



**INSTITUTO POLITÉCNICO NACIONAL**  
**CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN**  
**Laboratorio de Reconocimiento de Patrones**



# **RECUPERACIÓN DE IMÁGENES SOBRE LA BASE DEL CONTENIDO**

## **TESIS**

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

Presenta:

**M. en C. Alain César González García**

Directores:

**Dr. Juan Humberto Sossa Azuela**  
**Dr. Edgardo Manuel Felipe Riverón**

México, D. F., Septiembre 2007.



**INSTITUTO POLITÉCNICO NACIONAL**  
**SECRETARÍA DE INVESTIGACIÓN Y POSGRADO**

SIP-14

*ACTA DE REVISIÓN DE TESIS*

En la Ciudad de  México, D. F.  Siendo las  11:00  horas del día  06  del mes de  Febrero  de  2007  se reunieron los miembros de la Comisión Revisora de Tesis designada por el Colegio de Profesores de Estudios de Posgrado e Investigación del:

**Centro de Investigación en Computación**

para examinar la tesis de grado titulada:

**" RECUPERACIÓN DE IMÁGENES SOBRE LA BASE DEL CONTENIDO "**

Presentada por el alumno:

**GONZÁLEZ**

Apellido paterno

**GARCÍA**

Materno

**ALAIN CÉSAR**

nombre(s)

Con registro:

A	0	0	0	1	0	6
---	---	---	---	---	---	---

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

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

**LA COMISIÓN REVISORA**

Presidente

Secretario

Dr. Serguei Levachkins

Dr. Oleksiy Pogrebnyak

Primer vocal

Segundo vocal

Dr. Juan Humberto Sossa Azuela

Dr. Edgardo Manuel Felipe Riverón

Tercer vocal

Suplente

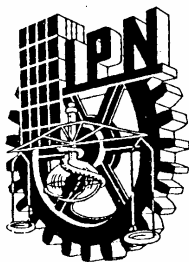
Dra. Patricia Rayón Villela

Dr. Ricardo Barrón Fernández

EL PRESIDENTE DEL COLEGIO

DR. HUGO CÉSAR COYOTE ESTRADA EN COMPUTACION

DIRECCION



**INSTITUTO POLITECNICO NACIONAL**  
**SECRETARÍA DE INVESTIGACIÓN Y POSGRADO**

*CARTA CESION DE DERECHOS*

En la Ciudad de México Distrito Federal el día 3 del mes Septiembre del año 2007, el que suscribe Alain César González García alumno del Programa de Doctorado en Ciencias de la Computación con número de registro A000106, adscrito al Centro de Investigación en Computación, manifiesta que es autor intelectual del presente trabajo de Tesis bajo la dirección del Dr. Juan Humberto Sossa Azuela y del Dr. Edgardo Manuel Felipe Riverón y cede los derechos del trabajo intitulado **RECUPERACIÓN DE IMÁGENES SOBRE LA BASE DEL CONTENIDO**, 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 directores del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección [alaing@sagitario.cic.ipn.mx](mailto:alaing@sagitario.cic.ipn.mx). Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

---

Alain César González García

## **Agradecimientos**

Gracias a todas las personas que de diferentes formas contribuyeron para que este trabajo fuera posible: mis directores de tesis Humberto Sossa y Edgardo Felipe por su paciencia y dedicación que me brindaron durante todo el desarrollo, Oleksy Pogrebnyak por sus atinadas observaciones, Adolfo Guzmán por ayudarme a seleccionar a la Recuperación de Información como área de investigación, Zvi Retchkiman y Ángel Kuri por la formación y conocimiento que compartieron conmigo, a mis sinodales por las sugerencias para mejorar este trabajo.

Gracias al Centro de Investigación en Computación del Instituto Politécnico Nacional y al Departamento de Ingeniería Eléctrica y Electrónica del Instituto Tecnológico de Toluca por las facilidades que me brindaron para realizar mis estudios doctorales.

Agradezco al CONACYT por el apoyo económico brindado como becario y a través del proyecto CONACYT número 46805.

Agradezco a la Secretaría de Investigación y Posgrado del Instituto Politécnico Nacional por el apoyo económico otorgado a través de los proyectos de investigación números 20060517 y 20071438.

# Resumen

Recuperación de imágenes significa buscar y recuperar información visual en forma de imágenes, dentro de una colección de bases de datos de imágenes, siendo una de sus áreas de investigación la organización y recuperación basada en el contenido, en términos de color y forma.

En este trabajo se desarrolla una simple pero efectiva metodología para la recuperación de imágenes de color sobre la base de su contenido. Se hace uso de los canales de color R, G y B para indexar las imágenes. Se prueba el desempeño en una red neuronal perceptron, entrenada con tres diferentes conjuntos de coeficientes ondeletas: dos globales y uno local.

El problema de recuperar imágenes desde una base de datos de imágenes se enfrenta de la siguiente forma: Previo al entrenamiento de la red neuronal, se obtiene una descripción basada en coeficientes ondeletas de cada imagen mediante la ondeleta Daubechies 4, con los que se entrena la red neuronal. Durante la recuperación, una imagen determinada se presenta a la red neuronal entrenada. El sistema responde con las imágenes más parecidas.

Para obtener los coeficientes de la transformada ondeleta se evalúan tres maneras diferentes: A partir de los tres canales de color RGB de la imagen completa, desde los histogramas de la ventana circular más grande dentro de cada uno de los canales de color RGB, y desde los histogramas de los canales de color RGB de sub-imágenes cuadradas en los que se ha dividido la imagen original.

Uno de los aspectos principales de nuestro enfoque es que no se requiere ninguna segmentación previa de la imagen. Durante el entrenamiento se presenta al sistema una imagen cuya clase se conoce de antemano. Se prueba el desempeño de la propuesta con bases de datos de imágenes del mismo tipo de objetos. Al final, se trabaja con diferentes tipos de objetos, por ejemplo, el sistema muestra imágenes de flores si la imagen de una flor se presenta al sistema, a pesar de que la base de datos podría contener no solamente imágenes de flores, sino también imágenes de personas, aviones, paisajes, ciudades, etc.

## Abstract

Image retrieval is referred to seek and recover visual information in form of images, within a collection of databases of images, being one of their investigation areas the organization and retrieval based on the content, in color and shape terms.

In this work we have described a simple but effective methodology for the retrieval of color images. We make use of the R, G and B channels of the color image for indexing. We test the performance of a neural network of perceptrons trained with three different wavelet-based describing features: two global and one local.

We face the problem of retrieving images from a database. During training a wavelet-based description of each image is first obtained using a Daubechies 4-wavelet transformation. Resulting coefficients are used to train a neural network (NN). During retrieval, a given image is presented to the already trained NN. The system responds with the most similar images.

Three different ways to obtain the coefficients of the wavelet transform are tested: From the entire image color channels, from the histogram of the biggest circular window inside the image color channels, and from the histograms of square sub-images in the image color channels of the original image.

Among the main features of our approach is that no previous segmentation of the object class is needed. During training we present to the system an object whose class is known beforehand.

Nowadays, we are testing the performance of our proposal with other databases, of the same kind of objects. At the end of our future research, we would like that the system shows photos of flowers if a photo of a flower is presented to the systems, when the database could contain not only images of flowers but also of people, airplanes, and so on.

# Contenido

<b>CAPÍTULO I</b> .....	1
<b>INTRODUCCIÓN</b> .....	1
<b>I.1. PLANTEAMIENTO DEL PROBLEMA</b> .....	2
<b>I.2. HIPÓTESIS DE TRABAJO</b> .....	4
<b>I.3. JUSTIFICACIÓN</b> .....	5
<b>I.4. OBJETIVOS</b> .....	5
<b>I.5. RESTRICCIONES</b> .....	6
<b>I.6. APORTACIONES</b> .....	6
<b>I.7. PUBLICACIONES HECHAS</b> .....	6
<b>I.8. ORGANIZACIÓN DE LA TESIS</b> .....	7
<b>CAPÍTULO II</b> .....	8
<b>PANORÁMICA DE TRABAJOS RELACIONADOS</b> .....	8
<b>II.1. LÍNEAS DE INVESTIGACIÓN</b> .....	8
<b>II.2. TRABAJOS RELACIONADOS CON LA FORMA Y EL COLOR</b> .....	11
<b>II.3. TRABAJOS RELACIONADOS CON LOS HISTOGRAMAS</b> .....	12
<b>II.4. TRABAJOS RELACIONADOS CON LAS ONDELETAS Y LAS REDES NEURONALES</b> ..	14
<b>II.5. ANÁLISIS DE LOS TRABAJOS</b> .....	15
<b>CAPÍTULO III</b> .....	17
<b>MARCO TEÓRICO DE REFERENCIA</b> .....	17
<b>III.1. LA TRANSFORMADA ONDELETA</b> .....	17
III.1.1 LA TRANSFORMADA ONDELETA CONTINUA.....	17
III.1.2 LA TRANSFORMADA ONDELETA DISCRETA.....	20
III.1.3 LA ONDELETA DAUBECHIES.....	21
III.1.4. MÉTODOS MULTI-RESOLUCIÓN.....	26
III.1.5. ALGORITMO DISCRETO DE LA TRANSFORMADA ONDELETA.....	31
<b>III.2 MODELOS DE COLOR</b> .....	32
<b>III.3 HISTOGRAMA</b> .....	33
<b>III.4 REGLA DEL VECINO MÁS CERCANO</b> .....	38

III.4.1 REGLA DE LOS VECINOS MÁS CERCANOS .....	38
III.4.2 DISTANCIA EUCLIDIANA.....	38
<b>III.5 REDES NEURONALES .....</b>	<b>39</b>
III.5.1 ELEMENTOS DE UNA RED NEURONAL ARTIFICIAL .....	39
III.5.2 TAXONOMÍA DE LAS REDES NEURONALES.....	41
III.5.3 PERCEPTRÓN MULTICAPA.....	43
III.5.4 ALGORITMO DE RUMELHART.....	43
III.5.5 GRADIENTE DESCENDENTE.....	44
III.5.6 ALGORITMO DE RETROPROPAGACIÓN.....	45
III.5.7 MÉTODOS DE ESTIMACIÓN .....	47
<b>CAPÍTULO IV .....</b>	<b>48</b>
<b>RECUPERACIÓN DE IMÁGENES MEDIANTE LA TRANSFORMADA ONDELETA Y LA REGLA DEL VECINO MÁS CERCANO.....</b>	<b>48</b>
<b>IV.1 INTRODUCCIÓN.....</b>	<b>48</b>
<b>IV.2 METODOLOGÍA.....</b>	<b>49</b>
<b>IV.3 NORMALIZACIÓN DEL MODELO DE COLOR.....</b>	<b>51</b>
<b>IV.4 AJUSTE DEL TAMAÑO DE LA IMAGEN .....</b>	<b>51</b>
<b>IV.5 CODIFICACIÓN DE LAS IMÁGENES .....</b>	<b>52</b>
IV 5.1 IMÁGENES COMPLETAS .....	53
IV 5.2. HISTOGRAMA DE UNA VENTANA CIRCULAR .....	55
IV 5.3. HISTOGRAMAS DE SUB-IMÁGENES CUADRADAS .....	57
<b>IV.6 ALMACENAMIENTO DE LOS VECTORES CARACTERÍSTICOS.....</b>	<b>59</b>
<b>IV.7 RECUPERACIÓN MEDIANTE LA REGLA DEL VECINO MÁS CERCANO.....</b>	<b>59</b>
<b>CAPÍTULO V.....</b>	<b>60</b>
<b>RECUPERACIÓN DE IMÁGENES MEDIANTE LA TRANSFORMADA ONDELETA Y LA CLASIFICACIÓN CON REDES NEURONALES.....</b>	<b>60</b>
<b>V.1 METODOLOGÍA .....</b>	<b>60</b>
<b>V.2 CONVERSIÓN DE IMÁGENES A PATRONES .....</b>	<b>61</b>
<b>V.3 ARQUITECTURA DE LA RED NEURONAL .....</b>	<b>62</b>
<b>V.4 ENTRENAMIENTO DE LA RED NEURONAL.....</b>	<b>63</b>
<b>CAPÍTULO VI .....</b>	<b>65</b>



<b>PRUEBAS Y RESULTADOS DE LA RECUPERACIÓN DE IMÁGENES.....</b>	<b>65</b>
<b>VI.1 RESULTADOS EXPERIMENTALES MEDIANTE EL VECINO MÁS CERCANO .....</b>	<b>66</b>
VI.1.1 CRITERIO DE EFICIENCIA .....	66
VI.1.2 EXPERIMENTO 1. IMÁGENES DE ESTUDIO EN TONOS DE GRIS .....	68
EFICIENCIA .....	69
VI.1.3 EXPERIMENTO 2. IMÁGENES DE MEMORAMA.....	69
EFICIENCIA .....	71
VI.1.4 EXPERIMENTO 3. IMÁGENES DE AVIONES EN COLOR.....	71
EFICIENCIA .....	72
VI.1.5 EXPERIMENTO 4. INSTRUMENTAL QUIRÚRGICO .....	73
EFICIENCIA .....	74
VI.1.6 EXPERIMENTO 5. RETINAS HUMANAS.....	74
EFICIENCIA .....	76
VI.1.7 EXPERIMENTO 6. 96 IMÁGENES DE 6 CLASES .....	76
EFICIENCIA .....	77
VI.1.8 EXPERIMENTO 7. 192 IMÁGENES DE 6 CLASES .....	77
EFICIENCIA .....	78
<b>VI.2 RESULTADOS EXPERIMENTALES MEDIANTE REDES NEURONALES.....</b>	<b>79</b>
VI.2.1 EXPERIMENTO 8. IMÁGENES DE ESTUDIO EN COLORES .....	80
EFICIENCIA .....	81
VI.2.2 EXPERIMENTO 9. IMÁGENES DE AVIONES EN TONOS DE GRIS.....	83
EFICIENCIA .....	83
VI.2.3 EXPERIMENTO 10. IMÁGENES DE AVIONES EN COLOR.....	85
EFICIENCIA .....	86
VI.2.4 EXPERIMENTO 11. INSTRUMENTAL QUIRÚRGICO .....	87
EFICIENCIA .....	88
VI.2.5 EXPERIMENTO 12. RETINAS HUMANAS.....	88
EFICIENCIA .....	89
VI.2.6 EXPERIMENTO 13. 96 IMÁGENES DE 6 CLASES .....	89
EFICIENCIA .....	90
VI.2.7 EXPERIMENTO 14. 192 IMÁGENES DE 6 CLASES .....	91

EFICIENCIA .....	92
<b>VI.3 ANÁLISIS DE LOS RESULTADOS .....</b>	<b>92</b>
<b>CAPÍTULO VII.....</b>	<b>93</b>
<b>SISTEMAS EN LÍNEA.....</b>	<b>93</b>
<b>VII.1 CIRES .....</b>	<b>93</b>
<b>VII.2 FIRE.....</b>	<b>99</b>
<b>VII.3 IRMA .....</b>	<b>101</b>
<b>VII.4 ANÁLISIS DE LOS RESULTADOS .....</b>	<b>105</b>
<b>CAPÍTULO VIII .....</b>	<b>106</b>
<b>CONCLUSIONES Y TRABAJOS FUTUROS .....</b>	<b>106</b>
<b>VIII.1 CONCLUSIONES.....</b>	<b>106</b>
<b>VII.2 TRABAJOS FUTUROS.....</b>	<b>107</b>
<b>REFERENCIAS .....</b>	<b>109</b>
<b>APÉNDICE A. CÓDIGO DEL ALGORITMO DE INTERPOLACIÓN BILINEAL .....</b>	<b>114</b>
<b>APÉNDICE B. CÓDIGO DEL ALGORITMO DE LA RED NEURONAL .....</b>	<b>115</b>

## Lista de Figuras

Fig. 1 Imagen-consulta. ....	4
Fig. 2 Imágenes recuperadas. ....	4
Fig. 3 Segmentación del plano tiempo-frecuencia de las Ondeletas. ....	18
Fig. 4 Función de escala Daubechies. ....	22
Fig. 5 Ondeleta Daubechies 4. ....	23
Fig. 6 Magnitudes de los filtros de la Ondeleta Daubechies 4. ....	25
Fig. 7 Fase de análisis del algoritmo piramidal de la codificación sub-banda. ....	28
Fig. 8 Fase de síntesis del algoritmo piramidal de la codificación sub-banda. ....	28
Fig. 9 Codificación sub-banda de una imagen de 256 x 256 a 32 x 32. ....	30
Fig. 10 Cubo de colores RGB. ....	33
Fig. 11 Imagen y su histograma. ....	34
Fig. 12 Imagen oscura y su histograma agrupado hacia los valores bajos. ....	35
Fig. 13 Imagen clara y su histograma agrupado hacia los valores altos. ....	35
Fig. 14 Imagen con contraste bajo y su histograma angosto. ....	36
Fig. 15 Imagen gris y su histograma en la parte media. ....	36
Fig. 16 Imagen con contraste alto y su histograma ancho. ....	37
Fig. 17 Histogramas de una imagen de color. ....	37
Fig. 18 Elementos de una neurona artificial. ....	40
Fig. 19 Capas de una Red Neuronal Artificial. ....	40
Fig. 20 Perceptrón Multicapa. ....	43
Fig. 21 Metodología para la recuperación de las imágenes. ....	50
Fig. 22 Interpolación bilineal. ....	52
Fig. 23 Extracción de los coeficientes de la imagen completa. ....	54
Fig. 24 Extracción de los coeficientes del histograma de una ventana circular. ....	56
Fig. 25 Extracción de los coeficientes de los histogramas de las sub-imágenes. ....	58
Fig. 26 Conversión de imágenes a patrones. ....	62
Fig. 27 Arquitectura de la Red Neuronal Artificial. ....	63
Fig. 28 Consultas con el número de imágenes parecidas. ....	66
Fig. 29. Imágenes de estudio en tonos de gris. ....	68
Fig. 30 Consulta con imágenes de estudio en tonos de gris. ....	69
Fig. 31 Imágenes de objetos de un memorama. ....	70

Fig. 32 Consulta con imágenes de un memorama.....	71
Fig. 33 Imágenes de seis diferentes clases de aviones.....	71
Fig. 34 Consulta con imágenes de las seis clases diferentes de aviones.....	72
Fig. 35 Imágenes de instrumental quirúrgico.....	73
Fig. 36 Consulta con imágenes de instrumental quirúrgico.....	74
Fig. 37 Imágenes de retinas humanas.....	75
Fig. 38 Consulta con imágenes de retinas humanas.....	76
Fig. 39 96 imágenes de 6 clases.....	76
Fig. 40 Consulta con 96 imágenes de 6 clases.....	77
Fig. 41 192 imágenes de 6 clases.....	78
Fig. 42 Consulta con 192 imágenes de 6 clases.....	79
Fig. 43 Imágenes de estudio en color.....	80
Fig. 44. Imágenes de estudio con diferentes rotaciones.....	80
Fig. 45 Consulta con imágenes de estudio en color.....	82
Fig. 46 Imágenes de las 6 clases de aviones, en tonos de gris.....	83
Fig. 47 Consultas con imágenes de las 6 clases de aviones.....	84
Fig. 48 Conversión de las imágenes de aviones a patrones.....	85
Fig. 49 Consulta de imágenes de aviones con RN.....	86
Fig. 50 Consulta de imágenes de instrumental quirúrgico con RN.....	87
Fig. 51 Consulta de imágenes de retinas humanas con RN.....	89
Fig. 52 Consulta de 96 imágenes de 6 clases con RN.....	90
Fig. 53 Consulta de 192 imágenes de 6 clases con RN.....	91
Fig. 54 Algunas imágenes de CIRES.....	94
Fig. 55 Comparación de la precisión de recuperación.....	95
Fig. 56 Cinco consultas en CIRES.....	96
Fig. 57 Cuatro consultas en CIRES.....	97
Fig. 58 Dos consultas en CIRES a clases específicas.....	98
Fig. 59 Imágenes disponibles para consultas en FIRE.....	99
Fig. 60 Resultados de consultas en FIRE.....	100
Fig. 61 Ejemplos de radiografías.....	102
Fig. 62 Consulta en IRMA.....	103
Fig. 63 Otra consulta en IRMA.....	104

## Lista de Tablas

Tabla 1. Panorámica de trabajos relacionados.....	15
Tabla 2. Conjuntos del Método $\pi$ .....	64
Tabla 3. Resultados de 25 consultas .....	67
Tabla 4. Eficiencia de los tres métodos .....	67
Tabla 5. Eficiencia del experimento 1: imágenes de estudio .....	69
Tabla 6. Eficiencia del experimento 2: con las imágenes de un memorama.....	71
Tabla 7. Eficiencia del experimento 3: imágenes de aviones.....	72
Tabla 8. Eficiencia del experimento 4: imágenes de instrumental quirúrgico .....	74
Tabla 9. Eficiencia del experimento 5: imágenes de retinas humanas .....	76
Tabla 10. Eficiencia del experimento 6: 96 imágenes de 6 clases .....	77
Tabla 11. Eficiencia del experimento 7: 192 imágenes de 6 clases .....	78
Tabla 12. Eficiencia del experimento 8: imágenes de estudio en color con RN.....	81
Tabla 13. Eficiencia del experimento 9: imágenes de aviones en tonos de gris con RN .....	83
Tabla 14. Eficiencia del experimento 10: imágenes de aviones con RN .....	86
Tabla 15. Eficiencia del experimento 11: imágenes de instrumental quirúrgico con RN .....	88
Tabla 16. Eficiencia del experimento 12: imágenes de retinas humanas con RN.....	89
Tabla 17. Eficiencia del experimento 13: 96 imágenes de 6 clases con RN.....	90
Tabla 18. Eficiencia del experimento 14: 192 imágenes de 6 clases con RN.....	92
Tabla 19. Clases de imágenes de CIRES.....	94
Tabla 20. Características visuales de FIRE .....	99
Tabla 21. Eficiencia del sistema IRMA.....	101
Tabla 22. Comparación con sistemas en línea.....	105

# CAPÍTULO I

## INTRODUCCIÓN

El procesamiento de información involucra a menudo el reconocimiento, almacenamiento, tratamiento y recuperación de imágenes e información multimedia.

Aunque una gran cantidad de información multimedia es generada de forma continua para una variedad de aplicaciones, los sistemas de información actuales son incapaces de tratar la información multimedia de una forma efectiva. La razón es que estos sistemas son diseñados fundamentalmente 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 [1].

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

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, compactación, procesamiento y recuperación de imágenes, así como 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 no existe dificultad en el ser humano para reconocer y recuperar datos multimedia, los sistemas de información actuales presentan varios problemas, debido a que en lo fundamental están basados en el tratamiento de información alfanumérica, algunas veces expandidos con herramientas de desarrollo gráfico y con simples técnicas de diagramación y dibujo [2]. Hasta ahora, no hay muchos sistemas que han acometido las tareas de reconocimiento y recuperación de imágenes exitosamente.

Mediante este proyecto se pretende desarrollar un sistema de información visual que utilice paradigmas orientados particularmente al procesamiento de la información visual en imágenes, específicamente mediante la organización y recuperación basada en el contenido, en términos de color y forma, y su clasificación mediante redes neuronales.

## I.1. Planteamiento del problema

En los sistemas de recuperación de imágenes, debe existir la capacidad de comparar dos imágenes de forma eficiente para determinar si tienen contenido similar en los términos de las características extraídas que en general representan información discriminante.

Una vez captada la imagen, es digitalizada formando una función bidimensional de intensidad luminosa  $F(x, y)$ , que puede considerarse como una matriz cuyos índices de fila y columna identifican un punto de la imagen, donde  $(x, y)$  representa las coordenadas espaciales y el valor de  $F$  en un punto cualquiera  $(x, y)$  es proporcional al brillo, ya sea en niveles de gris, o al color compuesto por las componentes RGB en ese punto denominado píxel.

Sea  $F(x, y)$ ;  $x, y = 1, 2, \dots, N$  un arreglo de píxeles de una imagen de dos dimensiones. Para las imágenes de color,  $F(x, y)$  denota el valor del color del píxel  $(x, y)$ . Si la información de color se representa en términos de los tres colores primarios RGB (rojo, verde y azul), la función-imagen puede ser escrita como en la ecuación 1:

$$F(x, y) = \{F_R(x, y), F_G(x, y), F_B(x, y)\} \quad (1)$$

En las imágenes en blanco y negro,  $F(x, y)$  denota el valor de la intensidad en la escala de grises del píxel  $(x, y)$ . Sea  $f$  un mapeo desde el espacio imagen hacia un espacio  $n$ -dimensional  $X = (x_1, x_2, \dots, x_n)$ ; esto se muestra en la ecuación 2:

$$f : F \longrightarrow X \quad (2)$$

Donde  $n$  es el número de características que se utilizan para representar a la imagen.

La diferencia vectorial entre dos imágenes  $F_1$  y  $F_2$ , puede ser expresada como la distancia  $D$  entre los vectores de características respectivos  $X_1$  y  $X_2$ .

Por lo tanto, el problema de la recuperación de imágenes puede ser propuesto como la aplicación del método de los  $K$ -vecinos más cercanos con el siguiente criterio: Dada una imagen-consulta  $P$ , recuperar una imagen  $M$  desde una base de datos de imágenes  $L$  tal que, la ecuación 3 exprese la condición de la distancia mínima entre la imagen-consulta y la imagen recuperada:

$$D(f(P), f(M)) \leq D(f(P), f(F)) \text{ para todo } F \in L, F \neq M \quad (3)$$

En general, la computadora es incapaz de reconocer objetos en una imagen, estén aislados o traslapados, presentándose siempre, en primera instancia, la necesidad de un proceso de segmentación. Como el reconocimiento de objetos no puede hacerse directamente, los objetos deben ser reconocidos indirectamente a través de sus propiedades, usando rasgos tales como el color, la textura y la forma, entre otros.

El problema del indexado o recuperación de imágenes almacenadas en bancos de datos de imágenes, se puede afrontar básicamente de dos formas. La primera supone que se cuenta con reconocedores poderosos para identificar los objetos en las imágenes. La segunda, en caso de que no se disponga de estos reconocedores, utilizaría rasgos de los objetos presentes en la imagen con las siguientes interrogantes:

¿Qué primitivas o rasgos usar para describir cada objeto?

¿Cómo convertir un objeto en rasgos para mediante estos descomponer una imagen en sus objetos?

¿Cómo estructurar el espacio o diccionario para organizar las imágenes descritas?

¿Cómo construir las consultas para extraer las imágenes del diccionario?

El problema se puede plantear de la forma siguiente: dado un conjunto de  $N$  imágenes,  $N \gg 10$ , almacenadas en un base de datos de imágenes, deseamos desarrollar una metodología para la recuperación automática de imágenes mediante consultas, es decir, recuperar un conjunto de imágenes parecidas y si la imagen-consulta esta en la base, también recuperarla. En general, la tarea estará orientada a recuperar las imágenes que contengan objetos tales como Automóviles, Hojas de árboles, Edificios, etc.

Para ilustrar lo anterior con un ejemplo, en la Figura 1 se presenta una imagen-consulta de un tipo específico de hojas de árbol, la cual se pretende recuperar; en la Figura 2 se presentan tres imágenes recuperadas, una de ellas es la imagen-consulta.



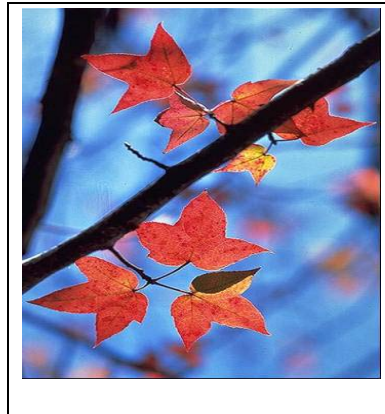


Fig. 1 Imagen-consulta.

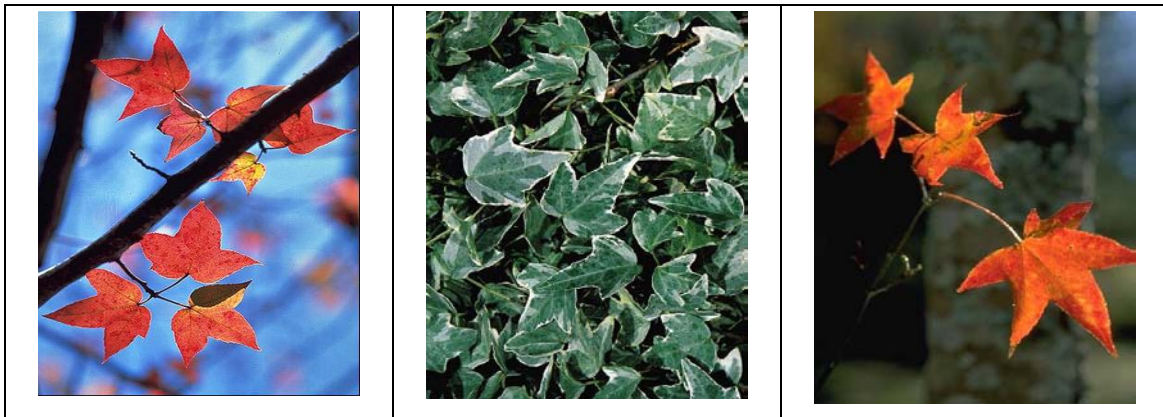


Fig. 2 Imágenes recuperadas.

## I.2. Hipótesis de trabajo

Se empiezan a desarrollar sistemas de recuperación de información visual como el denominado ImageScape, el cual es un sistema para encontrar imágenes sobre Intranets y la Internet. El sistema integra tecnologías tales como compresión de la base de datos de imágenes por cuantificación vectorial y árboles de decisión de  $k$  dimensiones, para búsquedas rápidas sobre espacios multidimensionales; además, permite consultas basadas en dibujos e iconos [3].

Actualmente se están investigando nuevos paradigmas para búsquedas de información visual, que incluyan iconos, dibujos y similitud entre imágenes. La futura investigación se enfocará en la fusión de técnicas de aprendizaje visual, tales como Redes Neuronales y Árboles de Decisión, entre otras, para resultar en la detección correcta de conceptos visuales, además de la indexación y recuperación de imágenes basada en su contenido en términos de color, forma, textura, etc.

Como ya se tienen algunos resultados teóricos y prácticos, según los trabajos relacionados que utilizan características visuales de bajo nivel tales como el color, la forma y la textura, además de la indexación de imágenes mediante ondeletas, se pretende obtener la estructura del espacio de imágenes, aplicar un clasificador mediante redes neuronales, procesar consultas y recuperar información visual en forma robusta y confiable sobre la base del contenido de las imágenes.

### **I.3. Justificación**

El principal objetivo que persigue una máquina de búsqueda es encontrar eficientemente textos en la Internet, ya sea en bibliotecas digitales, Intranets, en bases de datos de imágenes, etc., pero con poca o ninguna capacidad de encontrar imágenes o vídeos con contenido diverso.

Muchos usuarios de la Internet, editores de periódicos, o diseñadores profesionales de sitios Internet, necesitan en ocasiones encontrar imágenes a partir de unas pocas características globales. Con cientos de millones de sitios para buscar y con el 73 por ciento de la Internet dedicada a imágenes [3], encontrar la imagen exacta puede constituir una tarea muy compleja.

De aquí se observa la importancia de esta línea de investigación. La relevancia del problema y la dificultad para lograr de forma robusta y confiable los objetivos, radica en la gran cantidad y variedad de la información visual existente en la Internet, la que en general no está estructurada.

Por lo anterior, en el desarrollo de una metodología para la recuperación de imágenes en bancos de datos de imágenes, debe hacerse hincapié, primero en la estructuración del espacio de imágenes, y segundo, en las tecnologías de información a utilizar en el proceso de recuperación de las imágenes.

### **I.4. Objetivos**

El objetivo general y los específicos son los siguientes:

Objetivo general:

Desarrollar una metodología mediante la transformada Ondeleta para la recuperación de imágenes en bases de datos de imágenes, sobre la base de su contenido en términos de forma y color, y su clasificación mediante la regla de los vecinos más cercanos y las redes neuronales.

Objetivos específicos:

1. Probar con imágenes de estudio e imágenes complejas.
2. Realizar búsquedas a través de imágenes-consulta.
3. Desarrollar un algoritmo para la recuperación de imágenes mediante el análisis de características globales.

4. Desarrollar un algoritmo para recuperación de imágenes mediante el análisis de características locales.
5. Recuperar imágenes mediante el algoritmo del vecino más cercano.
6. Recuperar imágenes mediante redes neuronales.
7. Hacer las comparaciones pertinentes.

## **I.5. Restricciones**

Para el desarrollo de este trabajo se han impuesto las siguientes restricciones:

- Las imágenes reales son imágenes donde los objetos se tocan, se traslapan y son complejas. Se debe considerar como una limitación propia del proyecto, que los problemas básicos de reconocimiento de patrones no han sido aún resueltos para la imágenes complejas.
- Las formas de expresar las consultas será exclusivamente mediante imágenes-consulta, con el fin de eliminar cualquier posibilidad de error o ambigüedad que daría lugar a confusiones en los resultados.
- Los tipos de imágenes de la base de datos serán cada vez preestablecidos, por ejemplo: imágenes de automóviles, edificios, paisajes de bosques, montañas, aviones, carreteras, etc.

## **I.6. Aportaciones**

La idea de combinar diferentes tecnologías tales como la transformada ondeleta y redes neuronales con los algoritmos heurísticos para extraer coeficientes conforman una aportación original.

Acerca de los algoritmos heurísticos para obtener los coeficientes de la transformada ondeleta se evalúan tres maneras diferentes, dos de las cuales son aportaciones desarrolladas en este trabajo, las cuales consisten en:

- 1) Para calcular dichos coeficientes, se toman los histogramas de la ventana circular más grande dentro de cada uno de los canales de color RGB.
- 2) Consiste en obtener los coeficientes a partir de los histogramas de los canales de color RGB de sub-imágenes cuadradas en los que se ha dividido la imagen original.

## **I.7. Publicaciones hechas**

- A. C. González G., J. H. Sossa A. y E. M. Felipe R., "Images retrieval using Wavelets, Histograms and Subimages", 5ª Conferencia Internacional en Control, Instrumentación Virtual y Sistemas Digitales 2004, CIC - IPN, México., 403-411, 2004.

- A. González, H. Sossa, E. Felipe and O. Pogrebnyak., “Wavelet transforms and neural networks applied to image retrieval”. International Conference on Pattern Recognition (ICPR). Hong Kong, China, 909-912, 2006.
- A. González, H. Sossa, E. Felipe and O. Pogrebnyak., “Histograms, Wavelets and Neural Networks Applied to Image Retrieval”. 5<sup>th</sup> Mexican International Conference on Artificial Intelligence, Apizaco, Mexico., Lecture Notes in Artificial Intelligence, Springer 820-827, 2006.
- A. C. Gonzalez, H. Sossa and E. M. Felipe (2007). Image Retrieval based on Wavelet Computation and Neural Network Classification. 8th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 07), Santorini, Greece, June 6-8, 2007. ISBN 0-7695-2818-X.

## **I.8. Organización de la tesis**

Este documento está organizado de la siguiente forma: en el Capítulo 1 se presenta la recuperación de imágenes como problema a resolver. En el Capítulo 2 se describe su contexto a través del análisis del estado del arte. En el Capítulo 3 se expone el marco teórico de referencia de las alternativas para la recuperación de imágenes, entre ellas la transformada Ondeleta, así como la regla de los vecinos más cercanos y la teoría que sustenta las redes neuronales que se utilizan como clasificadores. En el Capítulo 4 se presentan tres alternativas de solución del problema planteado mediante la transformada Ondeleta Daubechies y la regla del vecino más cercano. En el Capítulo 5 se presenta una solución que incluye la clasificación mediante redes neuronales. En el Capítulo 6 se brindan las pruebas y resultados, que muestran la eficiencia de los métodos propuestos a través de los resultados logrados en varios experimentos. En el Capítulo 7 se presenta un análisis comparativo con tres sistemas en línea. En el Capítulo 8 se detallan las conclusiones, además de los trabajos futuros. Se presenta la relación de las referencias utilizadas en este documento. Finalmente los apéndices contienen el código fuente de los algoritmos de interpolación bilineal y red neuronal.

## CAPÍTULO II

# PANORÁMICA DE TRABAJOS RELACIONADOS

La recuperación de información visual es un área de investigación de la Tecnología de la Información. Tiene como propósito recuperar imágenes independientes o una secuencia de imágenes que son relevantes para una consulta dada, como una extensión de la recuperación de la información tradicional que incluye medios visuales. Los elementos visuales tales como el color, la textura y la forma de los objetos, además de su relación espacial, los significados de los objetos y las imágenes en las escenas, son usados como pistas para recuperar, desde una base de datos de imágenes, aquellas con contenido similar.

La tecnología emergente de la Multimedia y la posibilidad de compartir y distribuir datos de imágenes y video a través de redes de computadoras con un ancho de banda grande, hace evidente la necesidad de una herramienta para recuperar información visual confiablemente.

### **II.1. Líneas de investigación**

A continuación se presenta una lista de algunas líneas de investigación dentro de la recuperación de la información visual [2], [4]:

#### **1. Interfaces visuales para el acceso a bases de datos**

Ofrece la posibilidad de explorar espacios de información visual. Los atributos visuales son vinculados a funciones de distancia apropiadas; cada imagen se mapea en un punto de un espacio multidimensional; puntos cercanos pertenecen a imágenes similares; se combinan consultas de texto con información visual.

## **2. Modelos y estándares de representación**

Es necesario contar con modelos de datos y organizar estructuras de conocimiento para suministrar recuperación efectiva, ya que la recuperación de información visual se puede apreciar en varios niveles de abstracción, desde las pistas de bajo nivel, como lo es la recuperación en base a su contenido en términos de color, forma y textura, hasta el aspecto semántico.

El estándar MPEG-4 suministra soporte para composición de información audiovisual y representación de medios sintéticos en ambientes multimedia. MPEG-7 especifica un conjunto estándar de descriptores de información multimedia, además de estructuras de esquemas de descriptores y sus relaciones.

## **3. Modelos de bases de datos**

Con respecto a los modelos para almacenar información visual, varios sistemas utilizan archivos para almacenar imágenes y video. Con el sistema de archivos se tiene una conexión cerrada entre la aplicación, el modelo de almacenamiento y el método de acceso y se tiene una implementación sencilla, y con el modelo de datos administrado directamente por la aplicación.

Los sistemas de bases de datos ofrecen una separación entre la aplicación, el modelo de almacenamiento y el método de acceso, además de una interfaz para múltiples aplicaciones. Extensiones de bases de datos relacionales tienen definidos tipos de datos y operaciones con datos visuales; básicamente, se investiga en bases de datos orientadas a objetos.

## **4. Extracción automática de características de niveles bajo y medio**

Se lleva a cabo mediante el análisis de imágenes, que incluye la recuperación de imágenes basada en el contenido, en términos de color, forma y textura; además, el reconocimiento de patrones y la visión por computadora, son herramientas que suministran algoritmos para la extracción de las características.

## **5. Extracción automática de significado semántico**

Trata lo concerniente a las relaciones lógicas entre los objetos y el dominio que representa la imagen. Una perspectiva consiste en la identificación de primitivas semánticas: objetos, reglas, acciones y eventos, como abstracciones de signos semánticos.

Otra perspectiva es combinar asociaciones semánticas de signos visuales, por ejemplo, emociones e impresiones relacionadas con la presencia y combinación de colores, efectos inducidos por la sincronización y el orden espacial de los colores, etc. Además, usan el análisis semiótico para estudiar las relaciones de los sentidos con signos tales como cortes de cámara, colores, efectos de edición, ritmos, ángulos y líneas de tomas en una secuencia de video. Las características de bajo nivel son extraídas y trasladadas a semánticas de alto nivel mediante relaciones entre ellas.

## **6. Indexación de datos visuales**

Las estructuras de indexación están estrechamente relacionadas con la representación de los datos. Si los elementos visuales son caracterizados por vectores de rasgos, entonces se utiliza una estructura de indexación multidimensional. Varias estructuras de índice están enfocadas a ejecutar un filtrado en lugar de una clasificación, lo que es adecuado para la recuperación de la información visual. También es necesario cambiar funciones de similitud en tiempo real, asignando diferentes pesos a cada dimensión.

## **7. Modelos de similitud para características sensitivas**

Los modelos de similitud se basan en la evaluación de una función de distancia propia en una métrica del espacio característico. Se asume que un conjunto de características modela las propiedades sensitivas para poder ser representadas como puntos en un espacio característico adecuado. Además de la combinación actual de los modelos vectoriales y las distancias métricas, se sugiere investigar en modelos de similitud de juicio humano.

## **8. Subjetividad del usuario**

La recuperación se realiza a través de sesiones interactivas con el usuario. Las imágenes se obtienen como respuesta a una consulta previa con el optimismo de converger a una solución satisfactoria, pero cada nueva consulta por el mismo usuario empieza desde cero, por lo que se busca que el sistema tenga memoria para aprender y actualizarse durante las sesiones interactivas y se adapte a la subjetividad y a las necesidades del usuario.

## **9. Búsqueda y recuperación en la Internet**

La búsqueda, el acceso y la transmisión de información visual son tareas importantes en la Internet. La aparición de la Internet y el rápido decrecimiento del costo de los discos duros, han hecho posible el almacenamiento y la diseminación masiva de la información, mucho más de la que puede leer un ser humano en su vida. Los sitios Internet convencionales y las bibliotecas digitales, contienen archivos de documentos de lenguaje de marcaje de hipertexto, imágenes, video y otros archivos multimedia.

Una diferencia significativa es que las bibliotecas digitales contienen generalmente los archivos completos de una sola institución, por lo que se pueden organizar y estructurar para un acceso más rápido. Sin embargo, los sitios de la Internet convencional, contienen información de diferente índole, por lo que intentar organizarla y estructurarla es un trabajo prácticamente imposible; además, la búsqueda de imágenes y video presenta innumerables problemas que aún no han sido resueltos.

Existe la necesidad de un estándar para la representación de la información visual meta-data que permita interoperabilidad entre diferentes fuentes de información y sistemas de búsqueda en la red. Cuando se buscan imágenes y

videos existe un compromiso entre la calidad y el tiempo de respuesta. Algunas áreas de trabajo son:

- a) Organización jerárquica de conceptos asociados con el contenido visual, por ejemplo: el uso de características visuales simples como el color, la forma, la textura, la reducción de requerimientos computacionales para la indexación, etc.;
- b) Información resumida para la navegación y las consultas;
- c) Análisis y recuperación para trabajar en el dominio comprimido;
- d) Visualización con diferentes niveles de resolución.

### **10. Recuperación de imágenes tridimensionales y video**

Aplicaciones como la Biomedicina, el Diseño Asistido por Computadora, la Cartografía, la Restauración de Pinturas, los Catálogos de Arte, etc., se confinan a colecciones de imágenes tridimensionales y de video; se ofrecen pocos resultados en el modelado y la recuperación en estas áreas; además, tiene el video tridimensional como área incipiente de investigación.

La recuperación de información visual o recuperación de imágenes basada en su contenido ha tenido un amplio desarrollo dentro de la visión por computadora, gracias al auge de la multimedia, específicamente con el estudio de las técnicas de consultas a grandes bases de datos de imágenes mediante imágenes-consulta. El unir las características del color y la forma en el establecimiento de la similitud, frente a otros trabajos, permite recuperar imágenes similares a partir de consultas mediante una imagen-clave.

## **II.2. Trabajos relacionados con la forma y el color**

A continuación se presentan algunos trabajos desarrollados en los últimos años, que tienen una relación directa con el presente proyecto y con una o varias de las líneas de investigación anteriores, que han servido como base para su desarrollo.

Talagala et al., en [5], reportan un ejemplo de sistema para bibliotecas digitales que se denomina Thinker ImageBase, desarrollado en la Universidad de California, en Berkeley, para el Museo de Bellas Artes de San Francisco. Este sistema almacena 75000 imágenes con resolución de 3072 x 2048 píxeles, en un medio de almacenamiento de 3.2 terabytes para un cluster de 20 computadoras personales Intel P3 a 200 MHz, con 368 discos duros de 8 Gigabytes, interconectadas a través de switches Ethernet de 100 Mbps. La explosión en línea da lugar a dos preguntas: 1. ¿Cómo mantener el sistema de almacenamiento masivo? y 2. ¿Cómo se presentan los datos al usuario? Con respecto al sistema de almacenamiento se presenta un rendimiento de más de un 95% de disponibilidad en los accesos con éxito; la forma de presentar las imágenes al usuario es con la opción de diferentes resoluciones que van de 12.5% a 100%, según se necesiten las imágenes con mayor o menor detalle.



Ing - Sheen Hsieh y Kuo - Chin Fan, en [6], reportan del 93% al 97% de eficiencia en una base de datos de imágenes de 183 banderas de colores representadas en el modelo de color RGB, con 128 x 128 píxeles cada una y en formato BMP. Se presenta un sistema de recuperación de imágenes basado en regiones de color que usa propiedades geométricas. Las regiones de color se forman con una técnica de crecimiento de regiones que agrupa a los píxeles conectados que tienen el mismo color. De cada región de color se extraen dos características descriptivas: un grafo de relación espacial y los coeficientes descriptores de Fourier. Las regiones de color contienen información de color y relaciones topológicas de cada píxel en la imagen; de esta manera se incluyen las propiedades del color y las propiedades geométricas perceptibles de forma y relación espacial. Los grafos de relación espacial son utilizados para medir la distancia entre dos imágenes; los coeficientes descriptores de Fourier se usan para emparejar la forma geométrica de las regiones de color; estos utilizan puntos críticos que se detectan mediante la transformada Ondeleta de la función Gaussiana, que toma como entrada los puntos del contorno de las regiones y produce los puntos de inflexión como puntos críticos. Se obtiene un sistema invariante a la rotación, a la escala y a la traslación.

Para la recuperación de las imágenes sobre la base de la forma de los objetos, Adoram y Lew, en [10], usan el método de contornos activos adaptados al movimiento natural de las víboras; además, para mejorar dicho método aplican el algoritmo de Flujo del Vector Gradiente, ya que se adapta a concavidades en el objeto además de que no es necesario conocer a priori cuándo la víbora debe ser expandida o contraída para adaptarse al contorno. Primero realizan la segmentación de los objetos mediante contornos activos y después el cálculo de los momentos invariantes, los que comparan con un clasificador de distancia mínima con posibilidades de cambios en la rotación, escala, traslación y ruido.

### **II.3. Trabajos relacionados con los histogramas**

Liapis - Tziritas, en [7], presentan resultados con 210 imágenes de 384 x 256 píxeles del conjunto de datos de Corel Photo Gallery; reporta una eficiencia del 93% combinando características de textura y color, 52% cuando utiliza solo textura y 83% cuando utiliza solo el color; las características de textura son extraídas mediante Análisis de Marcos de Ondeletas Discretas con una descomposición completa en escala y orientación, que descompone la imagen en componentes ortogonales para obtener invariancia a la translación; los histogramas de las coordenadas de cromaticidad son usadas como características de color y utiliza la distancia de Bhattacharya como medida de similitud.

Trabajar en la recuperación de imágenes basada en su contenido en términos de la forma, implica la segmentación de los objetos, además de su tamaño, hechos estos que dificultan el procesamiento; González et al., en [8], presentan los momentos invariantes de Hu y Maitra utilizados para el emparejamiento u homogenización de las escenas. También Wu Yan et al., en [9], han desarrollado técnicas para la recuperación basadas en la forma en que utilizan los momentos

invariantes de Hu. Se definen algunas características que son invariantes a la traslación, rotación, cambio de escala y cambio en los niveles de gris; con la información de bordes obtenida mediante el operador de Sobel y el método de Otsu para obtener el umbral, combinado con los momentos invariantes de dimensión uno y los momentos centrales calculados mediante la ecualización del histograma, se obtienen buenos resultados.

El uso de características globales de la imagen para su recuperación sobre la base de su contenido, se presenta también en los trabajos de Jain - Vailaya [11], con una eficiencia en la recuperación de imágenes del 85%; el color en la imagen es representado mediante tres histogramas, uno por cada componente de color R (Rojo), G (Verde) y B (Azul). El método se enriquece con un esquema de intersección de histogramas normalizados; como esquema métrica se utiliza la distancia Euclidiana para la recuperación de imágenes basada en el color. Para describir la información de la forma contenida en una imagen se usa un histograma de dirección de los bordes, los cuales se obtienen con el operador de Canny, antes de utilizar la intersección de los histogramas para la recuperación basada en la forma. Los atributos de color y forma son integrados combinando valores de semejanza asociados a pesos que determinan el grado de discriminación del color y la forma.

El histograma es una representación de características útil, porque es un resultado estadístico que posee méritos de simplicidad, robustez y eficiencia; los histogramas de color contienen la información de la distribución de los colores en cada plano. Por lo tanto, las imágenes con pequeños cambios de traslación, rotación y distorsión, poseen una alta similitud en sus correspondientes histogramas de color. Algunas ideas sobre la codificación de imágenes mediante histogramas se presentan en el trabajo de Fan-Di Jou et al. [12] Donde se propone un algoritmo para emparejar histogramas de imágenes de gran tamaño con más de  $1024 \times 1024$  píxeles; el número de características y la resolución de cada característica se determinan por el tamaño del histograma; el algoritmo que se propone funciona de la misma forma con histogramas pequeños y grandes y se le puede aplicar a funciones de medida de similitud de histogramas tales como la función de intersección de histogramas, la norma  $L_1$ , la norma  $L_2$  y la prueba  $X^2$ . Con este algoritmo se obtiene un rendimiento del 93.5% en la recuperación de las imágenes.

Puzicha et al. [13] reportan un método para la recuperación de imágenes y una clasificación no supervisada mediante el agrupamiento de histogramas, con un rendimiento del 93.38%. Con respecto a la recuperación de imágenes basada en su contenido en términos de color, se incluye el indexado de color mediante la intersección de histogramas. Los histogramas de color son invariantes a la traslación y a la rotación, y si se trabaja con el histograma normalizado se obtiene invariancia a la escala. Una deficiencia de todos los histogramas globales, inclusive los de color, es que no indican en el espacio la localización de los píxeles de la imagen. Para subsanar esta debilidad se utilizan distribuciones locales de los coeficientes de Gabor y el agrupamiento de distribuciones de histogramas.

M. K. Mandal y T. Aboulnasr [14] reportan una eficiencia hasta del 87% mediante la transformación por ondeletas de los histogramas. Se reportan tres métodos basados en la extracción de coeficientes de ondeletas de los histogramas de las imágenes, para recuperar imágenes en base de su contenido en términos de textura, donde para obtener nueve canales de textura se sobre muestrean los coeficientes de ondeletas de las 9 sub-imágenes de tres niveles del esquema multi-resolución. Precisamente, la variación entre cada método consiste en reducir la complejidad del algoritmo original que considera todos los niveles de resolución de los coeficientes de ondeletas, al considerar como alternativas los tres principios siguientes: el primero consiste en seleccionar un solo nivel, el segundo en generar menos niveles de descomposición del algoritmo multi-resolución, y el tercero utiliza la resolución del histograma.

Worthington–Hancock en [15] presentan una estrategia de reconocimiento consistente en un método de bajo nivel que usa histogramas de atributos de la imagen derivados de sus hojas de mapas. Los atributos utilizados son: histogramas de escala de grises, histogramas de índice de forma, histogramas de ángulos de inclinación, histogramas de índices de forma y curvatura, histogramas de curvatura máxima y mínima, histogramas de curvatura media y Gaussiana, histogramas de co-ocurrencia de escala de grises e histogramas de co-ocurrencia de índices de forma. El objetivo principal es explorar cuándo la información topográfica derivada de la intensidad de imágenes 2D usando la forma de la sombra, puede ser usada para el reconocimiento de objetos 3D.

#### **II.4. Trabajos relacionados con las ondeletas y las redes neuronales**

Existen trabajos importantes en el campo de la recuperación de imágenes mediante ondeletas que han proporcionado los fundamentos adecuados para la creación de una aplicación basada en una red neuronal.

En lo que respecta a la extracción de características mediante ondeletas se presentan métodos de codificación en los trabajos de Wang-Bovik [16], quienes reportan una muy buena codificación en imágenes de 512 x 512 píxeles con una compresión de radio de 512:1 y una razón de bits por píxel de 0.015625, alcanzando una óptima percepción visual con una compresión de radio de 32:1 y una razón de 0.25 bits por píxel.

Sarlashkar et al. [17] para entrenar una red neuronal, ofrecen como resultado que las características extraídas usando ondeletas son invariantes a la escala en un 100% y también invariantes al ruido en niveles bajos, por ejemplo, con la adición de 20% de ruido aleatorio. Lo demuestra con imágenes de diferentes características muy populares en el procesamiento de imágenes (mandril, Lena y la imagen de una escena arbitraria) recuperando las imágenes en un 100% con un ruido aleatorio de menos del 20%; si se incrementa el ruido el rendimiento baja sustancialmente.

Traina et al. [18] presentan un sistema de recuperación de imágenes mediante transformadas ondeletas comparando sus características de textura, el sistema calcula la distancia Euclidiana entre la imagen consulta y todas las imágenes de la base de datos y presenta las 16 más cercanas. El sistema se entrena con 7 clases de 30 imágenes cada una, y se prueba con 290 imágenes que no están en la base de datos original, alcanzando una eficiencia del 85%.

Soo Beom Park et al. [20] reportan en su trabajo un rendimiento efectivo en la recuperación de imágenes del 81.7% en una prueba con 300 patrones de entrenamiento, y un rendimiento de 76.7% con 300 patrones de prueba con 10 imágenes de cada una de las 30 clases.

## II.5. Análisis de los trabajos

En la Tabla 1 se presenta un estudio comparativo de algunos desarrollos presentados en la panorámica de trabajos relacionados. Se observa la diversidad de metodologías utilizadas que tratan de reconocer la forma, el color y la textura. Las bases de datos más utilizadas son de unidades de cientos y van desde imágenes sencillas de caricaturas, banderas y marcas con eficiencias de recuperación muy altas pero cuando se utilizan imágenes complejas dicha eficiencia disminuye. Con respecto al año de desarrollo se observa un crecimiento en la complejidad de las imágenes junto con el incremento de la eficiencia obtenida. Claro esta que no son los únicos trabajos desarrollados pero son los que más se relacionan con la metodología propuesta para esta tesis.

Tabla 1. Panorámica de trabajos relacionados.

Metodología	Número de imágenes	Eficiencia en %	Año de desarrollo
Color mediante intersección del histograma y forma mediante la dirección de bordes [11]	400 imágenes de marcas	97	1996
Histogramas y calibración multiescala de la energía [13]	1000 imágenes	94	1998
Ondeletas y red neuronal [17]	60 imágenes populares	90	1998
Forma mediante contornos activos [10]	1500 caricaturas	80	1999
Bandas de escala de ondeletas de los histogramas [14]	220 imágenes	87	1999
Propiedades geométricas de regiones de color [6]	183 Banderas	93	2000
Forma mediante atributos del histograma [15]	560 imágenes de estudio	98	2000
Segmentación de regiones, textura mediante ondeletas y redes neuronales [18]	300 imágenes, 30 clases de 10 imágenes	77	2002

Tabla 1. Panorámica de trabajos relacionados (continuación)

Histogramas de color muy grandes [12]	1038 imágenes de marcas	94	2003
Color mediante histogramas de cromaticidad y textura mediante ondeletas [7]	210 imágenes de textura	93	2004
Agrupamiento basado en ondeletas [19]	3167 imágenes en 60 grupos	87	2005
Análisis de componentes principales de ondeletas [21]	1008 imágenes, 56 clases de 18 rostros de una sola persona	98	2006

Una vez observada una panorámica de trabajos relacionados, comprendemos que el estado del arte completo de las diferentes alternativas de la solución del problema de recuperación de imágenes en base a su contenido es muy amplio, por la diversidad de las líneas de investigación que participan en los diferentes trabajos presentados, pero queda claro que la metodología propuesta en este proyecto de tesis, que trata de recuperación de imágenes en base a su contenido en términos de forma y color, mediante ondeletas de los histogramas de ventanas circulares y de histogramas de sub-imágenes es original.

A continuación, en el Capítulo III, se presentan las bases teóricas de las diferentes áreas del conocimiento que participan en la solución particular que presentamos. Se comienza con la teoría de las transformadas ondeletas como herramienta matemática de la representación de la firma de una imagen de color a través de sus histogramas, se muestra el método multi-resolución porque es la esencia algorítmica que se utiliza. Para clasificar las firmas se utilizan dos opciones, el método del vecino más cercano y las redes neuronales.

## CAPÍTULO III

# MARCO TEÓRICO DE REFERENCIA

La recuperación de imágenes basada en el contenido posee la habilidad de recuperar imágenes 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 generación de la firma utilizando una transformación de la imagen para obtener un vector de características como una representación numérica simplificada, sirve para almacenar la firma en una base de datos y así acelerar el proceso de recuperación de imágenes, ya que la transformación caracteriza de forma efectiva propiedades globales de la imagen, tales como la forma y el color.

A continuación se presenta la teoría de la transformada Ondeleta, ubicada como la base matemática de la metodología a utilizar, que comprende la transformada Ondeleta y la Ondeleta Daubechies; además, también se valoran en la recuperación de imágenes los métodos multi-resolución, el modelo de color, la regla de los vecinos más cercanos y la teoría de las redes neuronales.

### III.I. La Transformada Ondeleta

La transformada Ondeleta, se utiliza para la localización de información en tiempo-frecuencia; básicamente tiene la capacidad de moverse de una escala a otra, por lo que es adecuada para el análisis de señales no estacionarias. La transformada Ondeleta consiste de un conjunto de funciones base que representan la señal en diferentes bandas de frecuencia, cada una con resoluciones distintas correspondientes a cada escala.

#### III.1.1 La Transformada Ondeleta Continua

La transformada Ondeleta, analiza la señal a diferentes frecuencias con diferentes resoluciones, es decir, cada componente espectral no es analizada con la misma

resolución. La transformada Ondeleta utiliza ventanas estrechas en frecuencias altas y ventanas anchas en bajas frecuencias.

El resultado es un análisis multi-resolución, en el cual la señal se analiza con diferentes resoluciones a diferentes frecuencias. De esta forma la resolución temporal y la resolución frecuencial varían en el plano tiempo-frecuencia sin violar el Principio de Incertidumbre de Heisenberg [23]. En la Transformada Ondeleta, cuando la frecuencia aumenta, la resolución temporal disminuye y cuando la frecuencia disminuye, la resolución temporal aumenta.

Debido a que la transformada Ondeleta incluye información relacionada con el tiempo y la frecuencia, la representación gráfica de esta transformada se realiza en un plano denominado plano tiempo-frecuencia, tal como se muestra en la Figura 3.

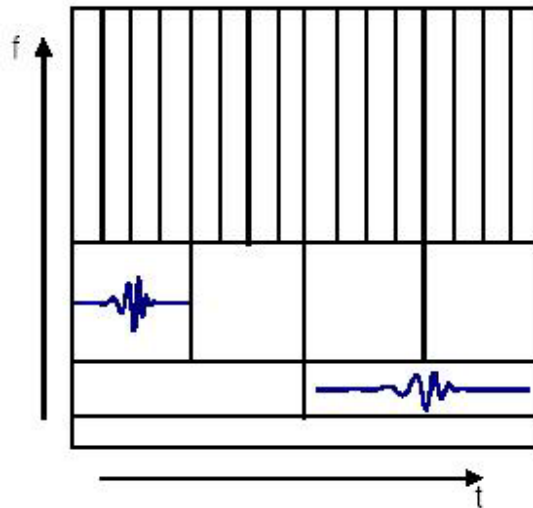


Fig. 3 Segmentación del plano tiempo-frecuencia de las Ondeletas.

Cada celda en la Figura 3 representa una porción igual del plano tiempo-frecuencia, pero dando diferentes proporciones al tiempo y a la frecuencia, aunque el área sea constante. A baja frecuencia, la altura del rectángulo es menor, lo que implica una mayor resolución de frecuencia; pero la anchura es grande, lo que corresponde con una pobre resolución temporal. A alta frecuencia la anchura del rectángulo decrece y la altura aumenta, i.e., la resolución frecuencial es menor y la resolución temporal es mayor. De esta forma se obtiene una resolución 'fina' en frecuencia y 'gruesa' en tiempo para señales de baja frecuencia y una resolución 'fina' en tiempo y 'gruesa' en frecuencia para señales de alta frecuencia.

La transformada Ondeleta se puede interpretar como una técnica de ventanas con tamaños variables que tienen bases de soporte compacto, donde el dominio es no-cero en un intervalo finito, y cero en el resto [21]. Al considerar una señal

continua  $f(t)$ , la definición de la transformada Ondeleta está dada por la ecuación 4.

$$C(a, b) = \int_{-\infty}^{\infty} f(t) \Psi_{a,b}(t) dt \quad (4)$$

La Transformada Ondeleta de la ecuación 4 depende de dos parámetros que cambian continuamente en el conjunto de los números reales correspondiendo a la escala y el tiempo respectivamente:

El parámetro  $a$ , escala, es definido como [1/frecuencia] y corresponde a la información frecuencial. El escalado es una operación matemática que dilata o comprime una señal. Las escalas pequeñas corresponden a señales comprimidas (alta frecuencia) y las escalas grandes a señales dilatadas (baja frecuencia). Si  $a > 1$  la señal se contrae y si  $a < 1$  la señal se expande.

El parámetro  $b$ , traslación, está relacionado con el desplazamiento de la ventana a lo largo del dominio temporal de la señal y por tanto este término corresponde a la información temporal de la señal.

La Transformada Ondeleta Continua proporciona una medida de la correlación o similitud entre la Ondeleta a diferentes escalas y la señal con la escala, o frecuencia, como medida de similitud entre ambas. La Transformada Ondeleta Continua se calcula cambiando la escala de la ventana de análisis, desplazándola en el tiempo, multiplicándola por la señal e integrando en el tiempo. Puesto que la transformación se consigue dilatando y trasladando la Ondeleta madre en forma continua, se genera información redundante.

Las funciones  $\Psi(t)$  se llaman ondeletas y se definen en la ecuación 5.

$$\Psi_{a,b}(t) = a^{-\frac{1}{2}} \Psi\left(\frac{t-b}{a}\right) \quad (5)$$

Las Ondeletas deben satisfacer la condición de ortogonalidad definida en la ecuación 6.

$$\Psi_{j,k}(t) \cdot \Psi_{m,n}(t) = \begin{cases} 1 & \text{si } j=m \text{ y } k=n \\ 0 & \text{en caso contrario} \end{cases} \quad (6)$$

Para que  $\Psi(t)$  pueda ser Ondeleta madre debe satisfacer las siguientes condiciones:

1. Tiene que ser continua e integrable.
2. Tiene que tener una Transformada de Fourier  $\Psi(w) = 0, \forall w < 0$ .



3. Tiene que cumplir que  $\Psi(t)$  tenga media cero, tal que

$$\int_{-\alpha}^{\alpha} \Psi(t) dt = 0 \quad \text{y por lo tanto la Ondeleta es una oscilación,}$$

$$\text{además: } \int_{-\alpha}^{\alpha} |\Psi(w)|^2 / w dw < \alpha$$

Se usan los términos de alta (baja) frecuencia para hacer referencia a los valores bajos (altos) del parámetro de escala. Esta relación inversa entre la escala y la frecuencia es bastante intuitiva; se puede pensar que una escala baja corresponde con una Ondeleta comprimida, donde los detalles cambian rápidamente, semejante al comportamiento de la señal en alta frecuencia.

La función original  $f$  puede reconstruirse a partir de su transformada Ondeleta continua por evaluación de la transformada ondeleta inversa:

$$f(t) = \frac{1}{C_{\psi}} \int_{-\alpha}^{\alpha} \int_{-\alpha}^{\alpha} \frac{1}{a^2} C(a,b) \psi_{a,b}(t) da db \quad (7)$$

Donde la constante  $C_{\psi}$  satisface la condición de admisibilidad dada por:

$$C_{\psi} = 2\pi \int_0^{\alpha} \frac{|\psi(w)|^2}{w} dw < \alpha \quad (8)$$

### III.1.2 La Transformada Ondeleta Discreta

La diferencia clave en el análisis discreto es que los parámetros de escala  $a$  y de translación  $b$  en la ecuación 5 no son continuos, sino enteros. En la mayoría de casos, la elección de  $a$  y  $b$  está limitada al siguiente conjunto de discretización:

$$(j, k) \in \mathbb{Z}^2; \quad a = 2^j; \quad b = k2^j = ka \quad (9)$$

Los índices  $a$  y  $b$  en  $\Psi_{a,b}(t)$  se reemplazan por  $j$  y  $k$ , respectivamente, así la ecuación 5 puede escribirse en formato discreto como:

$$\Psi_{j,k}[n] = 2^{\frac{-j}{2}} \Psi \left[ 2^{-j} n - k \right] \quad (10)$$

Para señales discretas, la transformada Ondeleta se muestra en la ecuación 11.

$$C [j, k ] = \sum_{n \in Z} f [n ] \Psi_{j, k} [n ] \quad (11)$$

Una base ondeleta la componen dos funciones: la función de escalado llamada ondeleta padre  $\varphi(n)$  y la ondeleta madre  $\Psi(n)$ . Las otras ondeletas se generan mediante la base por translación de la función escalado; y la dilatación y la translación de la ondeleta madre utilizando las relaciones:

$$\varphi_{j_0, k} [n] = 2^{\frac{-j_0}{2}} \Psi [2^{-j_0} n - k], \quad \Psi_{j, k} [n] = 2^{\frac{-j}{2}} \Psi [2^{-j} n - k], \quad j = j_0, j_0 + 1, \dots \quad (12)$$

La reconstrucción de la señal  $f(n)$  se puede desarrollar en series en las cuales se utilizan las funciones  $\varphi(n)$  y  $\Psi(n)$ .

$$f(n) = \sum_k c_k \varphi_{j_0, k} [n] + \sum_k \sum_{j=j_0}^{\alpha} d_{jk} \Psi_{j, k} [n] \quad (13)$$

En la ecuación 13, los coeficientes  $c_k$  se denominan coeficientes de aproximación en la escala  $j_0$  y los coeficientes  $d_{jk}$  representan el detalle de la señal a diferentes escalas.

Las dilataciones y las translaciones de la Ondeleta madre forman una base del espacio de detalle  $d_{jk}$  que se encuentra tras aplicaciones de filtros pasa altos, mientras que las funciones de escalado forman una base del espacio de las aproximaciones  $c_k$  mediante aplicaciones de filtros pasa bajos.

Las formas extensas están representadas por el primer término de la ecuación 13 con las funciones de escalado  $\varphi(n)$  que corresponde a las aproximaciones en el espacio. Las Ondeletas  $\Psi(n)$  representan oscilaciones localizadas en las posiciones próximas al punto  $k2^{-j}$  en las frecuencias cercanas a  $2^j$  y, así pues, presenta los detalles de la señal. Este tipo de análisis recibe el nombre de multiresolución o análisis multiescalar.

### III.1.3 La Ondeleta Daubechies

Un tipo de función base de ondeleta propuesta por Daubechies, se basa en que para cada entero  $r$ , la base ortonormal de  $L^2(\mathbb{R})$  se encuentra definida como se muestra en la ecuación 14 [24], [25]:

$$\Phi_{r, j, k} (x) = 2^{j/2} \Phi_r (2^j x - k) \quad j, k \in Z \quad (14)$$

En dicha ecuación la función  $\Phi_r$  en  $L^2(\mathbb{R})$  tiene la propiedad de que  $\Phi_r(x-k) | k \in \mathbb{Z}$  es una base secuencial ortonormal, donde  $j$  es la escala,  $k$  es la traslación y  $r$  es un filtro. Por lo tanto, el término  $f_j$  en la escala de  $2^{-j}$  de una función  $f \in L^2(\mathbb{R})$  se define por la ecuación 15:

$$f_j(x) = \sum_k \langle f, \Phi_{r,j,k} \rangle \Phi_{r,j,k}(x) \quad (15)$$

Los detalles o fluctuaciones para la ecuación 15 se definen en la ecuación 16:

$$d_j(x) = f_{j+1}(x) - f_j(x) \quad (16)$$

Para analizar los detalles en una escala determinada, se define una base ortonormal  $\Psi_r(x)$  con propiedades similares a las de  $\Phi_r$ . Donde conjunto de

funciones es ortonormal si  $\int_a^b \phi_k(t) \cdot \phi_l^*(t) dt = 0 \quad k \neq l$  y

$\int_a^b |\phi_k(t)|^2 dt = 1$  O, equivalentemente,  $\int_a^b \phi_k(t) \cdot \phi_l^*(t) dt = \delta_{kl}$  con  $\delta_{kl} = \begin{cases} 1 & \text{si } k = l \\ 0 & \text{si } k \neq l \end{cases}$

que es la función delta de Kroneker.

### Función de escala

La función de escala se muestra en la Figura 4

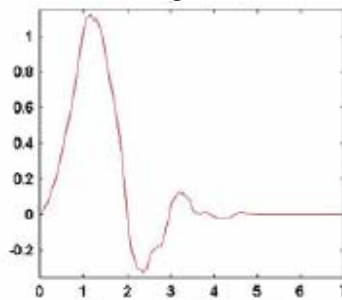


Fig. 4 Función de escala Daubechies.

La función de escala Daubechies se denota por  $\phi_3^D$ , se define por la relación doble escala:

$$\begin{aligned}\phi_3^D(x) &= \sum_{k=0}^3 p_k \phi(2x-k) \\ &= \frac{1+\sqrt{3}}{4} \phi_3^D(2x) + \frac{3+\sqrt{3}}{4} \phi_3^D(2x-1) + \frac{3-\sqrt{3}}{4} \phi_3^D(2x-2) + \frac{1-\sqrt{3}}{4} \phi_3^D(2x-3)\end{aligned}\quad (17)$$

Los valores diferentes de cero de la secuencia doble escala  $\{p_k\}$  son  $\{p_0, p_1, p_2, p_3\} = \left\{ \frac{1+\sqrt{3}}{4}, \frac{3+\sqrt{3}}{4}, \frac{3-\sqrt{3}}{4}, \frac{1-\sqrt{3}}{4} \right\}$ . Se observa que los coeficientes  $\{p_k\}$  tienen la propiedad  $p_0 + p_2 = p_1 + p_3 = 1$ . En general la secuencia doble escala para cualquier función de escalamiento tiene la propiedad:

$$\sum_k p_{2k} = \sum_k p_{2k+1} = 1 \quad (18)$$

### Ondeletas

La Figura 5, muestra la forma de onda de la Ondeleta Daubechies.

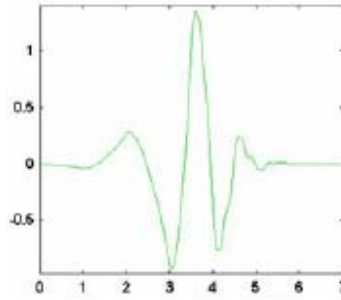


Fig. 5 Ondeleta Daubechies 4.

La relación doble escala para la Ondeleta Daubechies es de la forma:

$$\psi(x)_3^D = \sum_k q_k \phi_3^D(2x-k) \quad (19)$$

Donde:

$$q_k = (-1)^k \bar{p}_{-k+1} \quad (20)$$

Y  $\bar{p}_{-k+1}$  es el complejo conjugado de  $p_{-k+1}$ . Se observa que los coeficientes doble escala  $\{p_k\}$  son reales para  $\phi_3^D(x)$ , entonces se tiene que  $\bar{p}_{-k+1} = p_{-k+1}$ , luego los valores diferentes de cero de la secuencia doble escala  $\{q_k\}$  son:

$$\{q_{-2}, q_{-1}, q_0, q_1\} = \{p_3, -p_2, p_1, -p_0\} = \left\{ \frac{1-\sqrt{3}}{4}, -\frac{3-\sqrt{3}}{4}, \frac{3+\sqrt{3}}{4}, -\frac{1+\sqrt{3}}{4} \right\} \quad (21)$$

y la fórmula explícita para (19) es:

$$\psi(x)_3^D = \sum_{k=-2}^1 q_k \phi_3^D(2x-k) \quad (22)$$

$$= \frac{1-\sqrt{3}}{4} \phi_3^D(2x+2) - \frac{3-\sqrt{3}}{4} \phi_3^D(2x+1) + \frac{3+\sqrt{3}}{4} \phi_3^D(2x) - \frac{1+\sqrt{3}}{4} \phi_3^D(2x-1)$$

La característica especial de la familia Ondeleta Daubechies es la presencia de ortogonalidad en combinación con soporte compacto. Para lograr esto, la Ondeleta debe ser asimétrica y de baja regularidad.

### Relaciones de descomposición

Las relaciones de descomposición para la Ondeleta Daubechies están dadas por:

$$\phi_3^D(2x-l) = \sum_k [a_{l-2k} \phi_3^D(x-k) + b_{l-2k} \psi_3^D(x-k)], \quad l \in \mathbb{Z} \quad (23)$$

Donde:

$$a_{l-2k} = \frac{1-p}{2} p_{l-2k} \quad (24)$$

$$b_{l-2k} = \frac{1-q}{2} q_{l-2k} \quad (25)$$

Dado que solamente cuatro  $\{p_k\}$  y cuatro  $\{q_k\}$  son diferentes de cero, la relación de descomposición (23) se escribe de la siguiente manera:

$$\phi_3^D(2x-l) = \frac{1}{2} p_2 \phi_3^D(x - \frac{l-2}{2}) + \frac{1}{2} p_0 \phi_3^D(x - \frac{l}{2}) + \frac{1}{2} q_0 \psi_3^D(x - \frac{l}{2}) + \frac{1}{2} q_{-2} \psi_3^D(x - \frac{l+2}{2}), \quad l \text{ par} \quad (26)$$

$$\phi_3^D(2x-l) = \frac{1}{2} p_3 \phi_3^D(x - \frac{l-3}{2}) + \frac{1}{2} p_1 \phi_3^D(x - \frac{l-1}{2}) + \frac{1}{2} q_1 \psi_3^D(x - \frac{l-1}{2}) + \frac{1}{2} q_{-1} \psi_3^D(x - \frac{l+1}{2}), \quad l \text{ impar} \quad (27)$$

Además:

$$\phi_3^D(2x) = \frac{3-\sqrt{3}}{8} \phi_3^D(x+1) + \frac{1+\sqrt{3}}{8} \phi_3^D(x) + \frac{3+\sqrt{3}}{8} \psi_3^D(x) + \frac{1-\sqrt{3}}{8} \psi_3^D(x-1) \quad (28)$$

$$\phi_3^D(2x-1) = \frac{1-\sqrt{3}}{8} \phi_3^D(x+1) + \frac{3+\sqrt{3}}{8} \phi_3^D(x) - \frac{1+\sqrt{3}}{8} \psi_3^D(x) - \frac{3-\sqrt{3}}{8} \psi_3^D(x-1) \quad (29)$$

Los coeficientes de los pares de filtros ortonormales para la descomposición y reconstrucción en el caso de la función de la Ondeleta Daubechies 4 pueden ser expresados según [26] como:

$$G(n) = \left[ \frac{1-\sqrt{3}}{4\sqrt{2}}, -\frac{3-\sqrt{3}}{4\sqrt{2}}, \frac{3+\sqrt{3}}{4\sqrt{2}}, -\frac{1+\sqrt{3}}{4\sqrt{2}} \right] \quad H(n) = \left[ \frac{1+\sqrt{3}}{4\sqrt{2}}, \frac{3+\sqrt{3}}{4\sqrt{2}}, \frac{3-\sqrt{3}}{4\sqrt{2}}, \frac{1-\sqrt{3}}{4\sqrt{2}} \right] \quad (30)$$

$$\bar{G}(n) = \left[ -\frac{1+\sqrt{3}}{4\sqrt{2}}, \frac{3+\sqrt{3}}{4\sqrt{2}}, -\frac{3-\sqrt{3}}{4\sqrt{2}}, \frac{1-\sqrt{3}}{4\sqrt{2}} \right] \quad \bar{H}(n) = \left[ \frac{1-\sqrt{3}}{4\sqrt{2}}, \frac{3-\sqrt{3}}{4\sqrt{2}}, \frac{3+\sqrt{3}}{4\sqrt{2}}, \frac{1+\sqrt{3}}{4\sqrt{2}} \right] \quad (31)$$

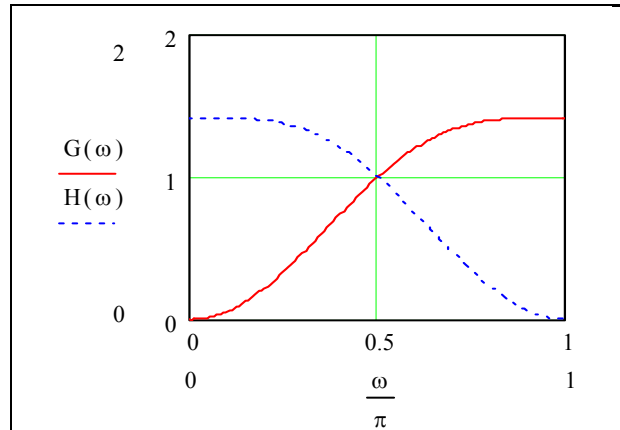


Fig. 6 Magnitudes de los filtros de la Ondelela Daubechies 4.

La Figura 6 muestra las magnitudes de la respuesta en frecuencia de los filtros de análisis y síntesis de la Ondelela Daubechies 4 dados por las ecuaciones (30) y (31). Se nota que las magnitudes de las respuestas en frecuencia de los filtros de análisis dados por la ecuación (30) son los mismos que las magnitudes de los filtros de síntesis de la ecuación (31). Esto es debido a la propiedad de ortonormalidad de la Ondelela Daubechies 4 y su función de escala asociada.

### ¿Por qué Daubechies 4?

La Ondelela Daubechies con  $r > 2$ , presenta una concentración de energía que preserva la tendencia de la información cuando se considera solamente como un filtro pasa bajas; en nuestro caso se trabaja con  $r = 4$ .

Se selecciono la Ondelela Daubechies 4 debido al buen rendimiento que presenta en el procesamiento de señales [46], [47] y [48]. Además muestra dos características favorables, que son buena precisión y costo computacional moderado, y que se explican a continuación:

El resultado de la transformada Ondelela produce un sub-muestreo alisado, que se calcula a partir de la función de escala referida como filtro pasa bajos y una versión sub-muestreada de la señal que refleja cambios entre elementos de la señal, calculada a partir de la función Ondelela referida como un filtro pasa altos.

Por ejemplo, la transformada Haar no tiene traslape entre pares sucesivos de sus funciones de escala y ondeleta, y la Ondelela Daubechies si tiene traslapos en sus dos funciones [25].

La función Ondelela Haar, o sea el filtro pasa alto produce un resultado que solo refleja la diferencia entre elementos impares y elementos pares, La diferencia entre un elemento impar y el elemento par sucesivo no se ven reflejados en los coeficientes de la banda calculada por medio de un simple paso del filtro pasa alto, aunque en pasos posteriores se reflejaran. En contraste existe un traslape entre filtros pasa altos sucesivos de la Ondelela Daubechies, lo que ocasiona que

cualquier cambio entre dos elementos sucesivos es reflejado en el resultado inmediato. Por lo anterior se afirma que la Ondeleta Daubechies es más precisa que la Ondeleta Haar, aunque su costo computacional es del doble [28].

Al buscar alternativas de Funciones Ondeletas se encontró el problema del costo computacional, ya que otras familias de Ondeletas, al manejar más de 4 coeficientes de los filtros, se incrementa demasiado el costo computacional y la precisión no se incrementa sustancialmente. Aseverando que la Ondeleta Daubechies es de buena precisión y costo computacional moderado.

### III.1.4. Métodos multi-resolución

El análisis multi-resolución o algoritmo piramidal, se desarrolló para descomponer señales de tiempo discreto. La idea es obtener una representación tiempo-escala de una señal discreta. En este caso, filtros con distintas frecuencias de corte son usados para analizar la señal en diferentes escalas. La señal se pasa a través de filtros pasa altas para analizar las componentes de alta frecuencia, y se pasa a través de filtros pasa bajas para analizar las componentes de baja frecuencia.

El proceso de descomposición comienza pasando la secuencia discreta correspondiente a la señal a través de un filtro pasa bajas de media banda con respuesta al impulso  $h[n]$ . El filtrado de la señal corresponde a la operación matemática de convolución de dicha señal con  $h[n]$ .

Este filtro elimina las componentes de frecuencia situadas por encima de la mitad del ancho de banda de la señal. La frecuencia más elevada que existe en la señal discreta es  $\pi$  radianes, si la señal se muestrea a la frecuencia de Nyquist, la cual es el doble de la frecuencia máxima que existe en la señal. Después de filtrar la señal mediante el filtro pasa bajas, la mitad de las muestras pueden ser eliminadas siguiendo la regla de Nyquist, ya que la señal tiene su frecuencia máxima en  $\pi/2$  radianes en lugar de tenerla en  $\pi$  radianes.

Simplemente desechando una muestra de cada dos, se submuestrea la señal por dicho factor, consiguiendo que la señal tenga la mitad de puntos. La escala de la señal es duplicada. Se debe tener en cuenta que el filtrado pasa bajas elimina las componentes de alta frecuencia, pero no modifica la escala; el submuestreo es el que puede cambiar la escala.

La resolución, por otro lado, está relacionada con la cantidad de información presente en la señal, y por tanto, es afectada por el proceso de filtrado. El filtro pasa bajas elimina la mitad de las componentes en frecuencia, lo cual puede ser interpretado como la pérdida de la mitad de la información. Por lo tanto, la resolución pasa a la mitad tras ser filtrada. Sin embargo, el submuestreo no afecta a la resolución, ya que este filtrado hace que la mitad de las muestras sean redundantes, por lo que una mitad de las muestras pueden ser eliminadas sin pérdida de información.

En la Figura 7 se observa el banco de filtros en octavas. Se muestra la fase de análisis, donde  $H$  representa un filtro media banda pasa bajas con respuesta al impulso y  $G$  un filtro media banda pasa altas; además, se aprecian independientemente las etapas correspondientes a la descomposición, mediante los filtros  $G$  y  $H$  y la reducción de la mitad de las muestras.

El contenido de baja frecuencia es la parte importante, ya que proporciona la identidad de la señal, que corresponde a las aproximaciones burdas de la señal. Sin embargo el contenido de alta frecuencia caracteriza el matiz de la señal, que corresponde a los detalles de la señal misma.

Como se aprecia en la Figura 7, en cada nivel la señal de entrada se pasa por los filtros  $H$  y  $G$ . Aunque el resultado del filtrado pasa altas constituye los detalles del nivel de descomposición que se está aplicando, este no es reexaminado y el resultado del filtrado pasa bajas se puede seguir descomponiendo, de manera que conforma la aproximación de la señal a cierto nivel.

La forma más compacta de describir el proceso de submuestreo, es la representación de los filtros en forma de operador. Así, para una secuencia  $f[n]$  que representa la señal discreta a ser descompuesta, los operadores  $H$  y  $G$  se definen según las expresiones siguientes:

$$(Hf)_k = \sum_k h[n - 2k]f[n] \quad (32)$$

$$(Gf)_K = \sum_K g[n - 2k]f[n] \quad (33)$$

Estas ecuaciones representan el filtrado de la señal mediante los filtros digitales  $h[n]$  y  $g[n]$ . El factor  $2k$  representa el submuestreo. Los operadores  $H$  y  $G$  corresponden a una banda en la descomposición multinivel conocida como ramificación o árbol de descomposición ondeleta. Así, la transformada puede ser resumida en una única línea como:

$$f \rightarrow (Gf, GHf, GH^2f, \dots, GH^{j-1}f, H^j f) = (d^{(j-1)}, d^{(j-2)}, \dots, d^{(1)}, d^{(0)}, c^0) \quad (34)$$

Donde a la parte derecha de la ecuación se le denomina coeficientes de detalle, excepto  $c^0$  que corresponde al coeficiente de la aproximación.

Los detalles y las aproximaciones se obtienen de forma iterativa como:

$$c^{(j-1)} = Hc^{(j)} \text{ y } d^{(j-1)} = Gd^{(j)} \quad (35)$$



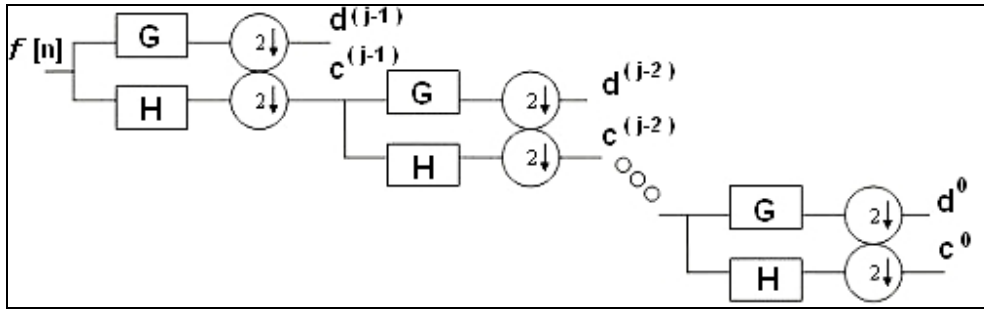


Fig. 7 Fase de análisis del algoritmo piramidal de la codificación sub-banda.

En la Figura 8 se muestra el proceso de reconstrucción, que es similar al de descomposición. La señal en cada nivel es interpolada por dos, pasada por dos filtros de síntesis representados por los operadores  $\bar{H}$   $\bar{G}$  (pasa altas y pasa bajas, respectivamente), cuyas respectivas salidas son sumadas. Por lo tanto, se definen los dos operadores adjuntos  $\bar{H}$  y  $\bar{G}$ , como:

$$(\bar{H}f) = \sum_K h[n-2k]f[n] \quad (36)$$

$$(\bar{G}f) = \sum_K g[n-2k]f[n] \quad (37)$$

La aplicación recursiva lleva a:

$$f = \sum_{j=0}^{n-1} (\bar{H})^j \bar{G}d^{(j)} + (\bar{H})^n c^{(0)} \quad (38)$$

Donde:

$$D_j = (\bar{H})^j \bar{G}d^{(j)} \quad \text{y} \quad C = (\bar{H})^n c^{(0)} \quad (39)$$

Se hará referencia a  $D_j$  como los detalles y a  $C$  como la aproximación.

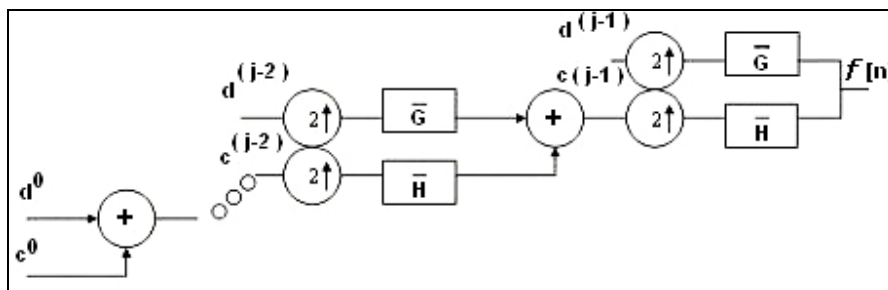


Fig. 8 Fase de síntesis del algoritmo piramidal de la codificación sub-banda.

Existen tres formas típicas de codificación de las imágenes: Transformadas o Descomposición lineal, Cuantificación y Codificación de la entropía; ahora se profundizará en la descomposición sub-banda mediante las transformadas y los métodos multi-resolución.

Las transformadas aplicadas a la descomposición sub-banda suministran una excelente compactación de energía y localización espacio-frecuencia de datos de imágenes sub-banda, ya que 1) existen diferencias significativas en la energía total entre diferentes sub-bandas y 2) dentro de cada sub-banda la energía tiende a ser agrupada espacialmente [28].

En los métodos multi-resolución se aplica la transformada Ondeleta para la recuperación de imágenes por medio de imágenes-consulta. Para codificar la imagen se extraen los coeficientes descomponiendo la imagen en bandas de baja frecuencia y bandas de alta frecuencia. La transformada Ondeleta es esencialmente una operación de convolución, que equivale a pasar la imagen a través de un filtro pasa bajas y un filtro pasa altas [27].

Debido a la normalización del espacio funcional en el diseño de la ondeleta base, los coeficientes en las bandas de frecuencia más bajas tienden a ser más dominantes y de magnitud mayor que los coeficientes de las bandas de frecuencia más altas [28].

Para obtener la información contenida en las imágenes, no bastan dos bandas de frecuencias baja y alta, sino se debe hacer una descomposición en más niveles para poder separar las características y así analizarlas independientemente. De aquí surge la idea de los filtros multiniveles, iterando el proceso de filtrado, es decir, aplicar sucesivamente el mismo procedimiento a las imágenes de salida de la etapa anterior hasta el nivel de precisión que se desee.

Esto da origen a la descomposición multinivel conocida como ramificación o árbol de descomposición ondeleta, Al ser descompuesta la imagen en una mayor cantidad de bandas de frecuencia, se posee información más detallada acerca de la misma, debido a lo cual esta metodología es conocida como multi-resolución.

La Figura 9 muestra la codificación sub-banda de una imagen o descomposición multinivel en aproximaciones, mediante la aplicación sucesiva de un filtro pasa bajas  $H$ . Se muestra la imagen con diferentes tamaños de píxeles para que se observe que la imagen no pierde su identidad con los tamaños 256 x 256, 128 x 128, 64 x 64 y 32 x 32 píxeles.

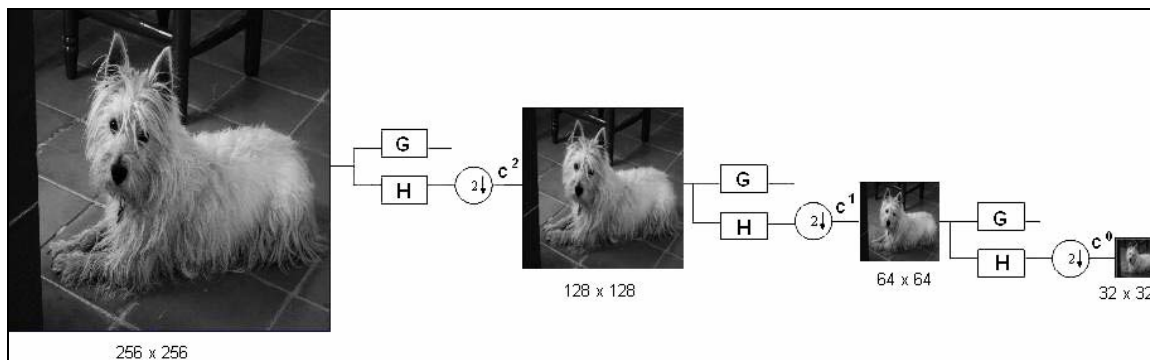


Fig. 9 Codificación sub-banda de una imagen de 256 x 256 a 32 x 32.

La codificación de imágenes sub-banda para imágenes naturales con regiones de actividad alta, tales como los bordes, son agrupados y separados desde los alrededores de las regiones de actividad baja; las regiones se pueden detectar mediante distribuciones con variancias grandes y pequeñas, respectivamente, asignando cuantificadores apropiados [17] [18].

Una descomposición multi-resolución permite obtener una representación invariante a la escala de una imagen. Los cambios de escala de una imagen son proporcionales a la distancia entre la escena y el centro óptico de la cámara. Cuando la escala de una imagen es cambiada, la interpretación de la imagen no tiene ningún cambio en su forma, tal y como se observa en la Figura 9.

La transformada Ondeleta presenta la característica favorable de preservar la forma a través de la localización espacial de los píxeles, presentando invariancia a la escala al pasar de un nivel de resolución a otro. Con ella se nota básicamente la misma distribución en las diferentes bandas de frecuencia, así como la misma forma con resoluciones distintas.

La apariencia de un objeto dentro de una imagen incluye características propias del objeto tales como la forma, el tamaño, el color y la textura, además de la ubicación, rotación y orientación del objeto en la misma imagen. El objeto y su fondo constituyen la apariencia de la imagen [55].

La recuperación de imágenes en base a su contenido busca y recupera imágenes similares en términos de su apariencia, o sea invariantes a la rotación, escala y translación.

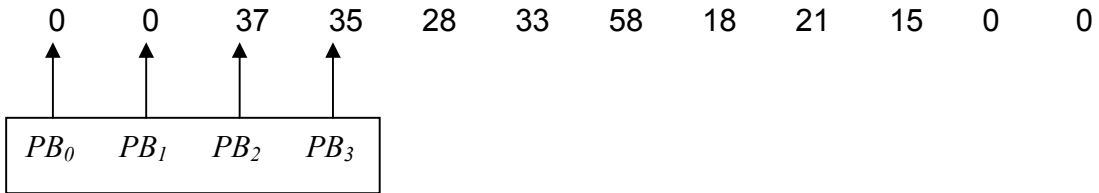
Si como técnica de modelado y recuperación de imágenes se utiliza la transformada Ondeleta para descomponer las imágenes en aproximaciones y detalles, las aproximaciones representan la componente principal de la imagen, entonces se conserva la apariencia, porque lo que se obtiene es la misma imagen pero a una escala menor, que sigue conservando la misma apariencia con cierto grado de fidelidad por el efecto de alizamiento, tal y como se observa en la Figura 9.

### III.1.5. Algoritmo discreto de la Transformada Ondeleta

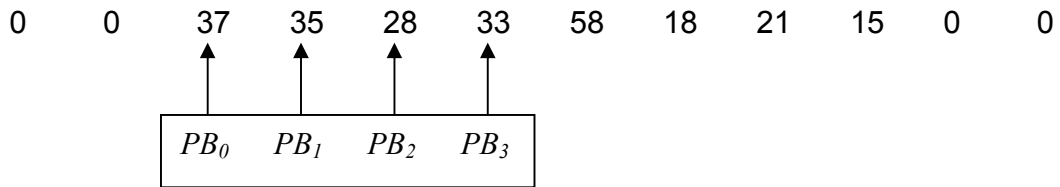
De acuerdo a la Ondeleta Daubechies 4, se aplica la convolución discreta de la imagen con los filtros pasa bajas que se definen como  $PB_n$ . Para ello, se agregan 2 ceros al inicio y al final del vector. Por ejemplo, si el vector de una fila o columna es:

37 35 28 33 58 18 21 15

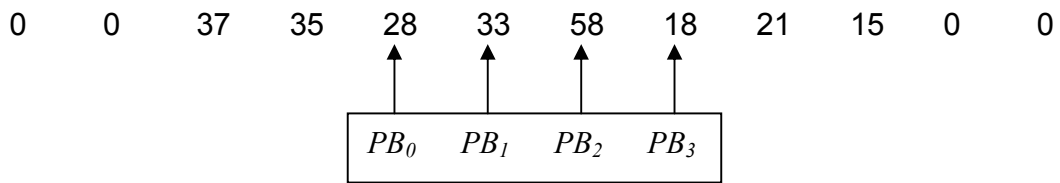
Quedaría para Daubechies 4 de la forma siguiente:



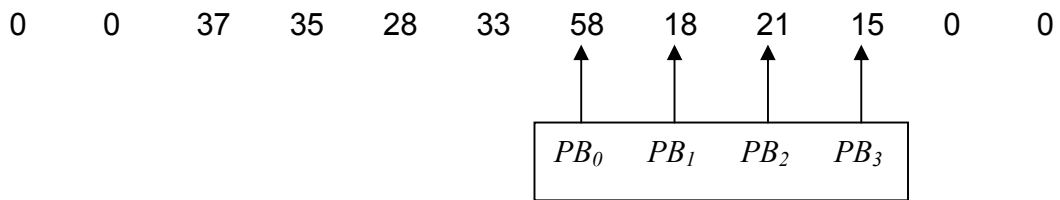
$(PB_0 \times 0) + (PB_1 \times 0) + (PB_2 \times 37) + (PB_3 \times 35)$  es el 1<sup>er</sup> Coeficiente de Aproximación.



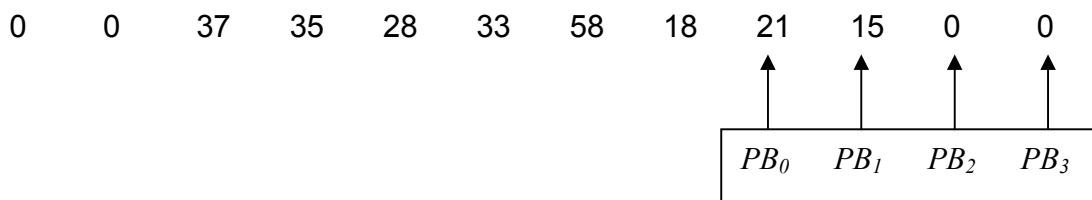
$(PB_0 \times 37) + (PB_1 \times 35) + (PB_2 \times 28) + (PB_3 \times 33)$  es el 2<sup>do</sup> Coeficiente de Aproximación.



$(PB_0 \times 28) + (PB_1 \times 33) + (PB_2 \times 58) + (PB_3 \times 18)$  es el 3<sup>er</sup> Coeficiente de Aproximación.



$(PB_0 \times 58) + (PB_1 \times 18) + (PB_2 \times 21) + (PB_3 \times 15)$  es el 4<sup>to</sup> Coeficiente de Aproximación.



$(PB_0 \times 21) + (PB_1 \times 15) + (PB_2 \times 0) + (PB_3 \times 0)$  es el 5<sup>to</sup> Coeficiente de Aproximación.

Así se obtienen 5 coeficientes de aproximación para las bajas frecuencias. El proceso anterior se lleva a cabo solo para el primer nivel; para más niveles se trabaja en forma sucesiva con los coeficientes que se obtienen en el nivel precedente. El sub-muestreo se logra al realizar el salto entre dos muestras como se observa en el proceso anterior.

Los coeficientes de detalle relacionados con las altas frecuencias, se obtienen en forma similar, pero con los 4 coeficientes correspondientes a los filtros pasa altas.

## III.2 Modelos de color

Se han postulado varias teorías para explicar el sistema de visión humano en lo que se refiere al color, desde los experimentos de Newton y Maxwell, hasta el clásico modelo tricromático de Thomas Young en 1802. El modelo tricromático asume que el ojo posee tres tipos de sensores que operan cada uno con diferentes longitudes de onda. Con los trabajos de P. K. Brown en 1964 [8], se estableció la evidencia fisiológica directa de la existencia de los tres tipos de sensores.

W. Frei en 1974 [8] propuso un modelo diferente con tres receptores, llamados conos de absorción rojo, verde y azul. Cada receptor tiene un espectro de sensibilidad, dos sensores espectrales sensitivos y un sensor que representa la absorción de pigmentos en la retina.

El modelo de Frei permite modelar matemáticamente el sistema de visión humano bajo condiciones ideales; sin embargo, la percepción del color es relativa y se conoce como adaptación cromática: "El color percibido por una distribución espectral de energía depende del área de visión y del estado de adaptación visual. Un espectador humano puede adaptarse notablemente bien a la iluminación que lo rodea, para captar la visión de una escena, y esencialmente normalizar la visión a algún equilibrio para captar el color de la escena".

Con la creación de los modelos de color cada vez más complejos y precisos, aparecieron los estándares de los modelos de color, los que fueron adoptados por los fabricantes de los monitores de computadora.

Entre los diversos modelos se encuentran los de percepción uniforme, por ejemplo  $L^*a^*b^*$ , y los modelos de percepción uniforme aproximada tal como el HSV (Matiz, Saturación y Valor), el HSI (Matiz, Saturación e Intensidad), HSB (Matiz, Saturación, Brillantez) y HSL (Matiz, Saturación y Luminosidad); a ambos tipos de modelos también se les llama independientes del dispositivo; la eficiencia de estos modelos es superior al modelo más comúnmente utilizado, que es el modelo de color RGB (Rojo, Verde y Azul), el cual no es de percepción uniforme [40].

Los modelos  $L^*a^*b^*$  y HSV, junto con sus modelos derivados se recomiendan para tratar imágenes en color, donde el color representa la característica discriminante

principal, por ejemplo, selección de frutas maduras: manzana, mango, etc., e igualación de color en pintura automotriz, entre otras aplicaciones [54].

Particularmente, el modelo RGB se convierte en el estándar para el despliegue de las imágenes de color en los monitores de color.

En este trabajo utilizamos las tres bandas de color de la imagen en el modelo de color RGB, porque como ya se mencionó es el estándar y el más utilizado.

En la Figura 10 se presenta el modelo RGB ideal representado en un sistema de coordenadas cartesianas normalizado. Las imágenes RGB están formadas por tres planos de imagen independientes, uno para cada color primario. El negro está en el origen y a través de una diagonal se disponen todos los grises hasta llegar al blanco. Todos los colores se representan en un cubo ideal normalizado que toma para el blanco las coordenadas (1, 1, 1). Los colores ciano, magenta y amarillo se encuentran respectivamente en los vértices opuestos a los denotados para los colores rojo, verde y azul, respectivamente [8].

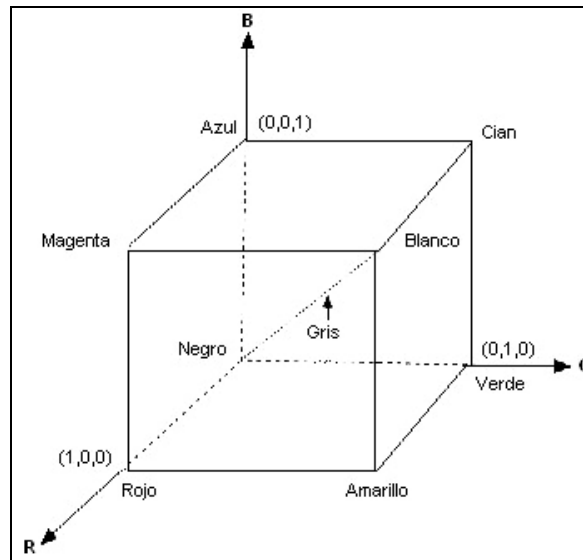


Fig. 10 Cubo de colores RGB.

### III.3 Histograma

El histograma de una imagen es una función discreta que representa el número de píxeles en la imagen en función de los niveles de intensidad. En otras palabras, el histograma de una imagen digital con niveles de gris en el rango  $[0, L - 1]$ , donde  $L$  es el número total de niveles de gris, es una función discreta de la forma que se expresa en la ecuación 40.

$$P(r_k) = n_k / n \quad (40)$$

Donde:  $r_k$  es el  $k$ -ésimo nivel de gris,  
 $n_k$  es el número de píxeles de la imagen con ese nivel de gris,  
 $n$  es el número total de píxeles de la imagen, y  
 $k = [0, 1, \dots, L - 1]$ .

En forma general, se dice que  $P(r_k)$  representa una estimación de la probabilidad de ocurrencia de un nivel de gris  $r_k$ .

La gráfica de un histograma se construye disponiendo los niveles de gris en el eje de las abscisas y las probabilidades correspondientes  $p(r_k)$  en el eje de las ordenadas. En la Figura 11 se presenta una imagen en niveles de gris y su correspondiente histograma.

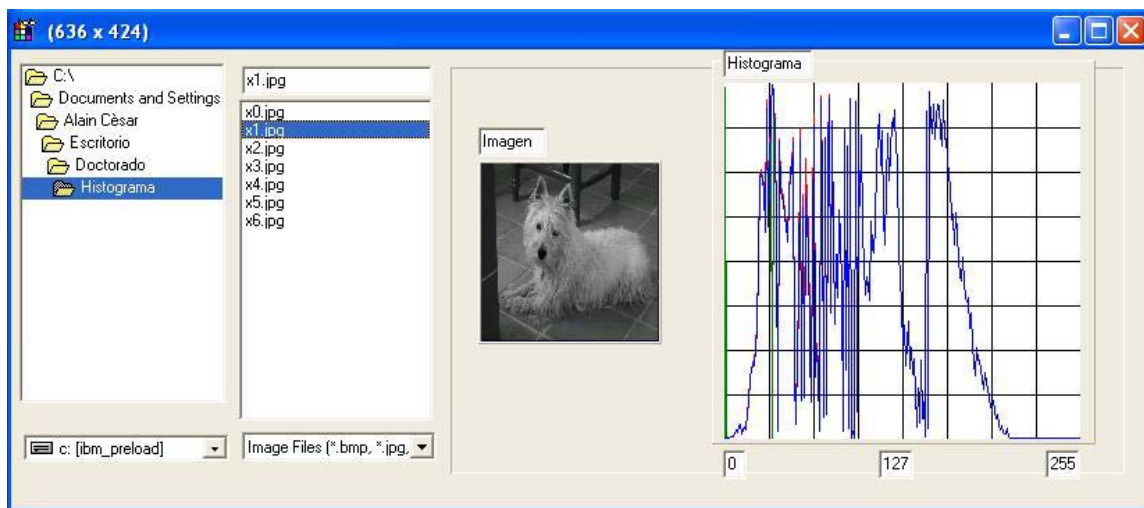


Fig. 11 Imagen y su histograma.

Una gráfica de esta función para todos los valores de  $k$ , da una descripción global de la apariencia de la imagen. A continuación se presentan como ejemplos algunos casos:

- **Imagen oscura**

Cuando el histograma muestra que los niveles de gris de los píxeles de la imagen están concentrados hacia la parte inferior del rango de valores, estos corresponden a una imagen con características oscuras, tal como se observa en la Figura 12

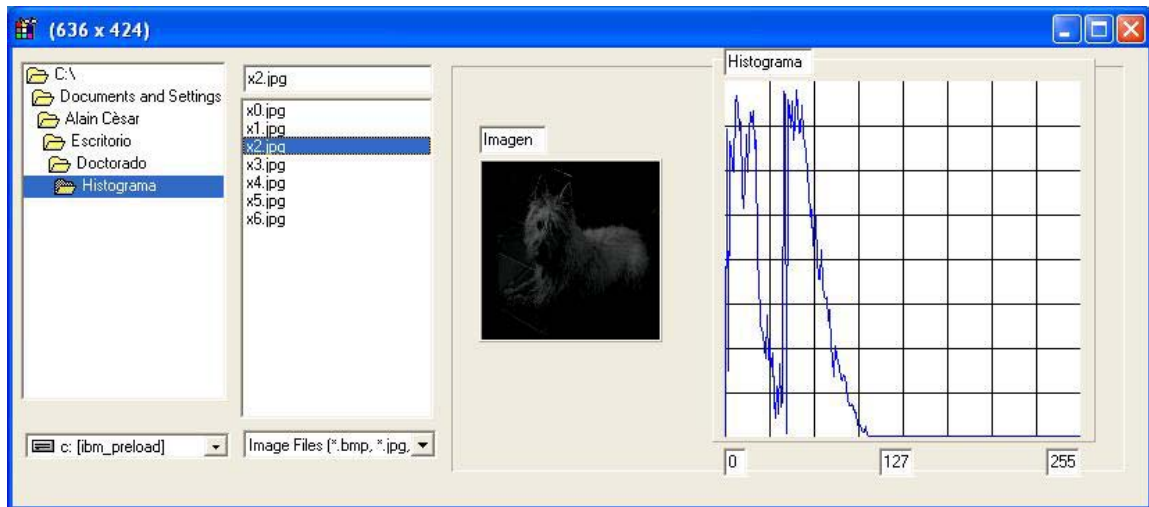


Fig. 12 Imagen oscura y su histograma agrupado hacia los valores bajos.

- **Imagen clara**

Cuando el histograma muestra que los niveles de gris de los píxeles de la imagen están concentrados hacia la parte superior del rango de valores, estos corresponden a una imagen con características claras como se muestra en la Figura 13.

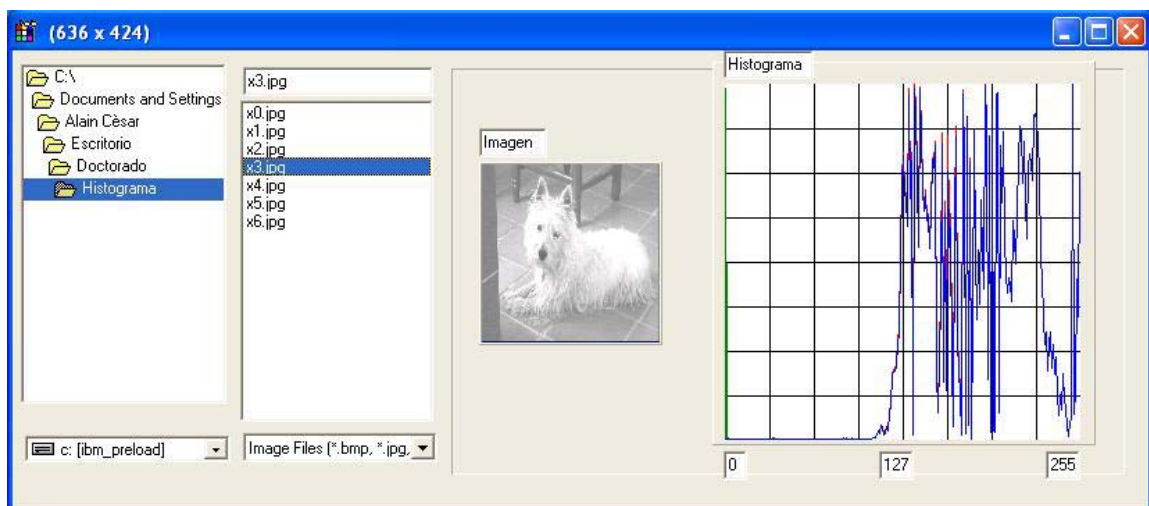


Fig. 13 Imagen clara y su histograma agrupado hacia los valores altos.

- **Imagen con contraste bajo**

En la Figura 14 se observa que cuando el histograma se muestra como una banda angosta, lo que indica un rango pequeño de valores, corresponde a una imagen



con bajo contraste. El contraste indica la relación de brillantez de los objetos de la imagen en su entorno.

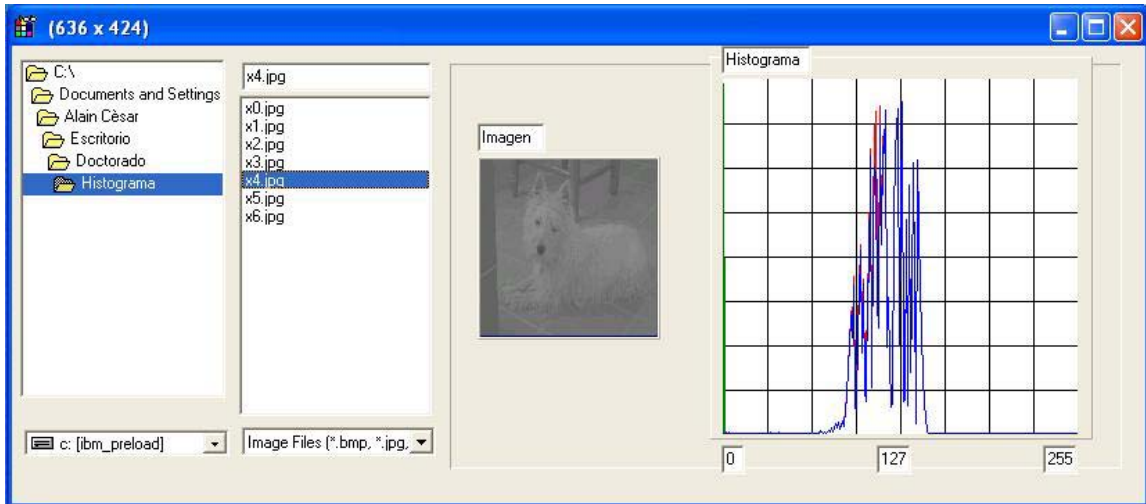


Fig. 14 Imagen con contraste bajo y su histograma angosto.

- **Imagen gris**

Cuando el histograma aparece con casi todos los valores en la parte media de la escala, la imagen aparecerá prácticamente gris con bajo contraste, como se observa en la Figura 15.

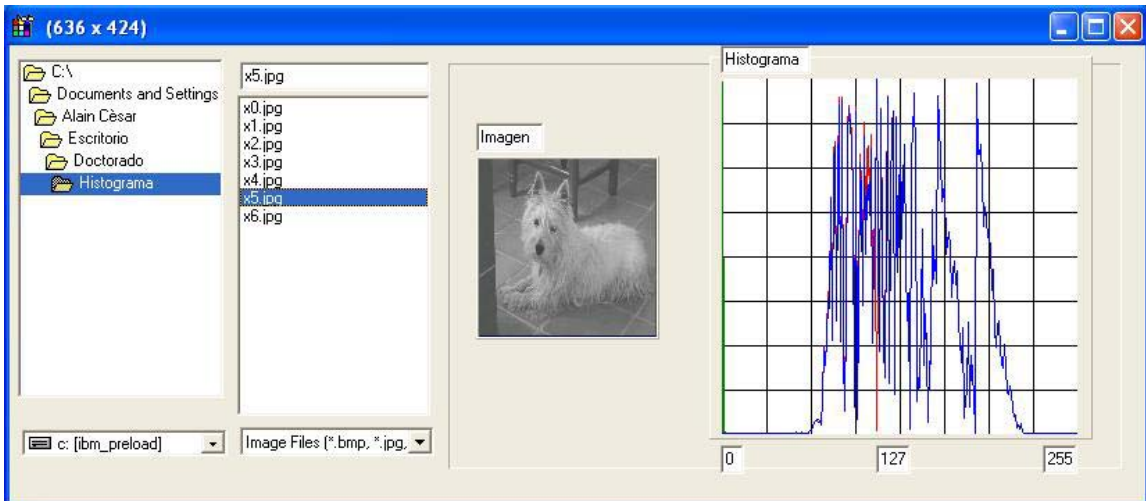


Fig. 15 Imagen gris y su histograma en la parte media.

- **Imagen con contraste alto**

Cuando el histograma muestra una anchura significativa, éste corresponde a una imagen altamente contrastada, tal y como se observa en la Figura 16.

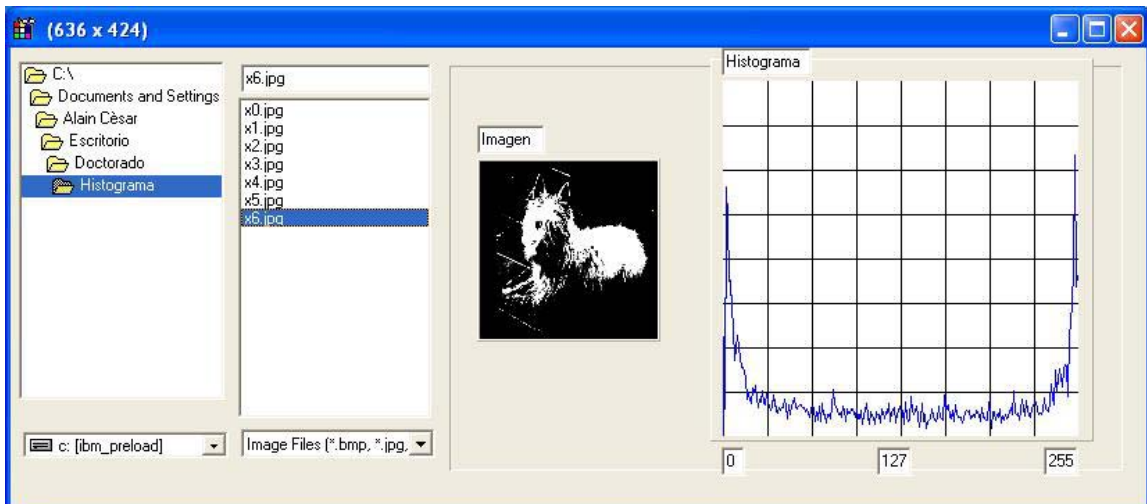


Fig. 16 Imagen con contraste alto y su histograma ancho.

La modificación global del histograma permite variar la distribución de los niveles de intensidad tanto en las imágenes en escala de grises, como en las imágenes multibanda en escala de colores; con ello básicamente se modifica el contraste de la imagen.

Para el modelo de color RGB (Red, Green, Blue), que es el estándar para el despliegue de las imágenes de color, existen tres histogramas independientes, uno para cada color Rojo, Verde y Azul, respectivamente; esto se observa en la Figura 17.

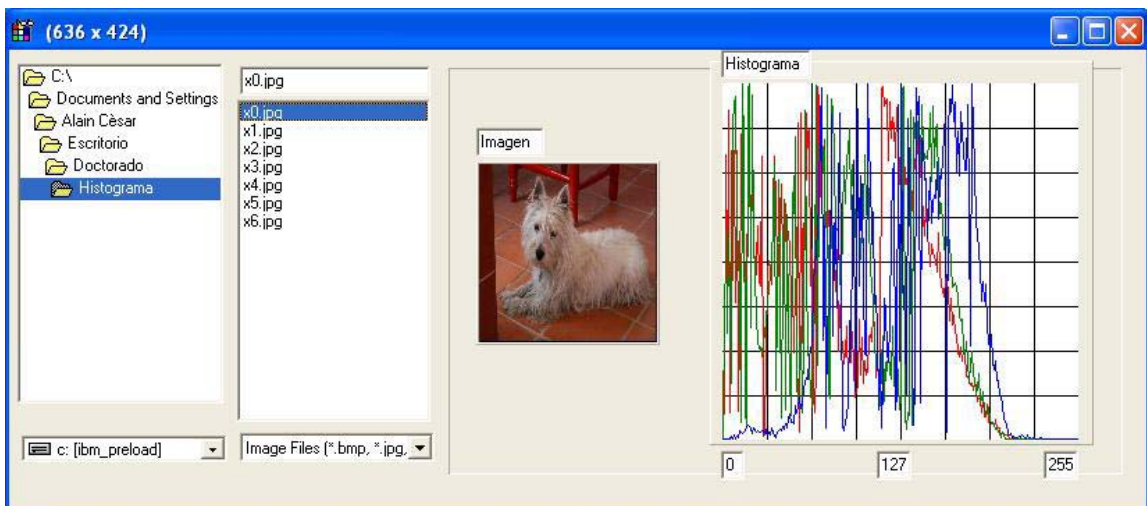


Fig. 17 Histogramas de una imagen de color.

### III.4 Regla del vecino más cercano

En 1967 Cover y Hart [30] estudiaron la Regla del vecino más cercano como una herramienta para la clasificación de patrones, con lo cual comenzó la historia de la regla destinada a ser uno de los métodos de aprendizaje supervisado de mayor popularidad. Tres años después, en 1970, Patrick y Fisher [30] propusieron una familia generalizada para la regla de los vecinos más cercanos, la cual se ha utilizado ampliamente debido a su poder y facilidad de implementación. En esencia, la regla debe cumplir con algunos requisitos tales como el tener una muestra de entrenamiento y una métrica para determinar la cercanía entre los patrones.

La regla no tiene una etapa de aprendizaje como las Redes Neuronales, ya que su aprendizaje consiste únicamente en cargar en memoria la muestra de entrenamiento [36]. El algoritmo se basa en el supuesto de que la clase de patrón a etiquetar sea la del prototipo más cercano a él. Las etapas básicamente son:

1. Buscar exhaustivamente el vecino más cercano al nuevo patrón, haciendo uso de la métrica euclidiana para cada patrón revisado de la muestra de entrenamiento.
2. Asignar al nuevo patrón la etiqueta de la clase de su vecino más cercano.

Como puede deducirse de esto, una de sus desventajas es la considerable carga computacional requerida para su funcionamiento, debido a la búsqueda secuencial que debe realizarse por toda la muestra para determinar el patrón más cercano.

#### III.4.1 Regla de los vecinos más cercanos

Procesamiento: Se ordenan los patrones de acuerdo a los valores de una de sus variables; se recomienda tomar aquella variable que tenga la mayor dispersión en sus valores [34] [35].

Búsqueda: Se examinan los patrones según el orden de sus distancias proyectadas en la variable escogida.

Cuando esta distancia proyectada resulta mayor que la distancia completa en toda la dimensión del patrón a su vecino más cercano dentro de los patrones ya examinados, no se necesita considerar más patrones. Los vecinos más cercanos del patrón son los ya encontrados.

En este método no se requiere el cálculo de ninguna distancia durante la etapa de procesamiento. Tampoco el proceso de búsqueda exige que la función distancia que se utilice satisfaga la desigualdad triangular.

#### III.4.2 Distancia Euclidiana

La distancia Euclidiana puede definirse de la siguiente forma [37], [38]:

Sean  $X = (x_1, x_2, \dots, x_d)$  y  $Y = (y_1, y_2, \dots, y_d)$  dos patrones a clasificar. La Distancia Euclidiana está dada por la ecuación 41.

$$DE(X, Y) = \sqrt{\sum_{i=1}^{i=d} (x_i - y_i)^2} \quad (41)$$

Donde:  $DE(X, Y)$  es la Distancia Euclidiana entre los patrones  $X$  y  $Y$ .

$x_i$  es el valor de la característica  $i$  del patrón  $X$ .

$y_i$  es el valor de la característica  $i$  del patrón  $Y$ .

$d$  es la dimensión del espacio de las variables.

$X$  es el patrón de la muestra de entrenamiento.

$Y$  es el nuevo patrón.

### III.5 Redes neuronales

Las Redes Neuronales Artificiales (RNA) son una forma de emular una característica propia de los humanos: la capacidad de memorizar y asociar los hechos. El hombre es capaz de resolver estas situaciones acudiendo a la experiencia acumulada. Así parece claro que una forma de aproximarse al problema consiste en la construcción de sistemas que sean capaces de reproducir esta característica humana. Las RNA no son más que un modelo artificial y muy simplificado del cerebro humano; es el ejemplo más perfecto del que se dispone, ya que es capaz de adquirir conocimientos a través de la experiencia y reutilizarlo en nuevas situaciones semejantes.

#### III.5.1 Elementos de una red neuronal artificial

Una neurona artificial es en esencia el elemento más simple de procesamiento. Se denomina procesador elemental o neurona  $J$  a un dispositivo simple de cálculo que, a partir de un vector de entrada  $X = [X_1, X_2, \dots, X_r]$  procedente del exterior o de otras neuronas, proporciona una única respuesta o salida. Normalmente tiene la representación que se muestra en la Figura 18 [41].

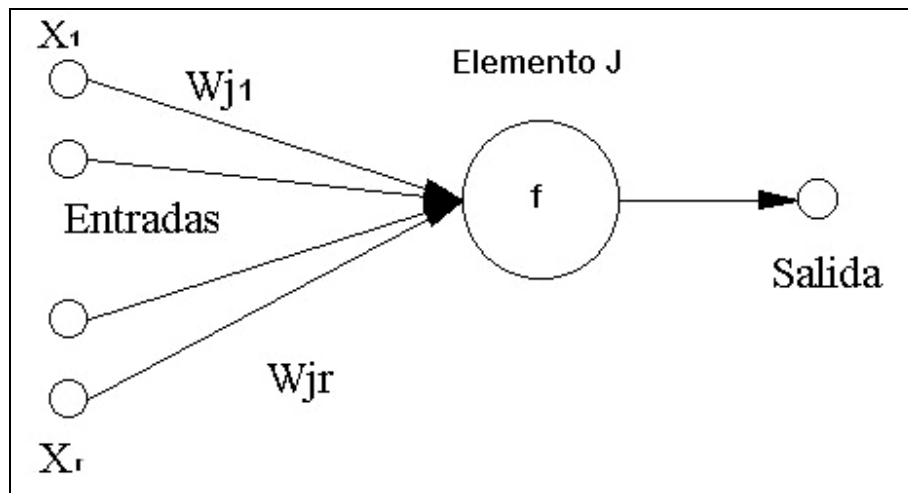


Fig. 18 Elementos de una neurona artificial.

1. **Unidad de proceso: La neurona artificial.** Existen tres tipos de unidades en cualquier sistema: de entrada, de salida y ocultas.

Como se puede ver en la Figura 19, las unidades de entrada reciben señales desde el entorno; las de salida envían la señal fuera de la red; y las unidades ocultas son aquellas cuyas entradas y salidas se encuentran dentro del sistema.

Se conoce como capa o nivel a un conjunto de neuronas cuyas entradas provienen de la misma fuente y cuyas salidas se dirigen al mismo destino.

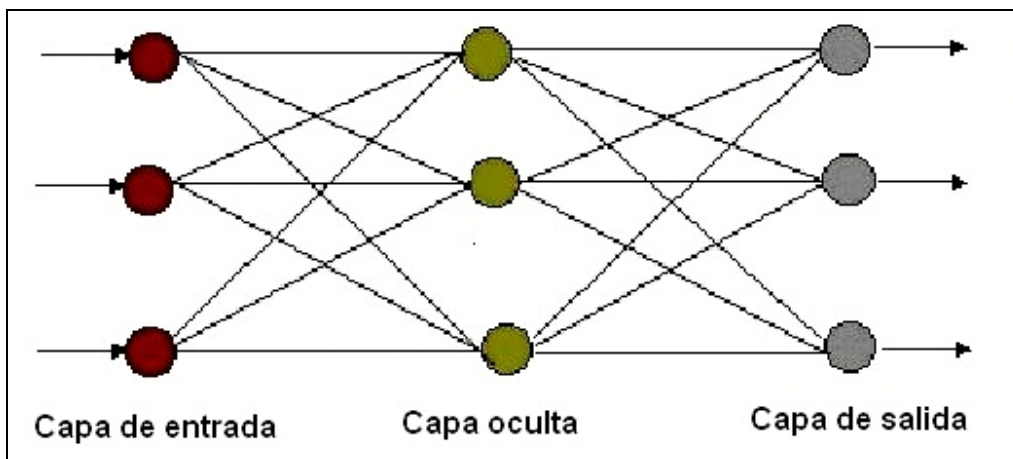


Fig. 19 Capas de una Red Neuronal Artificial.

2. **Estado de Activación.** Los estados del sistema en un tiempo  $t$  se representan por un vector  $A(t)$ . Los valores de activación pueden ser discretos o continuos; si son discretos, suelen tomar un conjunto discreto de valores binarios, así un estado activo se indicaría con un 1 y un estado pasivo por un

0; y si son continuos se considera un conjunto de estados de activación con valores entre  $[0,1]$ , o en el intervalo  $[-1,1]$ , que se obtienen de funciones.

3. **Función de Salida o de Transferencia.** Asociada con cada unidad hay una función de salida que transforma el estado actual de activación en una señal de salida.

Muy frecuentemente la función de salida es simplemente la identidad, de modo que el estado de activación de la neurona se considera como la propia salida. Existen cinco funciones de transferencia típicas que determinan distintos tipos de neuronas: 1.- Función Escalón (de umbral), 2.- Función Lineal, 3.- Función Mixta, 4.- Función Sigmoidal y 5.- Función Gaussiana

4. **Conexiones entre neuronas.** Las conexiones que unen a las neuronas que forman una RNA tienen asociado un peso  $w_{ji}$ , (ver la Figura 18), que es el que hace que la red adquiera el conocimiento.

Se considera que el efecto de cada señal es aditivo, de tal forma que la entrada neta que recibe una neurona es la suma del producto de cada señal individual por el valor de la sinapsis que conecta ambas neuronas y es lo que se conoce como red de propagación.

Se utiliza una matriz  $W$  de todos los pesos. Si  $w_{ji}$  es positivo, indica que la relación entre las neuronas es excitadora, es decir, siempre que la neurona  $i$  esté activada, la neurona  $j$  recibirá una señal que tenderá a activarla. Si  $w_{ji}$  es negativo, la sinapsis será inhibitoria. En este caso si  $i$  está activada, enviará una señal que desactivará a  $j$ . Finalmente si  $w_{ji}$  es 0, se supone que no hay conexión entre ambas.

### III.5.2 Taxonomía de las redes neuronales

Existen cuatro aspectos que caracterizan a una red neuronal: su topología, el mecanismo de aprendizaje, el tipo de asociación realizada entre la información de entrada y de salida, y la forma de representación de esta información [42].

1. **Topología de las Redes Neuronales.** La arquitectura de las Redes Neuronales Artificiales consiste en la organización y disposición de las neuronas formando capas. En este sentido, los parámetros fundamentales de la red son: el número de capas, el número de neuronas por capa, el grado de conectividad y el tipo de conexiones entre las neuronas.

*Redes de tipo Monocapa.* Se establecen conexiones laterales, cruzadas o auto recurrentes entre las neuronas que pertenecen a la única capa que constituye la red. Se utilizan en tareas relacionadas con lo que se conoce como auto-asociación; por ejemplo, para generar la información de entrada a partir de la retroalimentación que se le presenta a la red de forma incompleta o distorsionada.

*Redes de tipo Multicapas.* Son aquellas que disponen de conjuntos de neuronas agrupadas en varios niveles o capas. Una forma de distinguir la capa a la que pertenece la neurona, consiste en fijarse en el origen de las señales que recibe a la entrada y el destino de la señal de salida.

Cuando ninguna salida de las neuronas constituye una entrada de neuronas del mismo nivel o de niveles precedentes, la red se describe como de propagación hacia adelante. Cuando las salidas pueden ser conectadas como entradas de neuronas de niveles previos o del mismo nivel, incluyéndose ellas mismas, la red es de propagación hacia atrás.

**2. Mecanismo de Aprendizaje.** El aprendizaje puede ser comprendido como la modificación del comportamiento, inducido por la interacción con el entorno y como resultado de experiencias. El aprendizaje es el proceso por el cual una RNA modifica sus pesos en respuesta a la información de entrada, dado que el conocimiento se asocia a los pesos de las conexiones entre las neuronas.

Los cambios que se producen durante el proceso de aprendizaje se reducen a la destrucción, modificación y creación de conexiones entre las neuronas. La creación de una nueva conexión implica que el peso de la misma pasa a tener un valor distinto de cero; una conexión se destruye cuando su peso pasa a ser cero. Se puede afirmar que el proceso de aprendizaje ha finalizado, es decir, que la red ha aprendido, cuando los valores de los pesos permanecen estables para toda entrada posterior.

Un criterio suele considerar dos tipos de reglas de aprendizaje: las de aprendizaje no supervisado y las correspondientes a un aprendizaje supervisado. Estas reglas dan pie a una de las clasificaciones de las RNA: redes neuronales con aprendizaje supervisado y redes neuronales con aprendizaje no supervisado. La diferencia fundamental entre ambos tipos estriba en la existencia o no de un agente externo (supervisor) que controle el aprendizaje de la red [42].

**3. Tipo de asociación entre las informaciones de entrada y salida.** Las Redes Neuronales Artificiales son sistemas que almacenan cierta información aprendida; esta información se registra de forma distribuida en los pesos asociados a las conexiones entre neuronas de entrada y salida.

Existen dos formas primarias de realizar esta asociación de entrada / salida.

La primera es denominada heteroasociación, y se refiere al caso en el que la red aprende parejas de datos  $[(A_1, B_1), (A_2, B_2) \dots (A_n, B_n)]$ , de tal forma que cuando se presente cierta información de entrada  $A_i$ , deberá responder generando la correspondiente salida  $B_i$ .

La segunda se conoce como auto asociación, donde la red aprende cierta información  $A_1, A_2 \dots A_n$ , de tal forma que cuando se le presenta la información de entrada, realizará auto-correlación, respondiendo con el dato almacenado más parecido al de la entrada.

3. **Representación de la información de entrada y salida.** Las RNA pueden también clasificarse en función de la forma en que se representa la información de entrada y de las respuestas o datos de salida. Así, en un gran número de redes, tanto los datos de entrada como los de salida son de naturaleza analógica. Cuando esto ocurre, las funciones de activación de las neuronas serán también continuas, del tipo lineal o sigmoideal.

Otras redes sólo admiten valores discretos o binarios de entrada, generando también unas respuestas en la salida de tipo binario. En este caso, las funciones de activación de las neuronas son de tipo escalón. Existe también un tipo de redes híbridas en las que la información de entrada puede ser en forma de valores continuos, aunque las salidas de la red son discretas.

### III.5.3 Perceptrón Multicapa

El Perceptrón Multicapa es una Red Neuronal Artificial compuesta de varias capas de neuronas. Se entrena mediante un paradigma de red supervisado; el ajuste de los parámetros se realiza por el método del Gradiente Descendente a través del conocido algoritmo de retropropagación. Una red Perceptrón Multicapa típica es como la que se muestra en la Figura 20; contiene una capa de entrada, una de salida y puede estar formada por varias capas ocultas; generalmente solo se utiliza una capa oculta.

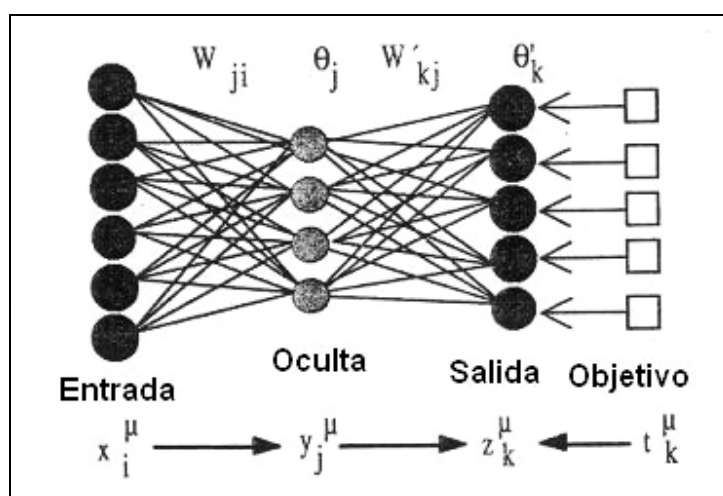


Fig. 20 Perceptrón Multicapa

### III.5.4 Algoritmo de Rumelhart

El algoritmo de Rumelhart consiste básicamente en el aprendizaje de un conjunto predefinido de pares de entrada-salida en dos fases:

Se aplica un patrón de entrada a la red, que estimula a las neuronas de la primera capa, el cual se va propagando por las capas superiores intermedias hasta llegar a



la capa de salida, donde la red genera una respuesta; se compara el resultado obtenido en las neuronas de salida con el valor deseado y se calcula un valor de error para cada neurona de salida.

A continuación, partiendo de la capa de salida, estos errores se transmiten hacia atrás, hacia la capa intermedia, que reciben el porcentaje de error aproximado; este proceso se repite para todas las capas de la red. Basándose en el valor de error recibido se ajustan los pesos de conexión de cada neurona, con la finalidad de que la siguiente vez que se presente el mismo patrón de salida esté más cercano al deseado.

### III.5.5 Gradiente Descendente

El Gradiente Descendente trata de encontrar los pesos que mejor se ajustan a los ejemplos y es la base del algoritmo retropropagación. El error se puede expresar mediante el cálculo de las diferencias de error al cuadrado, al que se le denomina error cuadrático medio y que se determina mediante la Ecuación 42.

$$E = \frac{1}{2} \sum (t_i - o_i)^2 \quad (42)$$

Donde:  $t_i$  es la salida deseada.

$o_i$  es la salida de la red.

$E$  es el error cuadrático medio.

Lo que se quiere es determinar el vector de pesos que minimice el error  $E$ . Esto se logra alterando los pesos en la dirección que produce el máximo descenso en la superficie del error; la dirección de cambio se obtiene mediante el Gradiente. El Gradiente especifica la dirección que produce el máximo incremento, por lo que el mayor descenso es el negativo en esa dirección; la regla de actualización de los pesos es entonces la que se muestra en la Ecuación 43.

$$W_{t+1} = -\eta \nabla_w E_p + W_t \quad (43)$$

Donde:  $\nabla_w$  es el incremento de los pesos.

$\eta$  es la razón de aprendizaje.

$t$  es el número de iteraciones.

### III.5.6 Algoritmo de retropropagación

Una red Perceptrón Multicapa tiene la forma que se muestra en la Figura 27 y cuenta con los siguientes elementos, a saber:

Los índices  $i, j, k$  denotan las neuronas ubicadas en las capas  $i, j, k$ .

$net$  es la sumatoria de los pesos.

$W$  es el vector de pesos.

$F(net_j)$  es la función sigmoideal.

$\theta$  es el BIAS.

$\eta$  es la razón de aprendizaje.

$\alpha$  es el momento.

La  $net$  de entrada de un nodo en la capa  $j$  está dada por la Ecuación 44, donde se obtiene la  $net_j$ , es decir, la suma de las entradas provenientes de otras neuronas.

$$net_j = \sum w_{ji} o_j \quad (44)$$

Posteriormente a la  $net$  se le aplica la función de activación  $f$  tal como se muestra en la Ecuación 45.

$$o_j = f(net_j) \quad (45)$$

La cual es una función *sigmoideal* representada por la Ecuación 46.

$$o_j = \frac{1}{1 + e^{-\frac{net_j + \theta_j}{\theta_j}}} \quad (46)$$

Para los nodos de la capa  $k$  la  $net$  de entrada de un nodo se representa por la Ecuación 47 y la salida  $O$  del nodo  $k$  por la Ecuación 48, respectivamente.

$$net_k = \sum w_{kj} o_j \quad (47)$$

Y la salida  $O$  del nodo  $k$  por la Ecuación 41, respectivamente.

$$o_k = f(\text{net}_k) \quad (48)$$

Dado que la idea central del algoritmo está en calcular los errores para las unidades de las capas ocultas, a partir de los errores de las unidades de la capa de salida, al propagarse capa tras capa hacia la de entrada se calcula el error cuadrático medio por patrón mediante la Ecuación 49, y la suma cuadrática de los errores con la Ecuación 50.

Para un determinado conjunto de entrenamiento, este error representa una medida del aprendizaje de la RNA, denominada función de costo; el objetivo es minimizar dicha función ajustando los pesos.

$$E_p = \frac{1}{2} \sum (t_{pk} - o_{pk})^2 \quad (49)$$

$$E = \frac{1}{2p} \sum_p \sum_k (t_{pk} - o_{pk})^2 \quad (50)$$

Rumelhart, Milton y Williams [30] sugirieron en 1986 que la actualización de los pesos se modificara mediante la inclusión de un término  $\alpha$  denominado momento; la modificación de los pesos fue entonces como la que se muestra en la Ecuación 51.

$$\nabla w_{ji}^{(t+1)} = \eta(\delta_j o_i) + \alpha \nabla w_{ji}^{(t)} \quad (51)$$

Es importante notar que cuanto más pequeño sea el parámetro de aprendizaje  $\eta$  más pequeños serán los cambios de los pesos de la red de una iteración a la siguiente; por otra parte, si este parámetro es demasiado grande con el fin de acelerar la convergencia hacia un error mínimo, se puede provocar una inestabilidad en el funcionamiento de la red, la que se traduce en un comportamiento oscilante en el entorno del mínimo buscado; por otro lado, para asegurar la convergencia deberá mantenerse el intervalo del momento entre 0 y 1.

La red de Retropropagación puede realizar el entrenamiento de 2 formas diferentes:

- *Modo en línea.* Donde la actualización de los pesos se realiza tras la presentación de cada ejemplo de entrenamiento.
- *Modo por lote.* Donde la actualización de los pesos de la red se efectúa tras la presentación de todos los ejemplos de entrenamiento que constituyen un ciclo de entrenamiento denominado época.

En cuanto a la inicialización de los pesos, con el fin de evitar la saturación de las neuronas, éstos suelen ser inicializados de manera aleatoria dentro de un intervalo de valores entre 0 y 1, o en algunos casos entre -0.5 y 0.5.

### III.5.7 Métodos de Estimación

Existen varios métodos para estructurar las muestras de control y de entrenamiento, encaminados todos a la estimación futura del error de los clasificadores. Entre ellos están los de Validación Cruzada, que se basan en la creación de más de un conjunto de datos para evaluar la eficiencia del clasificador. Existen variantes de este método, una de ellas es el Método  $\pi$  (PI) [43].

El Método  $\pi$  realiza la distribución de los patrones contenidos en la muestra de entrenamiento en forma aleatoria, sin reemplazo, en  $V$  conjuntos disjuntos  $C_1$  y  $C_2 \dots C_V$ , de un tamaño similar, realizándose el entrenamiento de la forma siguiente: Se toma el primer conjunto como la muestra de control y el resto de los conjuntos como la muestra de entrenamiento; se repite el proceso con el segundo conjunto y así sucesivamente para todos los conjuntos creados.

Una vez concluido el marco teórico de referencia para esta investigación, en el capítulo IV se presenta la idea principal de esta alternativa que consta de la recuperación de imágenes mediante la firma de las mismas, constituida por el vector característico de la transformada Ondeleta inmersa en el método multi-resolución y los algoritmos heurísticos de histogramas de ventanas circulares y sub-imágenes.

## CAPÍTULO IV

# RECUPERACIÓN DE IMÁGENES MEDIANTE LA TRANSFORMADA ONDELETA Y LA REGLA DEL VECINO MÁS CERCANO

En este capítulo se presentan tres métodos de recuperación de imágenes mediante la Ondeleta Daubechies 4 en el espacio de color RGB, dos de ellos a partir de características globales y uno a partir de características locales.

El primer método global extrae los coeficientes ondeletas de la imagen original en forma directa; el segundo método global extrae los coeficientes ondeletas del histograma de una ventana circular inscrita en la imagen.

El método basado en características locales, extrae los coeficientes ondeletas de los histogramas de sub-imágenes disjuntas distribuidas de manera uniforme en toda la imagen.

### IV.1 Introducción

Los sistemas de procesamiento de información actuales a menudo necesitan involucrar el almacenamiento y la recuperación de imágenes con información visual. Lo importante dentro de un proceso de recuperación de la información visual contenida en una imagen, es que la información visual recuperada sea la misma, o al menos, que el sistema brinde un grupo de imágenes con información similar a la deseada [1].

La recuperación de imágenes basada en el contenido posee la habilidad de recuperar imágenes utilizando como llave de búsqueda una imagen. Se trata de buscar en una base de datos de imágenes las  $n$  imágenes más parecidas a una imagen-consulta.

El esquema de generación de la firma mediante una transformación de la imagen para obtener un vector de características que sea una representación numérica simplificada, sirve para almacenar la firma en una base de datos y así acelerar el proceso de recuperación de las imágenes, ya que la transformación caracteriza de forma efectiva las propiedades globales de la imagen, entre otras el color y la forma [2].

La presencia de colores en una imagen posee un alto significado en el proceso de indexado y recuperación. La representación del color que se utiliza es el modelo RGB, que es el modelo de color más simple y más ampliamente utilizado por los dispositivos de visualización.

Como una característica global, el histograma de los planos de colores de una imagen representa la distribución del nivel de color en dichos planos; además, los histogramas de los planos de color son invariantes a la traslación y a la rotación, y si son normalizados, también son invariantes a la escala [13].

Por ser una característica global, el histograma de los planos de color no puede indicar la característica local referida a la localización espacial de los objetos de diferentes colores presentes en la imagen.

La organización de la búsqueda de las imágenes y su recuperación sobre alguna base robusta, constituye un área de investigación de extrema importancia. Sin embargo, este proceso depende directamente de la forma en que se procede sobre la imagen para lograr una eficiencia aceptable.

En este trabajo, la recuperación de las imágenes se realiza comúnmente dentro de las contenidas en una colección de bases de datos de imágenes.

A su vez, la organización de la búsqueda está basada en la codificación del contenido de la imagen mediante la Ondeleta Daubechies 4. Como técnica de recuperación, se utiliza la regla del vecino más cercano; y las relaciones espaciales entre los coeficientes calculados se realiza según la distancia Euclidiana.

## **IV.2 Metodología**

La Figura 21 muestra un resumen de la metodología general utilizada para la recuperación de las imágenes de una base de datos de imágenes, la cual consta de dos etapas.

Como se observa en la Fig. 21, en la primera etapa de procesamiento de la base de datos de imágenes, una vez desplegada una imagen, se utiliza la interpolación bilineal para ajustar el tamaño de la misma a 256 x 256 píxeles. Después, se utiliza el algoritmo discreto de la transformada Ondeleta para codificar la imagen en tres vectores de acuerdo a los tres métodos concebidos para ello en este trabajo, a saber, directamente a partir de la imagen completa, del histograma de una ventana circular inscrita en la imagen y de los histogramas de las sub-imágenes cuadradas en que se ha dividido la imagen original; finalmente se

almacenan los tres vectores y se procede con la siguiente imagen, hasta terminar de procesar el total de imágenes disponibles en la base de datos de imágenes.

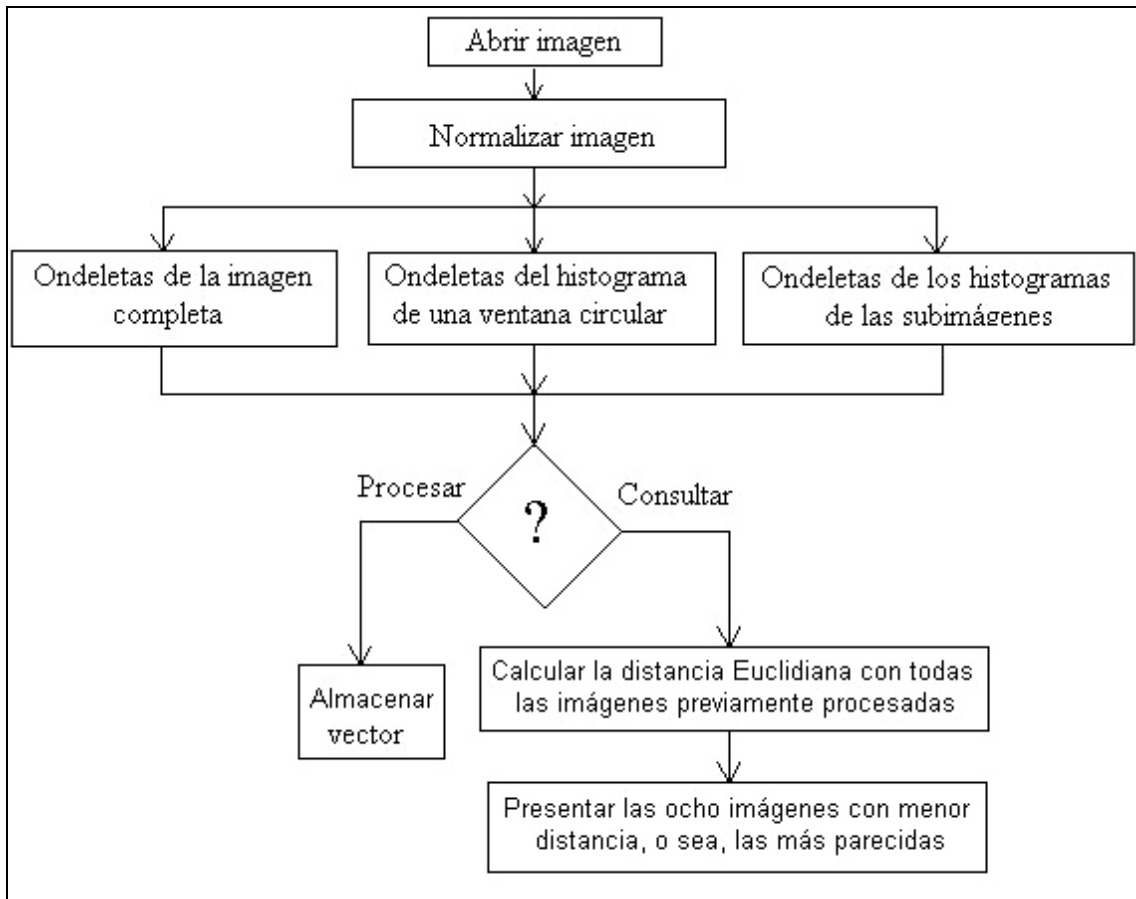


Fig. 21 Metodología para la recuperación de las imágenes.

La Fig. 21 muestra que para la segunda etapa, que es la de consulta, se procede de forma similar, esto es, se despliega la imagen-consulta, se ajusta el tamaño de la imagen a 256 x 256 píxeles, se codifica la imagen en tres vectores de acuerdo a los tres métodos mencionados, pero en lugar de almacenar esos tres vectores, se aplica la regla del vecino más cercano, específicamente se calcula la distancia euclidiana con respecto a todos los vectores de las imágenes previamente procesadas, a partir de lo cual se muestran las ocho que presentan la menor distancia por ser las más parecidas. Ahora explicaremos en detalle lo que en cada etapa se realiza.

### IV.3 Normalización del modelo de color

Debido a que las imágenes en diferentes bases de datos pueden tener diferentes formatos gráficos y ser de diferentes tamaños, se deben normalizar el modelo de color y ajustar el tamaño de la escala para que la comparación entre las imágenes sea consistente.

Las imágenes pueden estar almacenadas, entre otros, en los siguientes formatos gráficos: BMP, JPG, PNG o ICO. El modelo se normaliza cuando todas las imágenes se despliegan en el modelo de color RGB de 24 bits, con 8 bits para cada uno de los colores Rojo, Verde y Azul, para tener acceso directo a los píxeles y proseguir con ajuste del tamaño de la escala.

### IV.4 Ajuste del tamaño de la imagen

El propósito fundamental de ajustar el tamaño de las imágenes es el de poder utilizar la transformada Ondeleta con un tamaño uniforme de los vectores característicos. Las imágenes se pueden ajustar a cualquier tamaño, pero es preferible hacerlo sobre la base de que la cantidad de píxeles por fila y columna sea una potencia de dos [27].

Para ajustar el tamaño de las imágenes, se hace un submuestreo de la imagen de entrada sobreponiendo una rejilla, por ejemplo, de 256 x 256 puntos. Esto dará lugar a un píxel en la imagen de salida por cada punto de intersección de la rejilla. Para ello la imagen de entrada es muestreada en cada punto de la rejilla para determinar las componentes RGB del píxel en la imagen de salida. Cuando los puntos de intersección de la rejilla no coinciden en posición con los centros de los píxeles, los valores RGB del punto de la rejilla son determinados por interpolación bilineal a partir de las componentes de los colores de los píxeles adyacentes.

Las transformaciones geométricas pueden ser clasificadas como lineales y no lineales. Las transformaciones lineales o afines incluyen la traslación, la rotación y la escala, o combinaciones de éstas. Las transformaciones geométricas no lineales o deformables producen distorsiones de la imagen. Un acercamiento a esto puede ser implementado como una transformación geométrica lineal de escala [27].

Las ecuaciones 52 y 53 son para calcular la nueva posición de un píxel durante un acercamiento.

$$x' = a(x - x_0) + x_0 \quad (52)$$

$$y' = a(y - y_0) + y_0 \quad (53)$$

Donde  $(x', y')$  es la nueva posición para el píxel  $(x, y)$ ,  $a$  es el factor de acercamiento, y  $(x_0, y_0)$  es la coordenada del centro del acercamiento. Sin



embargo, las nuevas posiciones calculadas generalmente no caen en posiciones que brinden coordenadas enteras. También una práctica consiste en aplicar las ecuaciones 52 y 53 en reversa. Para cada píxel en la nueva imagen de salida, la posición original del píxel es calculada mediante las ecuaciones 54 y 55.

$$x = (x' + (a - 1)x_0) / a \quad (54)$$

$$y = (y' + (a - 1)y_0) / a \quad (55)$$

Sin embargo, las posiciones calculadas por las ecuaciones 54 y 55 tampoco llegan generalmente a posiciones que brinden coordenadas enteras. Para obtener el nuevo valor del píxel se utiliza comúnmente la interpolación bilineal.

La interpolación bilineal hace uso de los cuatro vecinos más cercanos de la posición calculada para obtener un promedio ponderado, lo cual se muestra en la Figura 22. La implementación del algoritmo de interpolación bilineal se muestra en el Apéndice A.

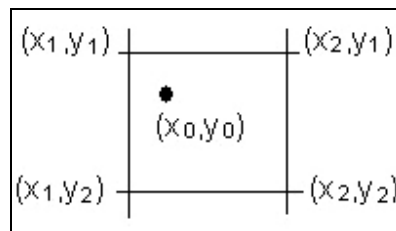


Fig. 22 Interpolación bilineal

## IV.5 Codificación de las imágenes

En este trabajo se utilizan características basadas en forma y color para la recuperación de las imágenes. Las características en una imagen de color en cada uno de los tres componentes R, G, y B se codifican mediante tres métodos diferentes:

- Imágenes completas,
- Histograma de una ventana circular, y
- Histogramas de sub-imágenes.

Los histogramas de niveles de gris y de color son invariantes a la traslación y a la rotación, pero como éstos no indican la localización espacial de los píxeles de la imagen por ser globales, se toma una combinación de cada uno de los tres métodos mencionados anteriormente con multi-resolución para tomar en cuenta la característica de la forma.

## IV 5.1 Imágenes completas

Se pretende indexar las imágenes teniendo en cuenta su contenido en términos de forma y distribución de los colores. Para recuperar imágenes, se propone utilizar métodos multi-resolución y la transformada Ondeleta para la codificación sub-banda, y así codificar la imagen en vectores característicos [39] [40].

El proceso de indexado aplica una transformada Ondeleta Daubechies 4 a la imagen para cada una de sus tres componentes de color del espacio RGB. Los coeficientes de ondeletas de las bandas de frecuencia más bajas son almacenados como vectores característicos [50].

En las matrices creadas con los coeficientes de la transformada Ondeleta, las submatrices de la esquina superior izquierda representan las bandas de frecuencia más bajas de la imagen 2-D, y las submatrices de la esquina inferior derecha representan las bandas de frecuencia más altas [27].

Generalmente los coeficientes de la transformada Ondeleta de las bandas de frecuencia más bajas representan las configuraciones de los objetos presentes en las imágenes, y las bandas de frecuencia más altas representan las variaciones de textura.

Los coeficientes de la transformada Ondeleta Daubechies 4 representan la semántica de una imagen, es decir, la distribución, tamaños y formas de los objetos y la variación local de los colores de los objetos y del fondo.

La transformación de un archivo de imagen en el dominio del tiempo al dominio de la frecuencia, se consigue mediante una operación de convolución entre la imagen y la transformada Ondeleta. Puesto que la transformada Ondeleta es una transformación lineal, la transformada de una señal bidimensional se puede calcular al convolucionar primero en dirección horizontal y luego en dirección vertical, o viceversa [17] [18].

En este método se parte de una imagen normalizada de tamaño 256 x 256 píxeles [50].

En el primer nivel se aplica el algoritmo discreto de la transformada Ondeleta en la dirección horizontal, con lo que se obtiene una imagen de tamaño 128 x 256 píxeles, ya que se diezman las filas.

Seguidamente se aplica el algoritmo discreto de la transformada Ondeleta en la dirección vertical para obtener una imagen de tamaño 128 x 128 píxeles.

Se procede sucesivamente de la misma forma en los siguientes niveles para obtener imágenes de tamaños 64 x 64, 32 x 32, 16 x 16, 8 x 8 y 4 x 4. Al hacerse esto en los tres planos de colores RGB, se obtienen 16 coeficientes por cada plano de color RGB [11]. De aquí resulta un vector de características con 48 coeficientes que conforman el vector característico de la imagen. El proceso completo se muestra en la Figura 23.

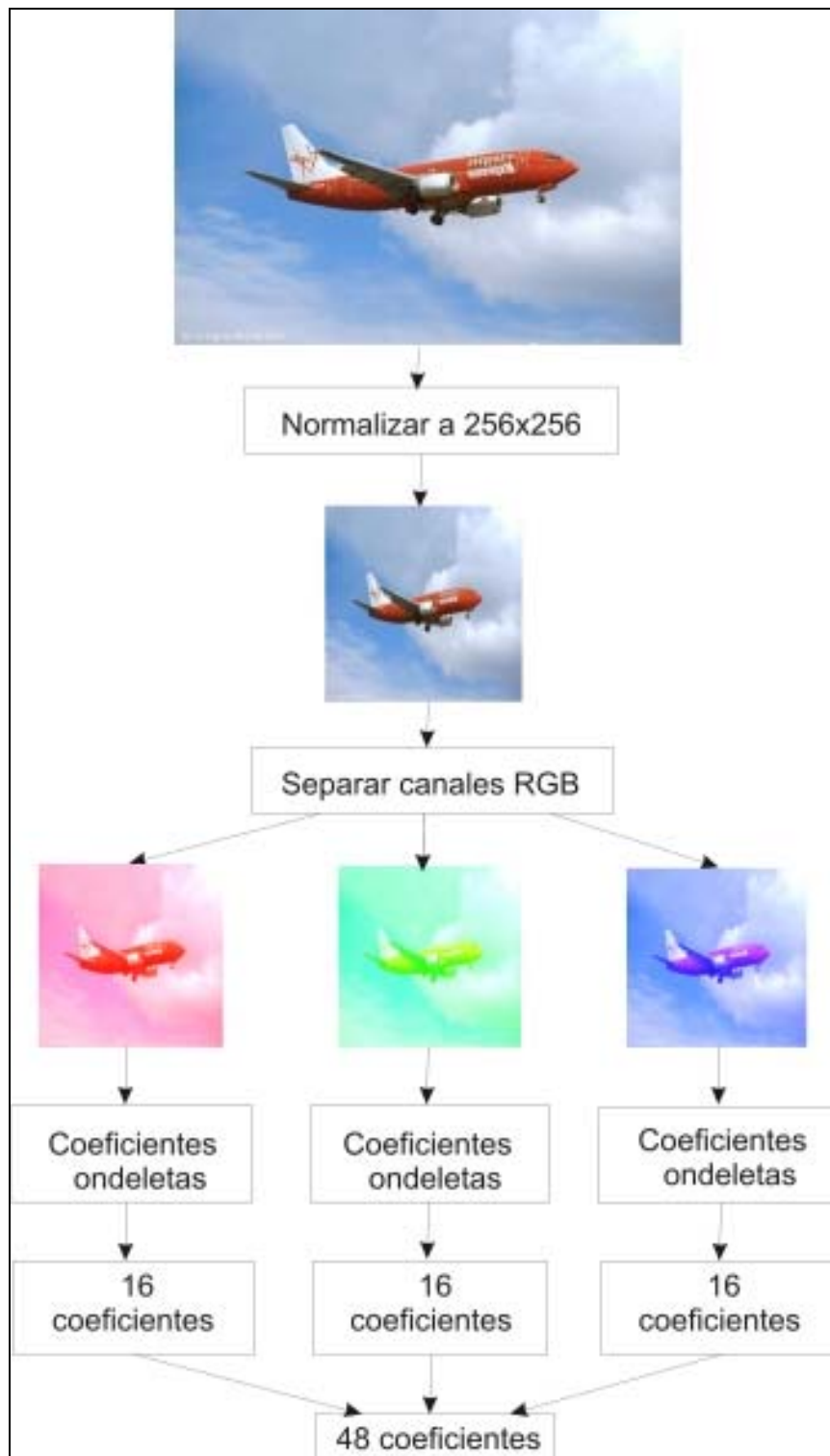


Fig. 23 Extracción de los coeficientes de la imagen completa.

## **IV.5.2. Histograma de una ventana circular**

El histograma de una imagen digital consiste en un diagrama de barras. Se utilizan como abscisas los niveles de la imagen en cada plano y como ordenadas el número de píxeles para cada plano de color RGB.

El histograma global de una imagen digital proporciona información sobre la misma, la que se reduce a la información de la frecuencia de ocurrencia de los diferentes niveles de colores que la componen; en ese sentido se pierde toda la información espacial de la imagen; se conoce la distribución de los niveles de color de cada píxel, pero se ignoran sus coordenadas correspondientes [8].

En la Figura 24 se representa el proceso para extraer los coeficientes de la transformada Ondeleta. En el método que se expone se codifican las imágenes en forma global para buscar la invariancia respecto a la rotación [13].

Primero se inscribe en la imagen normalizada de tamaño 256 x 256 píxeles una ventana de forma circular de diámetro 256 píxeles [51].

A continuación, se procede a calcular los histogramas de la información presente dentro de la ventana circular en los tres planos RGB. Después se normalizan los valores de los tres histogramas.

Se procede a aplicar el algoritmo discreto de la transformada Ondeleta Daubechies 4 a los 256 valores de los histogramas.

Lo anterior se hace sucesivamente para pasar la imagen primero de 256 a 128 píxeles de diámetro y después de 128 a 64, de 64 a 32 y de 32 a 16, para así contar finalmente con 16 coeficientes de ondeletas por cada plano de color RGB. De aquí resulta un vector de características de 48 coeficientes que conforma el vector característico de la imagen. El proceso completo se presenta en la Figura 24.

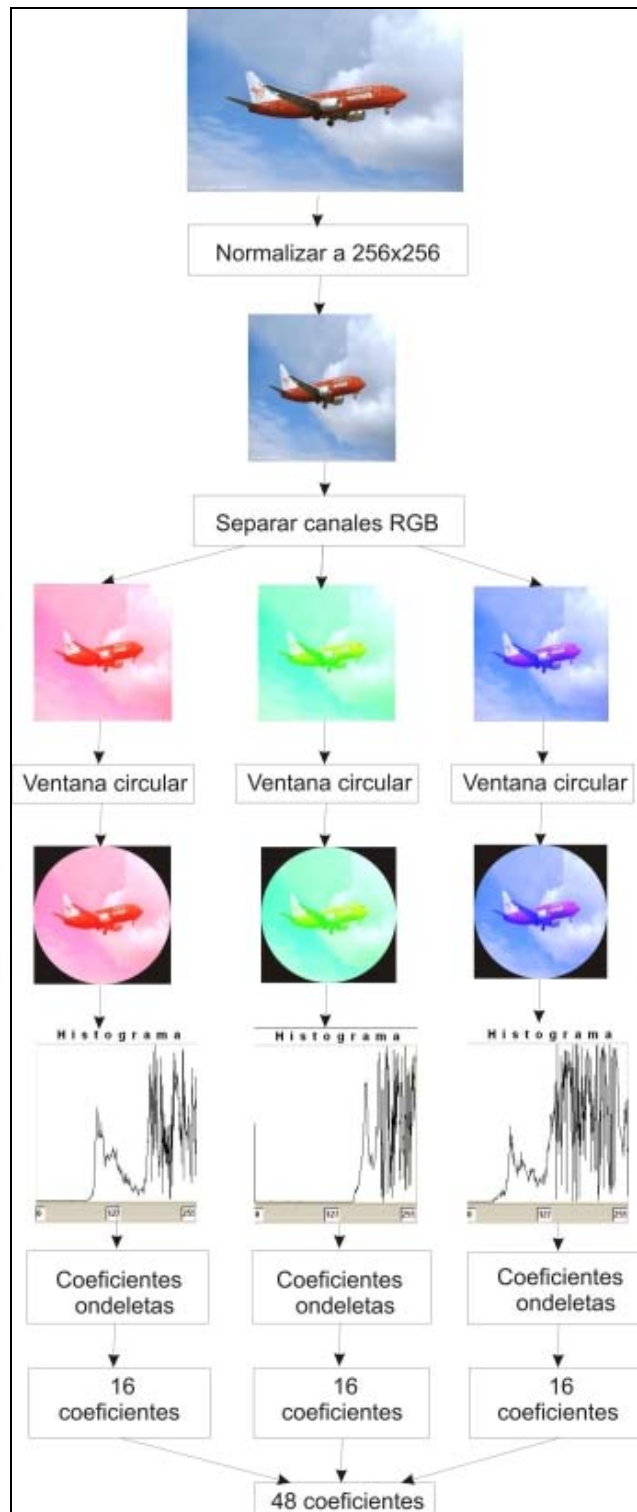


Fig. 24 Extracción de los coeficientes del histograma de una ventana circular.

### **IV.5.3. Histogramas de sub-imágenes cuadradas**

El histograma se considera una técnica de indexación de imágenes basada en el dominio del píxel, y la extracción de los coeficientes de ondeletas se considera como una técnica basada en el dominio de la compresión[31].

La baja complejidad del algoritmo de la convolución compuesto de multiplicaciones y sumas, que se utiliza en la extracción de coeficientes de ondeletas, se considera una buena característica [32] [33].

Al obtener el histograma del contenido de la imagen en cada plano de color y mediante la extracción de coeficientes de ondeletas con el algoritmo multi-resolución, se combinan técnicas basadas en el color y la forma [14].

En la Figura 25 se representa el método que analiza la imagen a partir de características locales [52].

Primero se divide la imagen de tamaño 256 x 256 píxeles en sub-imágenes de 64 x 64 píxeles, de forma tal que se cubre toda la imagen con 16 sub-imágenes disjuntas.

Después se calculan los histogramas de estas sub-imágenes, con lo que se obtienen 16 histogramas de 256 valores.

Seguidamente se aplica sucesivamente el algoritmo discreto de la transformada Ondeleta Daubechies 4 para obtener 128 valores a partir de los 256 valores originales, después se pasa a 64, de 64 a 32, de 32 a 16, de 16 a 8 de 8 a 4 y de 4 a 1.

Finalmente, se obtiene un coeficiente de ondeletas por cada sub-imagen, por lo que se logran en total 16 coeficientes de ondeletas por plano de color RGB, lo que da un total de 48 coeficientes que conforman el vector característico de la imagen.

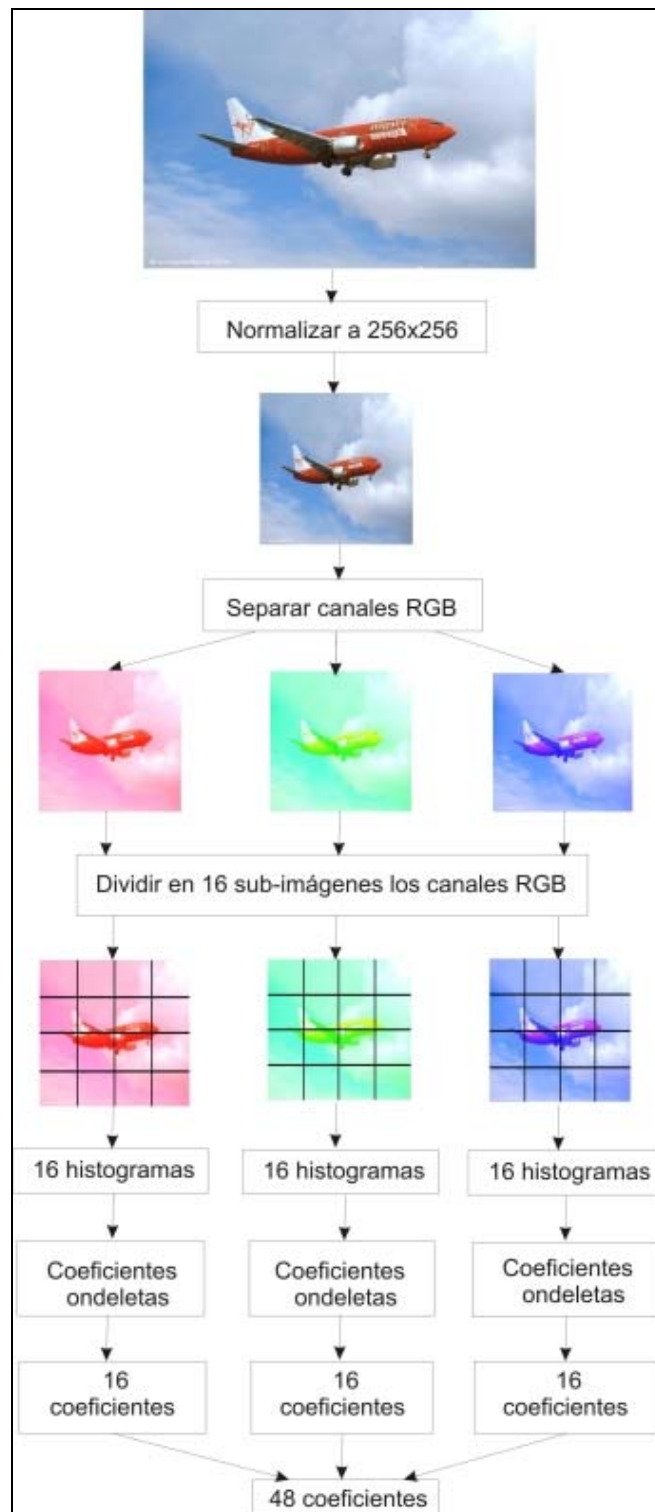


Fig. 25 Extracción de los coeficientes de los histogramas de las sub-imágenes.

## **IV.6 Almacenamiento de los vectores característicos**

Después de aplicar a una imagen los tres métodos descritos en el subepígrafe IV.5. Codificación de imágenes, se procede al almacenamiento de los tres vectores característicos o firmas de las imágenes en tres archivos secuenciales, donde cada registro corresponde a una firma, para integrar la base de datos de firmas de las imágenes de los tres métodos. Posteriormente se continúa con todas y cada una de las imágenes.

## **IV.7 Recuperación mediante la regla del vecino más cercano**

Para cada uno de los tres métodos descritos en el subepígrafe *IV.5 Codificación de las imágenes*, la recuperación es hecha mediante la regla del vecino más cercano a los coeficientes de ondeletas de la imagen-consulta, sobre la base de los coeficientes de ondeletas de todas las imágenes de la base de datos de imágenes que se obtiene con el procedimiento de almacenamiento (subepígrafe IV.6).

Finalmente, sobre la base de la distancia Euclidiana, la cual se utiliza como métrica para calcular la distancia entre los vectores característicos, se comparan los vectores característicos de las imágenes previamente procesadas con el vector característico de la imagen-consulta. Las imágenes con distancias euclidianas más pequeñas, en ese orden, son seleccionadas para mostrarlas como las imágenes que más se asemejan a la imagen consulta.

En el Capítulo 5 se presenta una alternativa donde también se utilizan los algoritmos heurísticos presentados para extraer las firmas de las imágenes y donde se clasifica mediante una combinación del algoritmo del vecino más cercano y el de las redes neuronales.



## CAPÍTULO V

# RECUPERACIÓN DE IMÁGENES MEDIANTE LA TRANSFORMADA ONDELETA Y LA CLASIFICACIÓN CON REDES NEURONALES

En este capítulo se describe la metodología general para la recuperación de la información visual con base en su contenido, al implementar la clasificación mediante la combinación de las redes neuronales y del método del vecino más cercano, sobre la base de la distancia Euclidiana entre los vectores de coeficientes de ondeletas de baja frecuencia que brinda la transformada Ondeleta de las imágenes.

### V.1 Metodología

El procedimiento para recuperar imágenes consiste de dos etapas[44] [45] [49]:

Primera etapa: Entrenamiento de la red neuronal con los siguientes pasos.

- 1.- Convertir las imágenes a patrones.
- 2.- Asignar los patrones a clases específicas.
- 3.- Entrenar la red neuronal.

Segunda etapa: Recuperación de las imágenes mediante una imagen-consulta.

- 1.- Convertir la imagen-consulta a patrón.
- 2.- Clasificar con la red neuronal para determinar la clase a la que pertenece el patrón de la imagen-consulta.

3.- Consultar la base de datos de imágenes únicamente en la clase reportada por la red neuronal, para encontrar los patrones más cercanos con el método del vecino más cercano basado en la distancia Euclidiana y así obtener las imágenes más parecidas a la imagen-consulta.

## V.2 Conversión de imágenes a patrones

Después de aplicar cada uno de los tres métodos descritos en el subepígrafe IV.5. *Codificación de las imágenes*, sobre cada uno de los tres planos de color RGB de la imagen seleccionada, en la Figura 26 con el botón Procesar, se procede al almacenamiento de los 16 coeficientes de baja frecuencia de la transformada Ondeleta de los tres planos, para dar un total de 48 valores reales con una precisión de 10 cifras decimales, los que junto con la especificación de la clase a la que pertenece la imagen, se conforma la codificación de un vector firma para la entrada de la red neuronal con entrenamiento supervisado.

Un ejemplo de los 16 coeficientes de ondeletas se presenta a continuación para cada plano de una imagen:

### Plano Rojo (R)

0.0000000000,0.0039062500,0.0000000000,0.0556640625,0.0595703125,0.0810546875,0.2470703125,0.3349609375,0.2441406250,0.3320312500,0.9492187500,0.4736328125,0.0830078125,0.0498046875,0.0400390625,0.0390625000

### Plano Verde (G)

0.0000000000,0.0039062500,0.0000000000,0.0000000000,0.0566406250,0.0722656250,0.2333984375,0.1884765625,0.3710937500,0.1855468750,0.1142578125,1.0000000000,0.6513671875,0.0556640625,0.0566406250,0.0410156250

### Plano Azul (B)

0.0000000000,0.0039062500,0.0000000000,0.0000000000,0.0146484375,0.0908203125,0.2304687500,0.1298828125,0.3193359375,0.2832031250,0.1708984375,0.2314453125,1.0000000000,0.3056640625,0.0546875000,0.0322265625

Al concatenar los tres vectores en el orden RGB, se conforma un único vector de 48 valores reales; además, se añaden los seis valores (0.9, 0.1, 0.1, 0.1, 0.1, 0.1) que indican, por ejemplo que la imagen pertenece a la clase uno, para así obtener un vector de 54 valores.

Así se procede para cada imagen de la muestra de entrenamiento, ver Figura 26, y de cada una de ellas obtener su respectivo patrón y conformar la base de datos de patrones para su posterior consulta.

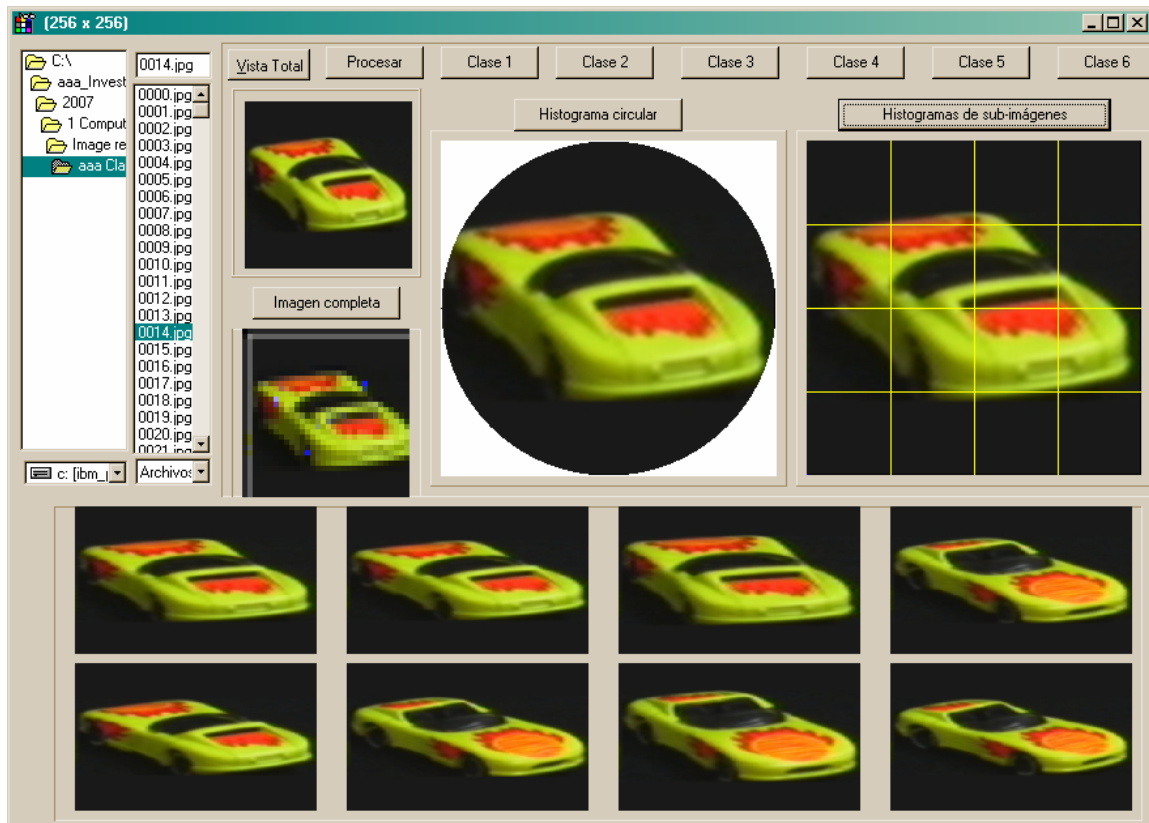


Fig. 26 Conversión de imágenes a patrones.

### V.3 Arquitectura de la red neuronal

Como se puede ver, en la Figura 27 se presenta la arquitectura de la red neuronal artificial compuesta de tres capas:

48 neuronas en la capa de entrada,

49 neuronas en la capa oculta y

6 neuronas en la capa de salida.

Los 48 valores de entrada son para representar el vector firma de la imagen; las 49 neuronas en la capa oculta se crean siguiendo la recomendación citada en [43]; y las 6 neuronas de la capa de salida representan las 6 clases de imágenes consideradas.

En el Apéndice B se muestra el código para implementar la red neuronal.

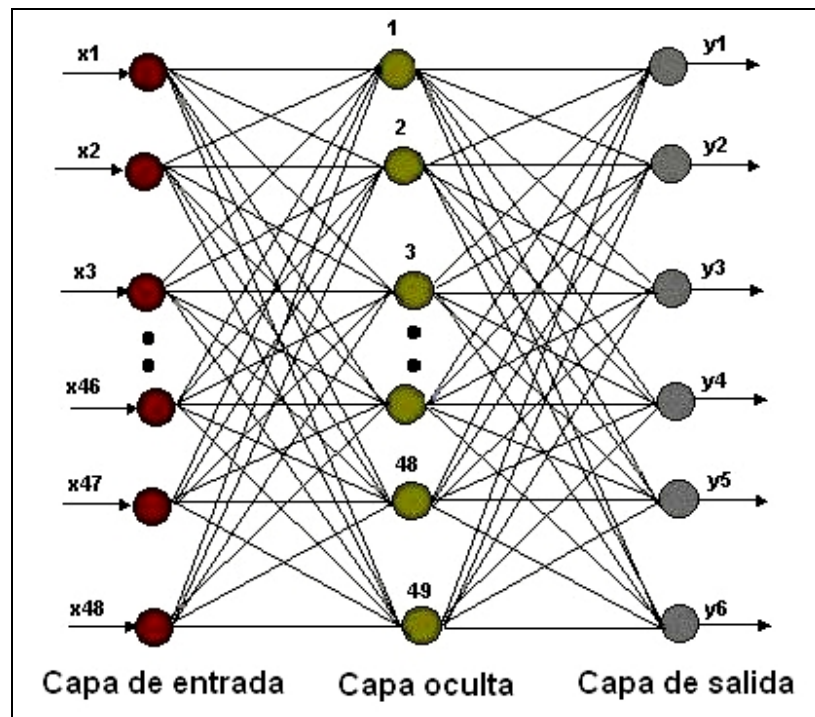


Fig. 27 Arquitectura de la Red Neuronal Artificial.

#### V.4 Entrenamiento de la red neuronal

Para el entrenamiento de la red neuronal se utiliza el Método  $\pi$ , el cual realiza la distribución de los patrones contenidos en la muestra de entrenamiento de un tamaño específico de  $X$  imágenes, por ejemplo 24 imágenes, de forma aleatoria y sin reemplazo, en 5 conjuntos disjuntos  $C_1, C_2, C_3, C_4$  y  $C_5$ . De un total de 120 imágenes, cada conjunto de 24 imágenes contiene cuatro imágenes de cada una de las seis clases, realizándose el entrenamiento de la forma siguiente:

Se toma el primer conjunto como muestra de control y el resto de los conjuntos como muestra de entrenamiento, y se repite el proceso para todos los conjuntos creados como se observa en la Tabla 2.

Cuando se utilizan las redes neuronales para recuperar imágenes, existen dos etapas importantes, la primera que es la del entrenamiento de la red neuronal mediante las firmas de las imágenes, y la segunda etapa, que es la de la recuperación de imágenes desde un banco de imágenes previamente procesado.

Tabla 2. Conjuntos del Método  $\pi$

Muestra de Control	Muestra de Entrenamiento
$C_1$	$C_2, C_3, C_4$ y $C_5$
$C_2$	$C_1, C_3, C_4$ y $C_5$
$C_3$	$C_1, C_2, C_4$ y $C_5$
$C_4$	$C_1, C_2, C_3$ y $C_5$
$C_5$	$C_1, C_2, C_3$ y $C_4$

En el Capítulo 6 se presentan los experimentos hechos y los resultados obtenidos, básicamente de la segunda etapa denominada “Recuperación de imágenes mediante una imagen-consulta”, con diferentes bases de datos de imágenes, desde aquellas con imágenes de estudio de un solo objeto en un fondo homogéneo contrastante, hasta con imágenes complejas de objetos y animales en su medio natural.

## CAPÍTULO VI

# PRUEBAS Y RESULTADOS DE LA RECUPERACIÓN DE IMÁGENES

El propósito de este trabajo es contar con un sistema de recuperación de imágenes robusto y eficiente que sea insensible a variaciones en la orientación de la imagen y a la posición, forma, tamaño y color de los objetos que contiene. Un sistema eficiente de recuperación de imágenes sobre la base del contenido debe tener las siguientes características: *Exactitud, Estabilidad y Velocidad*.

El sistema de recuperación debe ser exacto, esto es, las imágenes recuperadas deben parecerse a la imagen-consulta, y si la imagen-consulta está en la base de datos de imágenes, se debe recuperar ésta como la más parecida. Se clasifica una recuperación como exacta si para una imagen-consulta, en la percepción de un humano, el sistema recupera las imágenes más similares de las contenidas en la base de datos de imágenes.

La estabilidad es el grado de robustez que presente el sistema de recuperación a todos los tipos de consulta, o sea, que el sistema debe responder en todos los casos al variar las condiciones de cada imagen, debiendo mantener sus características de invariante a la rotación, a la traslación y al escalamiento.

La integración de las características de exactitud y estabilidad se pueden medir dentro del porcentaje de eficiencia del sistema de recuperación de imágenes.

En cuanto a la velocidad, el sistema de recuperación debe funcionar en tiempo real con bases de datos de imágenes de cientos o miles de imágenes. Para su comprobación, el tiempo total promedio que toma la recuperación de las imágenes debe ser medido, en nuestro caso la velocidad no se tomo en cuenta.

## VI.1 Resultados experimentales mediante el vecino más cercano

A continuación se brindan los resultados experimentales mediante el vecino más cercano.

### VI.1.1 Criterio de eficiencia

Las pruebas se basan en la recuperación de una imagen-consulta con las restantes de la base de datos de imágenes, donde se buscan y se extraen las 8 imágenes que más se le parezcan visualmente, sobre la base de sus correspondientes vectores característicos, tal como se observa en la Figura 28 donde se presentan 4 consultas.

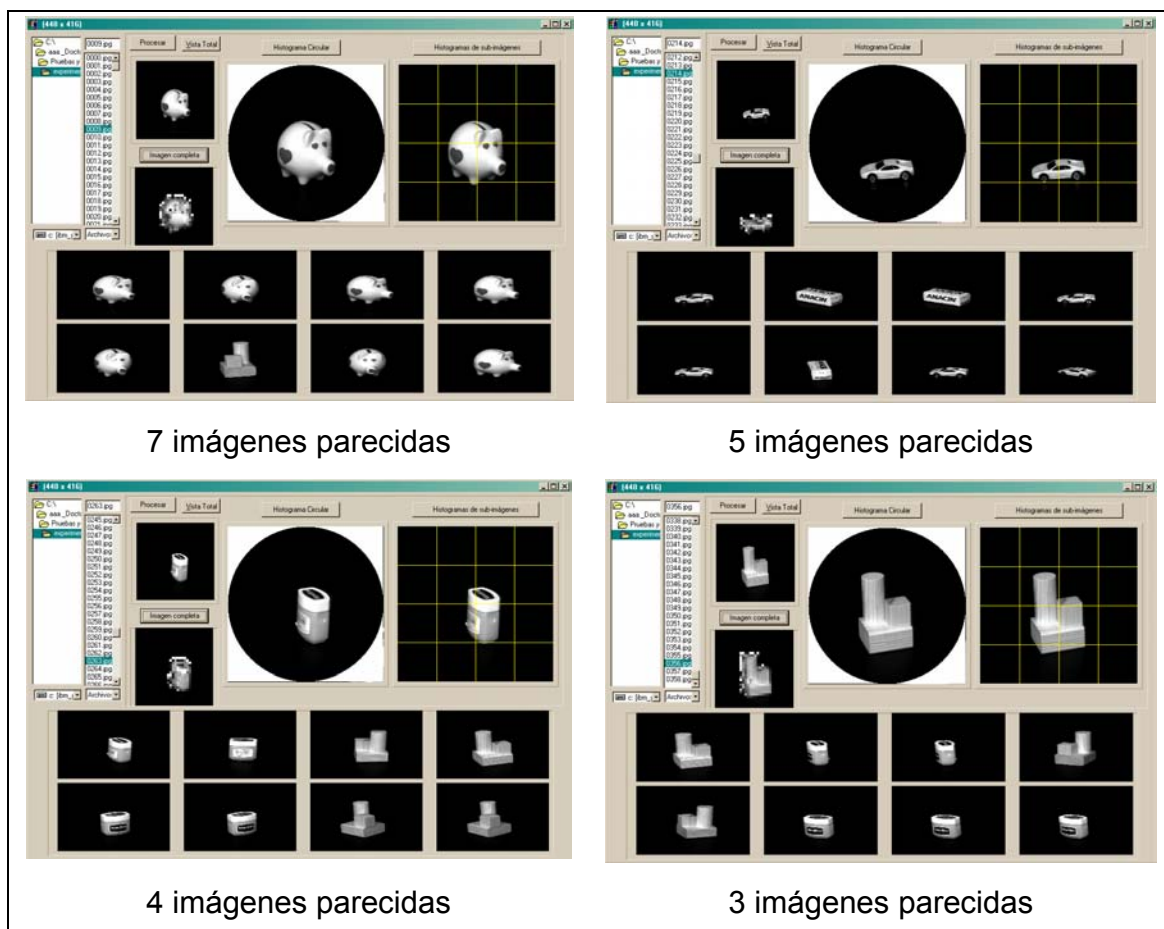


Fig. 28 Consultas con el número de imágenes parecidas.

El usuario del sistema determina visualmente el grado de parecido entre la imagen-consulta y las 8 imágenes dadas por el sistema para cada uno de los tres métodos (columnas 2, 4 y 6 de la Tabla 3).

A partir de ahí se determinan los aciertos de cada método que dan el porcentaje de la eficiencia en la recuperación en cada consulta (columnas 3, 5 y 7 de la Tabla 3).

Posteriormente se calcula el promedio de los porcentajes de eficiencia de las 25 consultas, para obtener así el porcentaje de eficiencia de cada método, expresado en la Tabla 4.

Tabla 3. Resultados de 25 consultas

Consultas	(A)	%	(B)	%	(C)	%
1	2	0.25	4	0.5	5	0.625
2	1	0.125	4	0.5	5	0.625
3	3	0.375	4	0.5	5	0.625
4	2	0.25	5	0.625	5	0.625
5	2	0.25	5	0.625	4	0.5
6	1	0.125	4	0.5	6	0.75
7	3	0.375	3	0.375	7	0.875
8	3	0.375	4	0.5	7	0.875
9	3	0.375	5	0.625	7	0.875
10	3	0.375	3	0.375	8	1.0
11	3	0.375	5	0.625	4	0.5
12	2	0.25	4	0.5	4	0.5
13	2	0.25	4	0.5	4	0.5
14	3	0.375	4	0.5	5	0.625
15	1	0.125	4	0.5	5	0.625
16	2	0.25	5	0.625	4	0.5
17	3	0.375	5	0.625	5	0.625
18	3	0.375	7	0.875	7	0.875
19	4	0.5	7	0.875	7	0.875
20	5	0.625	6	0.75	8	1.0
21	4	0.5	4	0.5	8	1.0
22	4	0.5	4	0.5	8	1.0
23	5	0.625	7	0.875	6	0.75
24	4	0.5	3	0.375	8	1.0
25	3	0.375	5	0.625	8	1.0
Promedios		0.355		0.575		0.75

Tabla 4. Eficiencia de los tres métodos

Método	EFICIENCIA
(A) Coeficientes de la imagen completa	0.355
(B) Coeficientes del histograma de una ventana circular	0.575
(C) Coeficientes de los histogramas de las sub-imágenes	0.75



## VI.1.2 Experimento 1. Imágenes de estudio en tonos de gris

La base de datos utilizada fue una parte de la Biblioteca de la Universidad de Columbia [56]. Se trata de 360 imágenes en niveles de gris de 5 diferentes objetos reales aislados, los que se muestran en la Figura 29, donde cada objeto aparece dispuesto en 72 orientaciones diferentes y con fondo homogéneo contrastante.



Fig. 29. Imágenes de estudio en tonos de gris.

La Figura 30 muestra la pantalla de la interfaz gráfica con el resultado de una consulta.

Se realizó el 100% de las consultas, calculándose la eficiencia total mediante el promedio de los resultados de todas ellas.

La eficiencia final de la recuperación de las imágenes para los tres métodos evaluados se muestra en la Tabla 5, calculada sobre la base del número de imágenes que resultaron semejantes a la imagen-consulta (ver subepígrafe VI.1.1).

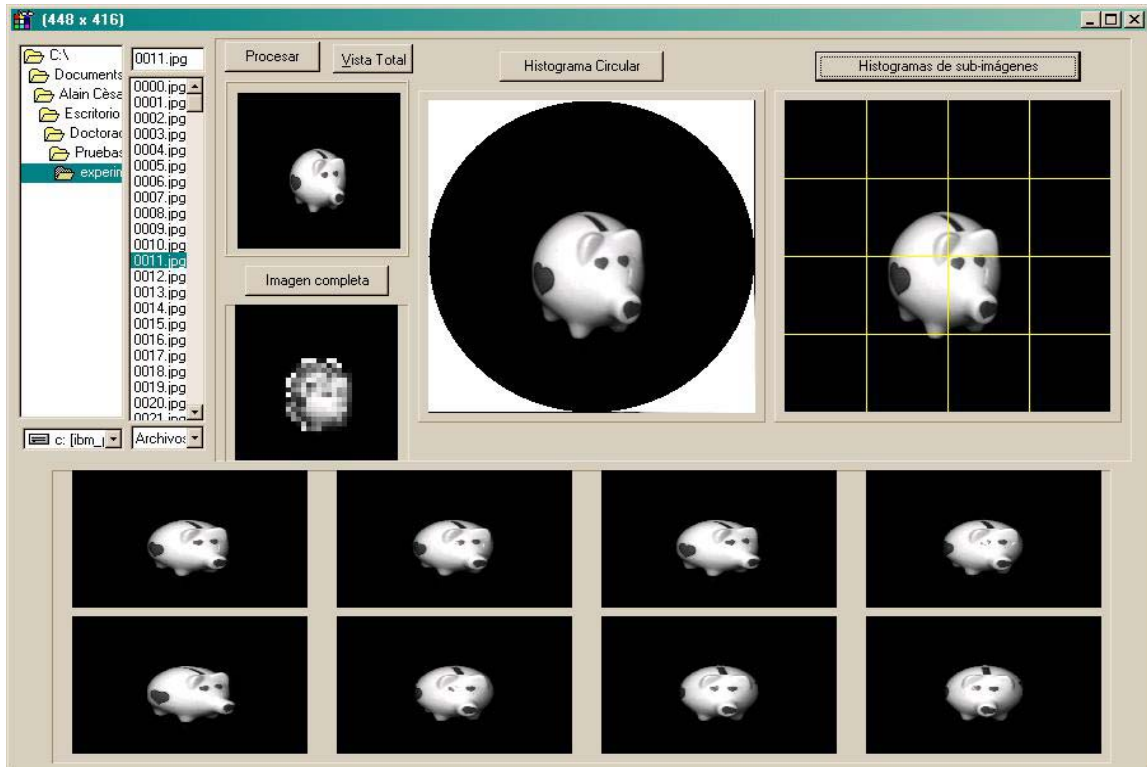


Fig. 30 Consulta con imágenes de estudio en tonos de gris.

Tabla 5. Eficiencia del experimento 1: imágenes de estudio

Método	EFICIENCIA
Coeficientes de la imagen completa	0.76
Coeficientes del histograma de una ventana circular	0.90
Coeficientes de los histogramas de las sub-imágenes	0.96

### VI.1.3 Experimento 2. Imágenes de memorama

La base de datos utilizada contiene imágenes de un memorama, como las imágenes de los objetos que se presentan en la Figura 31. Contiene 180 imágenes en colores de 36 diferentes objetos reales aislados, en 5 formas diferentes dispuestas en diferentes orientaciones y con fondo homogéneo contrastante.

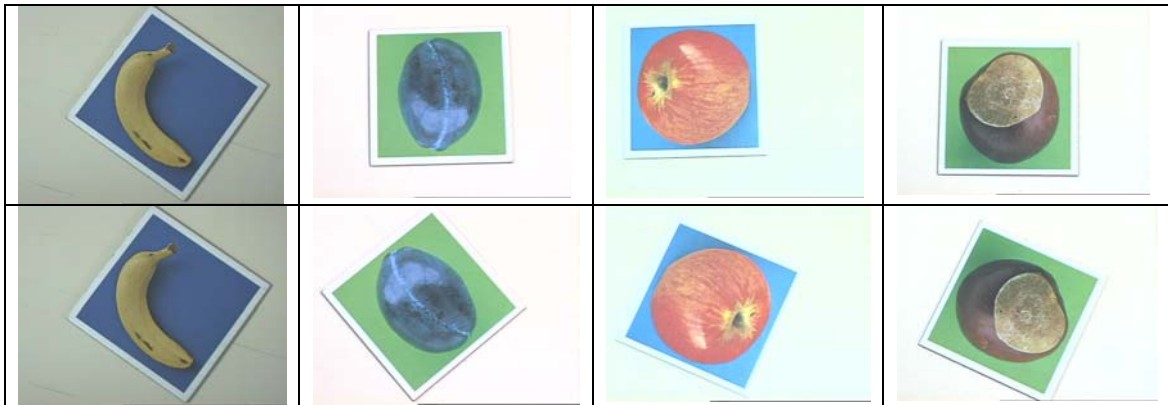


Fig. 31 Imágenes de objetos de un memorama.

La Figura 32 muestra la pantalla de la interfaz gráfica con el resultado de una consulta.

Se realizó el 100% de las consultas, calculándose la eficiencia total mediante el promedio de los resultados de todas ellas.

La eficiencia final de la recuperación de las imágenes para los tres métodos evaluados se muestra en la Tabla 6, calculada sobre la base del número de imágenes que resultaron semejantes a la imagen-consulta (ver subepígrafe VI.1.1).

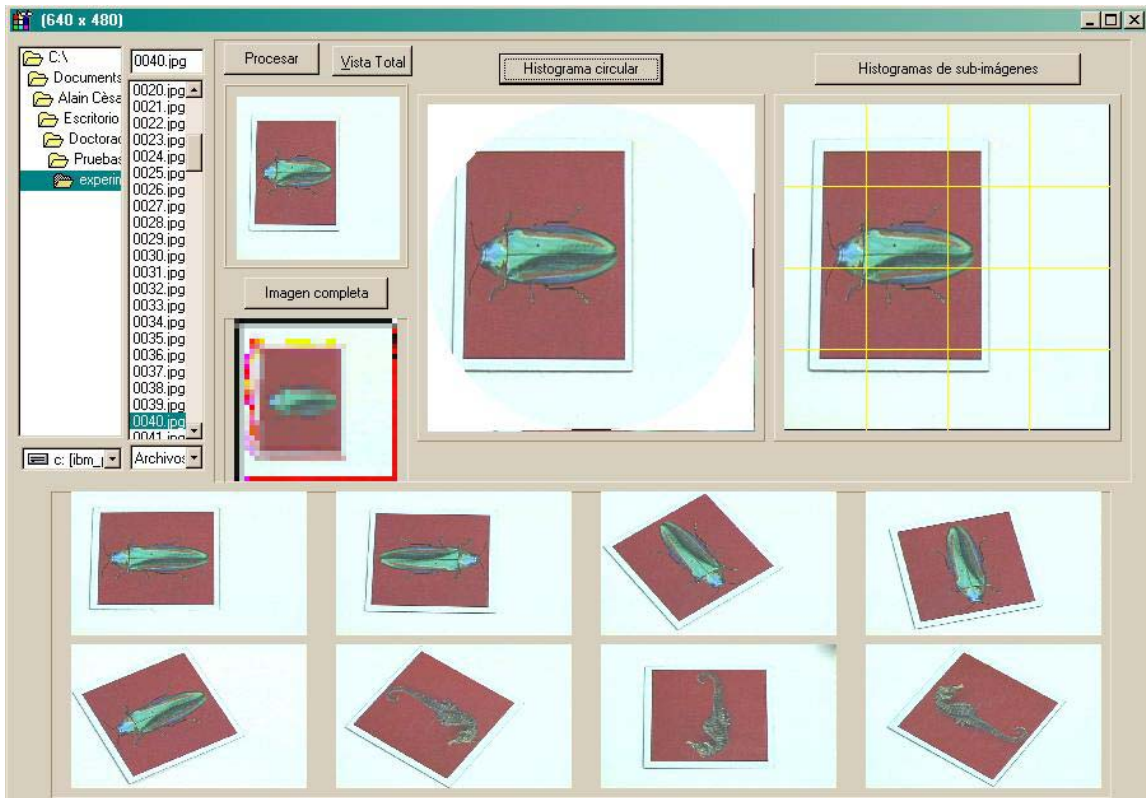


Fig. 32 Consulta con imágenes de un memorama.

Tabla 6. Eficiencia del experimento 2: con las imágenes de un memorama

Método	EFICIENCIA
Coeficientes de la imagen completa	0.46
Coeficientes del histograma de una ventana circular	0.88
Coeficientes de los histogramas de las sub-imágenes	0.90

### VI.1.4 Experimento 3. Imágenes de aviones en color

La base de datos utilizada fue la de la Biblioteca de “Aircraft-Images” [57]. En la Figura 33 se muestran las seis clases de imágenes de aviones de que consta. La base de imágenes consiste en 1068 imágenes en colores de aviones con fondo complejo, todos diferentes entre sí.



Fig. 33 Imágenes de seis diferentes clases de aviones.

La Figura 34 muestra la pantalla de la interfaz gráfica con el resultado de una consulta.

Se realizaron 300 consultas, calculándose el promedio de los resultados de eficiencia de todas ellas.

La eficiencia final de la recuperación de las imágenes para los tres métodos evaluados se muestra en la Tabla 7, calculada sobre la base del número de imágenes que resultaron semejantes a la imagen-consulta (ver subepígrafe VI.1.1).

Tabla 7. Eficiencia del experimento 3: imágenes de aviones

Método	EFICIENCIA
Coeficientes de la imagen completa	0.64
Coeficientes del histograma de una ventana circular	0.84
Coeficientes de histogramas de sub-imágenes	0.85

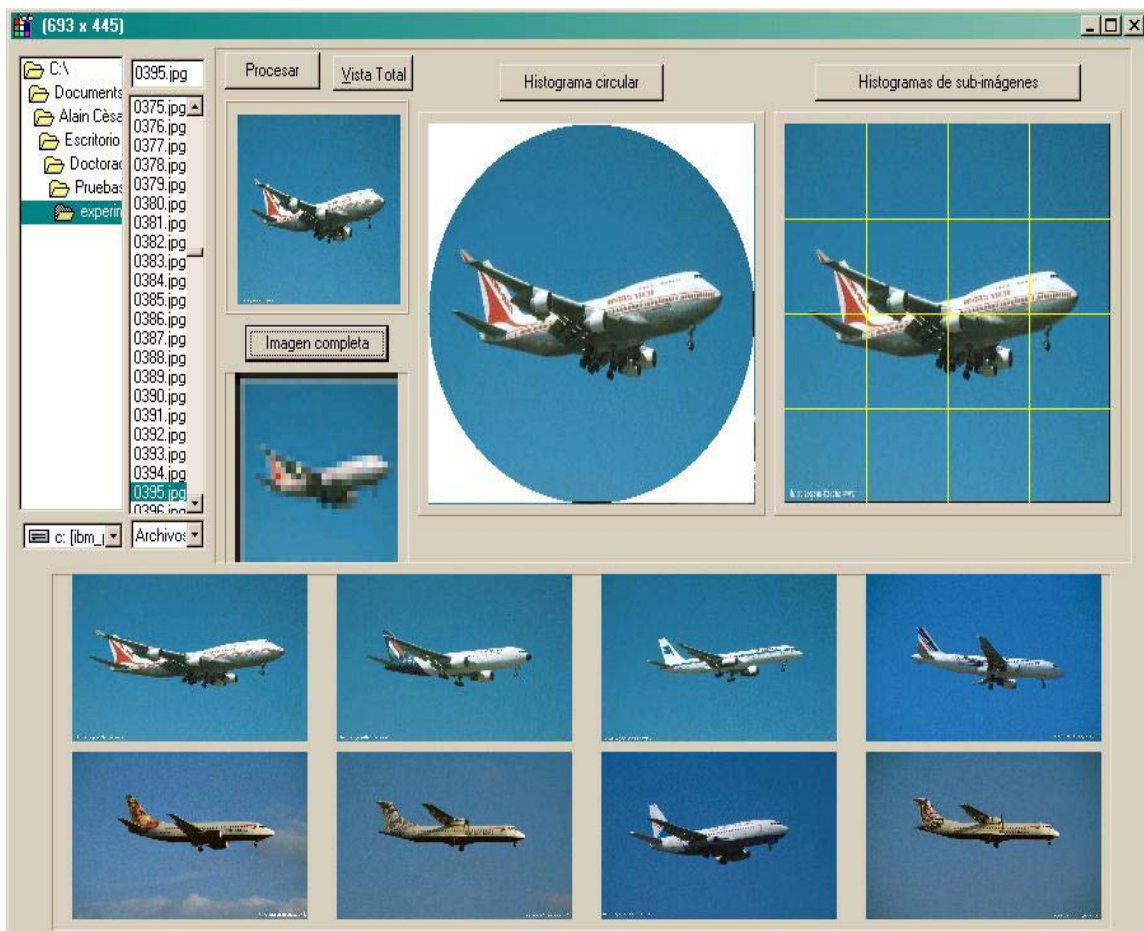


Fig. 34 Consulta con imágenes de las seis clases diferentes de aviones

### VI.1.5 Experimento 4. Instrumental quirúrgico

La base de datos utilizada contiene imágenes de instrumental quirúrgico. Las imágenes de los objetos se presentan en la Figura 35. Contiene 60 imágenes de 6 diferentes objetos reales aislados, dispuestos en diferentes orientaciones y con fondo contrastante.



Fig. 35 Imágenes de instrumental quirúrgico.

La Figura 36 muestra la pantalla de la interfaz gráfica con el resultado de una consulta.

Se realizó el 100% de las consultas, calculándose la eficiencia mediante el promedio de los resultados de todas ellas.

La eficiencia final de la recuperación de las imágenes para los tres métodos evaluados se muestra en la Tabla 8, calculada sobre la base del número de imágenes que resultaron semejantes a la imagen-consulta (ver subepígrafe VI.1.1).

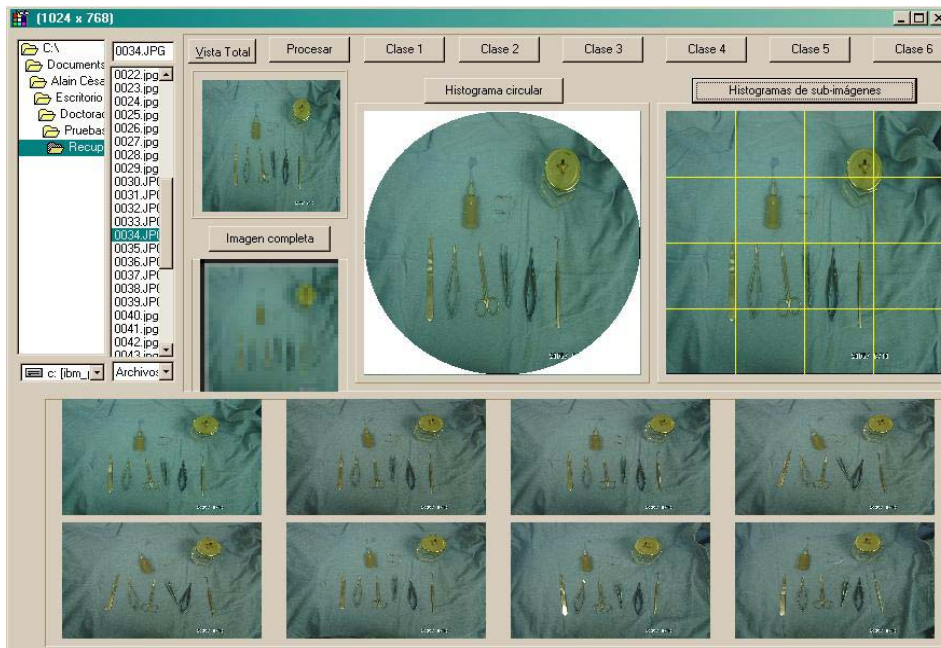


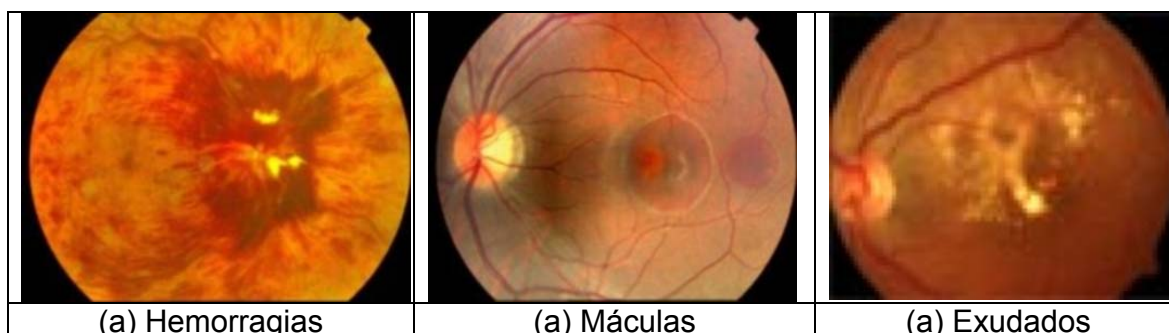
Fig. 36 Consulta con imágenes de instrumental quirúrgico.

Tabla 8. Eficiencia del experimento 4: imágenes de instrumental quirúrgico

Método	EFICIENCIA
Coefficientes de la imagen completa	0.56
Coefficientes del histograma de una ventana circular	0.82
Coefficientes de los histogramas de las sub-imágenes	0.80

### VI.1.6 Experimento 5. Retinas humanas

La base de datos utilizada contiene imágenes en colores de retinas humanas. Las imágenes de los objetos se presentan en la Figura 37. Contiene 76 imágenes de diferentes objetos reales aislados de 6 clases diferentes: normales y con 5 tipos de enfermedades distintas.



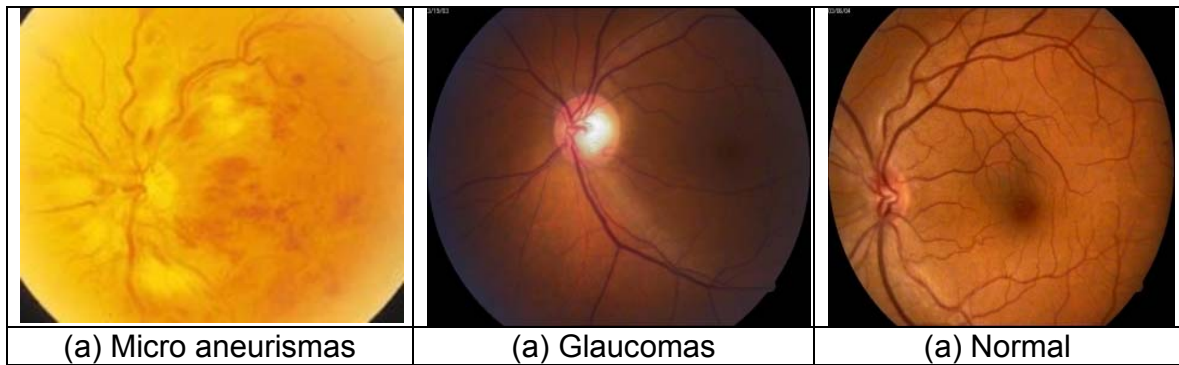


Fig. 37 Imágenes de retinas humanas

La Figura 38 muestra la pantalla de la interfaz gráfica con el resultado de una consulta.

Se realizó el 100% de las consultas, calculándose la eficiencia mediante el promedio de los resultados de todas ellas.

La eficiencia final de la recuperación de las imágenes para los tres métodos evaluados se muestra en la Tabla 9, calculada sobre la base del número de imágenes que resultaron semejantes a la imagen-consulta (ver subepígrafe VI.1.1).

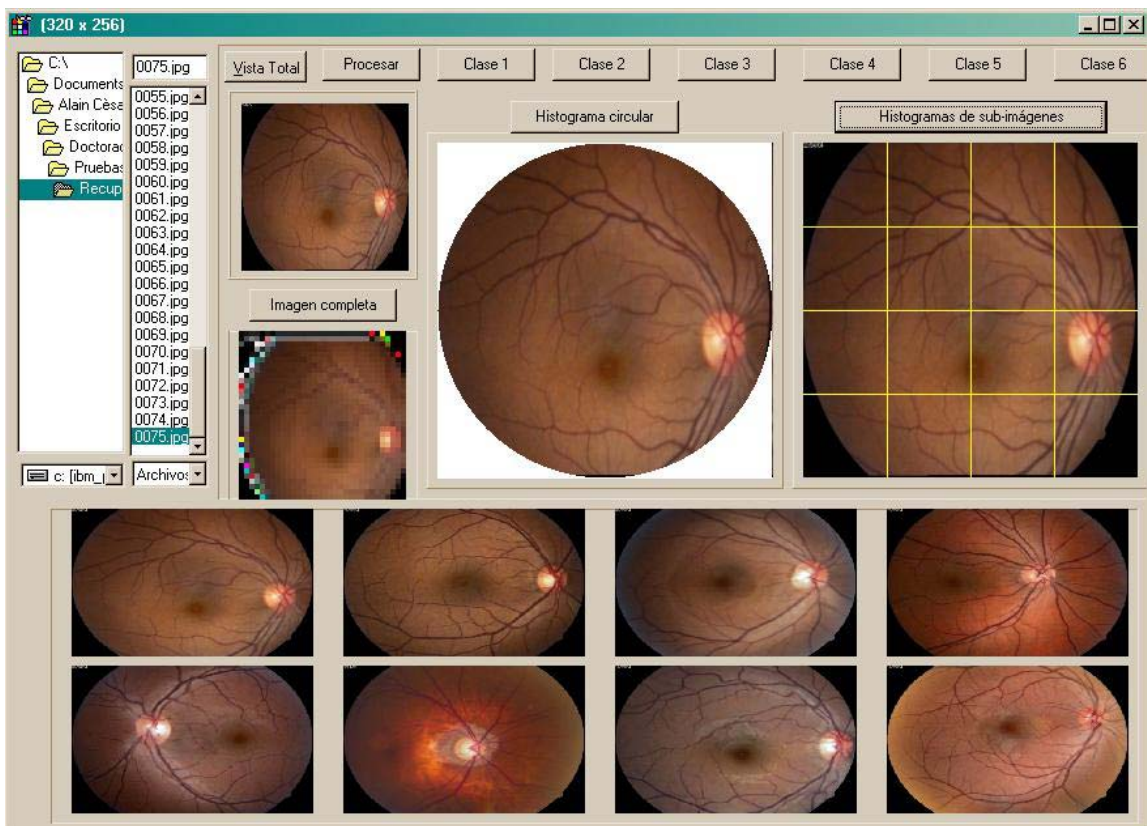




Fig. 38 Consulta con imágenes de retinas humanas

Tabla 9. Eficiencia del experimento 5: imágenes de retinas humanas

Método	EFICIENCIA
Coeficientes de la imagen completa	0.62
Coeficientes del histograma de una ventana circular	0.89
Coeficientes de los histogramas de las sub-imágenes	0.87

### VI.1.7 Experimento 6. 96 imágenes de 6 clases

La base de datos utilizada contiene imágenes en colores de 6 clases diferentes de objetos. Las imágenes de los objetos se presentan en la Figura 39. Contiene 96 imágenes de diferentes objetos en su entorno natural.

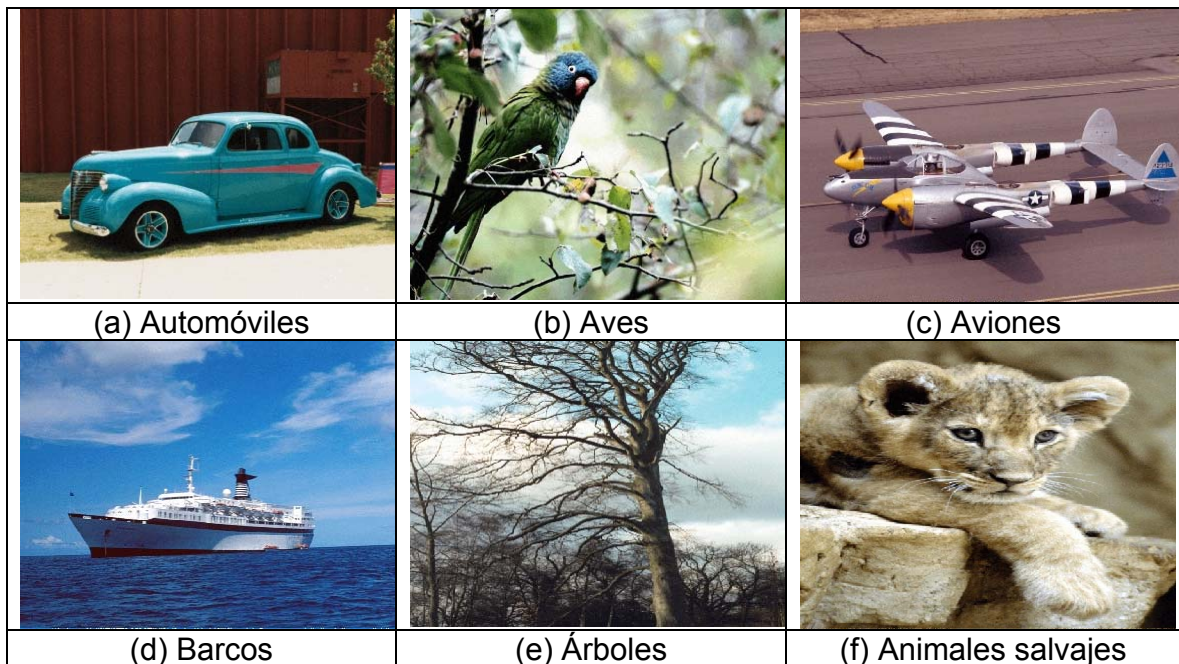


Fig. 39 96 imágenes de 6 clases

La Figura 40 muestra la pantalla de la interfaz gráfica con el resultado de una consulta.

Se realizó el 100% de las consultas, calculándose la eficiencia mediante el promedio de los resultados de todas ellas.

La eficiencia final de la recuperación de las imágenes para los tres métodos evaluados se muestra en la Tabla 10, calculada sobre la base del número de imágenes que resultaron semejantes a la imagen-consulta (ver subepígrafe VI.1.1).

Tabla 10. Eficiencia del experimento 6: 96 imágenes de 6 clases

Método	EFICIENCIA
Coeficientes de la imagen completa	0.48
Coeficientes del histograma de una ventana circular	0.62
Coeficientes de los histogramas de las sub-imágenes	0.64

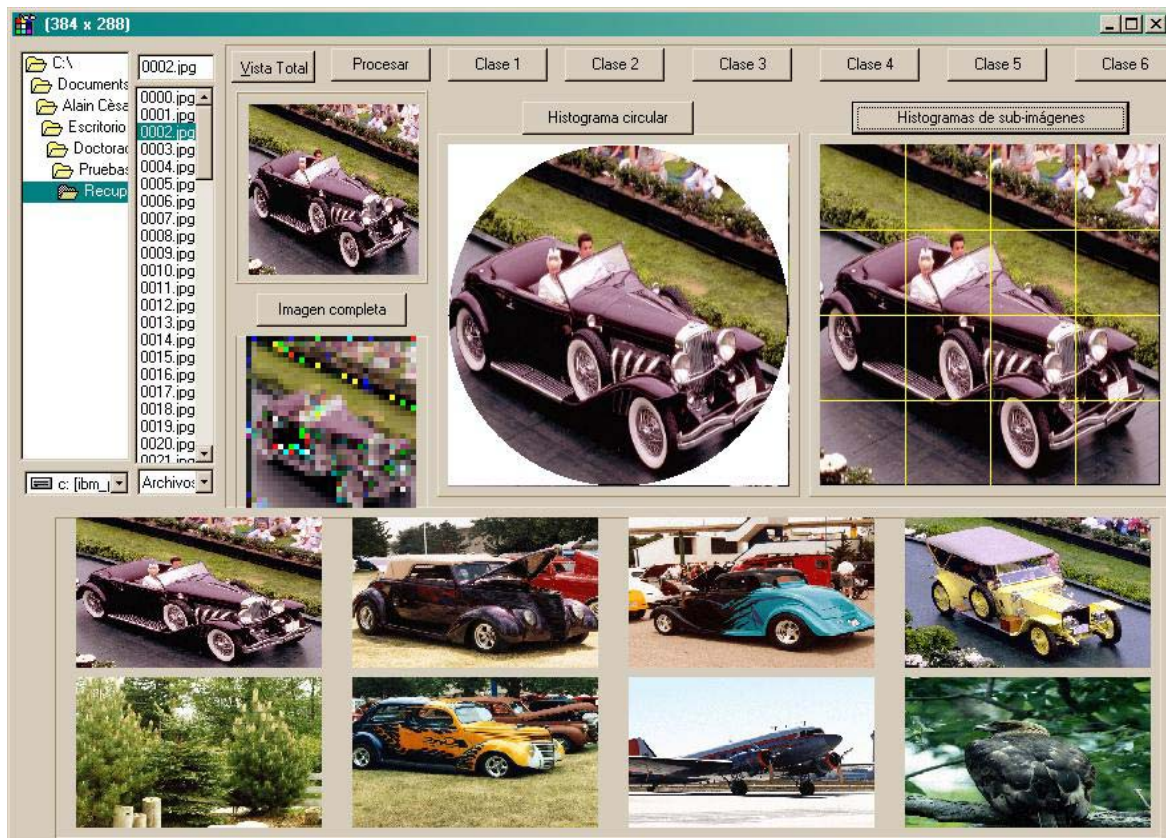


Fig. 40 Consulta con 96 imágenes de 6 clases

### VI.1.8 Experimento 7. 192 imágenes de 6 clases

La base de datos utilizada contiene imágenes en colores de 6 clases de paisajes. Las imágenes de los paisajes se presentan en la Figura 41. Contiene 192 imágenes de diferentes paisajes.

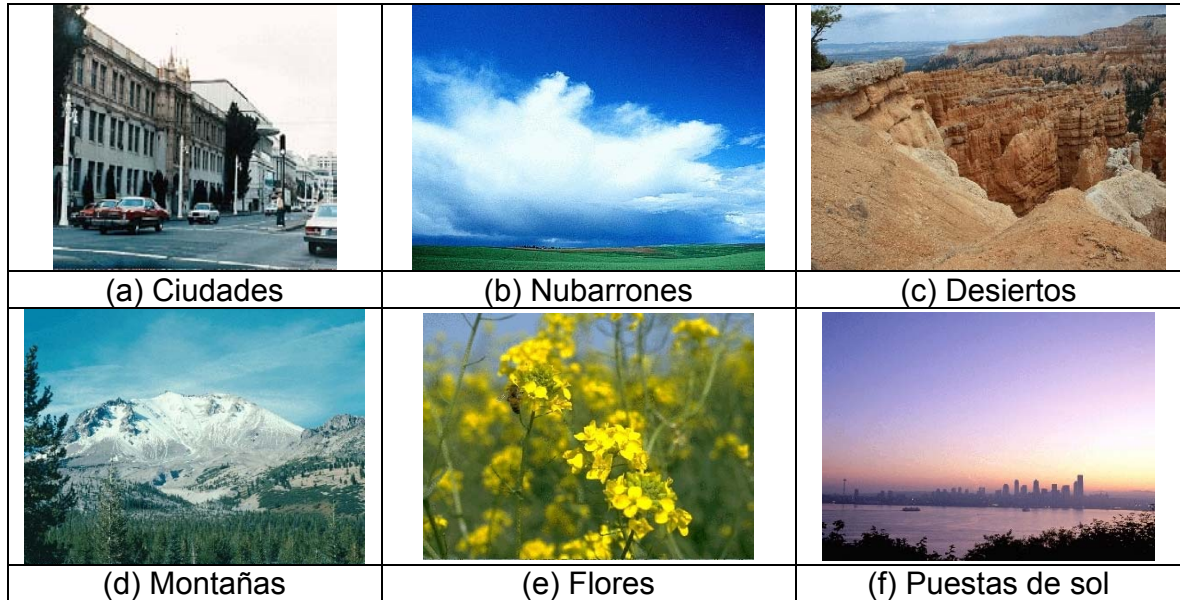


Fig. 41 192 imágenes de 6 clases

La Figura 42 muestra la pantalla de la interfaz gráfica con el resultado de una consulta.

Se realizó el 100% de las consultas, calculándose la eficiencia mediante el promedio de los resultados de todas ellas.

La eficiencia final de la recuperación de las imágenes para los tres métodos evaluados se muestra en la Tabla 11, calculada sobre la base del número de imágenes que resultaron semejantes a la imagen-consulta (ver subepígrafe VI.1.1).

Tabla 11. Eficiencia del experimento 7: 192 imágenes de 6 clases

Método	EFICIENCIA
Coeficientes de la imagen completa	0.52
Coeficientes del histograma de una ventana circular	0.60
Coeficientes de los histogramas de las sub-imágenes	0.62

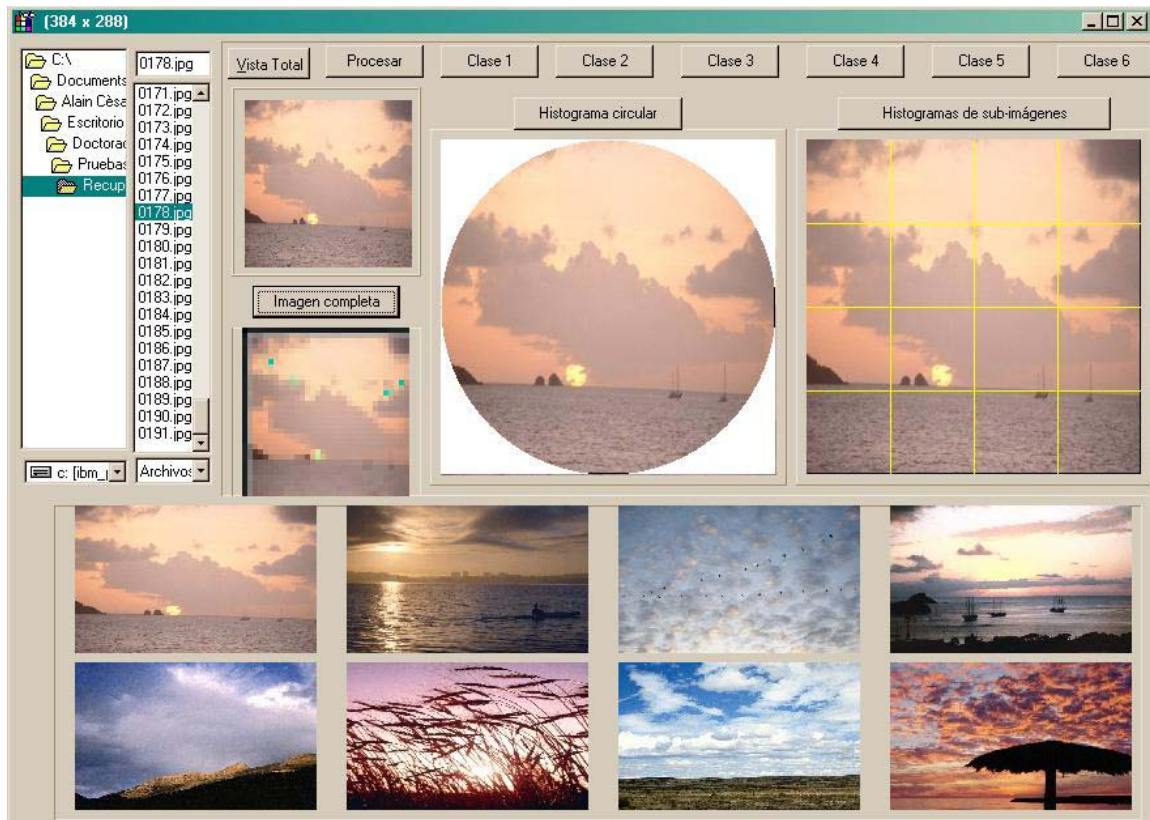


Fig. 42 Consulta con 192 imágenes de 6 clases

## VI.2 Resultados experimentales mediante redes neuronales

Se desarrollan tres procesos: Primero se convierten las imágenes a sus vectores característicos; segundo se entrena la red neuronal con los vectores característicos como entradas; y tercero, se procede a la recuperación de las imágenes, para que a partir de ahí se determinen los por cientos de error que dan un índice de la eficiencia en la recuperación.

### VI.2.1 Experimento 8. Imágenes de estudio en colores

La base de datos utilizada fue una parte de la de la Biblioteca de la Universidad de Columbia [56]. Se trata de 432 imágenes en color de 6 diferentes objetos aislados, los que se muestran en la Figura 43, donde cada objeto aparece dispuesto en 72 orientaciones diferentes y con fondo homogéneo contrastante.

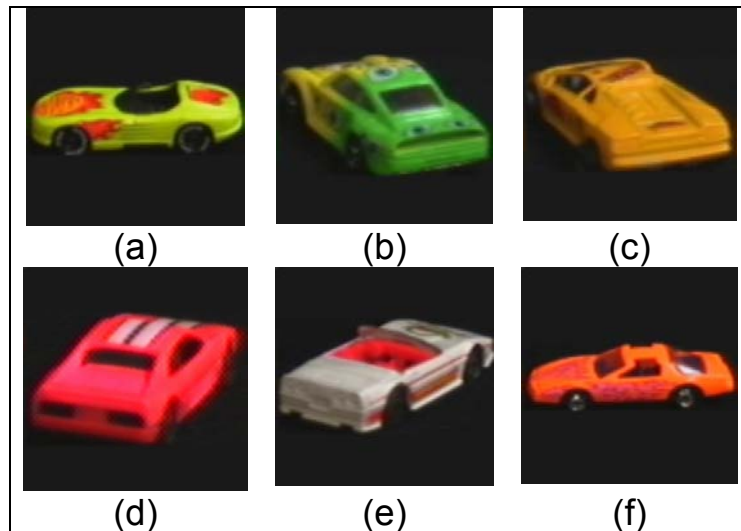


Fig. 43 Imágenes de estudio en color.

En la Figura 44 se muestran algunas imágenes del mismo objeto en diferentes orientaciones, se observa que el objeto se encuentra rotado ciertos grados con respecto a la primera imagen.

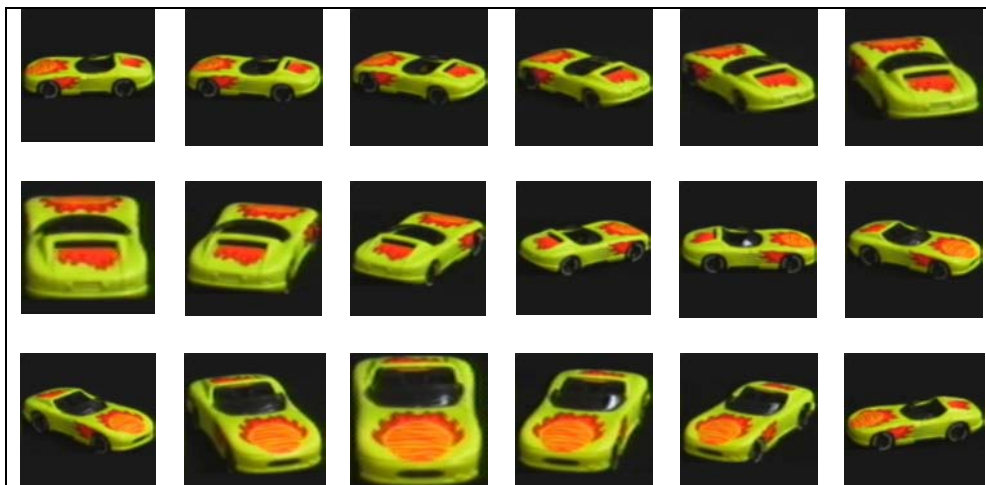


Fig. 44. Imágenes de estudio con diferentes rotaciones.

La Figura 45 muestra la pantalla de la interfaz gráfica con el resultado de 6 consultas.

Se realizó el 100% de las consultas, calculándose la eficiencia total mediante el promedio de los resultados de todas ellas.

La eficiencia final de la recuperación de las imágenes para los tres métodos evaluados se muestra en la Tabla 12, calculada sobre la base del número de imágenes que resultaron semejantes a la imagen-consulta (ver subepígrafe VI.1.1).

Tabla 12. Eficiencia del experimento 8: imágenes de estudio en color con RN

<b>Método</b>	<b>EFICIENCIA</b>
Coeficientes de la imagen completa	0.84
Coeficientes del histograma de una ventana circular	0.96
Coeficientes de los histogramas de las sub-imágenes	0.96

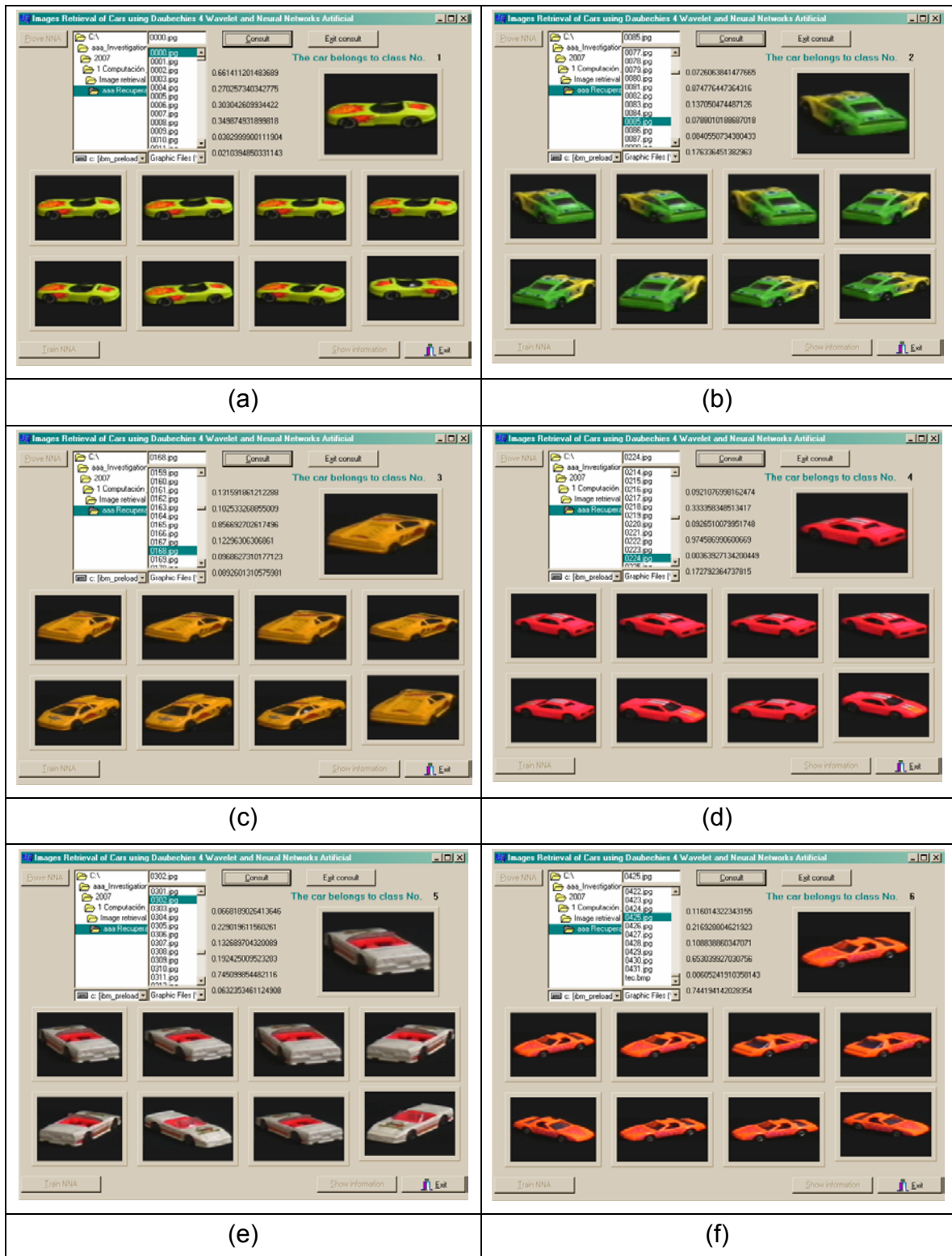


Fig. 45 Consulta con imágenes de estudio en color.

## VI.2.2 Experimento 9. Imágenes de aviones en tonos de gris

La base de datos utilizada fue la de la Biblioteca de "Aircraft-Images" [57]. En la Figura 46 se muestran las 6 clases de imágenes de aviones de que consta. La base de imágenes consiste en 1068 imágenes en tonos de gris con fondo complejo, todos los aviones diferentes entre sí.

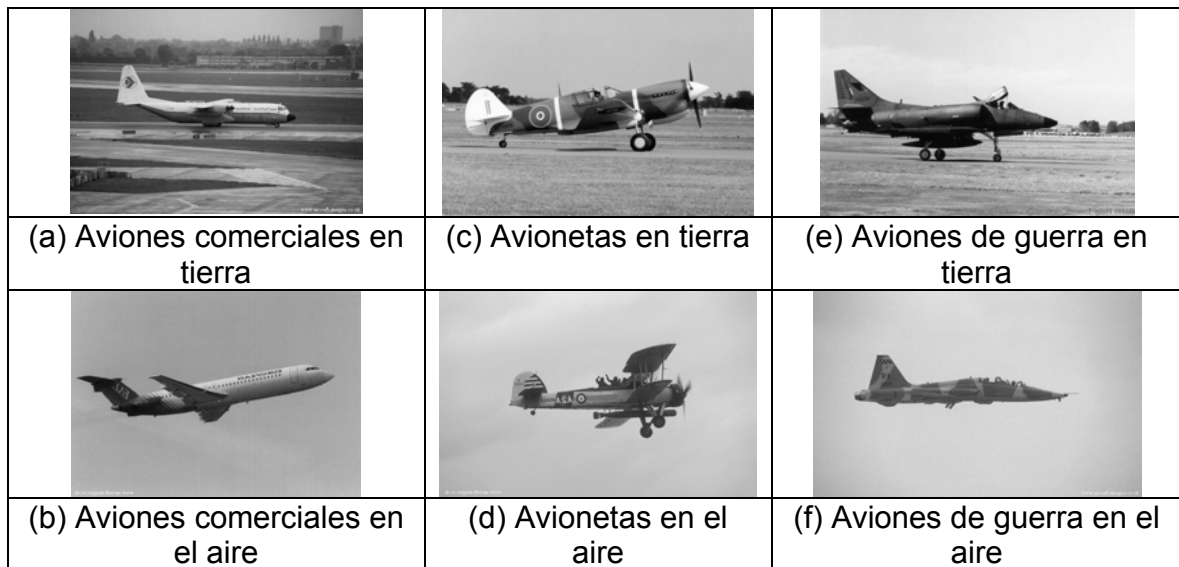


Fig. 46 Imágenes de las 6 clases de aviones, en tonos de gris.

La Figura 47 muestra la pantalla de la interfaz gráfica con el resultado de 6 consultas. Se realizaron las consultas, calculándose el promedio de los resultados de eficiencia de todas ellas. La eficiencia final de la recuperación de las imágenes para los tres métodos evaluados se muestra en la Tabla 13, calculada sobre la base del número de imágenes que resultaron semejantes a la imagen-consulta (ver subepígrafe VI.1.1).

Tabla 13. Eficiencia del experimento 9: imágenes de aviones en tonos de gris con RN

Método	EFICIENCIA
Coeficientes de la imagen completa	0.60
Coeficientes del histograma de una ventana circular	0.78
Coeficientes de histogramas de sub-imágenes	0.72



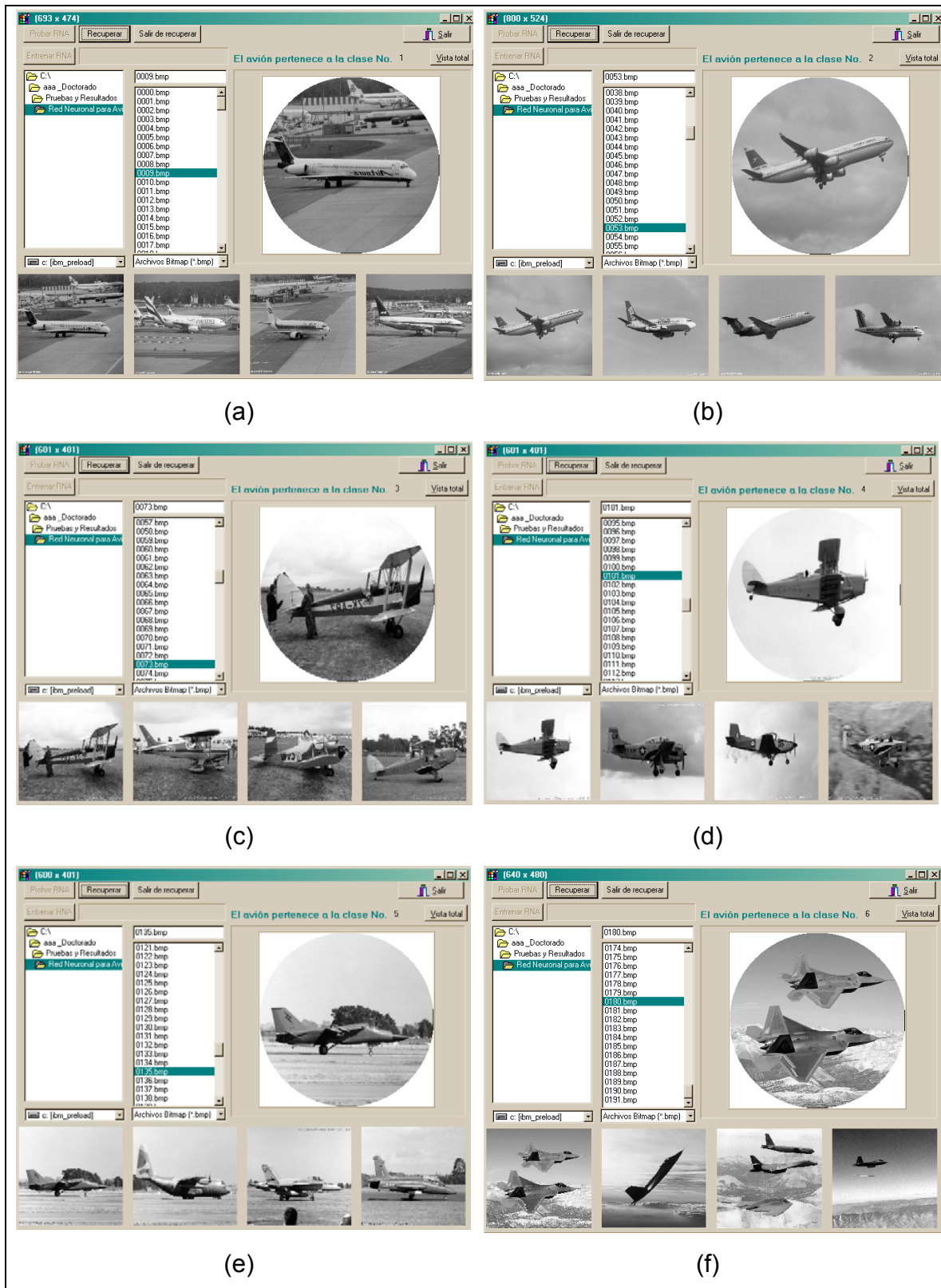


Fig. 47 Consultas con imágenes de las 6 clases de aviones.

### VI.2.3 Experimento 10. Imágenes de aviones en color

El problema específico es recuperar las imágenes de aviones de una de las seis clases siguientes: aviones comerciales en tierra, aviones comerciales en el aire, aviones militares en tierra, aviones militares en el aire, avionetas en tierra y avionetas en el aire, que se muestran en la Figura 33. La base de datos utilizada fue la de la Biblioteca de “Aircraft-Images” [57], consistente en 1068 imágenes de aviones en exteriores, todos diferentes entre sí.

Se aplica la metodología descrita en el capítulo IV, que consiste de dos etapas:

#### Primera etapa:

Conversión de las imágenes a patrones. El objetivo es asignar una clase específica a cada imagen y obtener los vectores característicos de todas las imágenes, mediante los tres métodos descritos en el subepígrafe IV.5 *Codificación de las imágenes*. Un ejemplo se muestra en la Fig. 48.



Fig. 48 Conversión de las imágenes de aviones a patrones.

#### Segunda etapa:

En la Figura 49 se muestra una consulta después de entrenar la red neuronal. Se utiliza el método de extracción de coeficientes a partir del histograma de una ventana circular, descrito en el subepígrafe IV.5.2.

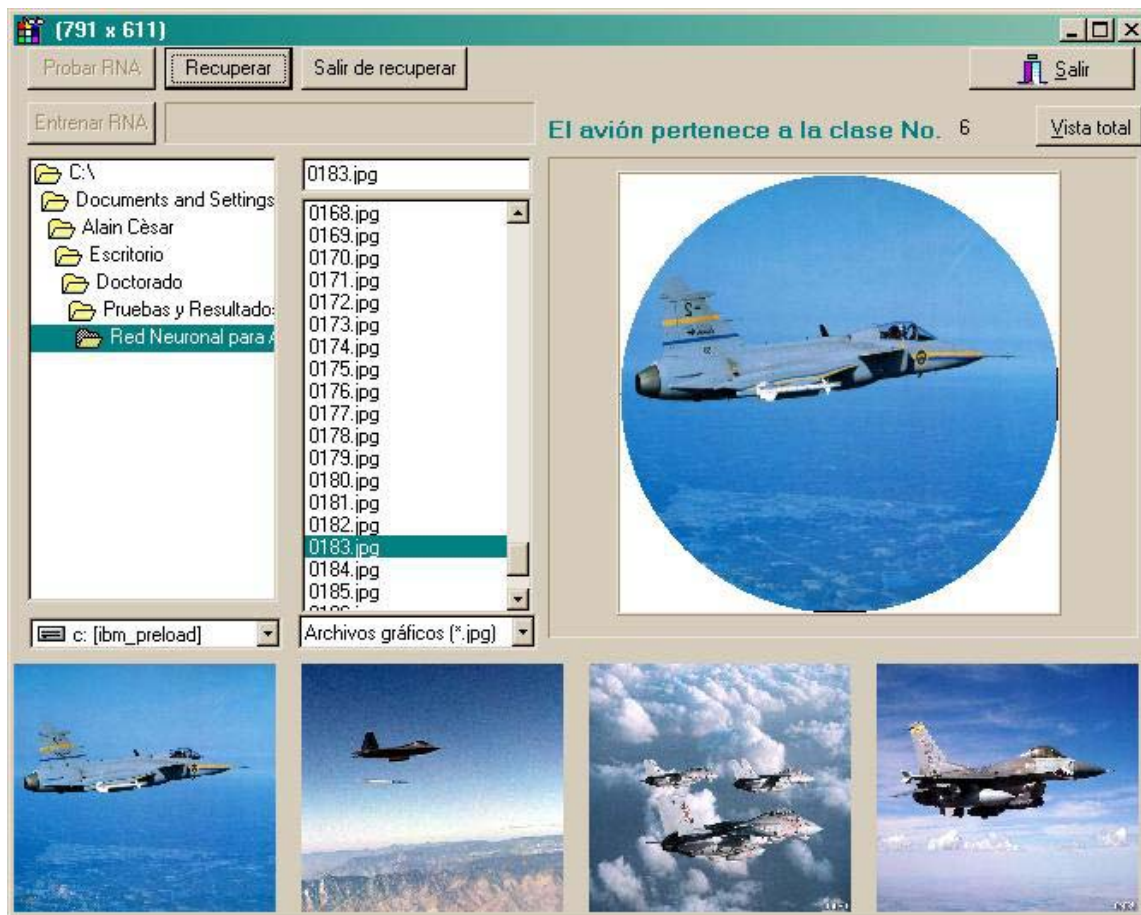


Fig. 49 Consulta de imágenes de aviones con RN.

La eficiencia final de la recuperación de las imágenes para los tres métodos evaluados, calculada sobre la base del número de imágenes que resultaron semejantes a la imagen-consulta, se muestra en la Tabla 14.

Tabla 14. Eficiencia del experimento 10: imágenes de aviones con RN

Método	EFICIENCIA
Coefficientes de las imágenes completas con la RN	0.72
Coefficientes del histograma de una ventana circular con la RN	0.86
Coefficientes de histogramas de sub-imágenes con la RN	0.88

## VI.2.4 Experimento 11. Instrumental quirúrgico

Se aplica la metodología descrita en el capítulo IV, que consiste de dos etapas.

**Primera etapa:** Conversión de las imágenes a patrones

Se obtienen los vectores característicos de todas las imágenes mediante los tres métodos descritos en el subepígrafe *IV.5 Codificación de las imágenes*, para posteriormente asignar una clase específica a cada imagen.

**Segunda etapa:** Consulta

En la Figura 50 se muestra el resultado de una consulta, previo entrenamiento de la red neuronal.

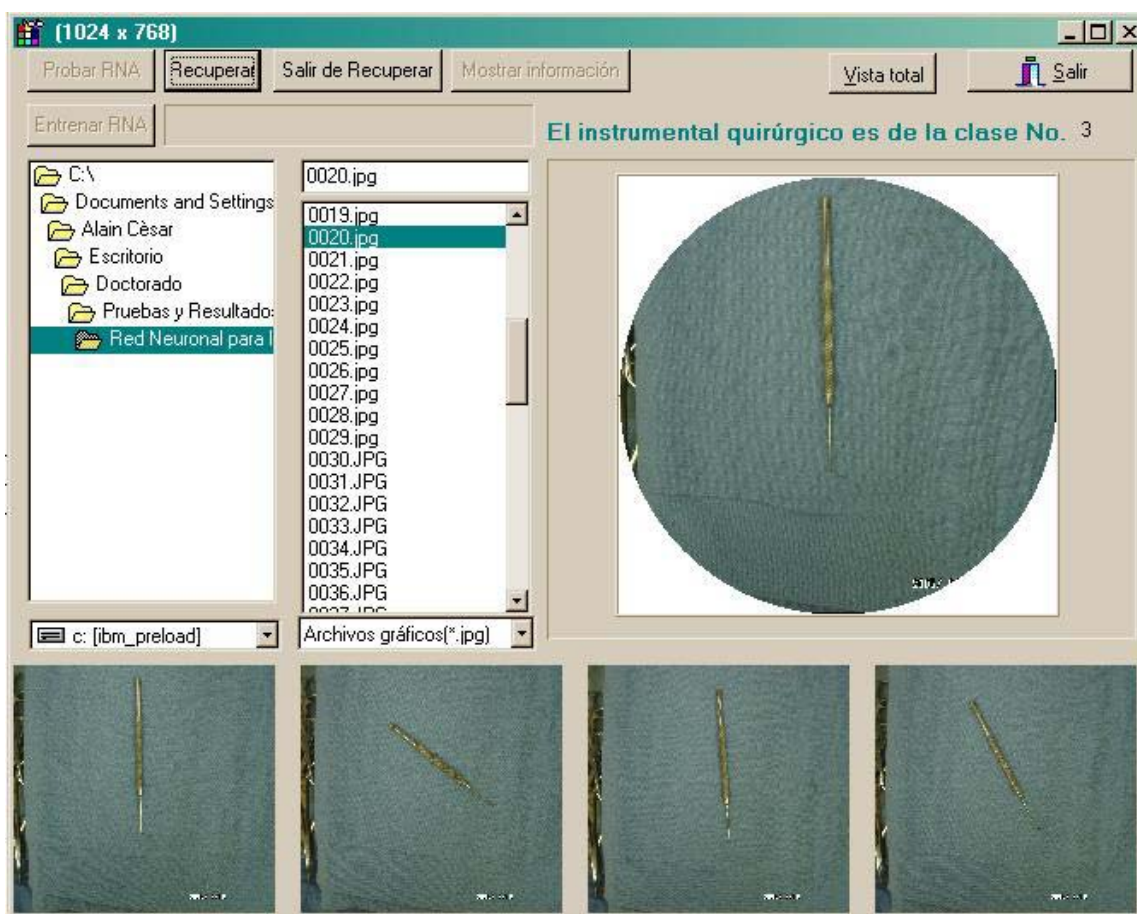


Fig. 50 Consulta de imágenes de instrumental quirúrgico con RN.

Se realizó el 100% de las consultas, calculándose la eficiencia mediante el promedio de los resultados de todas ellas. La eficiencia final de la recuperación de las imágenes para los tres métodos evaluados, calculada sobre la base del número de imágenes que resultaron semejantes a la imagen-consulta, se muestra en la Tabla 15.

Tabla 15. Eficiencia del experimento 11: imágenes de instrumental quirúrgico con RN

Método	EFICIENCIA
Coeficientes de las imágenes completas con la RN	0.88
Coeficientes del histograma de una ventana circular con la RN	0.94
Coeficientes de histogramas de sub-imágenes con la RN	0.92

### VI.2.5 Experimento 12. Retinas humanas

Se aplica la metodología descrita en el capítulo IV, que consiste de dos etapas.

#### Primera etapa:

Conversión de las imágenes a patrones, para asignar una clase específica a cada imagen y obtener los vectores característicos de cada una de ellas, mediante los tres métodos descritos en el subepígrafe *IV.5 Codificación de las imágenes*.

#### Segunda etapa:

En la Figura 51 se muestra el resultado de una consulta, previo entrenamiento de la red neuronal.

Se realizó el 100% de las consultas, calculándose la eficiencia mediante el promedio de los resultados de todas ellas.

La eficiencia final de la recuperación de las imágenes para los tres métodos evaluados, calculada sobre la base del número de imágenes que resultaron semejantes a la imagen-consulta, se muestra en la Tabla 16.

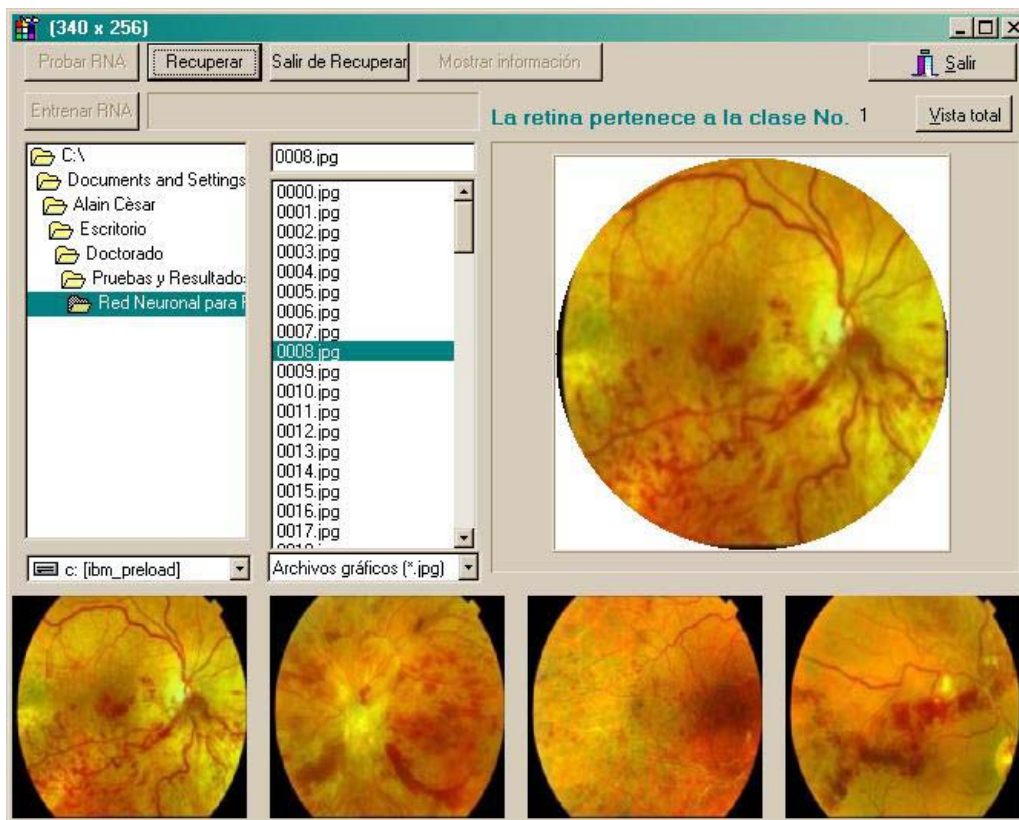


Fig. 51 Consulta de imágenes de retinas humanas con RN.

Tabla 16. Eficiencia del experimento 12: imágenes de retinas humanas con RN

Método	EFICIENCIA
Coeficientes de las imágenes completas con la RN	0.82
Coeficientes del histograma de una ventana circular con la RN	0.94
Coeficientes de histogramas de sub-imágenes con la RN	0.96

### VI.2.6 Experimento 13. 96 imágenes de 6 clases

Se aplica la metodología descrita en el capítulo IV, que consiste de dos etapas.

#### Primera etapa:

Conversión de las imágenes a patrones, para asignar una clase específica a cada imagen y obtener los vectores característicos de cada una de ellas, mediante los tres métodos descritos en el subepígrafe *IV.5 Codificación de las imágenes*.

#### Segunda etapa:

En la Figura 52 se muestra el resultado de una consulta, previo entrenamiento de la red neuronal.

Se realizó el 100% de las consultas, calculándose la eficiencia mediante el promedio de los resultados de todas ellas. La eficiencia final de la recuperación de las imágenes para los tres métodos evaluados, calculada sobre la base del número de imágenes que resultaron semejantes a la imagen-consulta, se muestra en la Tabla 17.

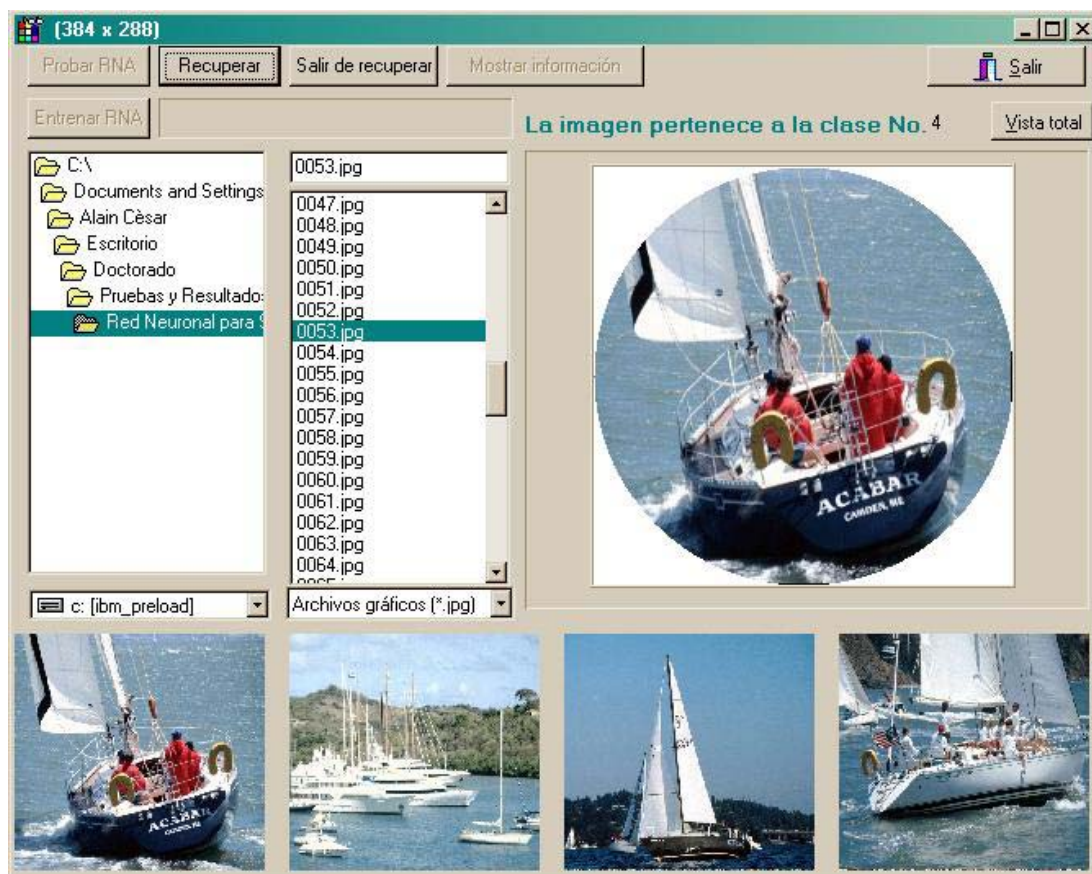


Fig. 52 Consulta de 96 imágenes de 6 clases con RN.

Tabla 17. Eficiencia del experimento 13: 96 imágenes de 6 clases con RN

Método	EFICIENCIA
Coeficientes de las imágenes completas con la RN	0.86
Coeficientes del histograma de una ventana circular con la RN	0.94
Coeficientes de histogramas de sub-imágenes con la RN	0.96

## VI.2.7 Experimento 14. 192 imágenes de 6 clases

Se aplica la metodología descrita en el capítulo IV, que consiste de dos etapas:

### Primera etapa:

Conversión de las imágenes a patrones, para asignar una clase específica a cada imagen y obtener los vectores característicos de todas las imágenes, mediante los tres métodos descritos en el subepígrafe *IV.5 Codificación de las imágenes*.

### Segunda etapa:

En la Figura 53 se muestra el resultado de una consulta, previo entrenamiento de la red neuronal.

Se realizó el 100% de las consultas, calculándose la eficiencia mediante el promedio de los resultados de todas ellas.

La eficiencia final de la recuperación de las imágenes para los tres métodos evaluados, calculada sobre la base del número de imágenes que resultaron semejantes a la imagen-consulta, se muestra en la Tabla 18.

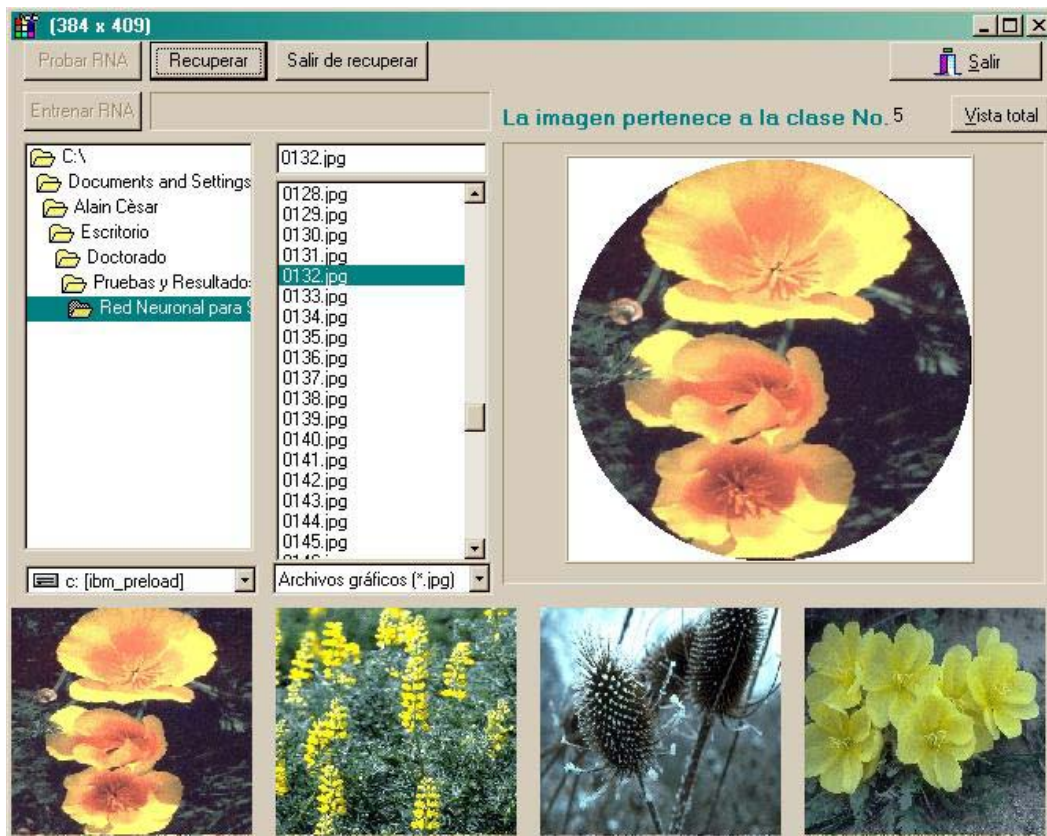


Fig. 53 Consulta de 192 imágenes de 6 clases con RN.



Tabla 18. Eficiencia del experimento 14: 192 imágenes de 6 clases con RN

Método	EFICIENCIA
Coeficientes de las imágenes completas con la RN	0.82
Coeficientes del histograma de una ventana circular con la RN	0.95
Coeficientes de histogramas de sub-imágenes con la RN	0.94

Una vez desarrolladas todas las pruebas descritas, se procede al análisis de los resultados, para posteriormente en el Capítulo 8 presentar las conclusiones alcanzadas con el desarrollo de la presente investigación.

### VI.3 Análisis de los resultados

- Es más alta la eficiencia con imágenes de estudio, que con imágenes complejas, hecho que comprueba una limitación inicial del proyecto, tal como se observa en las Tablas 5, 6 y 7. Con el mejor método que se denomina Coeficientes de histogramas de sub-imágenes se obtienen eficiencias de 0.96, 0.90 y 0.85 notándose la disminución de la eficiencia al crecer la complejidad de las imágenes.
- Tal como se observa en las Tablas 5, 6, 7, 8 y 9, el método denominado Coeficientes del histograma de una ventana circular, presenta una mejor eficiencia cuando se trata de imágenes con un gran contenido al centro de la imagen, por ejemplo objetos aislados como las retinas humanas e instrumental quirúrgico.
- Tal como se observa en las Tablas 5, 6, 7, 8 y 9, el método denominado Coeficientes del histograma de sub-imágenes, presenta una mejor eficiencia cuando se trata de imágenes complejas con un fondo heterogéneo, por ejemplo objetos en su medio ambiente como las imágenes de animales salvajes, aviones, automóviles y ciudades entre otras.
- Es más alta la eficiencia cuando se trabaja con imágenes en color que con las mismas imágenes en tonos de gris, tal como lo muestran las Tablas 13 y 14 de los experimentos 9 y 10 respectivamente.
- Se incrementa la eficiencia en la recuperación de imágenes basada en su contenido en términos de forma y color, en los métodos que utilizan el histograma como característica basada en el dominio del espacio mediante transformada Ondeleta cuando se combinan con la red neuronal, tal como se observa en las Tablas 7 y 14 que especifican eficiencias de la misma base de datos de imágenes de aviones, las Tablas 8 y 15 de la base de datos de imágenes de instrumental quirúrgico, las Tablas 9 y 16 de la base de datos de imágenes de retinas humanas, las Tablas 10 y 17 de la base de datos de 96 imágenes de 6 clases, y las Tablas 11 y 18 de la base de datos de 192 imágenes de 6 clases.

## CAPÍTULO VII

# SISTEMAS EN LÍNEA

A continuación se muestran tres sistemas en línea de recuperación de imágenes CIRES, FIRE e IRMA. La presentación de cada sistema incluye su descripción, algunas imágenes de su base de datos y ejemplos de consultas, el objetivo es comparar los resultados de los tres sistemas en línea y el desarrollado en este trabajo.

### VII.1 CIRES

CIRES [58] es un sistema en línea de recuperación de imágenes basado en su contenido, que combina principios de visión de alto nivel y de bajo nivel.

En el análisis de alto nivel utiliza organización perceptiva, inferencia y principios de agrupamiento para extraer información semántica que describa la estructura del contenido de una imagen.

En el análisis de bajo nivel emplea un modelo de canal de energía para describir la textura de la imagen y utiliza histogramas de color para mapear todos los colores en una paleta de colores fija. Un banco de Filtros de Gabor de pares simétricos es usado para extraer energía fraccional en varios canales espacio-frecuencia.

El sistema esta disponible para realizar consultas de imágenes que contienen objetos naturales como vegetación, árboles, cielo, etc., además de imágenes que contienen objetos hechos por el hombre tales como construcciones, torres, puentes, etc. La Base de datos que utiliza contiene imágenes de las 6 clases mencionadas en la Tabla 19 y en la Figura 54 se presentan algunas de estas imágenes para desarrollar consultas.

Tabla 19. Clases de imágenes de CIRES

Clase	No. de imágenes
Hecho por el hombre	1980
Aves	811
Insectos	1134
Mamíferos	2496
Flores	1161
Paisajes	2711
Total de imágenes	10,221

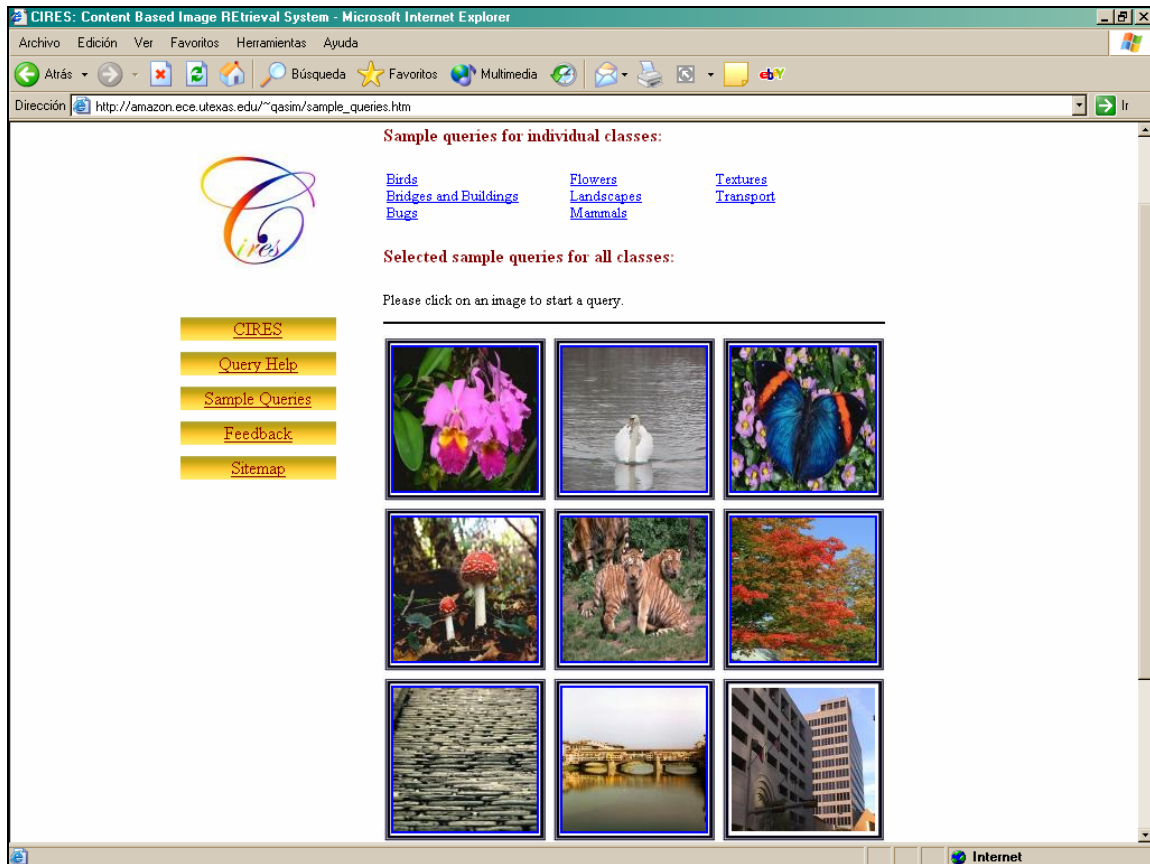


Fig. 54 Algunas imágenes de CIRES.

La Figura 55 presenta una consulta de una imagen que contiene una construcción o edificio, muestra las primeras 16 imágenes similares recuperadas por varios métodos, entre ellos:

- 1.- Estructura
- 2.- Color
- 3.- Textura
- 4.- Color y textura
- 5.- Estructura, color y textura.

Se observa en la Figura 55, que con la combinación de estructura, color y textura, se obtienen los mejores resultados. En la Figura 56 se observan 4 consultas y sus correspondientes resultados que muestran las primeras 16 imágenes recuperadas más similares, con la combinación de estructura, color y textura.

En CIRES 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.

En la Figura 57 se observa que el inciso (a) presenta mejor resultado, hasta la imagen número 9 en que falla la consulta; en el inciso (b) el sistema falla en la imagen 7.

La eficiencia de la combinación de color y textura con información de la estructura de la imagen queda demostrada en este sistema al observar la gráfica de la Figura 55, con el 77.4% de precisión en la recuperación como mejor valor en imágenes con objetos hechos por el hombre.

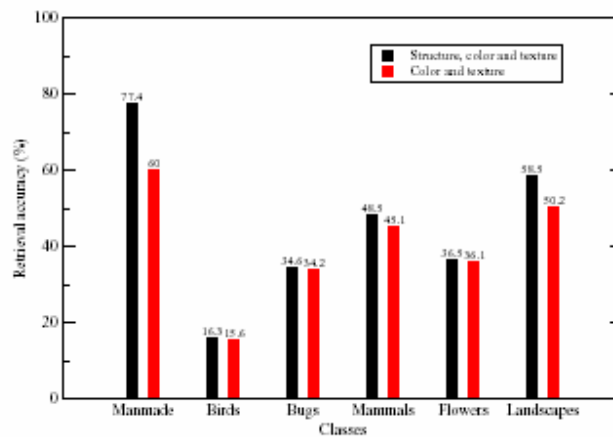
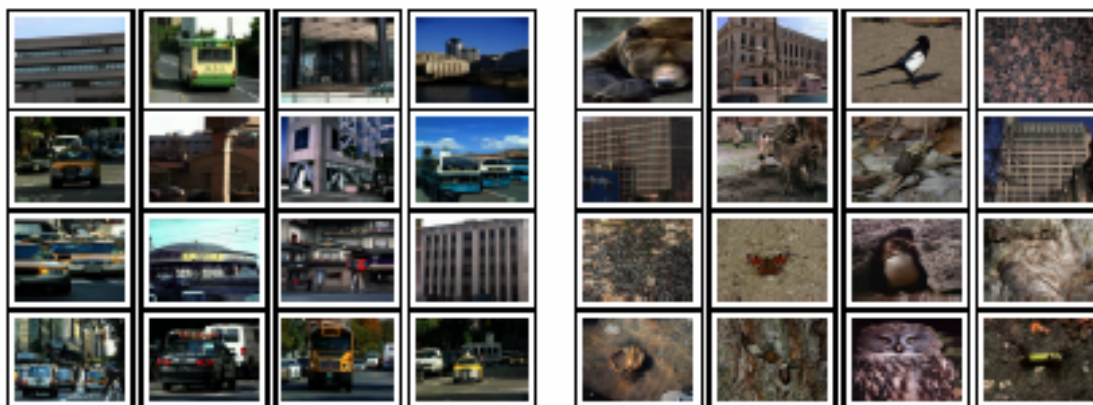


Fig. 55 Comparación de la precisión de la recuperación.



Query image: Building



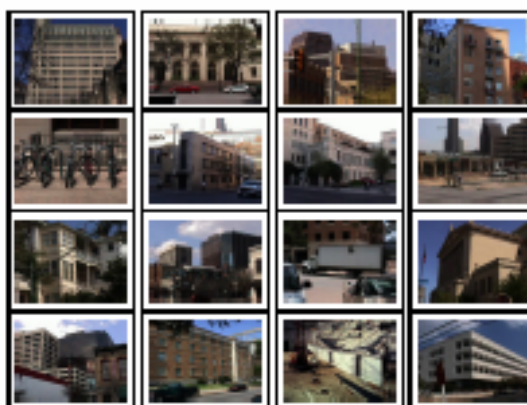
(a) Structure

(b) Color



(c) Texture

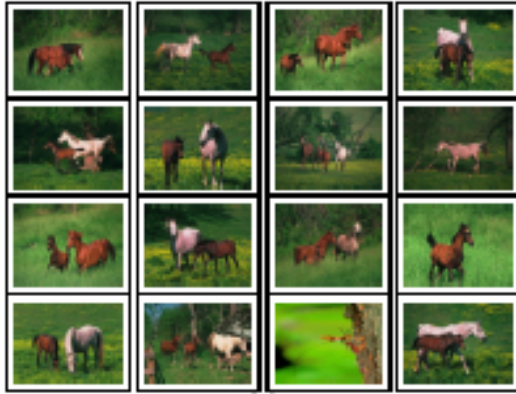
(d) Color + Texture (Equal weights)



(e) Structure + Color + Texture (Equal weights)

Fig. 56 Cinco consultas en CIRES.

  
 Query image: Horses  
 Weights:  $S = 0.33$ ,  $C = 0.33$ ,  $T = 0.33$ , L, A, B channels.



  
 Query image: Colorful vegetation  
 Weights:  $S = 0.33$ ,  $C = 0.33$ ,  $T = 0.33$ , L, A, B channels.



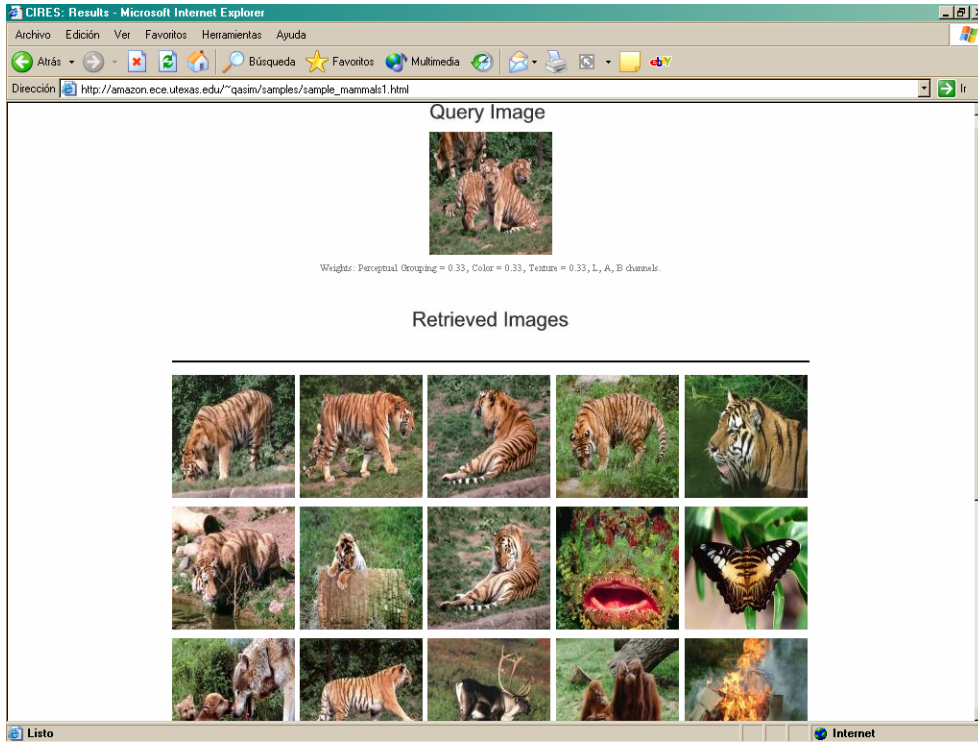
  
 Query image: Yellow cab  
 Weights:  $S = 0.33$ ,  $C = 0.33$ ,  $T = 0.33$ , L, A, B channels.



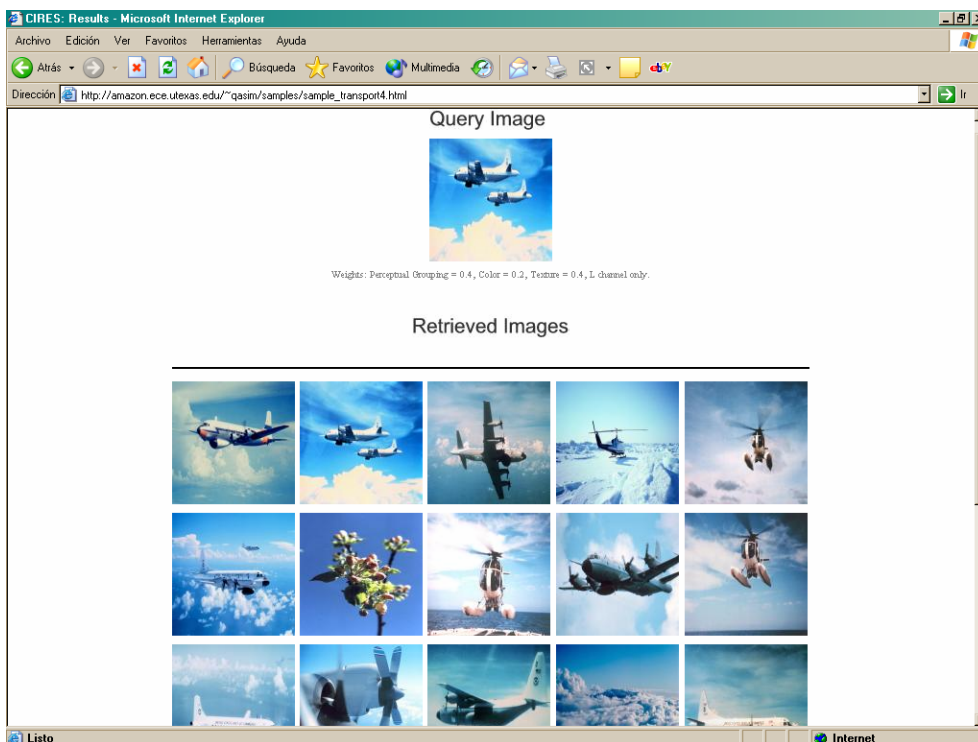
  
 Query image: Bridge  
 Weights:  $S = 0.4$ ,  $C = 0.3$ ,  $T = 0.3$ , L, A, B channels.



Fig. 57 Cuatro consultas en CIRES.



(a) Consulta de Mamíferos



(b) Consulta de Hecho por el hombre

Fig. 58 Dos consultas en CIRES a clases específicas.

## VII.2 FIRE

FIRE [59] es un sistema en línea de recuperación basado en su contenido, que tiene 7 características visuales disponibles para representar a una imagen y diversas medidas de expresar la distancia, una combinación ponderada de tales características admite consultas flexibles, la Tabla 20 presenta una lista de características visuales y sus medidas de distancia asociadas, una de las bases de datos de imágenes que se utilizó fue la de fotografías históricas de San Andrés y en la Figura 59 se presentan imágenes listas para realizar consultas en FIRE.

Tabla 20. Características visuales de FIRE

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

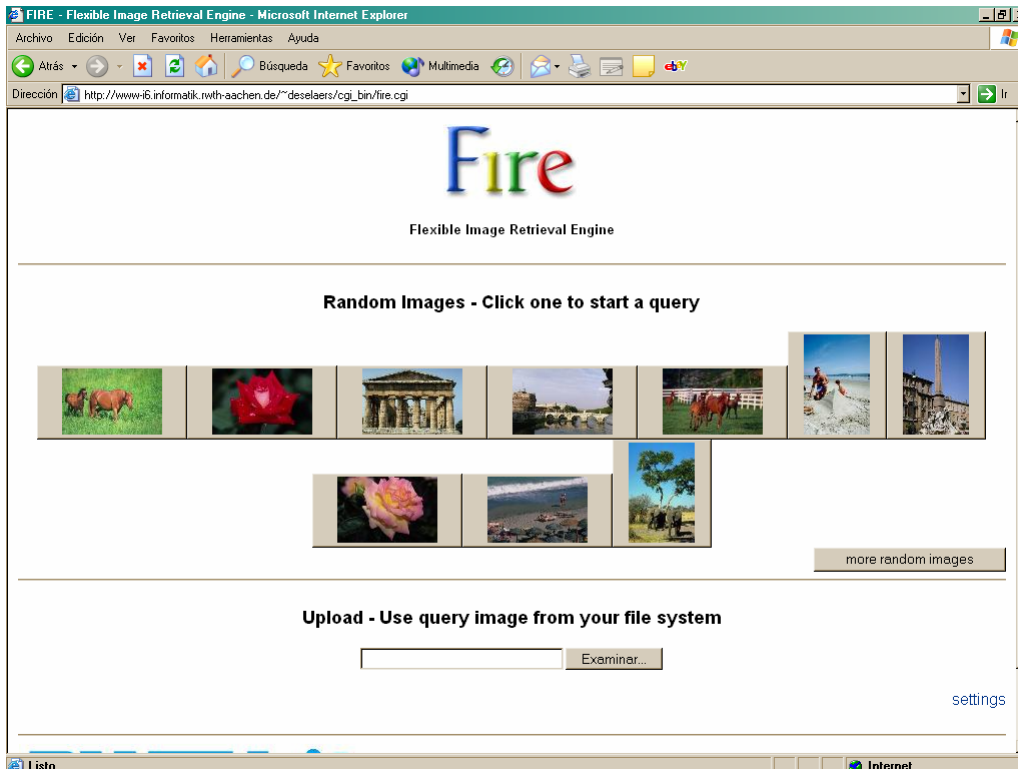
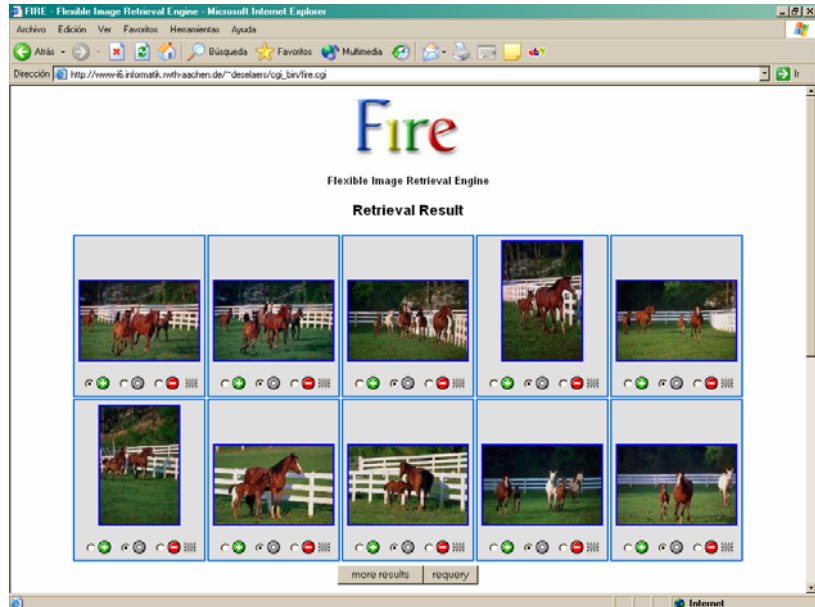


Fig. 59 Imágenes disponibles para consultas en FIRE.





(a)



(b)

Fig. 60 Resultados de consultas en FIRE.

La Figura 60 presenta dos consultas del sistema FIRE. Se observa en el inciso (b) que las características visuales no son suficientes para obtener buenos resultados. En CLEF 2004 se demuestra que FIRE obtiene mejores resultados en las consultas completamente automáticas cuando solo se utilizan características visuales; se obtiene una eficiencia de 0.394 al utilizar solo características visuales y 0.587 al combinar información en texto.

### VII.3 IRMA

IRMA **;** **Error! No se encuentra el origen de la referencia.** es un sistema en línea de recuperación de imágenes de radiografías. Primero calcula la probabilidad de que la imagen pertenezca a cada clase de imágenes mediante una versión extendida de la distancia tangente de Simard con un clasificador basado en la densidad del núcleo. Se propone un nuevo modelo de distorsión para radiografías que combinado con la distancia tangente resulta el cálculo de la distancia tangente distorsionada, obteniendo así los mejores resultados.

El objetivo es recuperar imágenes de las etapas de terapia del mismo paciente o también recuperar imágenes con diagnóstico similar en bases de datos de imágenes muy grandes; métodos de reconocimiento de patrones y análisis estructurado son utilizados para describir el contenido de una imagen en una firma característica.

Además de las consultas a bases de datos de imágenes, la clasificación automática y el indexado permiten simples inserciones de radiografías convencionales dentro del sistema, completando la información con fecha y nombre para el formato DICOM.

La base de datos de imágenes consiste de 110 de abdomen, 706 de manos, 103 de senos, 110 de cráneos, 410 de torsos y 178 de columnas vertebrales, sumando un total de 1617 imágenes de radiografías. La Figura 61 muestra unas imágenes de la base de datos con imágenes de radiografías escaladas a un tamaño común; se presentan imágenes de abdomen, mano, seno, cráneo, torso y columna vertebral.

La Tabla 21 presenta la eficiencia de cuatro medidas de distancia en dos métodos de clasificación. Se observa que el mejor resultado se obtiene con la distancia tangente distorsionada y la densidad de núcleo.

Tabla 21. Eficiencia del sistema IRMA

<b>Medida de la distancia</b>	<b>1-NN</b>	<b>Densidad núcleo</b>
Distancia Euclidiana	81.9	83.8
Distancia tangente	84.7	85.2
Modelo distorsionado	83.5	85.3
Distancia tangente distorsionada	85.8	87.5



Fig. 61 Ejemplos de radiografías.

La Figura 62 presenta la combinación de dos sistemas de recuperación de imágenes SPIRS e IRMA, que permiten la recuperación de imágenes en base a su contenido guiadas por formas locales de huesos individuales.

SPIRS es un sistema de recuperación de imágenes que cuenta con 17,000 radiografías de columnas vertebrales, diseñado por los Laboratorios Nacionales de Medicina del Instituto Nacional de la Salud de Estados Unidos.

En la Figura 62(a) se presentan 12 imágenes disponibles para ejecutar una consulta dentro de la base de las 17,000 imágenes. Se puede seleccionar la imagen para realizar una recuperación de forma visual, esto es, recorrer la base de datos de imágenes de 12 en 12, o mediante la inserción del número de identificación, caso en que el sistema nos muestra el conjunto de 12 imágenes donde se encuentra dicho número.

En la Figura 62(b) muestra el resultado de la recuperación: 12 imágenes de un total de 100 imágenes recuperadas, en un tiempo de 1.88 segundos; se observa una excelente eficiencia de recuperación.

La Figura 63(a) presenta imágenes para una consulta refinada extendida por medio de IRMA en una base de datos de 10,000 imágenes, donde cada imagen es representada por 1024 características globales.

La Figura 63(b) muestra el resultado de la recuperación en un tiempo de 1.84 segundos; se observa que en la recuperación, de las 12 imágenes 4 no corresponden con la imagen consulta, disminuyendo su eficiencia de recuperación.

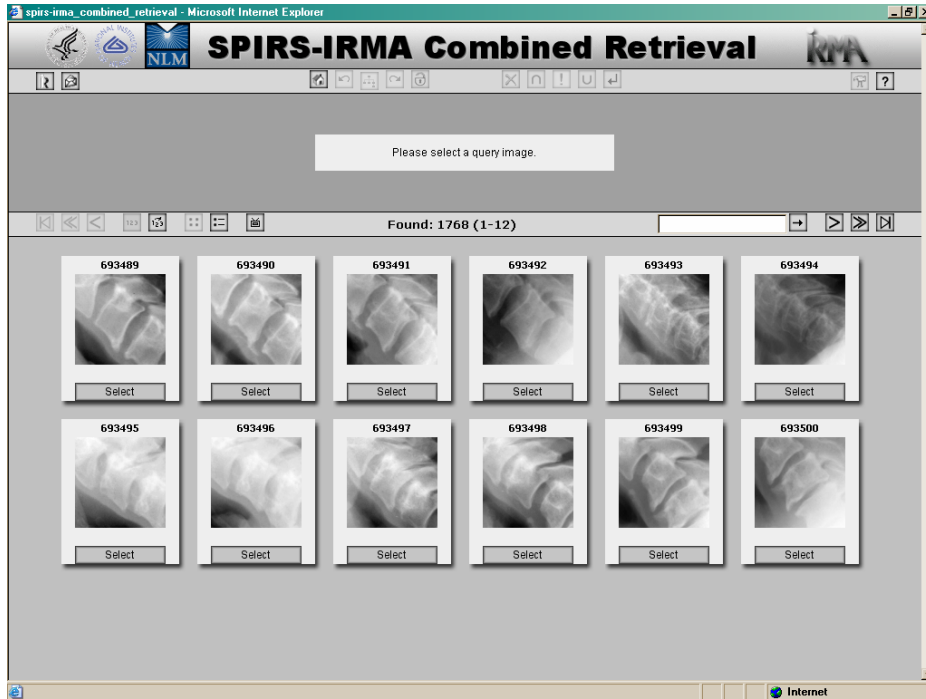


Fig. 62 (a) Recuperación de 12 imágenes

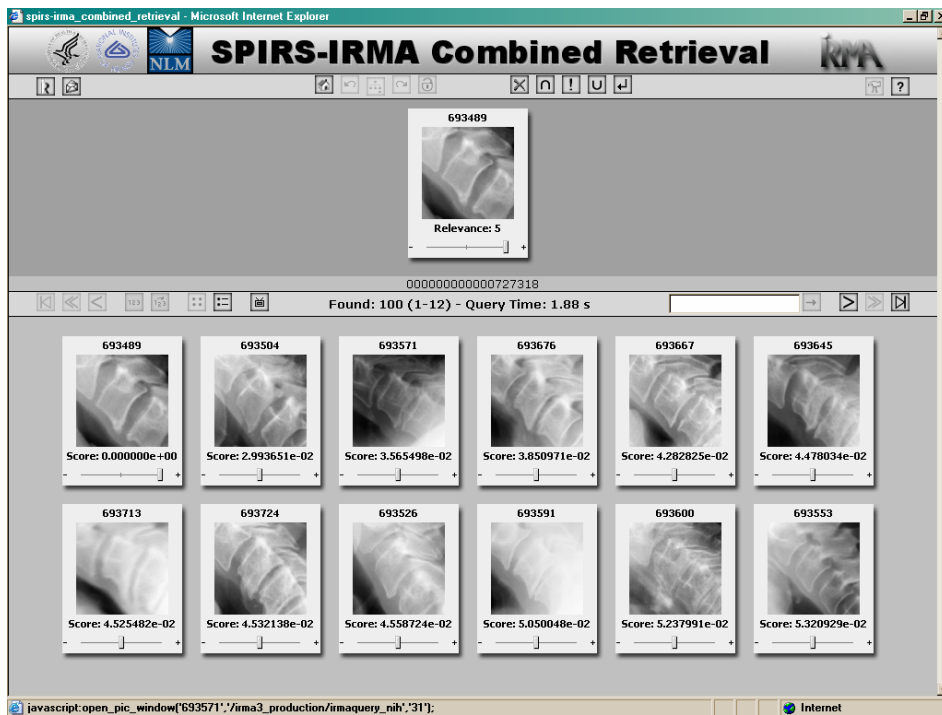


Fig. 62 (b) Tiempo para recuperar 12 imágenes de 100 imágenes

Fig. 62 Consulta en IRMA.



Fig. 63 (a) Imágenes a consultar.

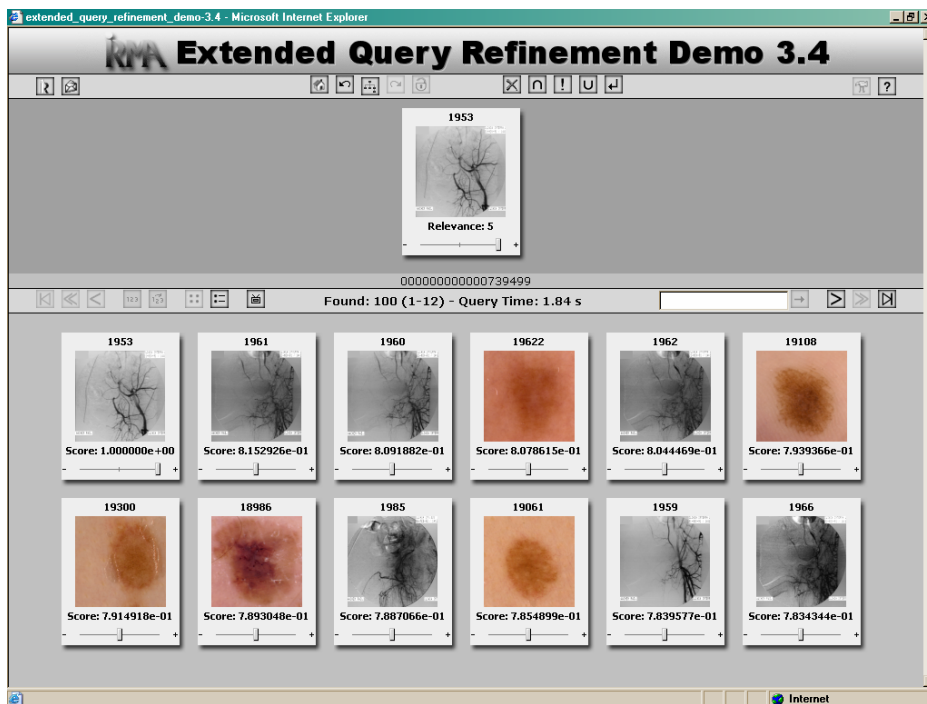


Fig. 63 (b) Imágenes resultantes de una consulta.

Fig. 63 Otra consulta en IRMA.

## VII.4 Análisis de los resultados

A continuación en la Tabla 22 se presenta un estudio comparativo de los sistemas en línea anteriores con el propuesto en este trabajo. Se observa la diversidad de metodologías utilizadas que tratan de reconocer la forma, el color y la textura. Las bases de datos utilizadas son de unidades de miles y contienen imágenes complejas; obviamente, estos no son los únicos trabajos desarrollados, pero sí son los que más se relacionan con la metodología propuesta en esta tesis.

Tabla 22. Comparación con sistemas en línea

Sistema en línea	Metodología	Número de imágenes	Eficiencia en %	Años de desarrollo	No. de publicaciones
CIRES [58] Content based Image Retrieval System	Organización perceptiva, inferencia y principios de agrupamiento, modelo de canal de energía, histogramas de color y un banco de Filtros de Gabor pares-simétricos.	10,221	77.4	1997 - 2005	14
FIRE [59] Flexible Image Retrieval Engine	Imagen escalada a 32 x 32, Imagen escalada a 32 x X, Descriptor global de textura, Histograma de textura Tamura, Histograma con núcleo monomial, Histograma con núcleo relacional y Color o tonos de gris.	10,000	58.7	2002 - 2006	7
IRMA <b>¡Error! No se encuentra el origen de la referencia.</b> Image Retrieval in Medical Applications	Recuperación de imágenes basada en la apariencia, cada píxel de la imagen es una característica. Se utiliza una imagen escalada a 32 x 32.	1,617	87.5	2002 - 2006	7
TESIS [52] Recuperación de imágenes en base a su contenido	Ondeletas de histogramas de una ventana circular y de sub- imágenes y clasificación mediante redes neuronales	1,068	96	2003 - 2006	5

## CAPÍTULO VIII

# CONCLUSIONES Y TRABAJOS FUTUROS

### VIII.1 Conclusiones

- Al comparar los resultados de la metodología desarrollada en esta tesis, con los presentados en la Tabla 1 *Panorámica de trabajos relacionados*, se observa que las diferentes metodologías incluyendo la nuestra, se basan en conceptos parecidos que consisten en extraer de las imágenes características de bajo nivel como la forma, el color y la textura, para así obtener una firma compacta de las imágenes, indexarlas y posteriormente recuperarlas. Además de un incremento en el número de imágenes y su complejidad, acompañados de una mejor eficiencia de recuperación en el desarrollo histórico, se puede observar que en la actualidad se trabaja con cientos de imágenes complejas con una eficiencia superior al 95%, hecho que cumplimos cabalmente.
- Al comparar el trabajo desarrollado en esta tesis, con los sistemas en línea de la Tabla 22 presentados en el capítulo 7, se observa que al incrementar el número de imágenes en unidades de miles, la eficiencia disminuye del 90% al 60% aproximadamente, aunque las metodologías utilizadas sean muy complejas y aparentemente robustas. Con respecto al número de años invertidos en el desarrollo de los sistemas, éste ha disminuido.
- De las dos conclusiones anteriores se observa que el problema de la recuperación de imágenes sigue siendo un problema abierto y que una de las alternativas actuales es incrementar la complejidad de las metodologías utilizadas para pasar de utilizar características de bajo nivel, tales como el color, la forma y la textura, a una combinación de estas con características de semántica de la imagen que describen su contexto y los objetos que contiene.

- El sistema desarrollado, aparte de extraer las características de las imágenes que conforman la información visual mediante ondeletas y de recuperar la información en forma de imágenes mediante el uso de redes neuronales, integra ideas propias para extraer los rasgos, tales como: el empleo de características globales a partir de los histogramas de los tres planos de color de la imagen completa, además de utilizar características de posición mediante histogramas de color de ventanas circulares, y mediante características locales expresadas en términos de histogramas de color de sub-imágenes cuadradas [51] [52]. El uso de estas técnicas proporciona un sistema robusto para la recuperación de imágenes sobre la base de su contenido.
- El objetivo general planteado se alcanzó en un 100%, ya que se logró una metodología mediante la transformada de ondeletas que posibilita la recuperación de imágenes en grandes bases de datos de imágenes, sobre la base de su contenido en términos de forma y color, además de lograr una clasificación mediante redes neuronales, según se observa en el Capítulo VI, *Pruebas y Resultados*.
- Con respecto a la Hipótesis planteada, se utilizó con éxito la indexación de las imágenes mediante ondeletas con características visuales de bajo nivel, como son la forma en una representación multi-resolución y el color como información que brinda el modelo de color RGB. Al aplicar un clasificador supervisado como lo es la red neuronal, se obtiene una estructura indexada del espacio de imágenes.
- Por los espacios de características creados, las imágenes pueden ser indexadas con el fin de ejecutar consultas por imágenes, a través de un directorio de imágenes con su espacio previamente estructurado y los tipos de imágenes del directorio previamente establecidos, por ejemplo: imágenes de automóviles, edificios, paisajes de bosques, montañas, aviones, carreteras, etc.
- Al utilizar una red neuronal para clasificar las imágenes en clases, y posteriormente aplicar la regla del vecino más cercano para encontrar las imágenes que más se parecen a la imagen de consulta sobre la base de la distancia Euclidiana, se logró que solo se presentaran imágenes de la clase que determinó la red neuronal, con lo cual se comprueba su alta eficiencia.

## VII.2 Trabajos futuros

- Evitar las variaciones obtenidas con los cambios en el tamaño del escalamiento de la imagen mediante el uso de la desviación estándar de los coeficientes de ondeletas.
- Evitar las variaciones obtenidas con los cambios de la rotación de la imagen mediante histogramas normalizados de ventanas circulares.



- Evitar las variaciones obtenidas con los cambios de la traslación de la imagen mediante algoritmos de emparejamiento de cadenas.
- Evitar las variaciones obtenidas con los cambios de la iluminación, el brillo y el contraste, mediante la normalización del histograma y el uso del modelo de color HSI.
- Implementación de algoritmos discretos de esquemas de Ondeletas Complejas, de Lifting, Adaptativas y Difusas, para mejorar la representación de las imágenes mediante los coeficientes de ondeletas.
- Implementación de los algoritmos de redes neuronales para la clasificación no supervisada.
- Quedan pendientes para futuros trabajos las consultas a través de texto, dibujos de bosquejos e iconos.

La lista anterior es solo un resumen de algunos trabajos específicos pendientes de desarrollar, sin olvidar las diferentes líneas de investigación de recuperación de información visual presentadas en la Sección *Panorámica de trabajos relacionados*.

## Referencias

- [1] Alberto Del Bimbo, *Visual Information Retrieval*, Morgan Kaufmann Publishers, 1999.
- [2] Clement Leung, *Visual Information Systems*, Springer, 1997.
- [3] Michael S. Lew, "Next -Generation Internet Searches for Visual Content", *Computer Information Retrieval*, Volume 33, Number 11, 46-53, Computer Society IEEE, 2000.
- [4] Alberto Del Bimbo, "A Perspective View on Visual Information Retrieval Systems", *Content-Based Access of Image and Video Libraries*, Workshop on Volume 21, 108 – 109, IEEE 1998.
- [5] Nisha Talagala, Satoshi Asami, David Patterson, Bob Futernick y Dakin Hart, "The Art of Massive Storage", *IEEE Computer Information Retrieval*, Volume 33, Number 11, 22-28, Computer Society, 2000.
- [6] Ing-Sheen Hsieh y Kuo-Chin Fan, "Color Image Retrieval Using Shape and Spatial Properties", *IEEE 15th International Conference on Pattern Recognition* Volume 1, 1023 – 1026, 2000.
- [7] Liapis, S.; Tziritas, G., "Color and texture image retrieval using chromaticity histograms and wavelet frames", *IEEE Transactions on Multimedia* Volume 6, Issue 5, 676 – 686, 2004.
- [8] Rafael C. Gonzalez, Richard E. Woods y Steven L. Eddins, *Digital Image Processing Using Matlab*, Pearson Prentice Hall, 2004.
- [9] Wu Yan, Ding Mingyue y Peng Jiaxiong, "Research on Using Moment Invariants in Scene Matching", *GMICV* 1995.
- [10] Meirav Adoram y Michael S. Lew, "IRUS: Image Retrieval Using Shape". *IEEE International Conference on Multimedia Computing and Systems* 2, 597-602, 1999.
- [11] Anil K. Jain y Aditya Vailaya, "Image Retrieval Using Color and Shape", *Pattern Recognition*, Vol. 29, No. 8, 1233-1244, Pergamon Press, Elsevier, 1996.
- [12] Fan-Di Jou, Kui-Chin Fan, Yang-Lang Chang, "Efficient matching of large-size histograms", *Pattern Recognition Letters* 25, Elsevier B. V, 277-286, 2004.
- [13] Jan Puzicha, Thomas Hofmann y Joachim M. Buhmann, "Histogram Clustering for Unsupervised Segmentation an Image Retrieval", *Pattern Recognition Letters*, Vol. 20, 899-909, Elsevier 1999.
- [14] M. K. Mandal y T. Aboulnasr, "Fast Wavelets Histogram Techniques for Image Indexing", *Computer Vision and Understanding* Vol. 75, Nos.1/2, 99-110, 1999.

- [15] Philip L. Worthington y Edwin R. Hancock, "Object Recognition Using Shape from Shading", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 5, 2001.
- [16] Zhou Wang and Alan Bovik, "Embedded Foveation Image Coding", IEEE Transactions on Image Processing, Vol. 10, No. 10, 2001.
- [17] Manish N. Sarlashkar, M. Bodruzzaman y M. J. Malkani, "Feature Extraction using Wavelet Transform for Neural Network based Image Classification", IEEE 0-7803-4547-9, 1998.
- [18] Traina, A.J.M.; Castanon, C.A.B.; Traina, C., Jr., "MultiWaveMed: a system for medical image retrieval through wavelets transformations", Computer-Based Medical Systems, 2003. Proceedings. 16th IEEE Symposium, 150 – 155, 2003.
- [19] Yuk Ying Cheng y Xiaoming Chen, "A study of clustering algorithm for wavelet-based image retrieval system", Consumer Communications and Networking Conference, 3rd CCNC, IEEE Vol. 2, 1322 – 1323, 2006.
- [20] Soo Beom Park, Jae Won Lee, Sang Kyon Kim, "Content-based image classification using a neural network", Pattern Recognition Letters, Elsevier B. Vol. 25, 287-300, 2004.
- [21] Jain, P.; Merchant, S.N., "Wavelet based multiresolution histogram for fast image retrieval", TENCON Conference on Convergent Technologies for Asia-Pacific Region Vol. 2, 581 – 585, 2003.
- [22] De Bianchi, M.F., Guido, R.C., Nogueira, A.L. and Padovan, P., "A Wavelet-PCA Approach for Content-Based Image Retrieval", System Theory, Proceeding of the Thrity-Eighth Southeastern Symposium Vol. 38, 425-428, 2006.
- [23] C. Sidney Burrus, Ramesh A. Gopinath y Hatitao Guo, *Introduction to Wavelets and Wavelets Transforms*, Prentice Hall, 1998.
- [24] B. W. Silverman y J. C. Vassilicos, *Wavelets*, Oxford University Press, 2000.
- [25] Stephane Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, Second edition 1999.
- [26] Michel Misiti, Yves Misiti, Georges Oppenheim y Jean-Michle Poggi, *Matlab Toolbox*, Matlab User's Guide, 1995.
- [27] James Z. Wang, *Integrated Region Based Image Retrieval*, Kluwer Academic Publishers, 2001.
- [28] L. Prasad y S. S. Iyengar, *Wavelets Analysis with Applications to Image Processing*, CRC Press, 1997.
- [29] Zixiang Xiong, Kannan Ramchandran y Michael T. Orchard, "Space-Frequency Quantization for Wavelet Image Coding", IEEE Transactions on Image Processing, Vol. 6, No. 5, 1997.

- [30] Andrew Internetb, *Statistical Pattern Recognition*, John Wiley and Sons, Second Edition 2002.
- [31] Byoung Chul Ko, Haesung Lee y Hyeran Byun, "Flexible Sub Blocks for Visual Information Retrieval", *Electronics Letters Online*, Vol. 36, No. 1, 254-25, IEE 2000.
- [32] Po-Cheng Wu, Liang-Gee Chen y Yeong-Kang Lai, "A Block Shifting Method for Reduction of Blocking Effects in Subband / Wavelets Image Coding", *IEEE Transactions on Consumer Electronics*, Vol. 44, No. 1, 1998.
- [33] Jia Wang, Wenjun Zhang y Songyu Yu, "Wavelet Coding Method Using Small Block DCT", *Electronics Letter Online* Vol. 37, No. 10, 2001.
- [34] Luisa Micó, Jose Oncina y Rafael C. Carrasco, "A Fast Branch & Bound Nearest Neighbor Classifier in Metric Spaces", *Pattern Recognition Letters* 17, 731-739, Elsevier 1996.
- [35] Simon Warfield, "Fast k-NN Classification for Multichannel Image Data", *Pattern Recognition Letters* 17, 713-721, Elsevier 1996.
- [36] Richard O. Duda, Peter E. Hart y David G. Stork, *Pattern Classification*, John Wiley & Sons, Second Edition 2001.
- [37] Louis Brasard, "Collection Space Distances", *Systems, Man and Cybernetics, Intelligent Systems for the 21st Century. IEEE International Conference on Volume 2, 1544 – 1549, 1995.*
- [38] Yoram Biberman, "A Context Similarity Measure", *Proceedings of Constraint Databases and Applications*, Springer 1995.
- [39] Bilge Günsel y A. Murat Tekalp, "Similarity Analysis for Shape Retrieval by Example", *IEEE 13th International Conference on Pattern Recognition Volume 2, 1996.*
- [40] Everest Mathias y Aura Conci, "Comparing the Influence of Color Spaces and Metrics in Content-Based Image Retrieval", *IEEE International Symposium on Computer Graphics, Image Processing, and Vision*, 371 – 378, 1998.
- [41] B. D. Ripley, *Pattern Recognition and Neural Networks*, Cambridge University Press, 1996.
- [42] Carl G. Looney, *Pattern Recognition using Neural Networks*, Oxford University Press, 1997.
- [43] Gasca A. Eduardo y Barandela A. Ricardo, "Algoritmos de aprendizaje y técnicas estadísticas para el entrenamiento del PM", *Proceedings del IV Simposio Iberoamericano de RP, La Habana, Cuba, 456-464, 1999.*
- [44] Peter McGuire y G. M. T. D`Eleuterio, "Eigenpixels and a Neural – Network Approach to Image Classification", *IEEE Transactions on Neural Networks*, Vol. 12, No. 3, 2001.

- [45] Nikos Papamarcos, Antonis E. Atsalakis y Charalampos P. Strouthopoulos, "Adaptive Color Reduction", IEEE Transactions on Systems, Man and Cybernetics – Parte B: Cybernetics, Vol. 32, No. 1, 2002.
- [46] Alain C. González G., Juan M. Hernández V., Alberto R. Hernández M. y Daniel González S., "Análisis de Señales del Subsuelo utilizando Wavelets", 4ª Conferencia Internacional en Control, Instrumentación Virtual y Sistemas Digitales 2002, CIC - IPN, IMP, CITIS - UAH, Pachuca México, 10-16, 2002.
- [47] A. C. Gonzalez G., A. R. Hernandez M., J. Moreno J., D. Gonzalez S., J. M. Hernandez V., "Harmonic Analysis of a Modulate Amplitude Signal using the Wavelet Transforms", 22<sup>nd</sup> IASTED International Conference Modeling, Identification And Control, MIC'2003, Innsbruck Austria., 304-308, 2003.
- [48] A. C. Gonzalez G., D. Gonzalez S., J. Moreno J., J. M. Hernandez V., A. R. Hernandez M., "Digital Processing of U Signal", 22<sup>nd</sup> IASTED International Conference Modeling, Identification And Control, MIC'2003, Innsbruck Austria. 309-312, 2003.
- [49] A. C. González G., R. Ibáñez C., J. H. Sossa A., J.A. Martínez N., "Reconocedor Óptico de Caracteres para Código Braille Estándar de 6 Puntos utilizando Redes Neuronales de Retropropagación", XXI Congreso Internacional de Ingeniería Electrónica Electro 2003, Instituto Tecnológico de Chihuahua, México., 79-84, 2003.
- [50] A. C. González G., J. H. Sossa A. y E. M. Felipe R., "Images retrieval using Wavelets, Histograms and Subimages", 5ª Conferencia Internacional en Control, Instrumentación Virtual y Sistemas Digitales 2004, CIC - IPN, México., 403-411, 2004.
- [51] A. González, H. Sossa, E. Felipe and O. Pogrebnyak., "Wavelet transforms and neural networks applied to image retrieval". International Conference on Pattern Recognition (ICPR). Hong Kong, China, 909-912, 2006.
- [52] A. González, H. Sossa, E. Felipe and O. Pogrebnyak., "Histograms, Wavelets and Neural Networks Applied to Image Retrieval". 5<sup>th</sup> Mexican International Conference on Artificial Intelligence, Apizaco, Mexico., Lecture Notes in Artificial Intelligence, Springer 820-827, 2006.
- [53] Zhiyong Zeng and Lihua Zhou., "A Novel Image Retrieval Algorithm Using Wavelet Packet Histogram Techniques", Systems and Control in Aerospace and Astronautics, 1st International Symposium, 1194-1197, 2006.
- [54] Utenpattanant, A.; Chitsobhuk, O. and Khawne, A., "Color Descriptor for Image Retrieval in Wavelet Domain", Advanced Communication Technology, The 8th International Conference Vol. 1, 818-821, 2006.
- [55] F. W. Mattern and J. Denzler, "Comparison of Appearance-Based Methods for Generic Object Recognition", Pattern Recognition and Image Analysis, Vol. 14, No. 2, 255–261, 2004.

[56] <http://www.cs.columbia.edu/CAVE/research/softlib/coil-100.html>

[57] <http://www.aircraft-images.co.uk>

[58] Qasim Iqbal and J. K. Aggarwal, "CIRES: A system for content-based retrieval in digital image libraries", Seventh International Conference on Control, Automation Robotics and Vision (ICARCV'02), Singapore, 205-210, 2002.

[59] Thomas Deselaers, Daniel Keysers, and Hermann Ney, "FIRE – Flexible Image Retrieval Engine: ImageCLEF 2004 Evaluation", C. Peters et al. (Eds.): CLEF 2004, LNCS 3491, Springer-Verlag Berlin Heidelberg, 688–698, 2005.

[60] Lehmann TM, Güld MO, Thies C, Fischer B, Spitzer K, Keysers D, Ney H, Kohlen M, Schubert H, Wein BB, "Content-based image retrieval in medical applications", *Methods of Information in Medicine*, 43(4), 354-361, 2004.

## Apéndice A. Código del algoritmo de interpolación bilineal

El código siguiente muestra el algoritmo para lograr la interpolación bilineal en una imagen digital. En él se realiza la interpolación de la imagen para lograr un tamaño de 256 x 256 píxeles.

```
Int Bilineal(BYTE Imagen[][256], float i, float j)
{
    int x1, y1, x2, y2;
    float Gray;
    x1 = (int) floor(i);
    x2 = x1+1;
    y1 = (int) floor(j);
    y2 = y1+1;
    Gray = (y2-j)*(x2-i)*Imagen[x1][y1] + (y2-j)*(i-x1)*Imagen[x2][y1]
    + (j-y1)*(x2-i)*Imagen[x1][y2] + (j-y1)*(i-x1)*Imagen[x2][y2];
    return int (Gray);
}
void Interpolacion (void)
{
    // calcula la imagen de salida
    for (y=0; y<=255; y++)
        for (x=0; x<=255; x++)
            {
                i = (x+(a-1)*x0) / a; // calcula la posición original
                j = (y+(a-1)*y0) / a;
                Salida[x][y] = Bilineal(Imagen, i, j);
            }
}
```

## Apéndice B. Código del algoritmo de la red neuronal

A continuación se procede a la implementación de la red neuronal, que recibe como argumentos sus dimensiones 48 neuronas en la capa de entrada, 49 en la capa oculta y 6 en la de salida nnwork **cerebro** (48, 49, 6);

```
// se crea la red neuronal.
nnwork::nnwork (int input, int hidden, int output)
{
    input_size = input;
    hidden_size = hidden;
    output_size = output;
    srand (time (0));
    hidden_nodes = new nlayer (hidden_size, input_size);
    assert (hidden_nodes);
    output_nodes = new nlayer (output_size, hidden_size);
    assert (output_nodes);
}
```

El estado de activación se representa mediante un conjunto de estados de activación, con un valor entre [0,1], o en el intervalo [-1,1], auxiliándose de una función sigmoidal. La función de salida o de transferencia que se utiliza es la función *sigmoidal* representada por la Ecuación 38.

A continuación se presenta el código de la función, indicándose en **negrita** los lugares donde se llama a dicha función

```
// Función de transferencia de los nodos.
float sigmoid (float data)
{
    return (1. / (1. + exp (-data)));
}

for (j = 0; j < hidden_size; j++) {
    sum = 0;
    // Cálculo del valor de salida
    for (i = 0; i < input_size; i++)
        sum += hidden_nodes -> nodes [j].weights [i] * data [i];
    hidden_nodes -> nodes [j].output = sigmoid (sum);
}

for (k = 0; k < output_size; k++) {
    sum = 0;
    // Cálculo del valor de salida de la función de transferencia
    for (j = 0; j < hidden_size; j++)
        sum += output_nodes -> nodes [k].weights [j] * hidden_nodes ->
```



```

    nodes[j].output;
    result [k] = sigmoid (sum);
}
}

```

Las conexiones que unen a las neuronas que forman una RNA tienen asociadas un peso; se presenta el código de la matriz  $W$  con todos los pesos  $w_{ji}$ .

```

// Actualización de los pesos en la capa de salida.
for (k = 0; k < output_size; k++)
for (j = 0; j < hidden_size; j++)
    output_nodes -> nodes [k].weights [j] += eta *
output_weight_delta [k] * hidden_nodes -> nodes [j].output;
// Actualización de los pesos en la capa oculta.
for (j = 0; j < hidden_size; j++)
for (i = 0; i < input_size; i++)
    hidden_nodes -> nodes [j].weights [i] += eta *
hidden_weight_delta [j] * data [i];

```

La función o regla de activación que se utiliza es la *sigmoidal*, La cual tiene básicamente el mismo código que la función de salida o transferencia.

La regla de aprendizaje que se encuentra inmersa en los pesos de las conexiones entre neuronas, implica cierto número de cambios en estas conexiones. En realidad, puede decirse que se aprende modificando los valores de los pesos de la red para minimizar el error.

A continuación se presenta el código.

```

// Error cuadrático medio
while (1) {
    run (data, output);
//primeramente se procesa la capa de salida en términos del error
for (k = 0, MSE = 0; k < output_size; k++) {
    output_weight_delta [k] = desired [k] - output [k];
    MSE += output_weight_delta [k] * output_weight_delta [k];
    output_weight_delta [k] *= output [k] * (1 - output [k]);
}
#ifdef WAFFLY
n++;
if (!(n % 10))
    cerr << "2*MSE (every 10th) currently is: " << MSE << endl;
#endif
if (MSE < MSE_max) break;
// Posteriormente se procesa la capa oculta en términos del error
for (j = 0; j < hidden_size; j++) {
for (k = 0, sum = 0; k < output_size; k++)
sum += output_weight_delta [k] * output_nodes -> nodes [k].weights[j];
hidden_weight_delta [j] = sum * hidden_nodes -> nodes [j].output * (1 -
hidden_nodes -> nodes [j].output);
}
}

```

```

// Actualización de los pesos en la capa de salida.
for (k = 0; k < output_size; k++)
  for (j = 0; j < hidden_size; j++)
    output_nodes -> nodes [k].weights [j] += eta *
      output_weight_delta [k] * hidden_nodes -> nodes [j].output;
// Actualización de los pesos en la capa oculta.
for (j = 0; j < hidden_size; j++)
  for (i = 0; i < input_size; i++)
    hidden_nodes -> nodes [j].weights [i] += eta *
      hidden_weight_delta [j] * data [i];
  }
delete [ ] output;
delete [ ] output_weight_delta;
delete [ ] hidden_weight_delta;
}

```