19: Escenas recuperadas dada una escena consulta de una montaña.

Figura 5.20: Escenas recuperadas dada una escena consulta de cielo/nubes.

Figura 5.21: Escenas recuperadas dada una escena consulta de una pradera.

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

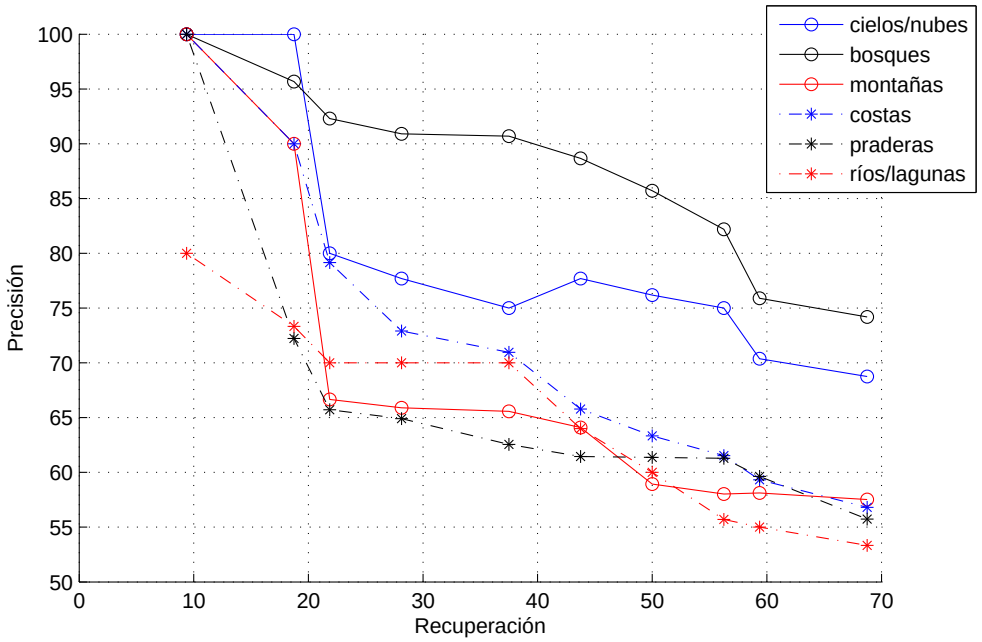


Figura 5.22: Eficiencia de nuestra propuesta, usando 10 clases y 700 imágenes de entrenamiento

En la figura (5.22) muestra que nosotros obtenemos **70.29 %** de eficiencia (ver tabla 5.3) mientras que en la propuesta de [46] y [44] obtienen 58 % cuando se aplica recuperación de imágenes a escenas de 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[46] y [44] obtienen 40 % cuando se aplica recuperación de imágenes a escenas de 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 [46] y [44]] obtienen 83 % cuando se aplica recuperación de imágenes a escenas de 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 [46] y [44] obtienen 30 % cuando se aplica recuperación de imágenes a escenas de 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 [46] y [44] obtienen **70 %** cuando se aplica recuperación de imágenes a escenas de 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 [46] y [44] obtienen **87** % cuando se aplica recuperación de imágenes a escenas de cielos/nubes (ver tablas 5.2).

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

Cuadro 5.2: Promedio de eficiencia para la metodología descrita en [46] y [44]

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

Cuadro 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 escenas de cielos/nubes y las montañas nuestra propuesta es un poco menor.

5.2. Identificación de la escena

A continuación se muestran los resultados experimentales, los cuales llamaremos: "Identificación de la escena consulta" A los diferentes tipos de escenarios naturales se

dividen en 2 grupos: El primer grupo contiene bosques, cielos/nubes y montañas. El segundo grupo está conformado por praderas, costas y ríos/lagos.

El procedimiento es como sigue: Se tomaron manualmente desde Internet 15 imágenes pertenecientes a cada tipo de escena de (ver figura, a continuación se aplica nuestra metodología para la recuperación de las imágenes para el total de las 90 imágenes, (es decir, 15 para cada una de los 6 tipos de escenarios), una vez que se realiza la recuperación de imágenes, se obtiene un vector indexado para cada una de estas 90 escenas consulta y seleccionamos las 2 clases (de las 10 posibles que se forman cuando se aplica el algoritmo de K-Medias) con la más alta probabilidad de ocurrencia. y el resultado queda como sigue:

- Grupo 1: ***Escenas de bosques.-***éstas están conformadas por las clases 8 y 9. ***Escenas de cielo y nubes.***-éstas están conformadas por la clase 2 y 5. ***Escenas de montañas.***-éstas están conformadas por las clases 1 y 5 (ver tabla 5.4 en la página siguiente).

- Grupo 2: ***Escenas de praderas***.-éstas están conformadas por las clases 4 y 7. ***Escenas de costas***.- éstas están conformadas por la clase 4 y 5. ***Escenas de ríos/lagos***.- éstas están conformadas por la clase 1 y 9 (ver tabla 5.5 en la página siguiente).

Para probar la eficiencia de la propuesta respecto a la identificación de la escena, se obtiene una matriz de confusión para cada grupo y así poder validar y clasificar el tipo de escena (ver tabla .

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

$$P = \frac{\text{Número de escenas relevantes identificadas}}{\text{Número total de escenas identificadas.}} x100\,\% \tag{5.3}$$

$$I = \frac{\text{Número de escenas relevantes identificadas}}{\text{Número total de escenas relevantes en la base de datos}} x100\,\% \tag{5.4}$$

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

Cuadro 5.4: Matriz de confusión para el grupo 1. Ésta muestra una eficiencia de 81,10 %(valor promedio).

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

Cuadro 5.5: Matriz de confusión para el grupo 2. Ésta muestra una eficiencia de 82.22 % (valor promedio).

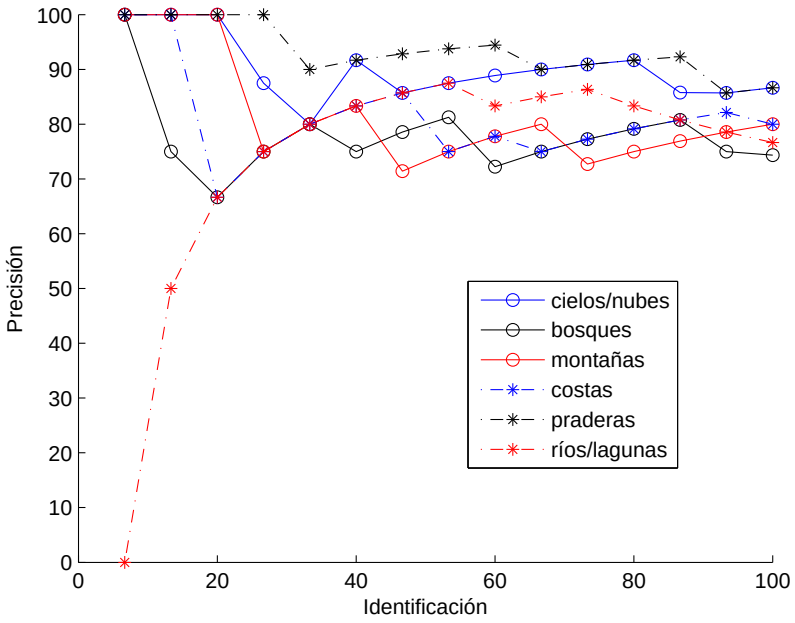


Figura 5.23: Eficiencia de nuestra propuesta de "Identificando la escena consulta".

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 %

Cuadro 5.6: Resultados obtenidos para "Identificando la escena consulta" (valores promedio obtenidos de la figura 5.23.

Capítulo 6

CONCLUSIONES Y TRABAJOS FUTUROS

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 escenarios naturales desde una base de datos de escenarios naturales. Como un resultado adicional, nuestra propuesta también permite la identificación de la escena 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 escenas: costas, ríos/lagos, montañas, bosques, praderas y cielos/nubes. Nuestra propuesta extrae desde cada tipo de escena vectores descriptores usando una combinación de puntos fijos y puntos aleatorios los cuales son seleccionados automáticamente. Se usa el algoritmo de K-Medias para formar inicialmente 6 clústeres y ahora 10 clústeres. Se usa un clasificador de 1-NN 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.

Durante la etapa de recuperación, el clasificador 1-NN ya está entrenado para recuperar desde la base de datos indexada las escenas mas similares dada una escena consulta. Los resultados experimentales que nuestra propuesta logra obtener mejores resultados que 4 métodos reportados en la literatura. Para poder validar los resultados obtenidos hacemos uso de de las mediciones de precisión 5.1 en la página 84 y recuperación 5.2 en la página 84. y para la identificación de las escenas consulta hacemos uso de las mediciones de precisión 5.3 en la página 97 e identificación 5.4 en la página 97.

También se probó nuestra metodología con la base de imágenes de la Catedral de Sacre Coeur (París) 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 escenas y para la identificación y clasificación de diferentes tipos de escena.

6.2. Trabajo actual y 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 clústeres. 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 escenas.

También 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.

6.3. Publicaciones realizadas.

- J.F. Serrano, J.H. Sossa, C. Avilés, R. Barrón, G. Olague, y J. Villegas, "Scene Retrieval of Natural Images", CIARP 2009, LNCS 5856, pp. 774-781,Guadalajara Jalisco, México.

- J. Félix Serrano, J. Humberto Sossa, Carlos Avilés, Juan Villegas. "Unsupervised Images Retrieval with Similar Lighting Conditions", artículo aceptado en el congreso de ICPR 2010, Estambul Turquía, 23-26 de Agosto de 2010.

.

Bibliografía

[1] 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.

[2] 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.

[3] 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.

[4] A.J.M. Content based image retrieval using approximate shape of objects. *17 th IEEE Symposium on Computer-based medical Systems (CBMS)*, 2004.

[5] Alain.C.Gonzalez-Garcia. *Image retrieval based on the contents.* PhD thesis, Center for Research in Computing (CIC)-IPN, Mexico DF, September 2007.

[6] Anna.Bosch, Xavier.Muñoz, and Robert.Marti. Which is the best way to organize/classify images by content? Technical report, Depertment of Electronics Informatics and Automaticas, University of Girona, Campus Montilivi, 2006.

[7] 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 Inteligence*, 22 No. 12:1349–1380, 2000.

[8] Barb and A.Shyu. Semantics modeling in diagnostec medical image databases using customized fuzzy membership functions. *FUZZY SYSTEMS. tHE 12 th IEEE INTERNATIONAL CONFERENCE*, 2, 2003.

[9] C.Avilés-Cruz. *Analyse de Texture par Statistiques D ' Ordre Supérieur: Caractérisation et Performances.* PhD thesis, Instituto Nacional Politécnico de Grenoble-Francia, 1997.

[10] C.C-Gotlieb and H.E.Kreyszig. Texture descriptors based on matrices. *Computer Vision, Graphics, and Image Processing*, 51, 1990.

[11] Darío.Maravall.Gómez-Allende. *Reconocimiento de formas y visión artificial.* Addison -Wesley Iberoamericana, 1997.

[12] M. Díaz and P. Sturm. Finding images with similar lighting conditions in large photo collections. *CIARP 2009. LNCS 5586, Springer*, pages 53–60, 2009.

[13] D.B.Russakof. Image similarity using mutual information. *ECCV*, 3, Prague Czech republic, 2004.

[14] P.R. Devijver and J. Kittler. *Pattern Recognition, A Statistical Approach.* Prentice Hall, New York, 1982.

[15] Richar O. Duda and P.E. Hart. *Pattern Classification ans scene Analysis.* Wiley, New York., 2000.

[16] E.G.Petrakis. Similarity searching in image databases. *IEEE Transactions on knowledge and Data Engineering, vol 14, pp 1187-1201, (2005)*, 14:1187–1201, 2005.

[17] 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.

[18] K. Fukunaga. *Introduction to statistical Pattern Recognition*. 1990.

[19] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1992.

[20] Woods Eddins Gonzalez. *Digital Image Processing using Matlab*. Prentice Hall, Boston, MA, USA, 2004.

[21] Gonzalo.Pajares.M. and Jesús.M.de.la-Cruz-García. *Visión por computadora (Imágenes Digitales y aplicaciones)*. Alfaomega Grupo Editor, S.A. de C.V., 2008.

[22] H.Tamura, S. Mori, and T.Yamawaki. Texture features corresponding to visual perception,. *IEEE Trans.on Sys. Man and Cyb . SMC-8(6)*, 1978.

[23] 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.

[24] J.H.Sossa-Azuela. *Rasgos descriptores para el reconocimiento de objetos*. 2006.

[25] Jensen J.R. *Introductory Digital Image Processing (Second Edition)*. Prentice Hall, 1996.

[26] J.Villegas-Cortez. Identificación de tipos de letra. Master's thesis, Universidad Autónoma Metropolitana (Unidad Azcapotzalco), Junio 2005.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] Marceau.D.J., Howarth., P.J.Dubois J.M., and Gratton.D.J. Evaluation of grey level co-ocurrency matrix method for land classification using spot imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 28,Num 4:513–519, 1990.

[32] M.Carlin. measuring the performance of shape similarity retrieval methods. *Computer vision and Image understanding, Vol 84, 2001*, 84, 2001.

[33] Presutti.M. *Co-currency Matrix in Multispectral Classification: Tutorial for Educators textural measures. The 4th day Educacao em Sensoriamento Remote Ambito not do Mercosul - 11 to August 13, 2004 - Sao Leopoldo RS. Brazil.*

[34] 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.

[35] Qasim.Iqbal and J.K.Aggarwal. Cires .ª system for content-based retrieval in digital image libraries". In *Seventh International Conference on Control, Automatization Robotics and Vision (ICARCV) Singapore pp 205-210*, 2002.

[36] R.Datta, D.Joshi, Jia.Li, and J.Z.Wang. Image retrieval: Ideas, influences, and trends of the new age . *ACM Computing Surveys, 40(2), paper 5,*, April 2008.

[37] 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.

[38] Y. Rui, Th.S.Huang, and Sh F.Chang. Image retrieval: Currente techniquies, promissing directions, and open issues. *Journal of Visual Communication and Image Representation 10, 39-62, 1999.*, 10:39–62, 1999.

[39] 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.

[40] 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.

[41] M. Stricker and M.Orengo. Similarity of color images, storage and retrieval for image and video databases. 1995. 1995.

[42] 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.

[43] Trujillo.Leonardo 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).

[44] J. Vogel. *Semantic Scene Modeling and Retrieval PhD Thesis.* PhD thesis, Swiss Federal Institute of technology Zurich. Zurich Germany, 2004.

[45] 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.

[46] J. Vogel and B. Schiele. Semantic modelling of natural scenes for content-based image retrieval. *Int. J of CV, Springer , 10.1007/s 11263-006-8614-1*, 2006.

[47] 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.

[48] J. Vogel and B. Schiele. Performance evaluation and optimization for content-based image retrieval. *Pattern Recognition*, 39(5):897–909, May 2006.

[49] 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.

[50] X.He, R.S.Zemel, and D Ray. Learning and incoporating top down cues in image segmentation. *Proc. IEEE CS Conference Computer Vision*, 1:338–351, 2006.

[51] 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.

[52] 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.

[53] 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.

.

.

.