



Casa abierta al tiempo
UNIVERSIDAD AUTÓNOMA METROPOLITANA
Unidad Iztapalapa

DIVISIÓN DE CIENCIAS BÁSICAS E INGENIERÍA

Identificación y Clasificación de Escenas por Causalidad de
Wiener-Granger

*Tesis para obtener el grado de Doctor en Ciencias
(Doctor en Ciencias y Tecnologías de la Información)*

M. en C. Cesar Benavides Alvarez

Asesores:

Dra. Graciela Román Alonso

Dr. Juan Villegas Cortez

Jurado Calificador:

Presidente: Dra. Elsa Rubio Espino

Secretario: Dr. Miguel Alfonso Castro García

Vocal: Dr. Marco Antonio Moreno Armendáriz

Vocal: Dr. Manuel Aguilar Cornejo

Vocal: Dr. Juan Villegas Cortez

Ciudad de México
Enero 2020



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA

ACTA DE DISERTACIÓN PÚBLICA

No. 00011

Matrícula: 2161801398

IDENTIFICACIÓN Y
CLASIFICACIÓN DE ESCENAS POR
CAUSALIDAD DE WIENER-GRANGER

En la Ciudad de México, se presentaron a las 10:00 horas del día 29 del mes de enero del año 2020 en la Unidad Iztapalapa de la Universidad Autónoma Metropolitana, los suscritos miembros del jurado:

DRA. ELSA RUBIO ESPINO
DR. MANUEL AGUILAR CORNEJO
DR. JUAN VILLEGAS CORTEZ
DR. MARCO ANTONIO MORENO ARMENDÁRIZ
DR. MIGUEL ALFONSO CASTRO GARCIA

Bajo la Presidencia de la primera y con carácter de Secretario el último, se reunieron a la presentación de la Disertación Pública cuya denominación aparece al margen, para la obtención del grado de:

DOCTOR EN CIENCIAS (CIENCIAS Y TECNOLOGIAS DE LA INFORMACION)

DE: CESAR BENAVIDES ALVAREZ

y de acuerdo con el artículo 78 fracción IV del Reglamento de Estudios Superiores de la Universidad Autónoma Metropolitana, los miembros del jurado resolvieron:

A probar

Acto continuo, la presidenta del jurado comunicó al interesado el resultado de la evaluación y, en caso aprobatorio, le fue tomada la protesta.



REVISÓ

MTRA. ROSALVA SERRANO DE LA PAZ
DIRECTORA DE SISTEMAS ESCOLARES

DIRECTOR DE LA DIVISIÓN DE CBI

DR. JESUS ALBERTO OCHOA TAPIA

PRESIDENTA

DRA. ELSA RUBIO ESPINO

VOCAL

DR. MANUEL AGUILAR CORNEJO

VOCAL

DR. JUAN VILLEGAS CORTEZ

VOCAL

DR. MARCO ANTONIO MORENO ARMENDÁRIZ

SECRETARIO

DR. MIGUEL ALFONSO CASTRO GARCIA

Agradecimientos

A mi esposa **Miriam Reyes**, por todo tu apoyo incondicional, sin tus consejos y regaños no lo hubiera logrado, por estar siempre a mi lado durante todo este periodo.

A mis padres **Alfredo** y **Francisca**, por apoyarme incondicionalmente, por su AMOR, aliento, esperanza y paciencia. Siempre pendientes de mí en todos los aspectos y sobre todo siempre creyeron en mí.

A mis hermanos **Benjamin** y **Alfredo**, que aunque no entienden mucho lo que hago, siempre me motivaron.

A mis asesores la **Dra. Graciela Román Alonso (UAM-I)** y el **Dr. Juan Villegas Cortez (UAM-A)**, por haberme transmitido sus conocimientos, por su paciencia, por la bonita amistad que surgió entre nosotros, por que siempre creyeron en mi desde el principio.

Al **Dr. Carlos Avilés Cruz (UAM-A)**, por las facilidades que me dio para trabajar dentro del laboratorio de Multimedia, de la UAM-Azcapotzalco y el apoyo con este trabajo de tesis, por su amistad y buenos consejos.

A los sinodales por aceptar revisar mi trabajo de investigación

- **Dra. Elsa Rubio Espino**
- **Dr. Marco Antonio Moreno Armendáriz**
- **Dr. Miguel Alfonso Castro García**
- **Dr. Manuel Aguilar Cornejo**
- **Dr. Juan Villegas Cortez**

Al **Dr. Francisco Fernandez de Vega** y al **Dr. Francisco Cháves de la O**, por recibirme en el Centro Universitario de Mérida, Universidad de Extremadura, y darme la oportunidad de trabajar dentro de su laboratorio de investigación.

Al **Dr. Leonardo Trujillo Reyes** por recibirme en el Instituto Tecnológico de Tijuana y darme la oportunidad de una estancia de investigación en su laboratorio Tree-LAB.

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

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

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

RESUMEN

El reconocimiento y la clasificación de imágenes se han utilizado ampliamente para la investigación en sistemas de visión artificial, en un esfuerzo por emular la visión humana en las máquinas. El objetivo de este trabajo es implementar una nueva estrategia llamada teoría de la causalidad de Wiener-Granger (CWG), como un nuevo paradigma de “leer” o “visualizar” una imagen digital, en una analogía de cómo las leemos los humanos de forma secuencial en líneas, con el fin de clasificar las imágenes contenidas en una base de datos. Esta estrategia se basa en representar las imágenes por medio de las características propias extraídas, vistas como texturas, haciendo uso de la metodología de recuperación de imágenes basada en contenido propio o content based image retrieval (CBIR), estas texturas serán analizadas por medio de la teoría de CWG con el fin de encontrar las relaciones causales entre las texturas, que representen un escenario en particular; posteriormente, se implementa un algoritmo genético (AG) para seleccionar las relaciones causales más relevantes de los tipos de escenarios (clases) en la base de datos, para encontrar un único patrón descriptivo para cada uno de ellos.

La metodología aquí propuesta consta de una etapa secuencial de extracción de características, una tarea de conformación de series temporales relacionadas a la imagen de estudio, una fase de estimación de causalidad y una selección de características de causalidad por medio del AG. Se implementó una etapa de clasificación y se utilizaron 700 imágenes de escenarios naturales para validar los resultados. Se desarrolló también la implementación del algoritmo propuesto dentro de un sistema de cómputo distribuido en paralelo reduciendo los tiempos de ejecución en un 84 % en comparación a la versión secuencial, logrando una eficiencia técnica del 100 % de reconocimiento y en pruebas de validación un 96 % de eficiencia por metodologías de resubstitución y validación cruzada respectivamente. Esta propuesta aporta una nueva perspectiva en el área de investigación del entendimiento y comprensión de las imágenes, y sus aplicaciones directas con aplicaciones a sistemas autónomos asistidos por visión.

Descriptor: Algoritmos evolutivos, Causalidad de Wiener Granger, Clasificación de escenas, Algoritmo k -medias, Clasificador k -NN, Cómputo paralelo.

Índice general

1. Introducción	1
1.1. Planteamiento del problema	3
1.2. Justificación	6
1.3. Propuesta	7
1.4. Objetivo general	9
1.5. Objetivos específicos	9
2. Estado del arte	10
2.1. Análisis de la metodología CBIR	10
2.2. Fundamentos de la teoría de Causalidad de Wiener-Granger	15
2.2.1. Causalidad en el dominio del tiempo: procesos estocásticos y modelo autorregresivo	18
2.2.2. Causalidad en el dominio de la frecuencia	20
2.2.3. Ejemplo computacional	22
2.3. Algoritmos evolutivos	22
2.4. Marco teórico del procesamiento digital de imágenes	25
2.4.1. ¿Qué es la visión?	25
2.4.2. Formación de imágenes en el ojo	26
2.4.3. Espectro visible para el ojo humano.	27
2.4.4. Imagen.	28
2.4.5. Píxel	28
2.4.6. Cuantización y muestreo de la imagen digital	29
2.5. Procesamiento digital de imágenes	31
2.5.1. Textura	33
2.6. Reconocimiento de patrones	34
2.6.1. Patrones	34
2.6.2. Elementos de un sistema de reconocimiento de patrones	35
2.6.3. Diseño de un sistema de reconocimiento de patrones	36
3. Metodología	38
3.1. Desarrollo de la arquitectura del sistema	38
3.2. Requerimientos computacionales	44
4. Algoritmo genético como soporte de la metodología CWG	46
4.1. Selección de relaciones causales por medio de un Algoritmo Genético	46
4.1.1. Codificación del individuo	46
4.1.2. Función de aptitud	48
4.1.3. Implementación del AG	50
4.2. Paralelización del algoritmo genético	52

5. Resultados	55
5.1. Resultados de clasificación	56
5.2. Rendimiento de la metodología paralela	63
5.3. Discusión de resultados	63
6. Conclusiones	71
Bibliografía	72
1. Glosario de términos	78
1.0.1. Información estadística de una imagen.	79
1.0.2. Acondicionado de la imagen.	80
1.0.3. Modelos de color	91
1.0.4. Teorema de Parseval	93
2. Cómputo paralelo	95
2.1. Introducción	95
2.1.1. SISD(Single instruction-Single data)	97
2.1.2. Multiprocesadores	97
2.1.3. Multicomputadoras	98

Índice de figuras

1.1.	Diagrama visión artificial	2
1.2.	Problema general de la recuperación de imágenes.	3
1.3.	Ejemplo de la iluminación en la imagen	4
1.4.	Representación de un píxel RGB y escala de Grises	5
1.5.	Diagrama a bloques para la recuperación de imágenes.	6
1.6.	Metodología propuesta aplicada para la clasificación de escenarios naturales.	8
1.7.	Ejemplo de las imágenes dentro de las bases de datos.	8
2.1.	Sistema CBIR escenarios naturales	11
2.2.	Metodología clásica de la clasificación de imágenes.	13
2.3.	Metodología CBIR propuesta por J. Vogel	14
2.4.	Transformada discreta de Fourier en el tiempo	22
2.5.	Un ejemplo simple dos modelos de procesos estocásticos	23
2.6.	Diagrama general que describe un proceso evolutivo.	24
2.7.	Diagrama de las partes del ojo humano.	26
2.8.	Representación óptica de un ojo mirando un objeto	27
2.9.	Espectro electromagnético	28
2.10.	Representación de una imagen digital	29
2.11.	Resolución espacial	30
2.12.	Resolución	30
2.13.	Representación discretización	31
2.14.	Elementos del análisis de imágenes	32
2.15.	Ejemplo de texturas.	34
2.16.	Elementos de un sistema de reconocimiento de patrones.	35
2.17.	Diseño de un sistema de reconocimiento de patrones.	36
3.1.	Ejemplo de selección de las texturas de referencia con segmentación manual	39
3.2.	Arquitectura del sistema de clasificación	40
3.3.	Etapas del acondicionamiento de la imagen	40
3.4.	Ejemplo de un emparrillado regular	41
3.5.	Ejemplo de la lectura temporal/secuencial de las vecindades de una imagen.	41
3.6.	Serie de tiempo / lectura secuencial de la imagen para el análisis de causalidad.	42
3.7.	Generación de la matriz de relaciones causales Π_s	43
4.1.	Proceso de generación de la matriz τ	47
4.2.	Construcción del genoma G_x	48
4.3.	Construcción de una imagen de consulta	49
4.4.	Diagrama general de flujo de la implementación del AG.	51
4.5.	Aplicación de los operadores genéticos.	51
4.6.	Arquitectura paralela diseñada	53

5.1. Generación del diccionario de texturas	55
5.2. Ejemplo de imágenes contaminadas con ruido	56
5.3. Ejemplo de imágenes transformadas para la prueba de rotación.	57
5.4. Representación gráfica de las relaciones causales para cada escenario natural.	59
5.5. Elementos recuperados en la prueba de resubstitución para la BDEN.	62
5.6. Resultados de clasificación para el AG	63
5.7. Medidas de eficiencia para los mejores individuos del AG	64
5.8. Imágenes recuperadas para elementos desconocidos en la base de conocimiento.	68
5.9. Resultados para las 50 imágenes desconocidas	69
5.10. Desempeño del AG para los escenarios naturales contenidos en la BDEN.	70
5.11. Rendimiento del algoritmo paralelo.	70
1.1. Análisis de píxeles en vecindad 4	78
1.2. Análisis de píxeles en vecindad 8	79
1.3. Histograma de imagen	79
1.4. Ruido gaussiano	81
1.5. Imagen con tipos de ruido sal y pimienta	82
1.6. Ventana	83
1.7. Ventana deslizante	84
1.8. Filtro promedio	85
1.9. Filtro gaussiano	85
1.10. Aplicación de filtro gaussiano	87
1.11. Filtro contra-harmónico	88
1.12. Filtro promedio geométrico	89
1.13. Filtro promedio-harmónico	89
1.14. Filtro YP	90
1.15. Filtro Laplaciano	90
1.16. Filtro diferenciador	91
1.17. Filtro enfatizador	92
1.18. Cubo RGB	93
1.19. Color HSI	93
1.20. Área bajo la curva de una señal	94
2.1. Imagen arquitecturas paralelas	97
2.2. Imagen SISD	97

Índice de tablas

2.1. Tabla de valores potencia de 2.	31
5.1. Rangos HSI para las texturas base de prueba.	55
5.2. Descripción de la BDEN utilizada para el trabajo de investigación.	56
5.3. Los valores obtenidos de la matriz τ	57
5.4. La matriz τ resultante de la selección manual de las relaciones causales más altas.	57
5.5. Mejores individuos resultantes de la evaluación del AG para cada escenario.	58
5.6. Valores finales de la matriz τ cuando se aplican los mejores individuos del AG.	58
5.7. Matriz de confusión para una prueba con 50 imágenes por escenario, usando la selección manual de relaciones causales.	58
1.1. Filtro gaussiano con ventana de 7x7.	87

Capítulo 1

Introducción

Uno de los desafíos que enfrentan los investigadores hoy en día es desarrollar un sistema de autenticación artificial que tenga capacidades de adquisición y procesamiento similares a las que poseen los humanos [28]. La visión artificial se define como la capacidad de una máquina para ver el mundo que lo rodea en una forma tridimensional a partir de un grupo de imágenes bidimensionales [57]. Dado que no existe un algoritmo efectivo que pueda reconocer completamente cualquier objeto que uno pueda imaginar en todo el entorno, la visión por computadora se considera aun un problema abierto. Para resolver este tipo de problemas se propone el desarrollo de sistemas de visión por computadora, los cuales están compuestos de diferentes etapas las cuales trabajan en conjunto para resolver un solo problema en particular [26].

Podemos decir que el reconocimiento automático de imágenes es uno de los problemas que pueden resolverse utilizando sistemas de visión por computadora. Los investigadores están ansiosos por desarrollar estos sistemas donde se han implementado diferentes técnicas para su mejora, como el aprendizaje automático ¹, el reconocimiento de patrones y los algoritmos evolutivos.

Para el ser humano ha sido uno de los retos más interesantes, el tratar de proveer a los sistemas artificiales de capacidades de adquisición y procesamiento parecidos a las que los seres humanos poseen [40]. Para que un sistema de visión artificial (ver Figura 1.1) pueda interactuar eficientemente con el medio que lo rodea de la misma manera que lo hacen las personas, es necesario que cuente con las capacidades adecuadas de captura y análisis automático de la información visual [57].

Cuando se habla del reconocimiento automático de objetos en imágenes digitales, el proceso se inicia con la captura 2D de una imagen en tres dimensiones, esta es procesada con el objetivo de mejorar su calidad y eliminar las posibles imperfecciones generadas al momento de ser capturadas; el siguiente paso consiste de separar los objetos en el fondo de la imagen a través de técnicas de segmentación, posteriormente la extracción de las características es la etapa donde los objetos son descritos por medio de su color, textura, geometría, etc; finalmente estas características son comparadas con las de otros objetos almacenados en la base de conocimiento. La comparación de estas características da como resultado una asignación del objeto a una clase correspondiente [26].

En la Figura 1.1 se muestra el diagrama a bloques de un sistema de visión artificial, en cada una de

¹N. del T. "Machine Learning"

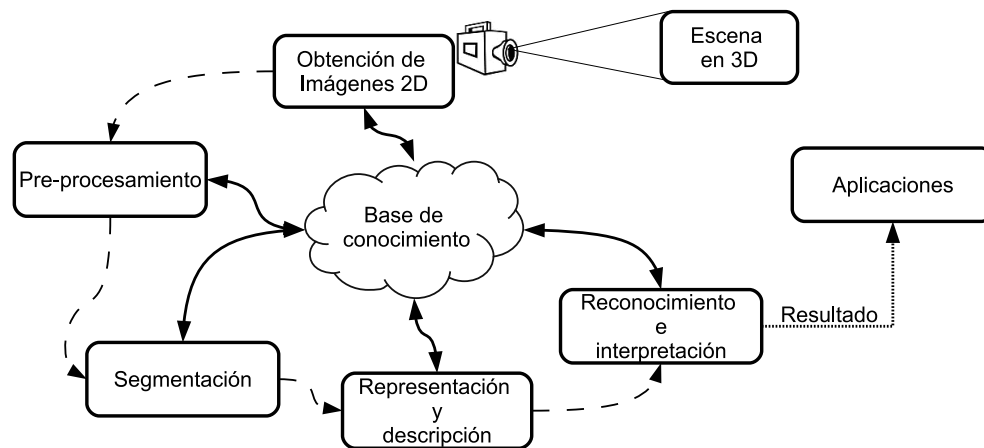


Figura 1.1: Diagrama a bloques del sistema de visión artificial.

las etapas es necesario tener un conocimiento previo, como puede ser, la cantidad de objetos a reconocer en la imagen, el tamaño de los objetos, sus características, su distribución, etc; con este conocimiento se pueden aplicar determinadas técnicas y algoritmos adicionales para un posible reconocimiento del objeto. Dada la descripción anterior, nos planteamos la pregunta: ¿Cómo es que una computadora logra realizar una gran cantidad de operaciones por segundo y no sea capaz de identificar una simple imagen e interpretarla como un escenario natural, una persona, un libro, etc? El enfoque dado a este trabajo consiste en utilizar las técnicas y herramientas computacionales con las cuales una computadora pueda reconocer los patrones contenidos en la imagen y, con ello realizar el proceso de identificación y clasificación.

Para las personas la tarea de reconocer y recuperar datos multimedia no presenta dificultad [44]. Los actuales sistemas de información presentan aun una serie de problemas debido a que están diseñados para procesar información alfanumérica, aunque algunas veces son ayudados con herramientas de desarrollo gráfico, y con simples técnicas de dibujo [33]. Por lo tanto, al día de hoy, se siguen desarrollando técnicas y sistemas automatizados enfocados hacia las tareas de identificación y clasificación de imágenes buscando desarrollar estas tareas de forma eficiente y lo más automatizadas posible.

En este contexto podemos decir que una de las tareas principales de un sistema automatizado de identificación o clasificación de imágenes es clasificar e identificar con éxito las imágenes a estudiar, y en particular en este trabajo, las imágenes de escenarios naturales².

Dicho lo anterior este trabajo presenta y propone un sistema de clasificación de imágenes de escenarios naturales haciendo uso de la recuperación de imágenes a partir de su contenido propio Content Base Image Retrieval (CBIR), para caracterizar la textura local dentro de un empujamiento regular, las características extraídas son representadas como series temporales; luego, las series temporales son analizadas por medio de la teoría de Causalidad de Wiener-Granger (CWG), la cual establece las relaciones causales que existen entre las texturas locales dentro de la imagen, y con esto se genera un patrón descriptivo a través de la fuerza de las relaciones causales para cada una de las clases o tipos de escenarios naturales contenidas en la base de datos. Los patrones generados a partir de las

²Se dice que una escena es natural si la imagen no tiene intervención o alteración por parte del ser humano.

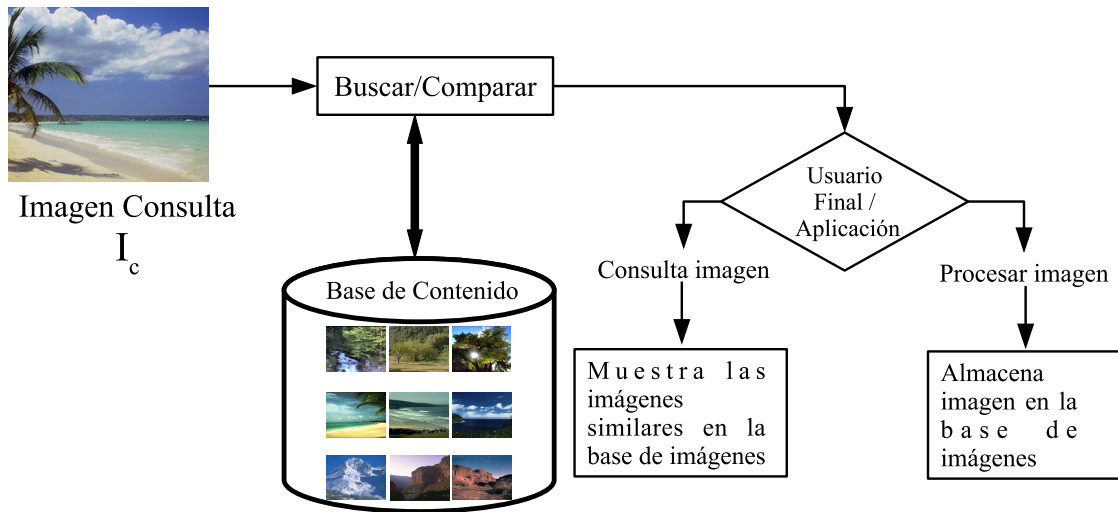


Figura 1.2: Problema general de la recuperación de imágenes.

relaciones causales son optimizados por medio de un algoritmo evolutivo, con el propósito de seleccionar las relaciones causales que contribuyan a discriminar de mejor manera una clase de la otra. El costo computacional de implementar estos algoritmos es muy alto, por ello se propone adicionalmente el uso de del cómputo paralelo para una pronta generación de resultados, por medio del modelo de programación de intercambio de mensajes (MPI).

Con la metodología propuesta se logra una perfecta clasificación de las imágenes para la prueba de resubstitución y hasta un 98 % de recuperación en las pruebas de ruido y rotación.

1.1. Planteamiento del problema

De manera general, el problema de la recuperación y clasificación de imágenes consiste en que dada una imagen consulta I_c , extraer de un repositorio de imágenes aquellas imágenes “similares” a I_c sobre la base de su contenido (ver Figura 1.2). Esto puede ser logrado al comparar la imagen consulta contra cada una de las imágenes del repositorio (base de imágenes).

Esta comparación se puede realizar píxel a píxel para determinar las imágenes más parecidas por medio de un análisis de distancia [40]. Sin embargo, en este trabajo de tesis la comparación se realiza al transformar la imagen consulta I_c en un descriptor de n rasgos. De esta manera se puede decir que dos imágenes son “*similares o parecidas*” si sus respectivos conjunto de vectores descriptores son parecidos o aproximados con respecto a una métrica dada. En los sistemas de identificación y clasificación de imágenes, debe existir la capacidad de comparar de manera eficiente dos imágenes para determinar si comparten contenido similar con respecto a las características extraídas; dichas características representan la información discriminante.

Desde este punto de vista, el problema de la identificación y clasificación de imágenes se puede plantear de la siguiente manera: una imagen digital constituye una función $f : U \times \mathbb{R}^2 \simeq \mathbb{R}$, la gráfica de la imagen I de f es el subconjunto de \mathbb{R}^3 que consiste de los puntos $(x, y, f(x, y))$, de los cuales el par ordenado (x, y) es un punto en U , y $f(x, y)$ es el valor en ese punto. Simbólicamente, la imagen

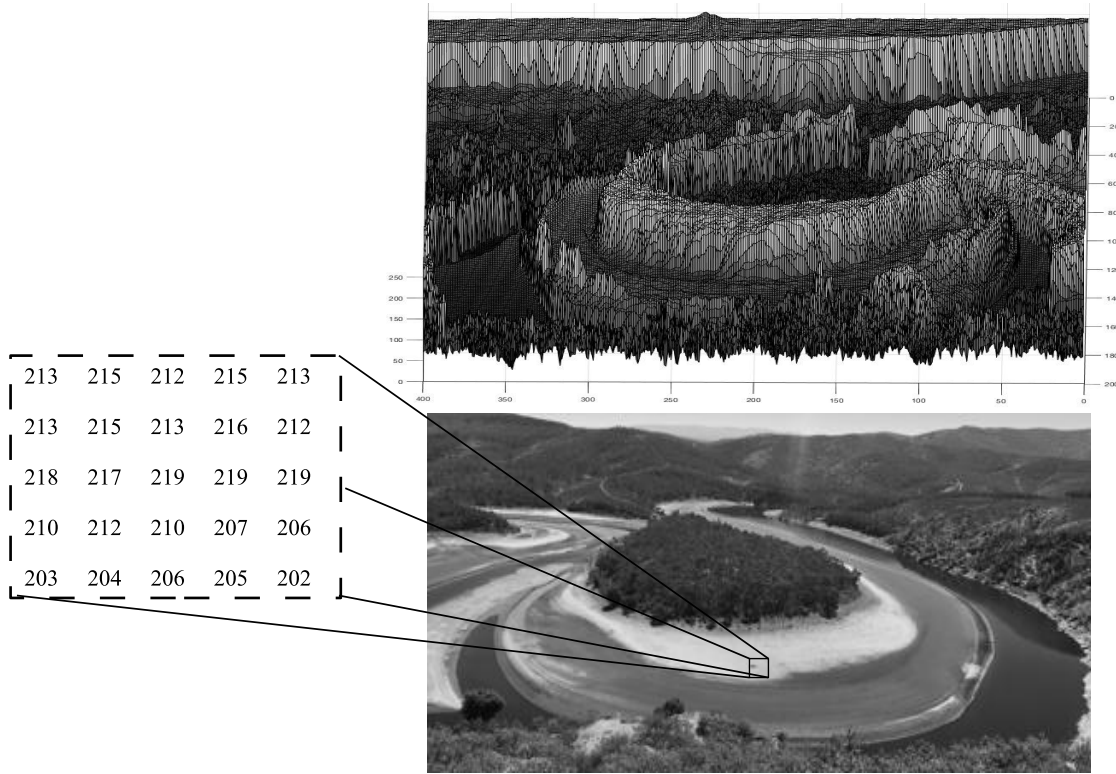


Figura 1.3: Ejemplo de la función que se produce físicamente por la luz que actúa en la escena, que se graba en la luz sensible material de la cámara.

$I = \{(x, y, f(x, y)) \in \mathbb{R}^3 \mid (x, y) \subset U\}$. La definición anterior considera que la función asociada a la luz que actúa en la escena u objeto está definida en general por las propiedades físicas, ópticas y geométricas que producen la experiencia perceptiva. Tenga en cuenta que esta definición abarca las asignaciones o transformaciones que actúan sobre la imagen en el tiempo y el espacio [28].

La función $f(x, y)$ la cual denota el valor de la intensidad del píxel en el punto (x, y) y se puede representar de manera gráfica como se observa en la Figura 1.3, el valor de un píxel específico en escala de grises se observa de manera representativa en la Figura 1.4(a). Para las imágenes en color, $f(x, y)$ denota el valor del color compuesto del píxel (x, y) en sus componentes de color rojo (red), verde (green) y azul (blue) (RGB), véase la Figura 1.4(b), la información en color se representa en términos de los tres colores primarios RGB [26], la función imagen se describe como se muestra en la Ecuación 1.1.

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

Sea F un mapeo desde el espacio imagen hacia un espacio $n - dimensional$, $X = \{x_1, x_2, \dots, x_n\}$ como se observa en la Ecuación 1.2:

$$F : f \rightarrow X \tag{1.2}$$

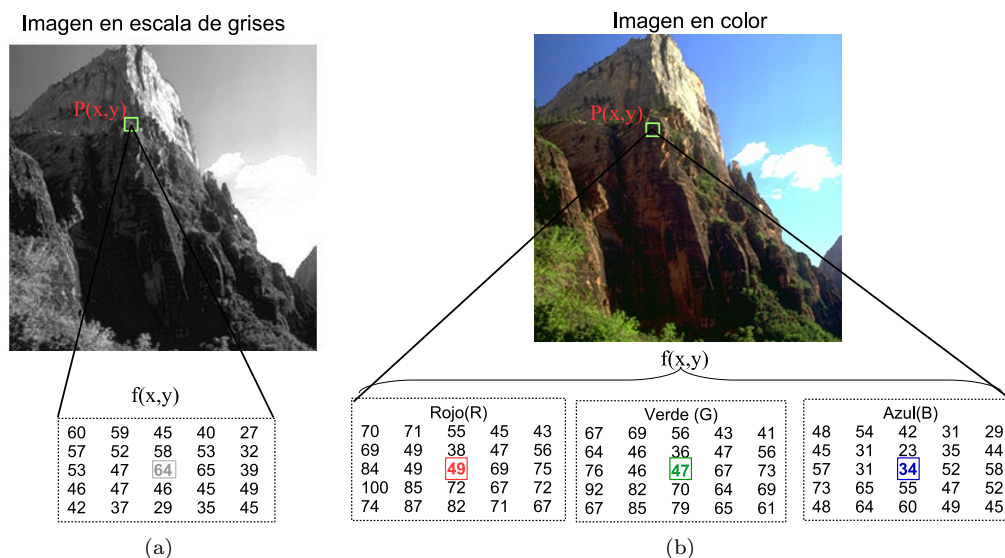


Figura 1.4: Representación de un píxel RGB y escala de Grises

En donde n es el número de características que se utilizan para representar a una imagen. La diferencia vectorial entre 2 imágenes f_1 , y f_2 se puede expresar como una distancia d , entre los respectivos vectores de características x_1 y x_2 . Dadas las ecuaciones anteriores, el problema de la recuperación y clasificación de imágenes con el criterio de la distancia mínima se puede proponer de la siguiente manera:

Dada una imagen consulta I_c , para recuperar una imagen f_r desde una base de datos de imágenes B , es necesario que se cumpla la Ecuación 1.3 con respecto a la distancia mínima entre la imagen consulta q y la imagen recuperada f_r .

$$d(F(I_c), F(f_r)) \leq d(f(I_c), F(f)) \quad (1.3)$$

$$\forall f \in B, f \neq f_r$$

En otros términos tenemos:

Dadas un grupo de p imágenes, $\{I_1, I_2, I_3, \dots, I_p\}$ con $p \gg 0$ pertenecientes a un conjunto heterogéneo B , extraer de B un subconjunto limitado de imágenes B_r , para una consulta dada I_c formulada en términos de un grupo de características. Véase Figura 1.5.

Podemos apreciar que la figura presenta una imagen consulta de una escena de montaña, la cual se pretende recuperar. Como resultado se extraen las tres imágenes más parecidas a la imagen de consulta, con ésta en el primer lugar gracias a la caracterización de la I_c como un patrón, y la comparación de ésta hacia la base de datos, siendo el resultado una interpretación visual de las imágenes más parecidas o asociadas a los patrones más cercanos en distancia al patrón de consulta.

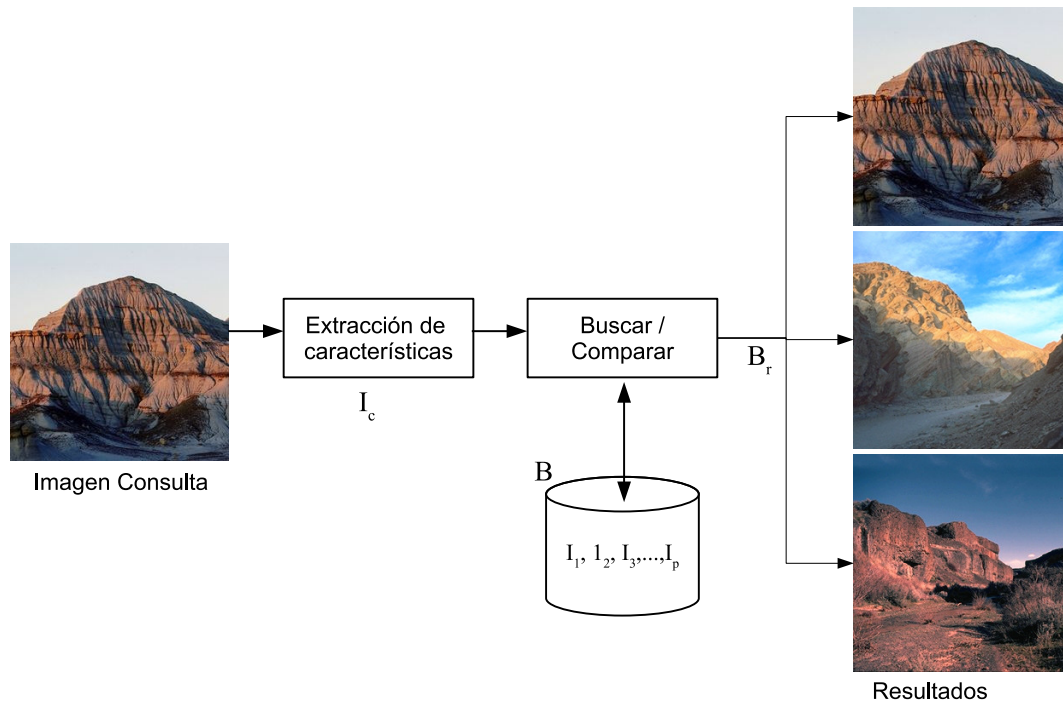


Figura 1.5: Diagrama a bloques para la recuperación de imágenes.

1.2. Justificación

Las imágenes se generan a un ritmo cada vez mayor por fuentes diversas tales como: satélites artificiales civiles y de defensa, vuelos de reconocimiento y vigilancia militar, dispositivos de captura de huellas digitales y captura de fotografías, experimentos científicos, imágenes biomédicas y sistemas de entretenimiento en el hogar por citar algunos.

Más de la mitad de la información en la Internet son imágenes, 85% de las cuales fueron tomadas con dispositivos móviles con una estimación final de 5 billones de imágenes reportadas hasta ahora [38]. Esta información se pretende organizarla, ordenarla y clasificarla de una manera automática.

Por lo tanto, sí se habla de una metodología capaz de diferenciar entre 10,000 clases de objetos diferentes, por decir algo, entonces hablamos de un problema de investigación aún abierto. De esta manera los sistemas de recuperación de imágenes se han venido desarrollando de manera amplia como un campo activo de investigación; se han implementado sistemas de recuperación por contenido mediante varias técnicas y enfoques; una de ellas es la metodología recuperación de imágenes por contenido propio o Content Based Image Retrieval (CBIR), que se explica con detalle en la siguiente sección.

Trabajos recientes han buscado la combinación de CBIR con diferentes técnicas para la generación automática de características por medio de la Programación Genética [51], para encontrar las mejores características que representen a una base de datos, con la idea de robustecer la técnica CBIR con alguna técnica que nos ayude a lograr clasificar mejor imágenes dentro de una base de datos.

En este trabajo, proponemos el uso de la extracción de características CBIR como entrada de un motor de análisis de causalidad de las texturas para caracterizar 6 tipos de escenarios, definiendo

manualmente un diccionario inicial conformado por 4 texturas base o fundamentales. En un futuro trabajo, se planifica que la conformación de este diccionario sea dinámico, considerando más texturas básicas y otros tipos de escenarios para mejorar el rendimiento de la clasificación.

1.3. Propuesta

En este trabajo, se desarrolla un sistema de recuperación de imágenes de escenarios naturales mediante la aplicación de la teoría de Causalidad de Wiener-Granger (CWG) [27] como una herramienta para analizar imágenes a través de la información del contenido propio de la imagen, sin el uso auxiliar de palabras descriptivas o datos dependientes de un idioma. Se identificaron las relaciones causales entre las texturas locales contenidas en una imagen, lo que proporciona una caracterización como un patrón descriptivo de un conjunto de escenas dentro de un conjunto de datos de imagen. La selección óptima de las relaciones de causalidad se llevó a cabo mediante la implementación de un algoritmo genético (AG) por medio de un proceso evolutivo.

Las etapas principales involucradas en el sistema propuesto y su desarrollo son las siguientes (Ver la Figura 1.6):

1. Lectura de la imagen: Primero, las imágenes se leen del repositorio y luego se aplica un cambio de espacio de color Rojo-Verde-Azul RGB, al Tono(Hue)-Saturación(Saturation)-Intensidad(Intensity) HSI.
2. Extracción de características: la extracción estadística de características CBIR se genera dentro de una vecindad en una cuadrícula de la imagen.
3. Conformación de series de tiempo³: las características numéricas de textura CBIR se organizan como una serie de tiempo, a través de una lectura secuencial para cada imagen contenida en el conjunto de datos.
4. Análisis de causalidad: el análisis CWG se aplica para calcular la matriz de relación causal entre diferentes características de textura.
5. Implementación del algoritmo genético: El AG se ejecuta para encontrar la caracterización óptima de las relaciones de causalidad que funcionan mejor para la recuperación de elementos naturales de las imágenes, las cuales tienen patrones de textura de causalidad similares para una escena en particular.

Este enfoque tiene como objetivo mejorar el proceso de optimización por medio de algoritmos evolutivos [14]. En este caso, dado que el AG muestra una implementación simple y rápida, se empleó para seleccionar las relaciones de las características estadísticas de textura local manejadas como series temporales, con el fin de caracterizar un escenario natural.

Se propone el uso de las siguientes bases de datos de imágenes ampliamente reconocidas por la comunidad científica del área de visión por computadora. En la Figura 1.7 se muestran ejemplos de las imágenes contenidas en cada una de ellas:

³Una serie temporal es un conjunto de observaciones secuenciales x_t , cada una de las cuales se registra en un tiempo específico t . [59]

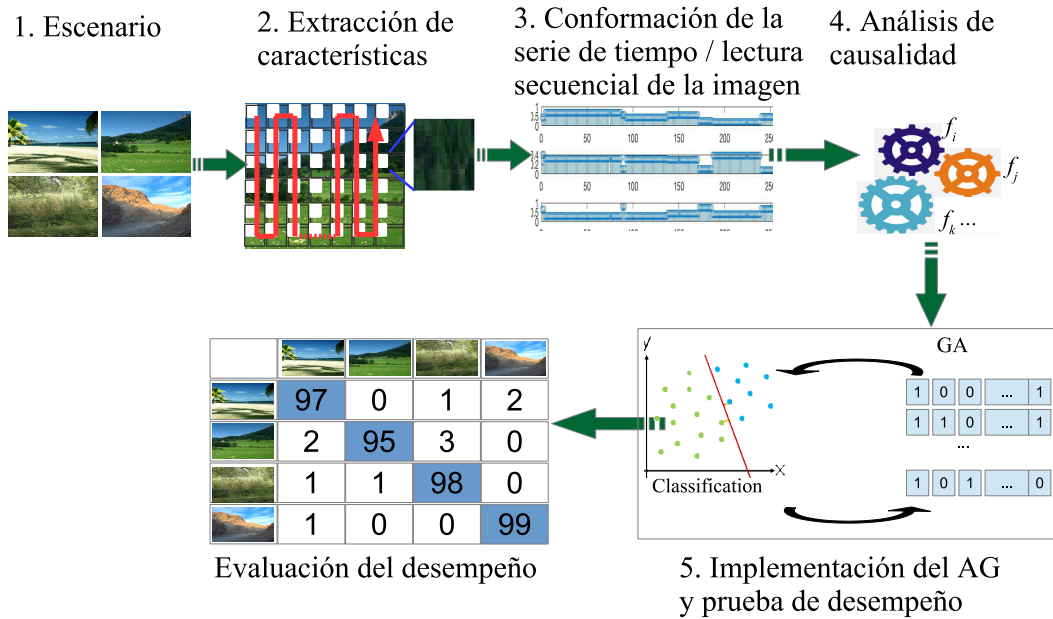


Figura 1.6: Metodología propuesta aplicada para la clasificación de escenarios naturales.

BD / clase	Bosques	Cielos	Costas	Montañas	Praderas	Ríos
Vogel and Shile						
Olivia and Torralba						

Figura 1.7: Ejemplo de las imágenes dentro de las bases de datos.

- Vogel and Shiele (V-S) [58], con 700 imágenes clasificadas como: 144 costa, 103 bosque, 179 montaña, 131 pradera, 111 río/lago y 32 cielo/nube.
- Oliva y Torralba (O-T) [50], con 1472 imágenes clasificadas como: 360 costa, 328 bosque, 374 montaña y 410 pradera.

Visualizando futuras implementaciones tales como un sistema autónomo de reconocimiento de escenas naturales montadas en un automóvil, que sería administrado por nuestra propuesta como el sistema autónomo [4, 6, 1], que reconoce escenarios naturales apoyando en la navegación de un automóvil autónomo o posiblemente un dron, con una seguridad del 100%, este sistema propuesto será un elemento importante en complementar la seguridad de los vehículos autónomos.

El resto del documento está organizado de la siguiente manera: el capítulo 2 presenta el estado del arte del problema del análisis de imágenes desde el criterio CBIR y la teoría CWG utilizada en

nuestro proyecto, así como el soporte teórico del modelo CWG a ser aplicado y los fundamentos básicos del procesamiento digital de imágenes, el reconocimiento de patrones y los algoritmos evolutivos; el capítulo 3, presenta la metodología propuesta para aplicar la teoría CWG en la caracterización de imágenes de escenarios naturales; el capítulo 4, explica nuestro enfoque de implementación del algoritmo genético para optimizar la selección de relaciones causales entre las texturas y la implementación de los algoritmos paralelos de nuestra propuesta para obtener una buena eficiencia al procesar una gran cantidad de imágenes; finalmente, los resultados y conclusiones se presentan en los capítulos 5 y 6, respectivamente.

1.4. Objetivo general

Obtener un modelo matemático basado en una lectura secuencial de la imagen, que haga uso de las ecuaciones de la teoría de Causalidad de Wiener-Granger, que sirva de apoyo para el análisis del contenido de imágenes digitales, a fin de lograr una correcta identificación y clasificación de escenas, dentro de un gran conjunto de imágenes de manera rápida y eficiente.

1.5. Objetivos específicos

- Aportar una nueva metodología a partir de un modelo matemático, para la aplicación de la causalidad en imágenes de escenarios naturales, para la clasificación de escenas.
- Proponer un sistema paralelo para agilizar el análisis de causalidad de las imágenes utilizando las unidades de procesamiento de un clúster.

Capítulo 2

Estado del arte

El problema de la clasificación y el reconocimiento de imágenes ha sido estudiado desde diferentes enfoques con el fin de apoyar la búsqueda visual para diferentes propósitos, en este capítulo se muestran los trabajos relacionados a los sistemas de clasificación de imágenes utilizando la recuperación de imágenes por contenido propio (CBIR), y la idea de complementarla con otras técnicas para tener mejores resultados a la hora de la clasificación. Se muestran también los fundamentos teóricos de la teoría de causalidad de Wiener-Granger, desarrollados a detalle para una mejor comprensión del trabajo realizado. Finalmente se se presenta el marco teórico del procesamiento digital de imágenes y el reconocimiento de patrones.

2.1. Análisis de la metodología CBIR

La implementación de un sistema de recuperación de imágenes basado en su contenido propio es una de las herramientas que nos ayuda a manejar la información contenida en las imágenes de manera eficiente. La Figura 2.1 muestra de manera gráfica la representación de un sistema CBIR. Este sistema ayudará a los usuarios a encontrar imágenes relevantes basadas en sus características de contenido propio o aquellas que visiblemente están relacionadas con ellas, desde nuestra percepción visual, tal sistema ayuda a los usuarios, incluso aquellos que no están familiarizados con la base de datos, a recuperar imágenes relevantes basadas en sus contenidos. Las áreas de aplicación en donde CBIR es una actividad principal son numerosas y diversas, como por ejemplo: galerías de arte y gestión de museos, teledetección y gestión de recursos de la tierra, sistemas de Información Geográfica, predicción del tiempo, telas y diseño de moda, etc.

Los enfoques previos para la recuperación basada en el contenido han tomado dos direcciones. En el primero, el contenido de la imagen se modela como un conjunto de atributos extraídos manualmente y manejados dentro del marco de los sistemas convencionales de administración de bases de datos. Las consultas se especifican utilizando estos atributos. La representación de imágenes basada en atributos implica un alto nivel de abstracción de imágenes. En general, cuanto mayor es el nivel de abstracción, menor es el alcance para plantear consultas apropiadas a la base de datos de imágenes. La recuperación basada en atributos es recomendada y avanzada principalmente por investigadores de bases de datos. El

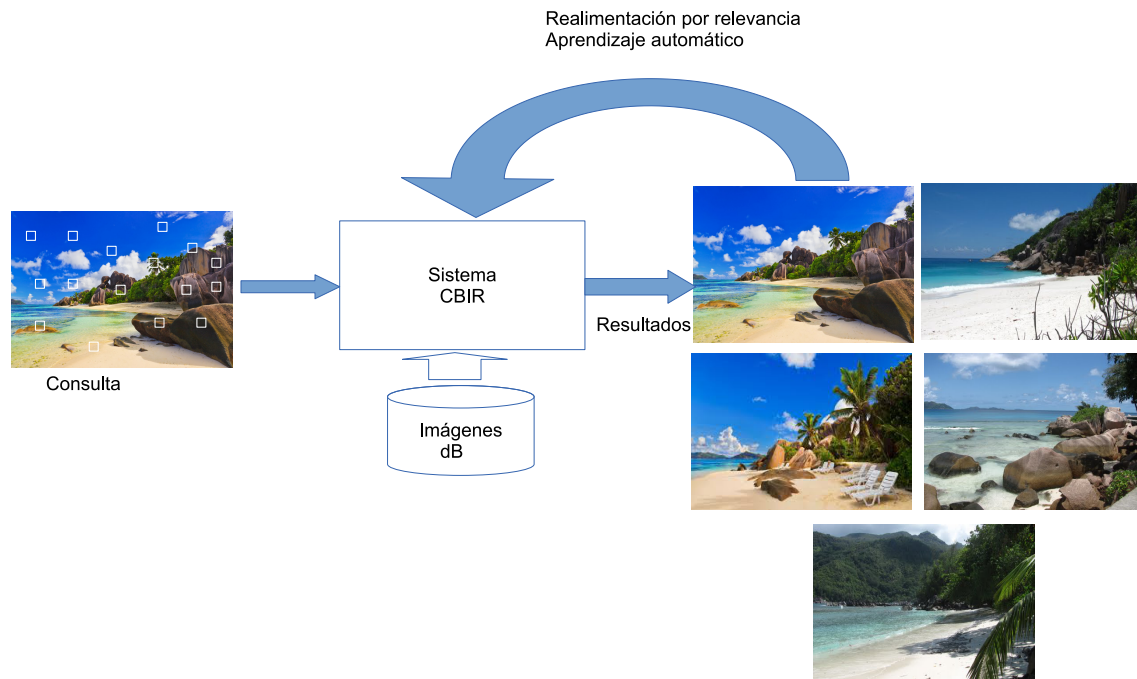


Figura 2.1: Sistema CBIR en escenarios naturales.

segundo enfoque depende de un subsistema integrado de extracción de características/reconocimiento de objetos para superar las limitaciones de la recuperación basada en atributos. Este subsistema automatiza la tarea de extracción de características y reconocimiento de objetos que ocurre cuando la imagen se inserta en la base de datos. Sin embargo, los enfoques automatizados para el reconocimiento de objetos son computacionalmente caros, difíciles y tienden a ser específicos del dominio. Este enfoque está avanzado principalmente por investigadores expertos en la interpretación de imágenes.

Como se presenta en [52], al principio de la década los 90's la técnica CBIR es reportada por primera vez por T.Kato [32]. El acrónimo CBIR fue acuñado para describir un sistema basado en el análisis de color y forma, que recuperaba imágenes de una base de datos. Sin embargo, es durante a principios de la década del año 2000, cuando esta técnica se orientó específicamente para recuperar automáticamente atributos y representar datos. Es importante mencionar que la técnica CBIR utiliza características de bajo nivel, tales como texturas, color y forma para representar imágenes. Por otro lado, un problema abierto y también un objetivo primordial es reconocer y clasificar las grandes cantidades de de imágenes que existen hoy en día. El mejor ejemplo de ello es el de Internet, el cual se ha convertido en un lugar de convergencia de todo tipo de imágenes (rostros, coches, dispositivos electrónicos, mapas, paisajes, flora, fauna, etc.). Independientemente del propósito del usuario, cada tipo de imagen tiene sus propios problemas específicos que resolver en términos de reconocimiento y clasificación. Por lo tanto, las técnicas CBIR están diseñadas para enfocarse en un tipo de imágenes, siendo los escenarios naturales uno de los más complicados, debido a la mezcla de formas y colores irregulares [29, 30, 52]. Tradicionalmente, el método para organizar las imágenes consiste en generar etiquetas asociadas a cada una de ellas. Así es como se crearon las bases de datos que contienen las consultas de etiquetas asociadas a cada imagen, teniendo cada imagen tantas etiquetas y conceptos como sea posible. Un

ejemplo de automatización de esta técnica está presente en los motores de búsqueda, a través del uso de la teoría de los autómatas. De esta manera, cuando se analiza una página web que contiene imágenes, se realiza un análisis estadístico sobre las palabras dentro de los textos de la página web, y automáticamente, las palabras más frecuentes en estos textos se asocian como etiquetas de las imágenes de la página web. Esta técnica ha sido intensa y ampliamente utilizada, pero sus resultados a veces son irrelevantes, es decir, el motor de búsqueda devuelve imágenes que no tienen ninguna relación con lo que se está buscando. Otra desventaja, en cuanto al etiquetado, consiste en el idioma de las etiquetas, ya que implica la traducción de las mismas a otros idiomas. Una institución que fue pionera en las aplicaciones CBIR fue IBM. Patentó el sistema Query By Image Content en 1995, y a lo largo de los años aparecieron diferentes sistemas. Entre ellos, Excel Photobook, desarrollado por MIT; Blobword, establecido por UC Berkeley; SIMBA; y FIRE. Con el auge y el uso masivo de imágenes digitales en Internet, aparecieron una variedad de motores de búsqueda de imágenes, tales como como Webseer en 1996 y Webseek en 1997. Al mismo tiempo, se produjeron importantes avances en la búsqueda en bases de datos relacionadas como IBM DB2 y Oracle, que dotaron a sus motores de productos con aplicación para facilitar la recuperación de imágenes por el contenido visual, acercando así el CBIR al campo de la industria.

Al día de hoy desarrollar sistemas CBIR que logren un grado razonable de independencia de dominio sigue en desarrollo, el principal problema para tener un sistema completamente automatizado es la extracción de características. Las características estudiadas al momento siguen siendo generadas a priori. Hay dos categorías principales de características: primitivas y lógicas. Las características de imagen primitivas o de bajo nivel, como los centroides y límites de los objetos, se pueden extraer de forma automática o semiautomática. Las características lógicas son representaciones abstractas de las imágenes en varios niveles de detalle. Algunas características lógicas pueden sintetizarse a partir de características primitivas, mientras que las características primitivas solo pueden obtenerse a través de una participación humana considerable. Las características lógicas denotan la semántica de dominio más profunda manifestada en las imágenes. Las características primitivas generalmente se derivan semiautomáticamente, mientras que las características lógicas se extraen manualmente o semiautomáticamente [5]. Finalmente, las consultas de las imágenes se procesan utilizando características primitivas y lógicas.

Previamente se desarrollaron un par de trabajos aplicando con éxito la técnica CBIR al problema de reconocimiento facial [9, 10]. A diferencia del reconocimiento facial, los escenarios naturales están compuestos por múltiples texturas, objetos en posiciones desconocidas y sus diferentes composiciones, hacen que el trabajo de clasificación de estas imágenes sea más difícil. Podemos decir que una sola característica no es suficiente para describir una imagen por completo, pero el uso o combinación de características como el color, la textura, la forma y la información espacial en la recuperación de imágenes puede no dar una buena discriminación en los resultados de recuperación. Con todo este estudio se ha logrado observar que existe una estrecha relación entre el color y la textura de una imagen.

La metodología CBIR se ha aplicado con éxito varias técnicas para el problema de reconocimiento facial [10, 2, 41, 42, 15]. Las soluciones se favorecen a través del control de la forma en que se obtienen las imágenes al determinar la cantidad de luz, la orientación, la distancia, etc. que permite obtener

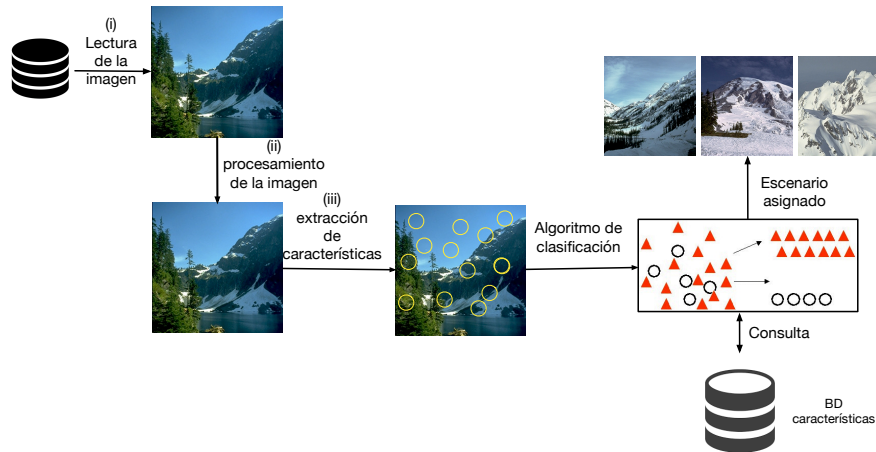


Figura 2.2: Metodología clásica de la clasificación de imágenes.

las imágenes de rostro ideales. Además, los puntos a identificar en una imagen de la cara son bien conocidos. Por otro lado, las múltiples texturas, los objetos en posiciones desconocidas y sus diferentes composiciones hacen que sea bastante difícil reconocer e identificar el tipo de escenario natural en una imagen o grupo de imágenes.

Una de las soluciones más recientes para la clasificación de escenarios naturales es el uso de la técnica de aprendizaje profundo (Deep Learning) [23], que consiste en un conjunto de redes neuronales conectadas entre sí en capas sucesivas, donde cada red de capa realiza una operación de convolución en la información de la capa anterior, como podemos ver en [47]. Esta metodología tiene la desventaja de requerir recursos computacionales de alto nivel (memoria y CPU) para la tarea de capacitación, a diferencia de la técnica CBIR que se puede implementar en sistemas con pocos recursos.

Cuando se usa CBIR para la clasificación de imágenes de escenarios, los descriptores significativos se determinan considerando que los atributos de textura propia de la imagen tienen una recuperación importante y efectiva. En este sistema, un usuario presenta una consulta de imagen, y el sistema devuelve las imágenes similares dentro de la base de datos. En la Figura 2.2 se muestra el diagrama general de un sistema de clasificación basado en CBIR de imágenes de escenarios naturales.

Uno de los primeros artículos que utiliza la metodología CBIR para la clasificación de escenarios naturales es J. Vogel [29], este trabajo define una cuadrícula regular de 10×10 en la imagen; de cada cuadrícula se abre una ventana de análisis (vecindad), ver Figura 2.3. La información local de textura, se extrae de una ventana y se compara con un diccionario de texturas base; entonces, el autor define un sistema de clasificación para el escenario natural. Un punto que debe mejorarse es la definición del diccionario de texturas base que se establece manualmente, incluidas solo las texturas típicas percibidas por los investigadores. Este enfoque obtiene un promedio de hasta 75 % de eficiencia para la prueba

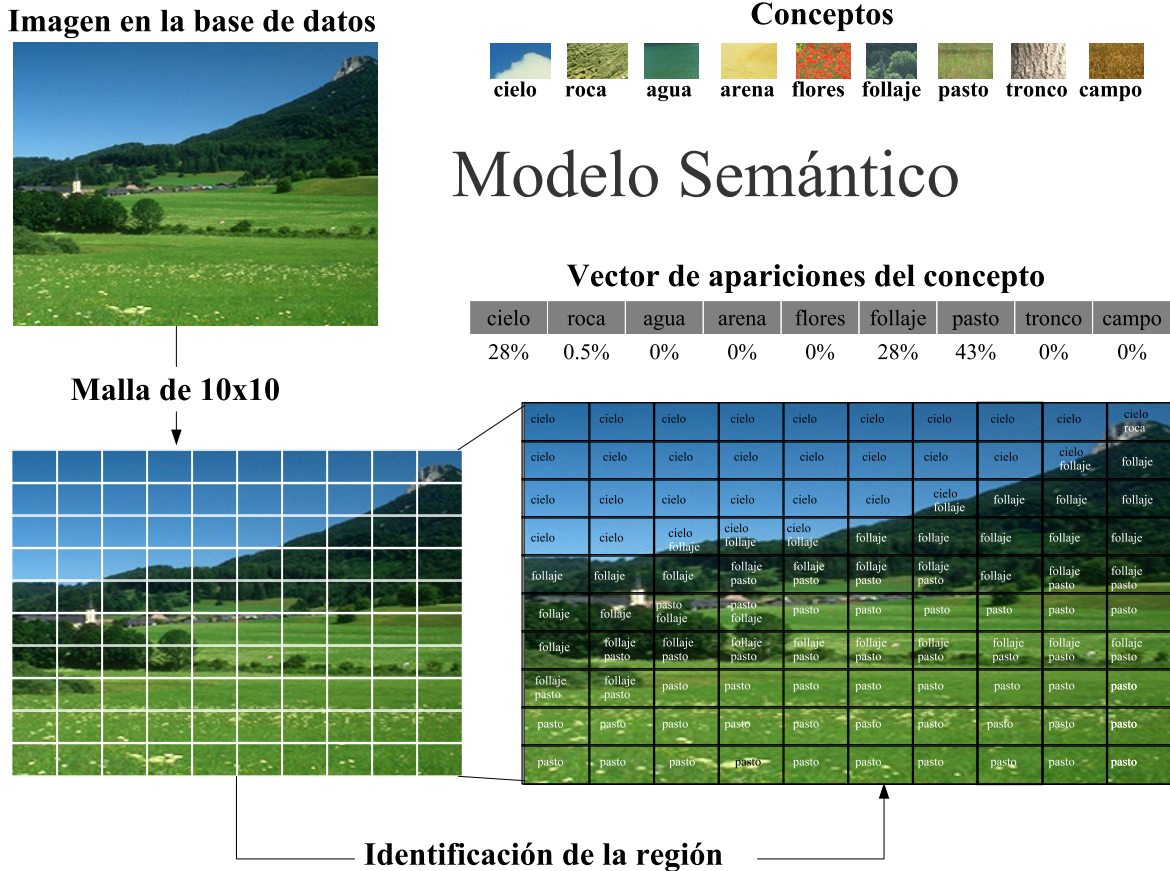


Figura 2.3: Metodología CBIR propuesta por J. Vogel [29].

de clasificación de validación cruzada.

A diferencia de una cuadrícula, en [30], se lanzan puntos al azar sobre la imagen y alrededor de cada punto se abre una ventana; Desde cada ventana, la información estadística local de la textura se extrae para agruparla de forma dinámica según un diccionario de texturas base. Las pruebas se realizan teniendo en cuenta que el diccionario generado obtiene un promedio de clasificación de 85 % de las bases de datos de escenarios naturales.

En [56], el enfoque CBIR se presenta para clasificar imágenes de paisajes naturales a través de la composición de características relevantes en relación con la textura como en [11], la forma y la distribución de la luminosidad.

Al ser una técnica de aprendizaje sin supervisión, CBIR todavía tiene algunas desventajas, ya que la información extraída solo se trata como un histograma que representa la composición de las texturas en un escenario. Esta forma de caracterizar un escenario no ha sido capaz de obtener más de 85 % de clasificación, por eso han surgido nuevas propuestas que utilizan metodologías híbridas para dar a CBIR una mayor solidez [37].

En [37, 61], los autores combinan la información CBIR con cierto contenido semántico introduciendo objetos conceptuales de alto nivel, tratando de vincular imágenes basadas en contenido con objetos

extraídos en su interior. Este trabajo obtuvo un porcentaje de clasificación de paisajes naturales no mayor que [29] y [30].

En este trabajo se presenta un método híbrido de tres componentes. Uno de ellos es CBIR, que genera la información sobre las características de textura locales de la imagen. A diferencia de realizar solo el manejo estadístico de las características obtenidas, en esta propuesta, el segundo componente, no utilizado aun por la literatura, el cual es responsable de aplicar un análisis de causalidad basada en la teoría de causalidad de Wiener Granger (CWG) para identificar las relaciones causales que existen dentro de las texturas básicas de un tipo de escenario natural. Dado que el componente de causalidad genera diferentes configuraciones de relaciones causales, el tercer componente consiste en un AG que permite la selección de la configuración que obtiene el mejor porcentaje de clasificación para cada escenario.

La visión computacional evolutiva (VCE) como área de investigación se está desarrollando actualmente en inteligencia artificial a través de dos áreas de trabajo, como la visión computacional y como la computación evolutiva. Comenzando desde un punto de vista práctico, VCE busca diseñar soluciones de software y hardware necesarias para resolver problemas de visión duros de computadora [28]. La computación bioinspirada dentro de la visión computacional contiene un conjunto de técnicas que se aplican con frecuencia a problemas difíciles de optimización. Su objetivo principal es generar soluciones formuladas de manera sintética y, con frecuencia, el proceso evolutivo artificial basado en la teoría evolutiva desarrollada por Charles Darwin es el aplicado [14].

2.2. Fundamentos de la teoría de Causalidad de Wiener-Granger

El problema de la definición de causalidad no es trivial para los sistemas complejos, donde a diferencia de los sistemas simples observados en el mundo todos los días, una comprensión intuitiva de causa y efecto es deficiente. En 1956 Norbert Wiener introdujo la noción de que una variable (o series de tiempo) que se podría llamar *causal* a otro, si la capacidad de predecir la segunda variable se ha mejorado, mediante la incorporación de información sobre la primera [49, 60]. Wiener sin embargo, carecía de una notación práctica para la aplicación de su idea. Dicha aplicación fue introducida en 1969 por el economista Clive Granger (1969), en el contexto de los modelos autorregresivos lineales de procesos estocásticos.

La teoría de causalidad hoy en día esta siendo aplicada a las en diferentes campos de las ciencias ya que nos es de utilidad para poder predecir la aparición de un evento a partir de un evento previo que se desencadenó en el pasado del cual ya tenemos información previa. La definición clásica de la causalidad se puede ver como:

Definición básica de Causalidad. “Para que un evento A cause un evento B , es necesario y suficiente que suceda A para que suceda B [43].”

El paradigma de *inferencia causal* se ha utilizado en diferentes campos de la ciencia, por ejemplo, en neurología, la teoría de la CWG [8] se usa para examinar áreas del cerebro y las relaciones causales entre ellas. El análisis de la CWG se realiza con los sensores [46, 43] y, últimamente, en imágenes de MRI [63, 21, 34], se esta utilizando la teoría de la CWG para el estudio de relaciones causales entre

áreas del cerebro que realizan actividades. Otros campos científicos en los que se ha aplicado la teoría de la CWG son el procesamiento de video para indexación y recuperación [18], procesamiento de video para identificación de personas masivas e identificación de vehículos [36, 53, 62] y análisis de escenarios complejos [20, 19]. En esta propuesta, por primera vez, la teoría CWG se aplica a la recuperación de elementos naturales y escenas naturales.

La idea básica de Causalidad Wiener-Granger (CWG) [8], o causalidad-G, es sencilla. Supongamos que tenemos dos variables $X1$ y $X2$, y tratamos de predecir $X1_{(t+1)}$ usando sólo los términos pasados de $X1$. También tratamos de predecir $X1_{(t+1)}$ usando términos anteriores de $X1$ y $X2$. Si la segunda predicción tiene significativamente más éxito, entonces el pasado de $X2$ contiene información útil para predecir $X1_{(t+1)}$ que no está en el pasado de $X1$. En este caso, $X2$ se dice causa en el sentido de Granger a $X1$. Obsérvese que estas variables pueden ser series de tiempo escalar, para lo cual utilizar el tipo normal (por ejemplo, $X1$) o de series de tiempo del topo vector, para los que usamos negrita (por ejemplo, $X1$). Debido a que los valores de una variable a la vez se predicen por valores de otras variables en los primeros tiempos, a menudo se dice que CWG depende de “precedencia temporal”. Sin embargo, no es suficiente simplemente que los acontecimientos en el resto de variables temporalmente preceden a eventos similares en la primera variable. Para que CWG sea significativa, la previsibilidad estadísticamente significativa debe ser establecida. En otras palabras, los valores distintos de cero para CWG por lo general se pueden obtener a partir de cualquier conjunto de series de tiempo, pero estos valores no tienen sentido a menos que se determine que no son estadísticamente significativos [8]. El enfoque de Granger, vía dos series de tiempo, es determinar series de tiempo por un factor predictivo de la otra, mediante el uso de modelos de regresión lineal. Más específicamente, para $X1$ y $X2$ de series de tiempo, si se puede predecir mejor $X2$ a la vez utilizando los valores de tiempo anteriores de $X2$ y $X1$, que por el simple uso de los valores anteriores de $X2$ solamente, entonces $X1$ contiene información útil para predecir $X2$. Aunque el mecanismo causal es desconocido, se dice que $X1$ en el sentido de Granger causa $X2$. Granger utiliza el modelado de series de tiempo autorregresivo lineal para formalizar la noción. Por lo tanto, se supone que $X1(t)$ y $X2(t)$ están modeladas por las sumas en los lados derechos de las ecuaciones 2.1 y 2.2:

$$X1(t) = \sum_{j=1}^p A_{11,j} X1(t-j) + \sum_{j=1}^p A_{12,j} X2(t-j) + E1(t) \quad (2.1)$$

$$X2(t) = \sum_{j=1}^p A_{21,j} X1(t-j) + \sum_{j=1}^p A_{22,j} X2(t-j) + E2(t) \quad (2.2)$$

donde:

$X1(t)$ y $X2(t)$ son los valores de predicción con propiedades estadísticas en el sentido amplio, p es el desfase de tiempo máximo en el modelo (llamado el orden del modelo), A es la matriz que contiene los coeficientes del modelo, $X1_{(t-j)}$ y $X2_{(t-j)}$ representan los valores pasados de $X1$ y $X2$, y $E1(t)$ y $E2(t)$ son los errores de predicción de los modelos (también llamadas innovaciones).

Si la varianza del error de predicción del modelo de $X2$ se reduce significativamente mediante la inclusión de los valores pasados de $X1$, en comparación cuando no se incluyen, entonces $X1$ causa en el sentido de Granger a $X2$. Si los supuestos de estocasticidad, estacionariedad en sentido amplio, y linealidad son válidas, entonces la hipótesis nula de ninguna relación significativa causalidad Granger se

puede probar con una prueba F. La prueba de causalidad de Granger [8] es una de hipótesis estadística para determinar si una sola serie de tiempo es útil en el pronóstico de otra. Por lo general las regresiones reflejan “simples” correlaciones. Se dice que una serie de tiempo $X1$ genera una causa-Granger $X2$ si se puede demostrar, usualmente a través de una serie de t -pruebas y F -pruebas en los valores retrasados de $X1$ (y también con valores retrasados de $X2$), estos valores de $X1$ proporcionan información estadísticamente significativa sobre los valores futuros de $X2$. Decimos que una variable $X1$ que evoluciona con el tiempo, la causalidad de Granger provoca otra variable $X2$ que evoluciona si las probabilidades del valor de $X2$ en función de sus propios valores pasados, y en los valores anteriores de $X1$ son mejores que las predicciones de $X2$ basándose sólo en sus propios valores pasados [27].

Granger define la relación de causalidad basándose en dos principios:

- La causa ocurre antes de su efecto.
- La causa tiene información única sobre los valores futuros de su efectos.

La solución a este problema requiere del procesamiento de un gran número de funciones relacionadas con la causalidad, sobre todo cuando se trabaja con más de dos señales.

El concepto de causalidad se abordará primeramente desde el plano matemático, con el fin de definir un modelo matemático para la identificación de escenarios naturales, para posteriormente pasar a implementar, con un abordaje en cómputo paralelo, hacia el análisis de imágenes en escenarios naturales. Se propone el uso de causalidad para determinar el tipo de escena y su clasificación, a partir del contenido propio de la imagen, dentro de subconjuntos de variables aleatorias (aproximación de las texturas), modelándolas como series de tiempo. El abordar como series de tiempo a las diferentes texturas presentes en un escenario natural, nos permitirá deducir sus componentes, para con ello poder predecir el escenario dentro de la imagen.

Debido a que los valores de una variable a la vez se predicen por los valores de otras variables en épocas anteriores, a menudo se dice que la CWG depende de la “precedencia temporal”. Sin embargo, no es suficiente simplemente que los eventos en las otras variables precedan temporalmente a eventos similares en la primera variable. Para que la CWG sea significativa, se debe establecer una predictibilidad estadísticamente significativa. En otras palabras, los valores distintos de cero para CWG generalmente se pueden obtener de cualquier conjunto de series de tiempo, pero estos valores no tienen sentido a menos que se determine que son estadísticamente significativos.

Grabaciones neurofisiológicas multielectrodos producen cantidades masivas de datos. El análisis de series temporales multivariantes proporciona el marco básico para el análisis de los patrones de interacciones neuronales en estos datos. Durante mucho tiempo se ha reconocido que las interacciones neuronales son direccionales. Ser capaz de evaluar la direccionalidad de las interacciones neuronales es por lo tanto una capacidad muy deseada para la comprensión de la naturaleza cooperativa de la computación neural [8]. Las investigaciones realizadas durante los últimos años ha demostrado que la CWG es una técnica clave para proporcionar esta capacidad [63].

En neurociencia, como en muchos otros campos de la ciencia y la ingeniería, las señales de interés se recogen a menudo en forma de series temporales simultáneas múltiples. Para evaluar la interdependencia estadística entre estas señales, se calcula las funciones de correlación cruzada en el dominio del tiempo y de las funciones de coherencia ordinarios en el dominio espectral. Sin embargo, en muchas

situaciones de interés, medidas simétricas como la coherencia ordinaria no son completamente satisfactorios, y se requiere una mayor disección de los patrones de interacción entre las señales grabadas a parcelar la conectividad funcional efectiva en las redes complejas. Trabajos recientes han empezado a considerar la influencia causal neural que puede ejercer una zona del cerebro sobre otra. La idea básica se remonta a Wiener que concibió la idea de que, si la predicción de una serie de tiempo podría mejorarse mediante la incorporación de los conocimientos de un segundo, entonces la segunda serie se dice que tiene una influencia causal sobre el primero. La idea de Wiener carece de la maquinaria para la aplicación práctica. Granger después formaliza la idea de predicción en el contexto de modelos de regresión lineal. En concreto, si la varianza del error de predicción autorregresiva de la primera serie de tiempo en el momento actual se reduce mediante la inclusión de mediciones anteriores de la segunda serie de tiempo, entonces la segunda serie de tiempo se dice que tiene una influencia causal sobre el primero. Los papeles de las dos series de tiempo se pueden invertir para abordar la cuestión de la influencia causal en la dirección opuesta. A partir de esta definición, es evidente que el flujo del tiempo juega un papel vital en permitir hacer inferencias acerca de las influencias causales direccionales de datos de series de tiempo. La interacción descubierta de esta manera puede ser recíproca o puede ser unidireccional.

Dos desarrollos adicionales de la idea de la causalidad de Granger son importantes. En primer lugar, para tres o más series de tiempo simultáneas, la relación causal entre dos cualquiera de la serie puede ser directa, puede ser mediada por una tercera, o puede ser una combinación de ambas. Esta situación puede abordarse mediante la causalidad de Granger condicional. En segundo lugar, la serie de tiempo natural, incluyendo los de la economía y la neurobiología, contiene aspectos oscilatorios en las bandas de frecuencia específicas. Por tanto, es deseable tener una representación espectral de la influencia causal. Importantes avances en esta dirección han sido realizadas por Geweke [24], quien encontró la técnica posición de una nueva serie de tiempo de descompresión que expresa el dominio del tiempo la causalidad de Granger en términos de su contenido de frecuencia.

A continuación, se establece el marco teórico de la CWG. Por simplicidad, la teoría se presenta solo para dos procesos aleatorios, que se puede extender a n -procesos. En nuestro enfoque, un proceso aleatorio corresponde a una lectura de señal asociada a un tipo de textura dentro de un escenario natural; por lo tanto, para el presente análisis, cada lectura de textura corresponde a un proceso estocástico representado por T_i , siendo i la textura i -ésima que tiene un comportamiento estocástico dispuesto en un escenario.

2.2.1. Causalidad en el dominio del tiempo: procesos estocásticos y modelo autorregresivo

A continuación de forma ilustrativa, se desarrolla un marco analítico para series de dos tiempos. El marco se puede generalizar para n -series asociadas a n -tiempos. Considere dos procesos estocásticos $X(t)$ y $Y(t)$, que son estacionarios de forma individual y conjunta. Bajo esta hipótesis, los procesos $X(t)$ y $Y(t)$ pueden representarse individualmente por sus modelos autorregresivos (VAR) como:

$$X(t) = \sum_{k=1}^{\infty} a_{1k} X(t-k) + \varepsilon_{1t}, \quad \text{con } \Sigma_1 = \text{var}(\varepsilon_{1t}), \quad (2.3)$$

$$Y(t) = \sum_{k=1}^{\infty} d_{1k} Y(t-k) + \eta_{1t}, \quad \text{con } \Gamma_1 = \text{var}(\eta_{1t}), \quad (2.4)$$

donde ε_{1t} y η_{1t} son ruidos de tipo gaussiano aleatorio con media cero y una desviación estándar dada, Σ_1 es la varianza de ε_{1t} , Γ_1 es la varianza de η_{1t} , a_{1k} y d_{1k} los coeficientes de regresión. Su modelo autorregresivo conjunto se define como:

$$X(t) = \sum_{k=1}^{\infty} a_{2k} X(t-k) + \sum_{k=1}^{\infty} b_{2k} Y(t-k) + \varepsilon_{2t}, \quad (2.5)$$

con $\Sigma_2 = \text{var}(\varepsilon_{2t})$, donde ε_{2t} es ruido gaussiano aleatorio con media cero y una desviación estándar dada, y Σ_2 es la varianza de ε_{2t} .

$$Y(t) = \sum_{k=1}^{\infty} c_{2k} X(t-k) + \sum_{k=1}^{\infty} d_{2k} Y(t-k) + \eta_{2t}, \quad (2.6)$$

donde $\Gamma_2 = \text{var}(\eta_{2t})$, Σ_2 y Γ_2 son las varianzas de los términos residuales, y a_{2k} , b_{2k} , c_{2k} , d_{2k} son los coeficientes de regresión. Sea Σ la matriz de varianza/covarianza de los términos residuales, definida como:

$$\Sigma = \begin{pmatrix} \Sigma_2 & \Upsilon_2 \\ \Upsilon_2 & \Gamma_2 \end{pmatrix} \quad (2.7)$$

donde Υ_2 es la covarianza entre ε_{2t} y η_{2t} ($\Upsilon_2 = \text{cov}(\varepsilon_{2t}, \eta_{2t})$). Si $X(t)$ y $Y(t)$ son independientes, entonces b_{2k} y c_{2k} son cero, por lo tanto $\Upsilon_2 = 0$, $\Sigma_1 = \Sigma_2$ y $\Gamma_1 = \Gamma_2$. De lo anterior, podemos definir la interdependencia estadística total entre $X(t)$ y $Y(t)$ como

$$F_{X,Y} = \ln \frac{\Sigma_1 \Gamma_1}{|\Sigma|} \quad (2.8)$$

donde $|\Sigma|$ es el determinante de Σ . Ecuación 2.8 es también definida como la causalidad del tiempo, y afirma que cuando ambas series de tiempo ($X(t)$ y $Y(t)$) son independientes, entonces $F_{X,Y} = 0$, y $F_{X,Y} > 0$ cuando ambas series de tiempo son dependientes. Como se aprecia en la Ecuación. 2.3, Σ_1 mide la precisión de la predicción autorregresiva de $X(t)$ basado en sus valores anteriores, mientras que Σ_2 en la Ecuación 2.5 representa la precisión de predecir $X(t)$ basada en sus valores anteriores de ambas $X(t)$ y $Y(t)$. De acuerdo a [27] y [49], si Σ_2 es menor que Σ_1 entonces $Y(t)$ se dice que tiene una influencia causal en $X(t)$. La cuantificación de la influencia causal se define por

$$F_{Y \rightarrow X} = \ln \frac{\Sigma_1}{\Sigma_2} \quad (2.9)$$

Se puede ver claramente que $F_{Y \rightarrow X} = 0$ cuando no hay influencia causal de $Y(t)$ a $X(t)$. La expresión $F_{Y \rightarrow X} > 0$ cuando hay una influencia causal.

2.2.2. Causalidad en el dominio de la frecuencia

Aquí se presenta el análisis de frecuencia de los modelos autorregresivos presentados en la subsección anterior. Comenzamos tomando el tiempo discreto de la transformada de Fourier (DTFT) de las ecuaciones. 2.5 y 2.6 para traducir datos de series temporales al dominio de frecuencia, lo que conduce a las siguientes expresiones:

$$X(w) - \sum_{k=1}^{\infty} a_{2k} X(w) e^{-jwk} - \sum_{k=1}^{\infty} b_{2k} Y(w) e^{-jwk} = E_x(w), \quad (2.10)$$

$$Y(w) - \sum_{k=1}^{\infty} d_{2k} Y(w) e^{-jwk} - \sum_{k=1}^{\infty} c_{2k} X(w) e^{-jwk} = E_y(w), \quad (2.11)$$

donde $X(w)$, $Y(w)$, $E_x(w)$, and $E_y(w)$ son los DTFT de $X(t)$, $Y(t)$, ε_{2t} y η_{2t} , respectivamente; j representa la unidad imaginaria, k es la variable de tiempo discreto, y w es la variable de frecuencia discreta. Después de agrupar términos comunes y hacer cuatro cambios variables, obtenemos:

$$X(w)a_2(w) + Y(w)b_2(w) = E_x(w), \quad (2.12)$$

$$X(w)c_2(w) + Y(w)d_2(w) = E_y(w), \quad (2.13)$$

donde $a_2(w) = 1 - \sum_{k=1}^{\infty} a_{2k} e^{-jwk}$, $b_2(w) = -\sum_{k=1}^{\infty} b_{2k} e^{-jwk}$, $c_2(w) = -\sum_{k=1}^{\infty} c_{2k} e^{-jwk}$, and $d_2(w) = 1 - \sum_{k=1}^{\infty} d_{2k} e^{-jwk}$.

Las Ecuaciones 2.12 y 2.13 pueden ser expresadas en forma vectorial como:

$$A(w) \cdot \begin{bmatrix} X(w) \\ Y(w) \end{bmatrix} = \begin{bmatrix} E_x(w) \\ E_y(w) \end{bmatrix} \quad \text{donde} \quad A(w) = \begin{bmatrix} a_2(w) & b_2(w) \\ c_2(w) & d_2(w) \end{bmatrix}.$$

Resolviendo para $X(w)$ y $Y(w)$ en el anterior sistema de ecuaciones tenemos:

$$\begin{bmatrix} X(w) \\ Y(w) \end{bmatrix} = A^{-1}(w) \cdot \begin{bmatrix} E_x(w) \\ E_y(w) \end{bmatrix}, \quad (2.14)$$

donde $A^{-1}(w)$ es la inversa de la matriz de $A(w)$. Porque $A^{-1}(w)$ expresa la relación entre las series de tiempo y sus respectivos residuos. También se conoce como la función de transferencia $H(w)$ del sistema lineal anterior. $A^{-1}(w)$ se puede encontrar usando cualquier método, y se puede expresar como:

$$A^{-1}(w) = H(w) = \begin{bmatrix} \frac{d_2(w)}{\det(A(w))} & \frac{-b_2(w)}{\det(A(w))} \\ \frac{-c_2(w)}{\det(A(w))} & \frac{a_2(w)}{\det(A(w))} \end{bmatrix}, \quad (2.15)$$

donde $\det(A(w))$ es el determinante de $A(w)$. Reescribiendo la Ecuación 2.14 en términos de $H(w)$ obtenemos:

$$\begin{bmatrix} X(w) \\ Y(w) \end{bmatrix} = H(w) \cdot \begin{bmatrix} E_x(w) \\ E_y(w) \end{bmatrix} \quad (2.16)$$

2.2.2.1. Matriz espectral $S(w)$

De las señales conjuntas $X(w)$ y $Y(w)$, la matriz espectral $S(w)$ se puede encontrar como:

$$S(w) = \begin{bmatrix} X(w) \\ Y(w) \end{bmatrix} \cdot \begin{bmatrix} X(w) \\ Y(w) \end{bmatrix}^* \quad (2.17)$$

donde (*) denota la transpuesta conjugada. Reescribiendo la Ecuación. 2.17 en términos de la Ecuación 2.16 tenemos,

$$S(w) = \left(H(w) \cdot \begin{bmatrix} E_x(w) \\ E_y(w) \end{bmatrix} \right) \cdot \left(H(w) \cdot \begin{bmatrix} E_x(w) \\ E_y(w) \end{bmatrix} \right)^* \quad (2.18)$$

Aplicando * la transpuesta conjugada al tamaño correcto, tenemos,

$$S(w) = \left(H(w) \cdot \begin{bmatrix} E_x(w) \\ E_y(w) \end{bmatrix} \right) \cdot \left(\begin{bmatrix} E_x(w) \\ E_y(w) \end{bmatrix}^* \cdot H(w)^* \right) \quad (2.19)$$

Entonces, agrupando los términos $E_x(w)$ y $E_y(w)$, tenemos,

$$S(w) = H(w) \cdot \left(\begin{bmatrix} E_x(w) \\ E_y(w) \end{bmatrix} \cdot \begin{bmatrix} E_x^*(w) & E_y^*(w) \end{bmatrix} \right) \cdot H(w)^*, \quad (2.20)$$

y multiplicando la parte central, se tiene:

$$S(w) = H(w) \cdot \left(\begin{bmatrix} E_x(w)E_x^*(w) & E_x(w)E_y^*(w) \\ E_y(w)E_x^*(w) & E_y(w)E_y^*(w) \end{bmatrix} \right) \cdot H(w)^* \quad (2.21)$$

Usando el teorema de Parseval, véase Apéndice 1.0.4, para la parte central, obtenemos,

$$S(w) = H(w) \cdot \Sigma \cdot H^*(w), \quad (2.22)$$

Donde $S(w)$ se puede expresar en términos de espectros automáticos para $X(w)$ y $Y(w) \rightarrow S_{XX}$ y S_{YY} , respectivamente, y la cruz de espectros S_{XY} y S_{YX} , como

$$S(w) = \begin{bmatrix} S_{XX} & S_{XY} \\ S_{YX} & S_{YY} \end{bmatrix} \quad (2.23)$$

De acuerdo a [16], la independencia entre $X(t)$ y $Y(t)$ puede ser expresada como

$$f_{X,Y}(w) = \ln \left[\frac{S_{XX}(w) \cdot S_{YY}(w)}{|S(w)|} \right], \quad (2.24)$$

donde $|S(w)| = S_{XX}(w) \cdot S_{YY}(w) - S_{XY}(w) \cdot S_{YX}(w)$ es el determinante de $S(w)$.

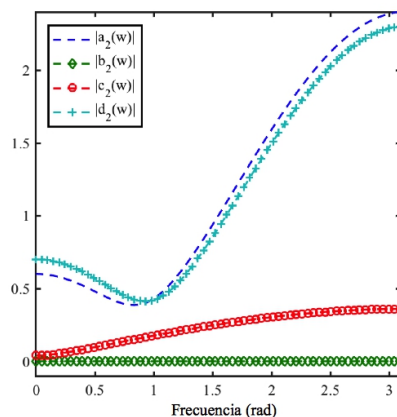


Figura 2.4: Transformada discreta de Fourier en el tiempo de coeficientes a_2 , b_2 , c_2 , y d_2 .

2.2.3. Ejemplo computacional

Para mostrar cómo calcular la función en la ecuación. 2.24, consideremos un ejemplo simple que consiste en dos procesos estocásticos modelados como:

$$\begin{aligned} X(t) &= 0.9X(t-1) - 0.5X(t-2) + \epsilon(t) \\ Y(t) &= 0.16X(t-1) - 0.2X(t-2) + 0.8Y(t-1) - 0.5Y(t-2) + \eta(t) \end{aligned} \quad (2.25)$$

donde $\epsilon(t)$ y $\eta(t)$ son dos procesos de ruido blanco correlacionados con matriz de covarianza

$\Sigma = \begin{pmatrix} 1 & 0.4 \\ 0.4 & 0.7 \end{pmatrix}$. Se puede observar una sola realización de los procesos estacionarios conjuntos en la Figura 2.5(a). Para calcular la interdependencia total $f_{X,Y}(w)$, necesitamos la matriz espectral $S(w)$ la cual depende de la función de transferencia $H(w)$, así, primero procedemos a calcular $H(w)$. Comparando Ecuación 2.25 con Ecuaciones 2.5 y 2.6, resulta que $a_{21} = 0.9$, $a_{22} = -0.5$, $b_{21} = 0$, $b_{22} = 0$, $c_{21} = 0.16$, $c_{22} = -0.2$, $d_{21} = 0.8$, $d_{22} = -0.5$. Una vez identificados los coeficientes del modelo VAR en el dominio del tiempo, procedemos a calcular su respuesta de frecuencia $a_2(w)$, $b_2(w)$, $c_2(w)$ y $d_2(w)$, como la respuesta al impulso de cuatro filtros cuya función de transferencia tiene los coeficientes $[1.00 \ -0.90 \ 0.50]$, $[0.000 \ 0.000 \ 0.00]$, $[0.00 \ -0.160 \ 0.20]$, y $[1.00 \ -0.800 \ 0.50]$, respectivamente. Tales respuestas de impulso se calcularon utilizando 1024 puntos en el rango $[0, \pi]$, y se pueden ver en la Figura. 2.4. Podemos calcular $H(w)$ como en la Ecuación 2.15, y su gráfica se muestra en la Figura 2.5(c). Ahora, $S(w)$ se puede calcular usando la ecuación. 2.24, de donde extrajimos $S_{XX}(w)$ y $S_{YY}(w)$, y su gráfica se muestra en la Figura 2.5(b); finalmente, $f_{X,Y}$ es calculada como se muestra en la Figura 2.5(d).

2.3. Algoritmos evolutivos

El llevar a las computadoras la famosa teoría de evolución de Darwin (publicada en *El origen de las especies basada en la selección natural* [12]) consiste en tratar de imitar, con programas de cómputo, la capacidad de una población de organismos vivos para adaptarse a su medio ambiente por medio de mecanismos de selección y reproducción. En los últimos cuarenta años varios métodos estocásticos

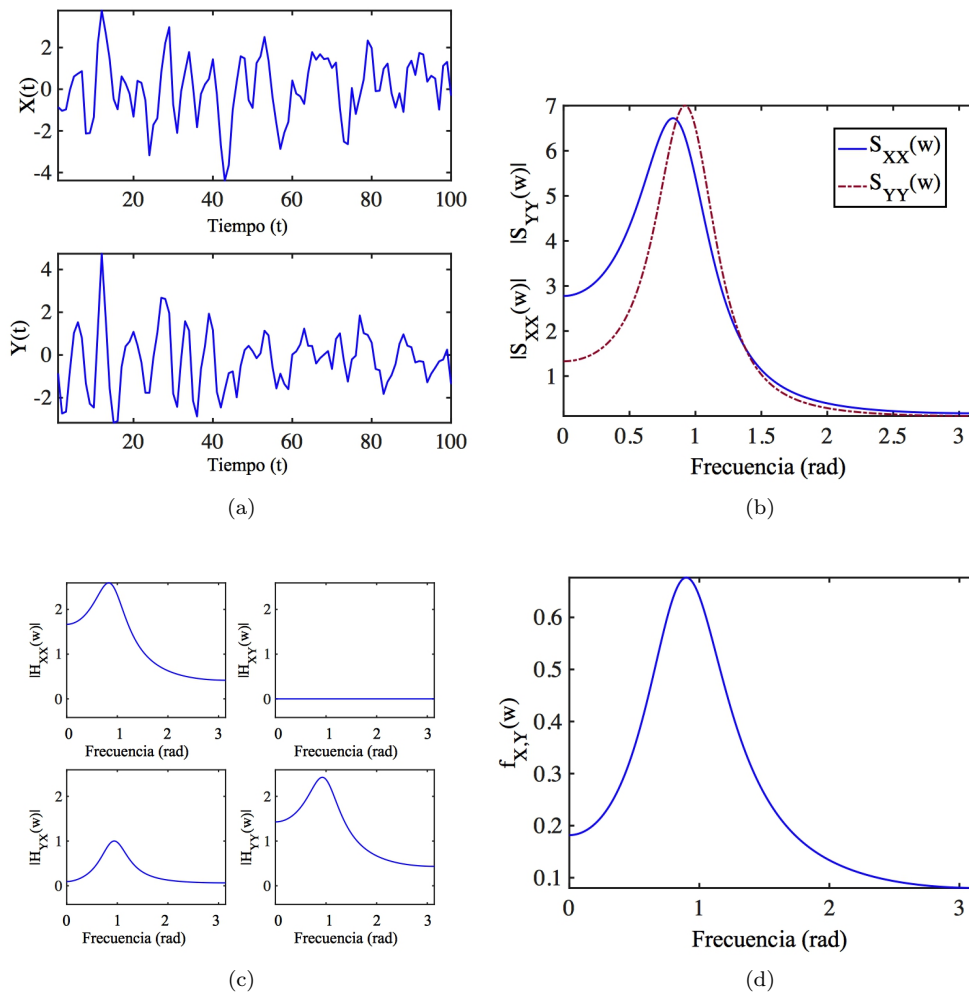


Figura 2.5: Un ejemplo simple que consta de dos modelos de procesos estocásticos, (a) Una sola realización de los procesos estacionarios conjuntos, (b) Espectros automáticos $S_{xx}(w)$ y $S_{yy}(w)$, (c) Función de transferencia $H(w)$ respuesta en el dominio de frecuencia, (d) Interdependencia total $f_{XY}(w)$ en el dominio de la frecuencia.

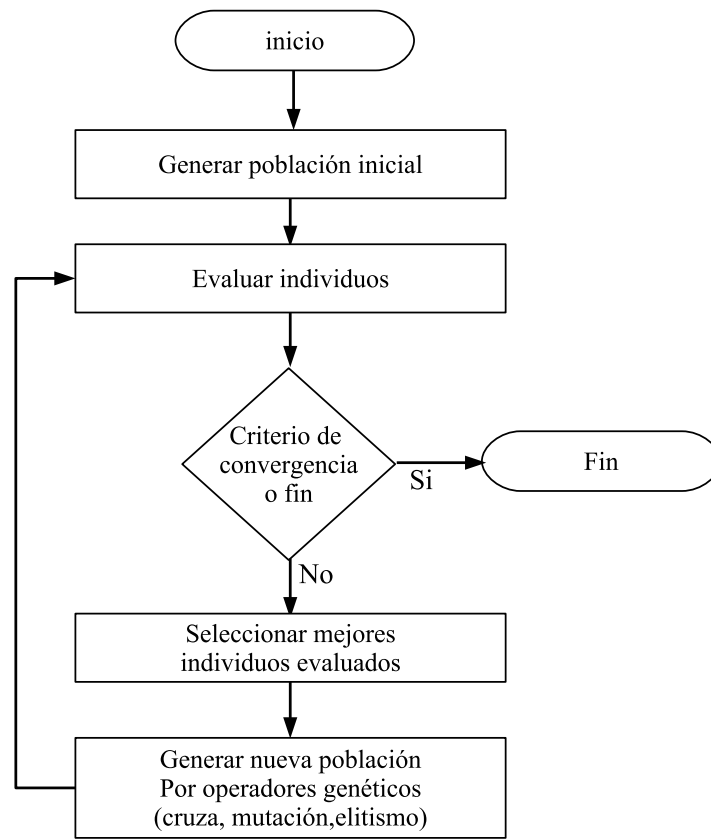


Figura 2.6: Diagrama general que describe un proceso evolutivo.

de optimización se han basado en este principio. *Darwinismo Artificial* o *algoritmos evolutivos* es el nombre común para estas técnicas, quizás la más comúnmente escuchadas sean *algoritmos genéticos*, *estrategias evolutivas* o *programación genética* [35, 25, 51, 45].

Los componentes comunes en estas técnicas son las poblaciones, que representan puntos muestra de un espacio de búsqueda, y que evolucionan bajo la acción de operadores estocásticos [54, 42, 47]. La evolución usualmente se organiza en generaciones y esto asemeja de una forma simple la genética natural, la Figura 2.6, muestra la generalidad de un proceso evolutivo. El motor de esta evolución está dada por:

1. Una selección, asociada a una medida de la calidad de un individuo respecto al problema a resolver.
2. Operadores genéticos, usualmente mutación y cruza o recombinación, que producen una nueva generación de individuos.

La eficiencia de un algoritmo evolutivo depende fuertemente de los parámetros establecidos: las poblaciones sucesivas (generaciones) tienen que converger hacia lo que se desea, esto es, hacia el óptimo global de una función de desempeño. Una gran parte de la búsqueda teórica en algoritmos evolutivos está dedicada al delicado problema de la convergencia, así también el tratar de focalizar qué tipo de problemas es fácil o difícil para un algoritmo evolutivo. La respuesta en teoría es que converge, pero

otras cuestiones prácticas que cuestionan, como la rapidez de convergencia, permanecen abiertas. Se puede ver que el interés en las técnicas evolutivas están suficientemente bien fundadas teóricamente, y de esto que justificadamente en los últimos cuarenta años se han podido aplicar con éxito en diferentes desarrollos experimentales [51, 31, 48].

Las técnicas evolutivas se suelen recomendar cuando los métodos clásicos y sencillos fallan para espacios de búsqueda muy grandes, variables mixtas, cuando se tienen varios óptimos locales, o cuando las funciones son muy irregulares. En otros problemas tales como los dinámicos o interactivos también pueden retomarse desde los algoritmos evolutivos, y finalmente estos métodos también pueden combinarse (hacerlos híbridos) con los métodos clásicos de optimización (e.g. gradiente descendente, búsqueda tabú)[39].

Se ha tenido mucho éxito en la implementación de estas técnicas en la industria [35, 25], en diversos y variados campos como lo es el dominio del análisis de imágenes [51, 42, 64, 31, 48, 47] y la visión robótica [45]. Algunos de estos trabajos son aplicados en algoritmos genéticos, y programación genética, con el fin de resolverlos de manera efectiva buscando soluciones dentro de un gran conjunto de posibilidades.

2.4. Marco teórico del procesamiento digital de imágenes

En esta sección se describen las bases del procesamiento digital de imágenes, y sus conceptos básicos que se manejan, así como lo que es el reconocimiento de patrones, estas técnicas son de las más importantes para el desarrollo de nuestro sistema de clasificación.

2.4.1. ¿Qué es la visión?

La visión la entendemos como el sentido de la vista, que se considera el más complejo de todos los sentidos, a pesar del conocimiento limitado sobre el funcionamiento interno del cerebro. Por lo tanto, la visión se considera como una de las fronteras finales de la ciencia, porque su comprensión se basa en descifrar cómo las células individuales dan lugar a operaciones complejas que crean toda la experiencia visual [28, 26, 40]. Por otro lado, el ojo es el órgano de la visión, o de la sensibilidad a la luz, que permite a los organismos asimilar información del medio ambiente. El ojo humano es un órgano que reacciona a la luz para varios propósitos; en combinación con el cerebro, se dice que exhibe un comportamiento consciente, debido a sus muchos componentes y funciones que permiten la conciencia del medio ambiente a través de la percepción y la visión de la luz con el objetivo de realizar tareas como el reconocimiento de objetos y la percepción de profundidad. Por lo tanto, el acto de ver comienza cuando el ojo enfoca una imagen de su entorno en una membrana sensible a la luz en la parte posterior del ojo, llamada retina. La Figura 2.7 proporciona un bosquejo simple de los componentes principales del ojo humano ¹.

Desde la antigüedad, los hombres han quedado asombrados por la impresión que deja el sentido de la vista, y el deseo de recrear ese proceso es indeleble en la historia de la humanidad. Hoy en día, tres escuelas diferentes de la comunidad científica investigan la visión. Los neurofisiólogos intentan comprender cómo funcionan los mecanismos sensoriales y neuronales de los sistemas biológicos. Los psicólogos perceptivos intentan comprender los problemas relativos a la conducta que rigen la tarea de

¹Imagen tomada de <https://tusdudasdesalud.com/vision/anatomia-ojo/las-partes-del-ojo>

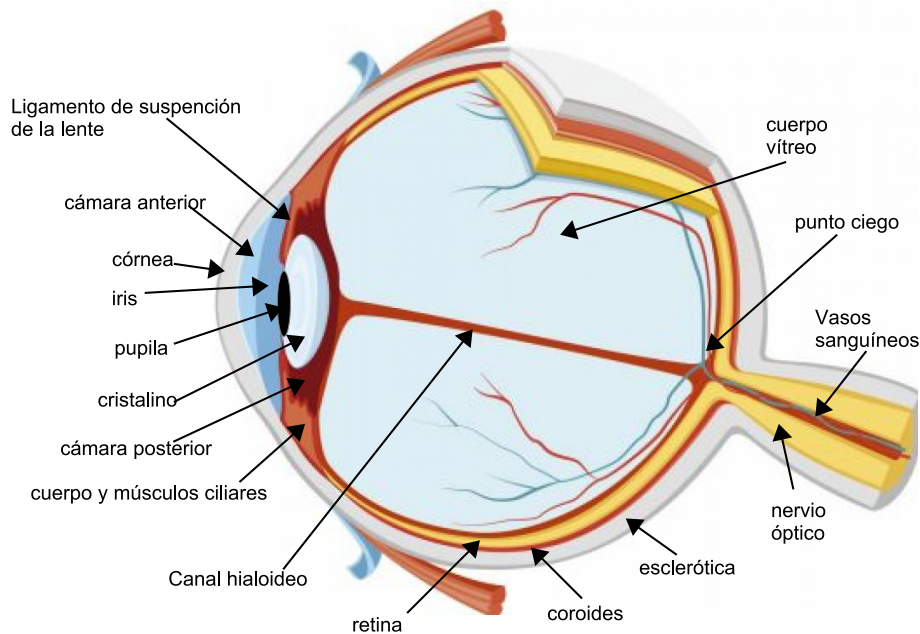


Figura 2.7: Diagrama de las partes del ojo humano.

la percepción, y los científicos de la visión por computadora investigan los problemas computacionales y algorítmicos asociados con la adquisición, procesamiento y comprensión de imágenes. El enfoque científico es importante desde la disciplina de las ciencias básicas, ya que implica las leyes de las matemáticas, la física y la biología, así como el conocimiento de su aplicación práctica. Sin embargo, el arte en el estudio de la visión no debe pasarse por alto. Los buenos resultados solo pueden producirse a partir de imágenes adecuadas, por lo que la pintura, la fotografía y la videografía son habilidades que no deben descartarse a la ligera [28, 57].

2.4.2. Formación de imágenes en el ojo

La formación de las imágenes en la retina se produce a partir de cuatro procesos: la refracción de la luz, la acomodación del cristalino, la constricción de la pupila y la convergencia de los ojos. La refracción de la luz es la desviación del rayo de luz cuando pasa de un medio a otro por diferentes densidades. En el ojo humano la luz tiene que pasar del aire a cuatro medios de diferentes densidades que son: la córnea, el humo acuoso, el cristalino y el humo vítreo antes de llegar a la retina. Cuando la imagen está lejos del observador, los cuatro puntos de refracción de la luz en el globo ocular funcionan en conjunto y tratan de llevar la imagen al centro o fovea de la retina, donde la agudeza visual es mayor. En el caso contrario, es el cristalino que se encarga de realizar estos procedimientos de refracción, cambiando su forma de cóncava a convexa según convenga. Este proceso que se da en el cristalino se llama acomodación. Los músculos del iris también cumplen la función de la formación de la imagen en la retina, pues son ellos quienes realizan la constricción de la pupila, para que los rayos de luz no estimulen súbita o prolongadamente a la pupila. La constricción de la pupila es el proceso simultáneo a la acomodación del cristalino. Finalmente, la convergencia de los ojos en el enfoque de

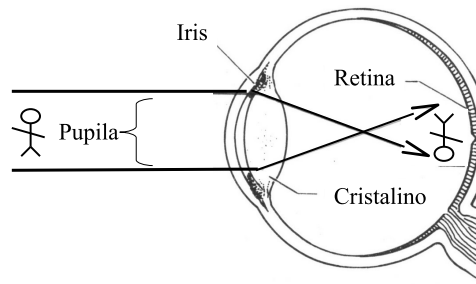


Figura 2.8: Representación óptica de un ojo mirando un objeto

un grupo de objetos o visión binocular, consiste en el movimiento de los globos oculares para que la luz proyectada por los objetos llegue a puntos idénticos en ambas retinas. Solo así, percibimos un objeto tridimensional y como un conjunto. Una vez dados estos cuatro procesos, la retina construye una imagen invertida Figura² 2.8, producto de la refracción de la luz en los medios transparentes. Sin embargo, nuestro cerebro aprende, desde el primer año de vida de los individuos, a interpretar la imagen en su localización y posición exacta.

La imagen formada en la retina produce una estimulación en los pigmentos de la cornea y los bastones y transforman la luz en impulsos nerviosos que las neuronas bipolares transmiten a su vez a las neuronas ganglionares, cuyos axones salen del globo ocular a través del nervio óptico. El nervio óptico es un conjunto de ramas nerviosas ubicadas en cada globo ocular. Estas ramas se toman en un punto del encéfalo denominado quiasma óptico; en el cual algunas de las fibras se cruzan al lado opuesto, mientras que otras no los hacen. Después de atravesar el quiasma, las fibras forman el tracto óptico (cintilla óptica), llegan al encéfalo y terminan en el tálamo; en este último, establecen sinapsis con neuronas de tercer orden, cuyos axones se dirigen en los centros visuales de los lóbulos occipitales de la corteza cerebral.

2.4.3. Espectro visible para el ojo humano.

El espectro visible de luz es el espectro de radiación electromagnética que es visible para el ojo humano. Va desde una longitud de onda de 400nm hasta 700nm . Además, también se conoce como: *el espectro óptico de la luz*.

Estas son entonces las ondas que componen lo que llamamos *luz visible*. Cuando estamos viendo un objeto, es porque ese objeto está siendo iluminado por la luz visible. Por otra parte, cuando vemos que el cielo es de color azul, que el pasto es de color verde o que el cabello de alguien es de color negro, es porque en ese momento estamos recibiendo diferentes longitudes de onda en la banda de los 400 nm y los 700 nm .

Aunque los seres humanos estemos limitados a la banda visible del espectro electromagnético (EM), las máquinas pueden percibir casi el espectro completo, desde los rayos gamma, a las ondas de radio, como lo podemos ver en la Figura³ 2.9.

²Imagen tomada de <http://wwwedplasticamayalen.blogspot.com/2015/10/la-camara-oscura-el-funcionamiento-de.html>

³Espectro electromagnético: <http://www.paz-digital.net/imagenes/haarp/haarp004.png>

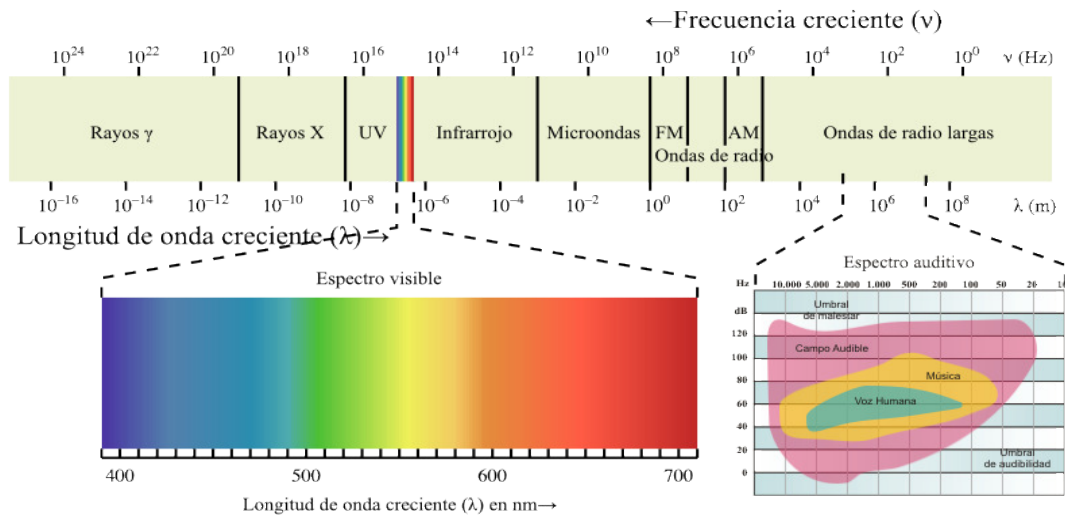


Figura 2.9: Espectro electromagnético

Los autores de muchos artículos y libros aun no se ponen de acuerdo para decir dónde termina el campo del Procesamiento Digital de Imágenes y dónde empiezan otros campos como el Análisis de Imágenes y la Visión por Computadora. La tarea de esta última disciplina es utilizar computadoras y dispositivos físicos para emular la visión humana, incluyendo el aprendizaje, hacer inferencias y actuar basándose en entradas visuales. La Visión por Computadora, es claramente, un área de la Robótica en la que los programas intentan identificar objetos representados en imágenes digitalizadas proporcionadas por cámaras de video, permitiendo así que los robots “vean” ⁴ [57].

2.4.4. Imagen.

Una imagen puede ser definida matemáticamente por una función bidimensional $f(x, y)$, donde x y y son coordenadas espaciales en un plano, y f es cualquier par de coordenadas es la intensidad o nivel de gris de la imagen en esa coordenada $\forall x, y \in R$ [26].

Imagen digital

Cuando x, y, y los valores de f son todas cantidades finitas, discretas, tal que $x, y \in \mathbb{N}$ y $f(x, y) \in \mathbb{N}$ si $0 \leq f \leq L - 1$, decimos que la imagen es una *imagen digital* como se ve en la Figura 2.10. Descrito por [26], una imagen digital se compone de un número finito de elementos, cada uno con un lugar y valor específicos.

2.4.5. Píxel

El píxel es la unidad de medida más pequeña de video que tiene una luminosidad y color específico [57]. Sus proporciones están limitadas por el número de líneas que componen la trama de barrido (el patrón de puntos que forman la imagen) y la resolución a lo largo de cada línea. En la forma más común de gráficas por computadora, los miles de diminutos píxeles que componen una imagen

⁴“Computer vision”. Encyclopedia Britannica.2019. Web. 19 Nov. 2019. <https://www.britannica.com/technology/computer-vision>



Figura 2.10: Ejemplo de imagen continua (izquierda) vs imagen digital (derecha), siendo ésta última una aproximación discreta en puntos dispuestos en formación matricial, y valores de frecuencia

individual se proyectan en una pantalla de visualización en forma de puntos iluminados que desde la aparecen como una imagen continua. Un haz de electrones crea la cuadrícula de píxeles mediante el trazado de cada línea superior a la línea de fondo. Un píxel también puede ser el elemento más pequeño de un dispositivo sensible a la luz, como las cámaras que utilizan dispositivos de carga acoplada⁵. El término píxel también conocido como elemento de imagen o “picture element”, y en otra partes es conocido por muchos como la “célula menor”.

2.4.6. Cuantización y muestreo de la imagen digital

1. *Resolución espacial.* Tiene que ver con el tamaño de la imagen Figura 2.11. Usualmente decimos que una imagen digital de tamaño $M \times N$ con L niveles de gris tiene una resolución espacial de $M \times N$ píxeles y una resolución de nivel de gris de L niveles [26]. e.g. Tenemos los estándares de tamaño en resolución de pantallas VGA (640×480), XGA (1024×768), etc. véase la Figura⁶ 2.12.
2. *Cuantización del tono.* Se refiere al más pequeño cambio discernible en nivel de gris aunque, medir los cambios discernibles en niveles de intensidad es un proceso altamente subjetivo véase la Figura 2.13. Este tiene que ver con la discretización del nivel de gris asociado a cada punto de la imagen. La potencia de 2 que determina la cuantización, el número de niveles de gris, es usualmente 8 bits, es decir, 256 diferentes niveles de grises. Algunas aplicaciones especializadas utilizan 16 bits, esto depende de la aplicación ya que si necesitamos obtener muchos rasgos o información, es necesario tener mayor detalle del objeto. Veamos la tabla 2.1:

⁵ “píxel”. Encyclopædia Britannica.2019. Web. 19 Nov. 2019. <https://www.britannica.com/technology/pixel>

⁶ Imagen de referencia tomada de <http://www.intermediarte.org/wp-content/uploads/2009/09/screenresolutions.jpgtext>

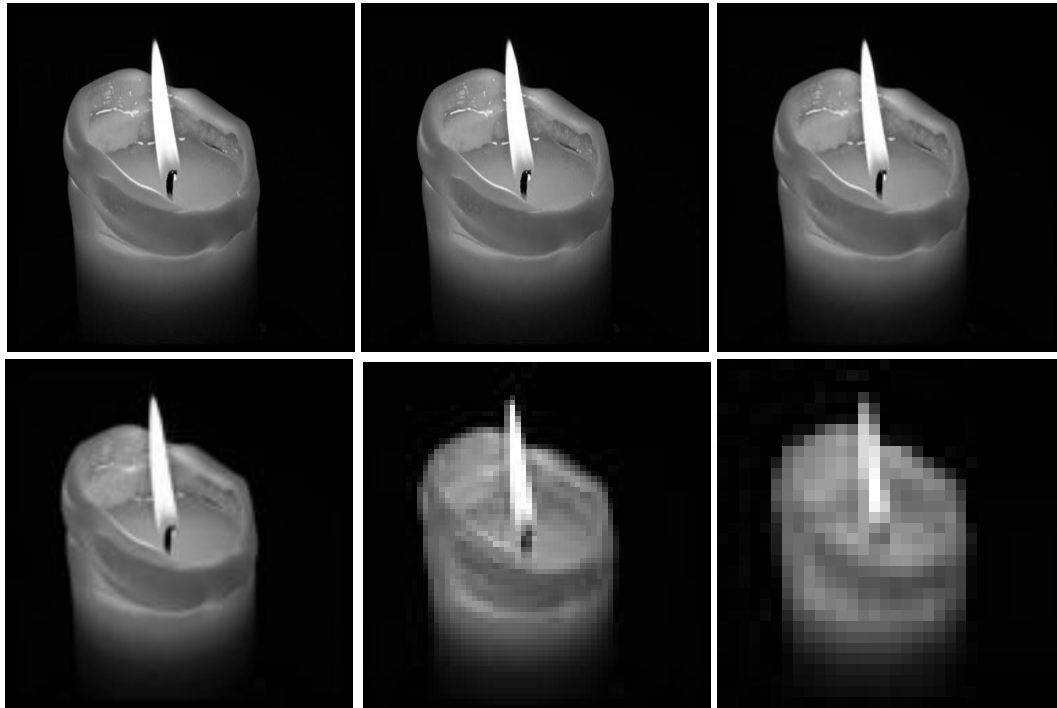


Figura 2.11: Imagen de 1024x1024 original y sus submuestreos de 512x512, 256x256, 128x128, 64x64 y 32x32.

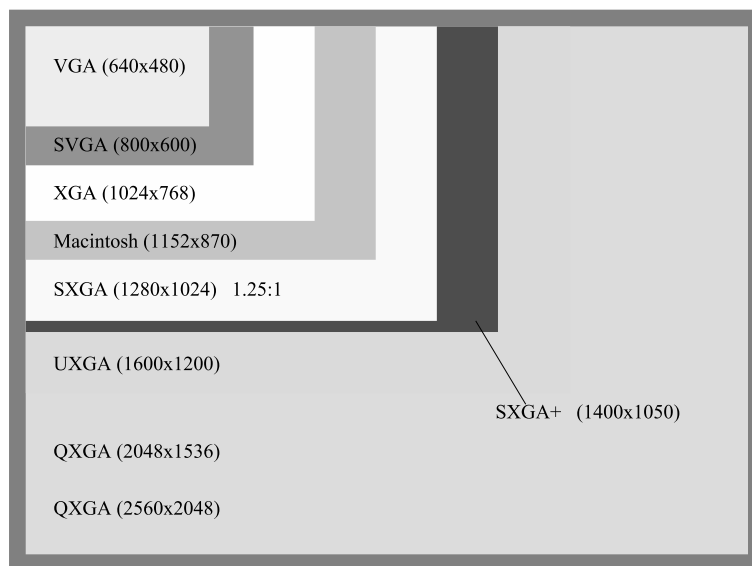


Figura 2.12: Tamaños estándar en resolución.

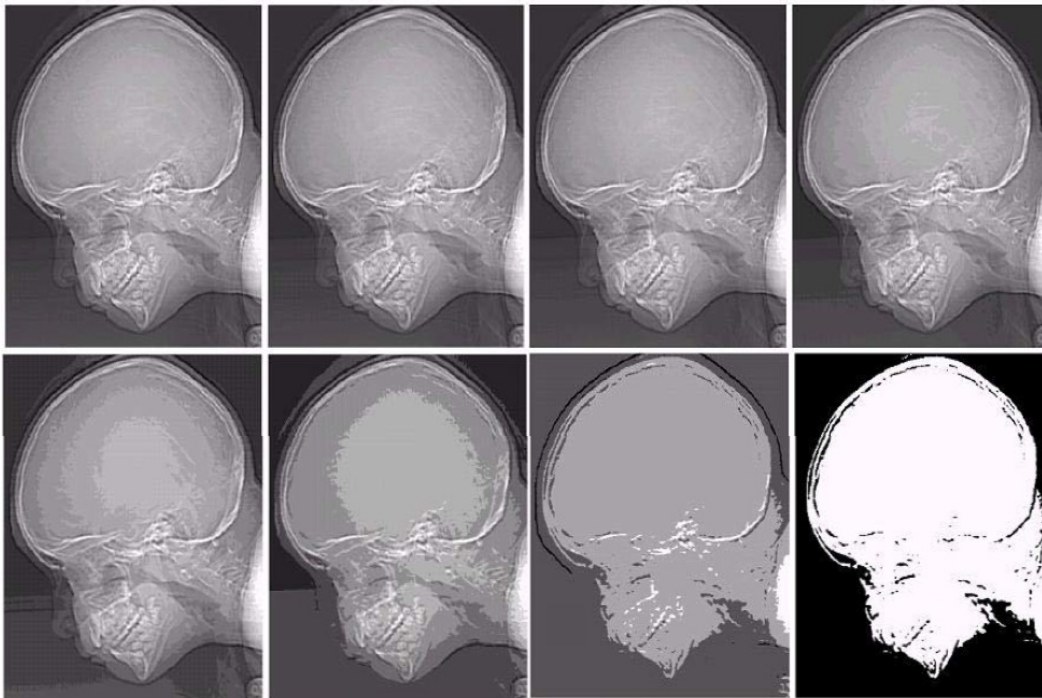


Figura 2.13: Imagen de 472 x 374 píxeles de resolución, en 256, 128, 64, 32, 16, 8, 4 y 2 niveles de gris [26].

Tabla 2.1: Tabla de valores potencia de 2.

Número de bits	potencia	valores
1	2^1	2
2	2^2	4
4	2^4	16
8	2^8	256

2.5. Procesamiento digital de imágenes

El procesamiento de imágenes es el conjunto de técnicas computacionales para el análisis, la mejora, la compresión y la reconstrucción de imágenes. Sus principales componentes son importadores, en el que se captura una imagen a través de la exploración o la fotografía digital; el análisis y la manipulación de la imagen, se realiza utilizando diversas aplicaciones de software especializados; y de salida (por ejemplo, a una impresora o monitor). El procesamiento de imágenes tiene amplias aplicaciones en muchas áreas, incluyendo la astronomía, la medicina, la robótica industrial, y la teledetección por satélites⁷.

Para el procesamiento digital de imágenes, la segmentación de imágenes es una de las principales problemáticas a resolver, permitiendo la extracción de contornos, texturas y regiones presentes y

⁷“image processing”. Encyclopedia Britannica.2019. Web. 19 Nov. 2019. <https://www.britannica.com/technology/image-processing>

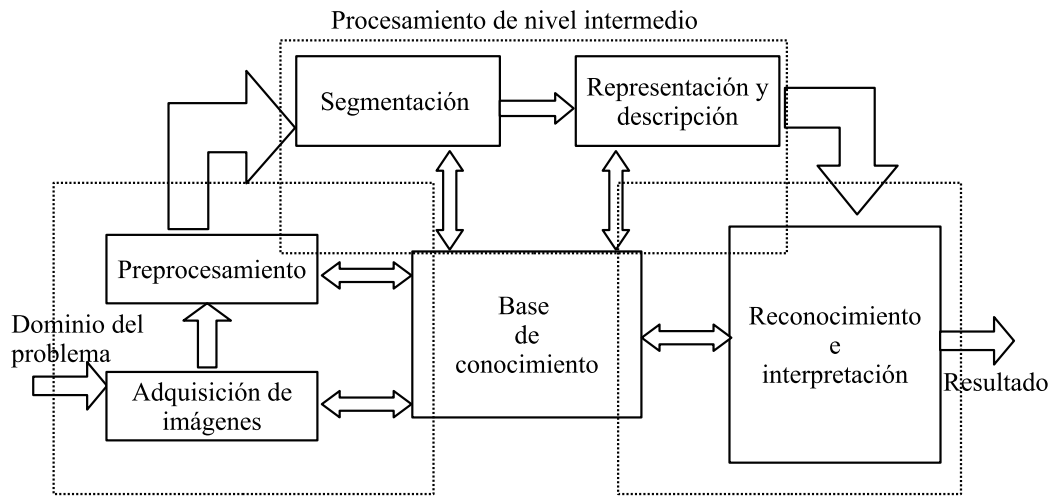


Figura 2.14: Elementos del análisis de Imágenes desde el PDI [26].

facilita etapas posteriores del procesamiento relacionadas con la representación e interpretación de la información contenida. Existen diferentes formas de realizar la segmentación de una imagen. El enfoque basado en reconocimiento de patrones consiste en el uso secuencial de:

- Técnicas de extracción de características, que identifican o describen numéricamente píxeles o regiones de la imagen.
- Técnicas de reconocimiento de patrones, que permiten separar regiones en la imagen, identificando las mismas según sus diferentes características.

Una gran cantidad de algoritmos ha sido desarrollada para la extracción de características y reconocimiento de patrones, con enfoques y criterios muy variados. Cada uno se basa en un determinado paradigma de análisis del problema y un conjunto de parámetros que lo definen. La gran cantidad de parámetros involucrados en el uso de cada técnica hace a menudo difícil la optimización del algoritmo que se desea utilizar. Frecuentemente, además, es necesaria la combinación de más de una técnica para una tarea específica de segmentación.

Nuestro sentido más avanzado es la vista, y no es sorprendente que las imágenes jueguen el papel más importante en la percepción humana de su entorno, y a partir de ello conforma muchas de sus tareas. La Figura 2.14 muestra un sistema del procesamiento digital del imágenes de propósito general propuesto en [26], el cual es el que se toma de referencia para todo el procesamiento de la imagen digital. A continuación vemos los bloques:

- **Adquisición de imágenes.** Generalmente incluye preprocesamiento e.g. cambiar el tamaño la imagen.
- **Preprocesamiento.** Después de que la imagen digital ha sido obtenida, el siguiente paso es el preprocesamiento. Su función es mejorar la imagen de manera que se incremente la oportunidad de éxito de los siguientes procesos. El pre-procesamiento casi siempre trata con técnicas para realzar el contraste y eliminar cualquier elemento de ruido.

- **Segmentación.** Definida de forma amplia, la segmentación particiona una imagen de entrada en sus partes constituyentes u objetos. En general, la segmentación automática es una de las tareas más difíciles en el procesamiento digital de imágenes. Un proceso de segmentación robusto lleva el proceso un largo camino hacia la solución satisfactoria de los problemas de imagen que requieren los objetos para ser identificados individualmente. Por otro lado, los algoritmos de segmentación débiles o críticos casi siempre garantizan el fracaso final. En general, cuanto más precisa es la segmentación, el reconocimiento más probable es tener éxito.
- **Segmentación y Descripción.** La segmentación y la descripción casi siempre siguen después de una etapa de segmentación, que por lo general son píxeles en formato RAW⁸, que constituye ya sea el límite de una región (es decir, el conjunto de píxeles que separan una región de imagen de otra) o todos los puntos en la misma región. Descripción, también llamada función de selección, se ocupa de extraer los atributos de la información, los cuales resultan en alguna información cuantitativa de interés o son básicos para la diferenciación de una clase de objetos de otro.
- **Reconocimiento e interpretación.** El reconocimiento es el proceso que etiqueta, o asigna un nombre, a un objeto basándose en la información que proveen sus descriptores. La interpretación involucra la asignación de significado a un conjunto de objetos reconocido. En términos de nuestro ejemplo, identificar un rostro de una persona, requiere la asociación de los descriptores para este *rostro* con la etiqueta *R1* por ejemplo.
- **Base de Conocimiento.** La base de conocimiento mantiene toda la información (conocimiento) que se tiene acerca del problema. Este conocimiento puede ser simple, teniendo solo las regiones de la imagen donde la información de interés se sabe que esta localizada, con lo que se reduce la búsqueda que guía el camino a esa información. Además de guiar la oración de cada modulo del proceso, la base del conocimiento también controla la interacción entre módulos. Como se representa con flechas de doble sentido en la Figura 2.14.

Es importante mencionar que los resultados del procesamiento pueden verse en la salida de cualquier paso de la Figura 2.14 descritas en [26]. También debe hacerse notar que no todas las aplicaciones del procesamiento digital de imágenes requieren la complejidad de interacciones de la Figura 2.14. Muchas aplicaciones caen fuera de este esquema. De hecho, no todos los módulos, son siempre necesarios. Por ejemplo, el realce de imágenes para interpretación visual humana rara vez va más allá de la etapa de pre-procesamiento. En general, las funciones de pre procesamiento que incluyen reconocimiento e interpretación están asociadas con aplicaciones de análisis de la imagen, en la cual el objetivo es extraer las características o la información necesaria de una imagen de manera automática o semiautomática.

2.5.1. Textura

La textura en este trabajo es una de las partes más importantes ya que dentro del análisis de imágenes digitales es necesario tener una idea de lo que significa la “textura”. Desde el PDI es un

⁸El formato de imágenes RAW (entiéndase como “formato de imagen sin modificaciones”) es un formato de archivo digital de imágenes que contiene la totalidad de los datos de la imagen tal y como ha sido captada por el sensor digital de la cámara fotográfica.

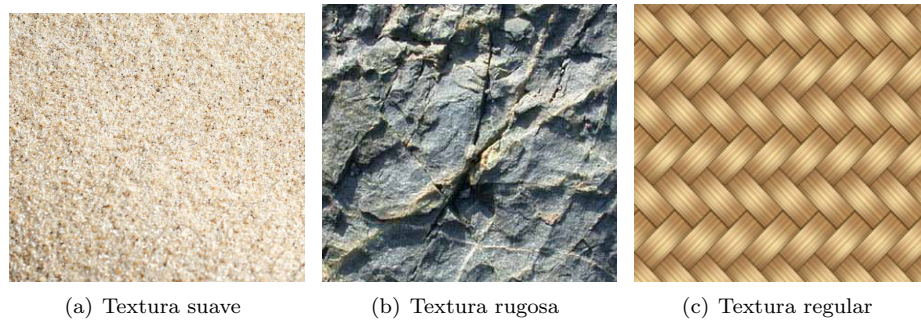


Figura 2.15: Ejemplo de texturas.

método descripción de regiones el cual consiste en la cuantificación de su contenido. Este descriptor proporciona intuitivamente medidas de propiedades tales como suavizado Figura 2.15(a), rugosidad Figura 2.15(b), regularidad Figura 2.15(c), entre otras. Existen tres métodos principales en el PDI para describir la textura de una región, los estadísticos, estructurales y espectrales. Las soluciones estadísticas nos dan características de textura como suavidad, rugosidad, granulosidad, y algunas más similares como se ve en la Figura 2.15. Las estructurales tratan de la composición de primitivas de imágenes, tales como la descripción de texturas basadas en líneas paralelas regularmente espaciadas. Finalmente las técnicas espectrales están basadas en las propiedades del espectro de Fourier y son utilizadas primordialmente para detectar periodicidad global de una imagen mediante la identificación de picos estrechos de alta energía del espectro [26].

2.6. Reconocimiento de patrones

El reconocimiento de patrones es la ciencia que se encarga de la descripción y clasificación (reconocimiento) de objetos, personas, señales, representaciones, etc [7, 13]. Esta ciencia trabaja con base en un conjunto previamente establecido de todos los posibles objetos (patrones) individuales a reconocer [22].

Con el paso de los años se han diseñado e implementado algoritmos para simular la habilidad humana de reconocer objetos. A la par con el desarrollo de métodos capaces de emular las diferentes formas de reconocimiento de patrones, ha evolucionado la construcción de sistemas automáticos dotados de cierta inteligencia, que procesan la información de los objetos. Las tareas en reconocimiento de patrones pueden ser descriptivas o predictivas. Las tareas descriptivas son para caracterizar el conjunto de datos de cierta clase y las tareas predictivas son para asignar una etiqueta clase a una muestra a partir de un conjunto de muestras cuya clase es conocida [22].

2.6.1. Patrones

La representación que describe a un objeto se le llama patrón [13]. Existen dos tipos de patrones: abstractos y concretos, los patrones abstractos son aquellos que no podemos visualizar por decir las ideas, pensamientos, estados de ánimo, etc., y los patrones concretos son aquellos que podemos visualizar, por ejemplo, símbolos, letras, fotografías, firmas, rostros, señales auditivas, etc. [7]. Un patrón esta

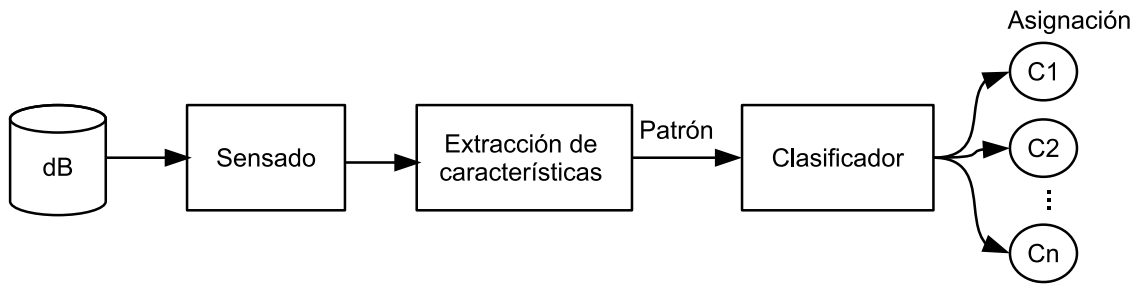


Figura 2.16: Elementos de un sistema de reconocimiento de patrones.

formado por un conjunto de n valores o medidas para las características o atributos que representan a un objeto. Dependiendo del dominio del problema, un patrón puede ser tratado como un vector de características en un espacio n -dimensional, como se puede ver la ecuación 2.26; o como una cadena de símbolos, como se puede ver en la ecuación 2.27, donde S es el conjunto de todas las cadenas formadas a partir de los símbolos. El conjunto de los valores posibles para las características de un patrón es el espacio de características o espacio de representación.

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_3 \end{bmatrix} \in X \subset \mathbb{R} \quad (2.26)$$

$$x = x_1, x_2, \dots, x_n \subset S \quad (2.27)$$

El reconocimiento de patrones se enfoca de dos maneras: supervisado y no supervisado. En el supervisado se cuenta con información de la categoría a la que pertenecen las muestras de un conjunto de entrenamiento. En el no supervisado no contamos con la información referente a las clases.

2.6.2. Elementos de un sistema de reconocimiento de patrones

En la Figura 2.16, se describen en general las etapas de un sistema de reconocimiento de patrones, el objetivo de estas etapas es ajustar el sistema para que sea capaz de clasificar señales u objetos de entrada en una de las clases predefinidas [7, 13]. Para ello deberá analizar un cierto número de características y para poder clasificar satisfactoriamente señales de entrada, es necesario un proceso de aprendizaje en el cual el sistema crea un modelo de cada una de las clases a partir de una secuencia de entrenamiento o conjunto de vectores de características de cada una de las clases.

Como puede verse en el diagrama de la Figura 2.16, el sistema de reconocimiento de patrones de manera general se divide en tres etapas:

1. *Sensado*: El objetivo de esta etapa es convertir y representar los datos obtenidos para su futura clasificación. Generalmente el costo de este depende de la calidad del sensor o electrodo ocupado,

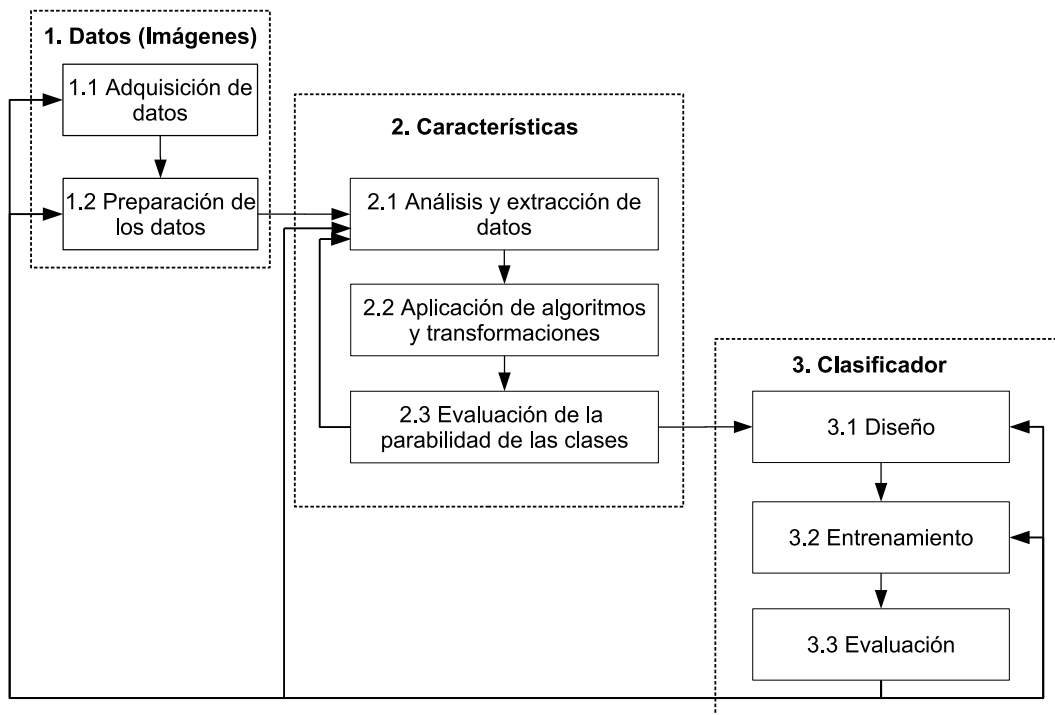


Figura 2.17: Diseño de un sistema de reconocimiento de patrones.

por lo tanto entre mayor calidad se tenga mayores características tendrá, por ejemplo menor ruido o mayor fiabilidad en las señales obtenidas.

2. *Extractor de características:* A partir de la señal obtenida por el sensor, a este se le extraerá información conocida como características relevantes y así eliminará redundancias e información irrelevante.
3. *Clasificador:* de acuerdo a las características extraídas en la etapa anterior, el clasificador asignará una categoría apropiada a los patrones obtenidos.

2.6.3. Diseño de un sistema de reconocimiento de patrones

En la Figura 2.17 se muestran las 3 etapas de diseño que incluye una serie de tareas que se deben realizar para su construcción. Contiene las etapas de: datos, características y clasificación.

El proceso inicia con la adquisición y preparación de los datos (tarea 1.1). En la tarea (1.2), preparación de los datos, se plantea una estrategia de pre-procesamiento, que puede ser de diferente índole dependiendo del tipo de patrones, con el objetivo de limpiar, seleccionar e integrar los datos para su procesamiento.

En la etapa de características se realiza primero la tarea (2.1), análisis de datos con el objetivo de plantear una estrategia de extracción de características. Posteriormente se aplican algoritmos de selección y transformación en la tarea (2.2), este procesamiento en el dominio de las características es con dos objetivos: incrementar la separabilidad de las clases y reducir la dimensión de los patrones. A

continuación se realiza la tarea (2.3), una evaluación de la separabilidad de las clases, en caso de no obtenerse los resultados deseados, podemos replantear la estrategia de formación de características o de procesamiento en el dominio de las características. La etapa termina cuando se consigue un conjunto de características con un criterio de separabilidad conveniente [22].

La última etapa es para la construcción del clasificador. En la tarea (3.1) se selecciona y se configura un algoritmo de clasificación. Posteriormente en la tarea (3.2), se entrena el clasificador, esta tarea depende del algoritmo seleccionado y consiste en utilizar un conjunto de datos, llamado conjunto de entrenamiento, para especificar el clasificador. Finalmente una evaluación del clasificador en la tarea (3.3), si el resultado de la evaluación no es satisfactorio, hay diferentes opciones de retorno en el proceso de diseño, se puede cambiar la estrategia de evaluación, el conjunto de datos de entrenamiento, el tipo de clasificador utilizado o regresar a las etapas anteriores para replantear la estrategia de formación de características o de adquisición y preparación de los datos.

Capítulo 3

Metodología

En este capítulo, describimos la metodología desarrollada para la técnica de CWG con ayuda de un AG aplicado a imágenes de escenarios naturales.

Para el uso de la metodología CBIR existen diferentes factores determinantes que deben tenerse en cuenta al extraer la información de las imágenes, como lo son luminosidad, la orientación, la escala, la homogeneidad, etc. La característica principal en nuestros patrones propuestos es la textura local, de modo que tratamos de crear un diccionario base para luego construir las series de tiempo a partir de la lectura de las imágenes y la comparación con de estas con el diccionario, con el cual se aplicó la teoría de CWG.

3.1. Desarrollo de la arquitectura del sistema

Para la construcción del diccionario se seleccionan un conjunto de k texturas sobre las imágenes a estudiar a las que llamaremos *texturas de referencia*. Las k texturas generadas representan partes de objetos dentro de las imágenes como cielo, nubes, vegetación, roca, etc., tratando de hacer una segmentación manual de las escenas como se muestra en la Figura 3.1.

Una vez obtenido el conjunto de las k texturas de referencia, los valores en el espacio de color HSI, de cada una de ellas se analizan para generar un rango de valores máximos y mínimos que les represente, estos valores nos ayudan a determinar los umbrales de comparación para las texturas de prueba de una imagen de consulta, tal como se muestra en el capítulo de pruebas en la Tabla 5.1.

La metodología propuesta para la identificación y clasificación de escenas por CWG se muestra en la Figura 3.2. A continuación se describen las etapas de la arquitectura.

1. **Base de datos de escenarios naturales (BDEN).** Representa el conjunto de las imágenes a considerar, contiene las imágenes de los escenarios naturales.
2. **Lectura de la imagen.** Se encarga de extraer las imágenes de la base de datos, que serán procesadas en formato de color RGB.
3. **Procesamiento de la imagen.** Realiza el pre-procesamiento de las imágenes, con un ecualizado del histograma y después, se aplica el algoritmo de corrimiento de media presentado en [55] para



Figura 3.1: Ejemplo de selección de las texturas de referencia con segmentación manual.

dejar la imagen como se muestra en la Figura 3.3, con esto la imagen queda preparada para ser utilizada en la siguiente etapa.

4. **Cambio de espacio de color RGB a HSI.** El espacio de color RGB no nos proporciona la información necesaria para la extracción de características ya que para el análisis del contenido visual buscamos caracterizar información no susceptible del color, de este modo pasamos al espacio de color HSI, tal que nos proporciona la información relacionada con la textura.
5. **Extracción de características.** Este bloque consta de tres etapas importantes:
 - **Emparrillado de la imagen.** Tomando como referencia el trabajo realizado en [29], se realiza un emparrillado regular de la imagen de 10×10 ventanas, para el análisis de textura CBIR; en nuestra propuesta usamos un emparrillado de $r \times c$ ventanas, tal que se tiene la posibilidad $r \neq c$, donde $c := \text{número de ventanas en horizontal (columnas)}$, y $r := \text{número de ventanas en vertical (filas)}$.
 - **Construcción de la vecindad.** En cada uno de los cuadros resultantes del emparrillado se extrae una vecindad de tamaño $p \times p$ píxeles, a partir del vértice superior izquierdo de cada región, como se observa en la Figura 3.4, tal que $p < r \wedge p < c$.
 - **Extracción de características CBIR.** La imagen es leída a partir de las vecindades de la siguiente forma: nos colocamos en la esquina superior izquierda de la imagen, se crea una vecindad de tamaño $p \times p$ píxeles desde la esquina superior derecha de la ventana y se extrae la información CBIR, y nos desplazamos siguiendo en orden vertical descendente de las vecindades, leyendo cada una de ellas. Una vez que se llega al final de los renglones, nos movemos a la vecindad derecha y subimos hasta el primer renglón, después nos movemos a la vecindad derecha y bajamos nuevamente, esta lectura se repite para toda la imagen hasta llegar a la última rejilla, como se muestra en la Figura 3.5. Cada lectura de la vecindad genera un patrón de tamaño 1×3 , i.e. una característica por canal de la imagen. Finalizada la extracción de características de todas las vecindades leídas en el orden establecido, se tendrá

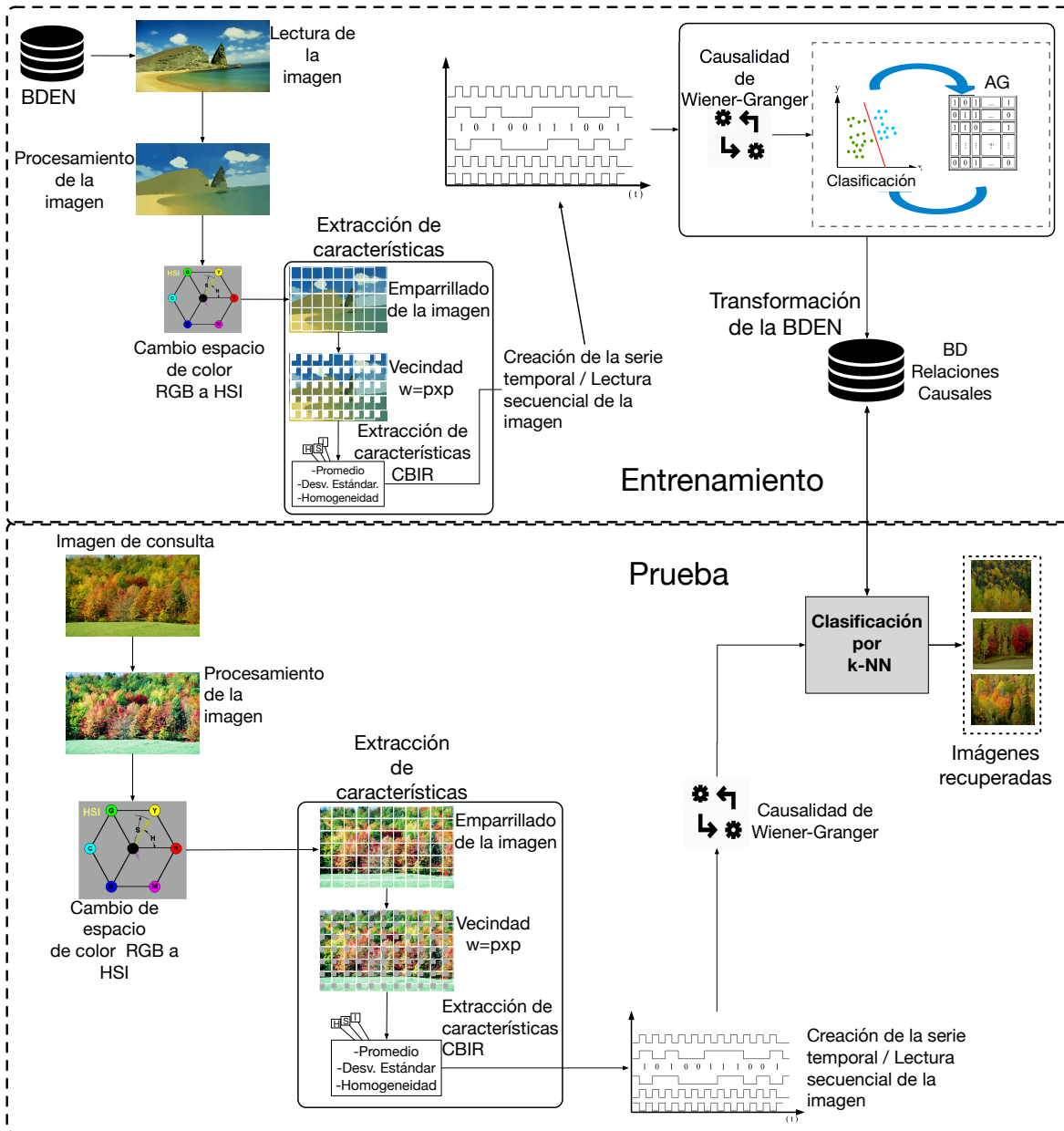


Figura 3.2: Arquitectura del sistema de clasificación.



Figura 3.3: Etapas del acondicionamiento de la imagen

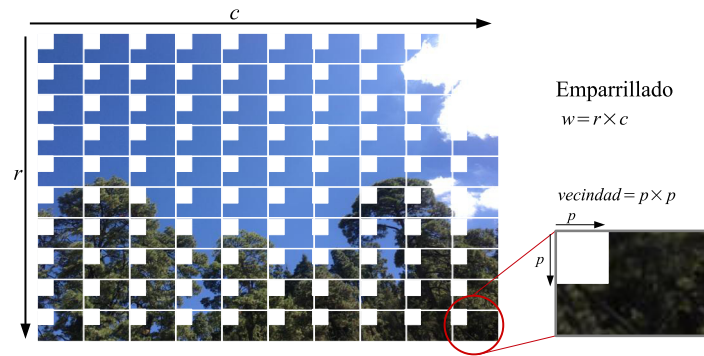


Figura 3.4: Ejemplo de un emparrillado regular $w = 10 \times 10$, con vecindades dentro de cada ventana de tamaño $vecindad = 10 \times 10$ píxeles.

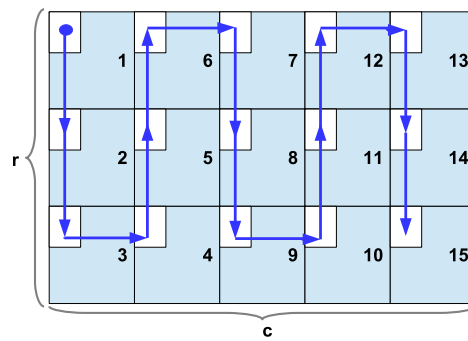


Figura 3.5: Ejemplo de la lectura temporal/secuencial de las vecindades de una imagen.

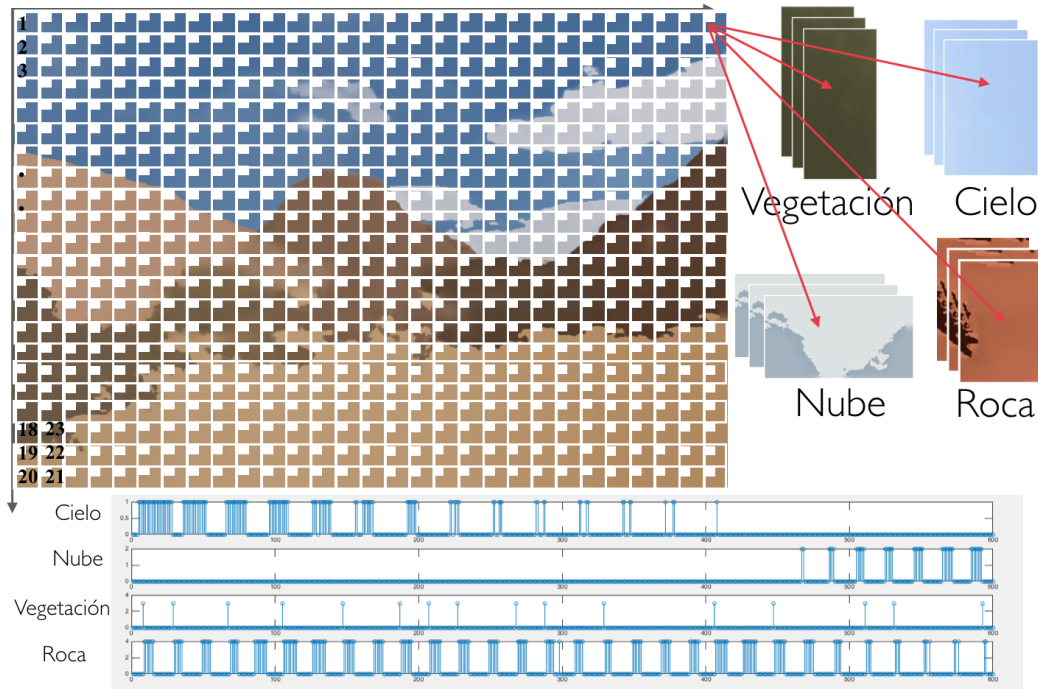


Figura 3.6: Serie de tiempo / lectura secuencial de la imagen para el análisis de causalidad.

una matriz M_s^i de tamaño $w \times 3$, donde $w = r * c$: el número de vecindades analizadas para cada i -ésima imagen de la clase C_s .

- 6. Generación de las series de tiempo / lectura secuencial de la imagen.** Por cada M_s^i del paso anterior, cada entrada de matriz se compara con las k -texturas del diccionario para construir una señal discreta T_s^i a través de la lectura secuencial de la imagen, definida como una matriz de tamaño $k \times w$, véase Figura 3.6.

En la comparación con el diccionario de texturas base, el valor 1 se asigna a la característica más cercana a la muestra y 0 a las demás, de acuerdo con los valores de umbral que caracterizan cada textura tal como se presentaron previamente. Después de procesar las entradas/muestras de todos los M_s^i , el conjunto de señales para cada escenario se almacena en FM_s , la matriz de series de tiempo correspondiente a una clase s que contiene imágenes de Img_s .

- 7. Análisis de causalidad de Wiener-Granger.** Cada matriz FM_s creada en el paso anterior se utilizó en el análisis CWG, para obtener las relaciones causales entre cada una de las texturas base. Se generó una matriz de relaciones de causalidad, η_s , relacionada con las imágenes de entrenamiento, como se muestra en la Figura 3.7; en esta matriz, los colores más oscuros representan relaciones causales más fuertes, estas relaciones se pueden representar a través de un diagrama de estado donde las líneas continuas representan solo las relaciones causales más fuertes. El análisis de causalidad se realizó por medio del toolbox de causalidad MVGC [3], el cual se invocó como una llamada de sistema externo.

Una vez que se ha realizado el análisis de causalidad para cada escenario C_s , obtenemos una

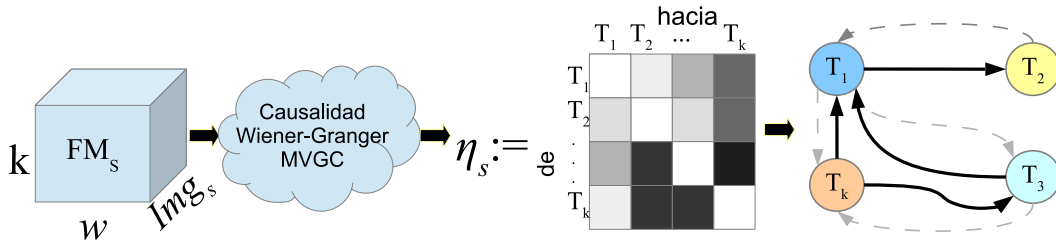


Figura 3.7: Generación de la matriz de relaciones causales Π_s obtenida por el toolbox MVGC.

matriz de relaciones de causalidad η_s de tamaño $k \times k$, con el total de las relaciones causales F_{T_i, T_j} de la textura $T_i \rightarrow T_j$ (como se indica en la ecuación (3.1)), de modo que si un valor de $F_{T_i, T_j} = 0$ significa que no hay una relación causal de la textura $i \rightarrow j$, y en la medida en que el valor aumenta con respecto a otros valores de η_s , decimos que la relación causal es significativa con respecto a los demás.

$$\eta_s = \begin{bmatrix} F_{T_1, T_1} & F_{T_1, T_2} & \dots & F_{T_1, T_k} \\ F_{T_2, T_1} & F_{T_2, T_2} & \dots & F_{T_2, T_k} \\ \vdots & \vdots & \vdots & \vdots \\ F_{T_k, T_1} & F_{T_k, T_2} & \dots & F_{T_k, T_k} \end{bmatrix} \quad (3.1)$$

Las matrices de causalidad η_s se normalizan de acuerdo con la suma total de sus valores, siendo $N_s = \sum_{i,j=1}^k F_{T_i, T_j}$, tal que η_s^N es la matriz normalizada del s -ésimo escenario, para $s = 1, \dots, C_s$, con C_s : el número de tipos de escenarios considerados, como dado en la ecuación (3.2). A partir de esta matriz resultante, los valores de la diagonal principal no se tienen en cuenta porque estos valores no generan fuerza en la relación de causalidad; Como se observa en la teoría, no existe una relación causal entre las mismas variables.

Al final, para C_s clases o escenarios, el *concentrado total de las matrices*, Γ , se define en la Ecuación (3.3).

$$\eta_s^N = \begin{bmatrix} F_{T_1, T_1} & F_{T_1, T_2} & \dots & F_{T_1, T_k} \\ F_{T_2, T_1} & F_{T_2, T_2} & \dots & F_{T_2, T_k} \\ \vdots & \vdots & \vdots & \vdots \\ F_{T_k, T_1} & F_{T_k, T_2} & \dots & F_{T_k, T_k} \end{bmatrix} * \frac{1}{N_s} \quad (3.2)$$

$$\Gamma = \cup_{l=1}^{C_s} \eta_l^N = \{\eta_1^N, \eta_2^N \dots \eta_{C_s}^N\} \quad (3.3)$$

Las matrices Γ son utilizadas como entradas y sirven como un patrón descriptivo para cada escenario o clase contenida en la base de datos.

8. **Selección de relaciones causales mediante algoritmo genético.** Para buscar las relaciones causales entre diferentes texturas que son más importantes o relevantes, para cada una de las escenas, esto se puede lograr de una manera simple eliminando las relaciones que tienen un valor

numérico menor que un umbral establecido previamente.

Sin embargo, una desventaja de este método es el establecimiento del umbral que se utilizará, porque no hay conocimiento de *a priori* del valor óptimo; Además, la complejidad aumenta cuando aumenta el número de texturas en el diccionario, junto con el número de clases e imágenes a examinar. Otro inconveniente de esta solución es que algunas de las relaciones débiles también podrían ser importantes para caracterizar un escenario. Por lo tanto, es necesario implementar una selección automática que discrimine las relaciones relevantes como un proceso de optimización combinatoria. Los algoritmos genéticos (AG) se han utilizado con éxito en varios problemas de visión por computadora junto con el procesamiento de imágenes digitales [48] y la clasificación [28, 14, 54, 31, 45]. En este trabajo, AG se muestra como una solución adecuada para la optimización requerida.

3.2. Requerimientos computacionales

La evaluación experimental de la propuesta se generó utilizando la potencia de la computadora clúster de doble núcleo de 19 procesadores. Cada procesador es una CPU Intel®Xeon®E5 – 2670 v3 2.30 GHz y 74 GB RAM.

La *B DEN* utilizada para la evaluación consta de los siguientes datos:

- Vogel and Shiele (V_S) [58], que incluye 6 paisajes con 700 imágenes clasificadas como: 144 costa, 103 bosque, 179 montaña, 131 pradera, 111 río/lago y 32 cielo/nube.
- Oliva y Torralba (O_T) [50], incluyendo 4 paisajes con 1472 Imágenes clasificadas como: 360 costa, 328 bosque, 374 montaña y 410 pradera.

Como se puede observar la cantidad de imágenes en la base de datos, el número de clases, el tamaño del emparrillado y el número de texturas base son cantidades importantes que se tienen que tomar en cuenta al momento de desarrollar este trabajo, ya que si se incrementa la cantidad de alguna de las variables mencionadas, el costo computacional se dispara, por esto es que debemos explotar todas las características computacionales que tengamos a nuestra disposición, es por eso que se propone el uso de una metodología de programación paralela. El enfoque paralelo desarrollado funciona en una arquitectura de memoria distribuida utilizando la biblioteca MPI; es decir, hay un conjunto de procesos sin memoria compartida, y estos procesos funcionan en paralelo, y la comunicación pasa por el intercambio de mensajes para determinar las relaciones causales relevantes de todo el escenario.

Cada proceso puede acceder a la *B DEN* para extraer y trabajar con el conjunto de imágenes correspondiente. La complejidad del algoritmo en esta propuesta se da para la ecuación (3.4).

$$O(N_{class} \times k \times CIP \times r \times c \times nImg \times t_{comp} \times WGC_{tb}) \quad (3.4)$$

donde:

- N_{clas} = Número de clases en la base de datos
- k = Número de texturas base

- CIP = Constante para cada procesamiento de imagen
- r = Cantidad de renglones en la malla
- c = Cantidad de columnas en la malla
- $nImg$ = Cantidad de imágenes por clase
- t_{comp} = Tiempo de comparación con las texturas base
- WGC_{tb} = Tiempo de análisis de causalidad

Por lo tanto surge la necesidad de implementar mecanismos de programación paralela para resolver el problema de una manera más eficiente, manejando grandes cantidades de información en un menor tiempo.

En el capítulo siguiente se mostrará la implementación del AG logrando el resultado óptimo, tras resolver el reto del cómputo masivo involucrado sin perder precisión en los resultados.

Capítulo 4

Algoritmo genético como soporte de la metodología CWG

En esta sección se desarrolla la implementación de la metodología planteada para la selección de las relaciones causales para cada uno de los escenarios naturales, por medio de un algoritmo genético (AG) y el desarrollo de una metodología de cómputo paralelo para agilizar el tiempo de búsqueda, de las mejores soluciones que cumplan la función de aptitud.

4.1. Selección de relaciones causales por medio de un Algoritmo Genético

Buscando el análisis de las matrices Γ generadas por el análisis CWG para encontrar las relaciones causales más significativas para un escenario, proponemos que cada matriz sea tratada con una implementación del AG. En esta sección proporcionamos la propuesta del AG en detalle.

En este enfoque, cada matriz $\eta_s^N \in \Gamma$ se expresa usando la representación vectorial, vea la Figura 4.1 partes (a) y (b); esto solo se logra concatenando las filas de la matriz η_s^N , entonces las entradas de las diagonales se eliminan como se muestra en la Figura 4.1 parte (c). En la Figura 4.1, parte (d), una reasignación de los valores después de ajustar la eliminación anterior. Esto proporciona un vector de índice continuo que tiene el tamaño de cada vector $1 \times (k^2 - k)$ por cada fila s -ésima, una por escenario. Siguiendo este proceso, finalmente, se crea la matriz τ , que contiene la conformación lineal de cada matriz η_s^N , con $s = \{1, 2, \dots, Cs\}$ en diferentes filas, como se muestra en la Figura 4.1 parte (e).

4.1.1. Codificación del individuo

Se conforma una representación binaria individual para un escenario, $\tau[i]$, para crear una matriz de tipo de filtro de tamaño $1 \times (k^2 - k)$ de ceros y unos, de modo que si se trata de una entrada o la relación causal se selecciona en esa matriz, se usa el valor 1 y 0 si no. Por lo tanto, tenemos Cs filas, una fila por escenario, se pretende que cada fila de la matriz de filtro pueda ser diferente de las otras

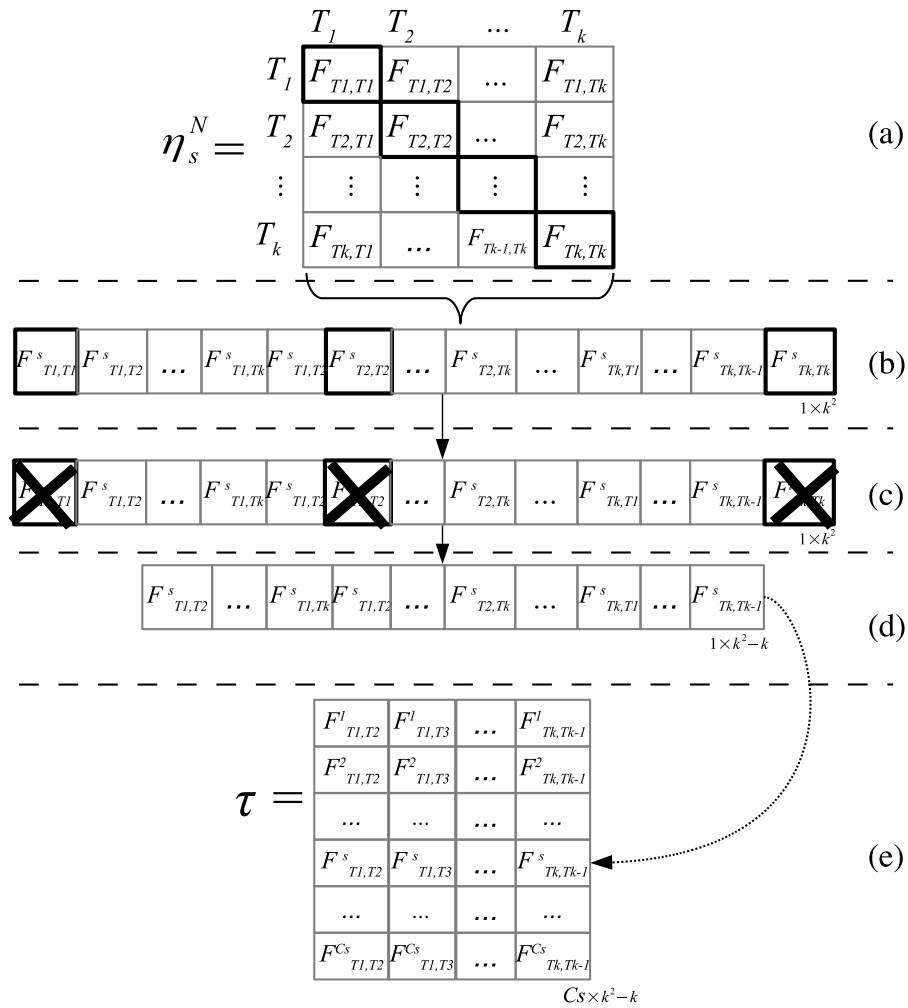
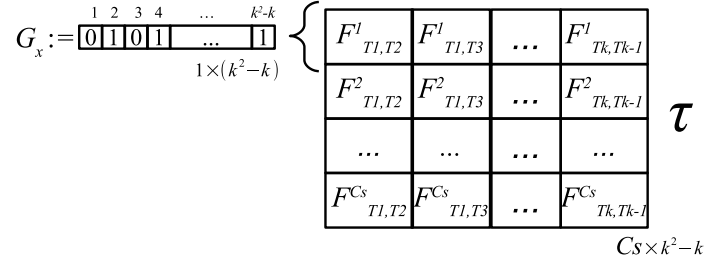


Figura 4.1: Proceso de generación de la matriz τ , para un escenario dado, para cada s -ésimo renglón $\in \tau$.


 Figura 4.2: Construcción del genoma G_x .

líneas, con el propósito de caracterizar cada tipo de escenario de una manera única. Luego es necesario aplicar un proceso automático para determinar qué valores de la matriz τ son características relevantes para distinguir las relaciones causales de cada escenario y, en función de este resultado, selecciona qué valores se eliminarán para número preestablecido de texturas. Con la selección de los valores causales más relevantes, se busca tener una clasificación por medio de un clasificador de distancia, hacia la matriz τ para cada una de las imágenes de consulta.

4.1.2. Función de aptitud

La evaluación de la aptitud de cada individuo se genera de la siguiente manera. Primero, la Ecuación (4.1) se aplica al individuo G_x , que representa una selección de relaciones de textura para el s -ésimo escenario en cuestión, usando la matriz τ en la Figura 4.2.

$$\rho_s^{G_x} = \frac{\prod_{l=1}^{k^2-k} G_x(l) * \tau_{s,l}}{\sum_{m=1}^{Cs} \prod_{l=1}^{k^2-k} G_x(l) * \tau_{m,l}}, \text{ tal que } G_x(l) \neq 0 \quad (4.1)$$

Así, $\prod G_x(l) * \tau_{s,l}$ se refiere al producto de las entradas τ , ubicadas en s -ésimo escenario (fila s) y columna l , especificando una relación causal, luego $G_x(l)$ es una entrada válida no nula del genoma. Por lo tanto, $\rho_s^{G_x}$ es la probabilidad total para el individuo G_x aplicado a todos los escenarios.

Con base en estos datos, por medio de la teoría de la probabilidad, el individuo G_x debe cumplir la condición: $\rho_s^{G_x} > \rho_j^{G_x}$, tal que $s \in \{1, 2, \dots, Cs\}$, $s \neq j$ y $1 \leq j \leq Cs$.

Es decir, las probabilidades de Cs correspondientes a cada escenario que evalúa el individuo G_x se obtienen con el cálculo de $\rho_j^{G_x}$. La ecuación (2) da el primer paso para el proceso de optimización, considerando la probabilidad máxima relacionada con las relaciones causales que mejor caracterizaron el s -ésimo escenario frente a los demás.

$$\bigvee_s^{G_x} = \begin{cases} \rho_s^{G_x} & \text{if } \rho_s^{G_x} = \text{Max}\{\rho_j^{G_x}\}_{j=1,\dots,Cs} \\ 0 & \text{if } \rho_s^{G_x} \neq \text{Max}\{\rho_j^{G_x}\}_{j=1,\dots,Cs} \end{cases} \quad (4.2)$$

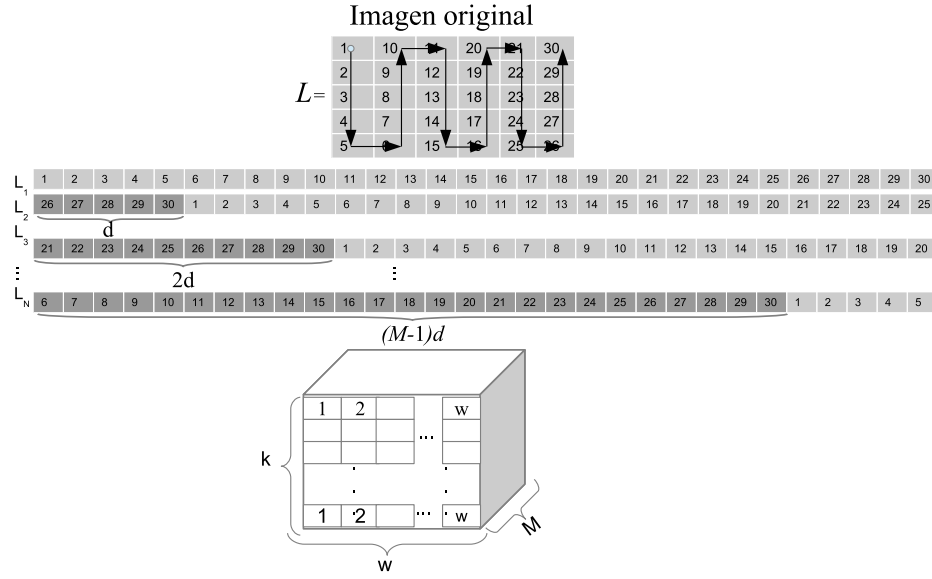


Figura 4.3: Construcción de una imagen de consulta con N muestras.

Entonces, la función de aptitud o “fitness”, $f_s(G_x)$, esta determinada por la Ecuación (4.3).

$$f_s(G_x) = \begin{cases} CP_s & \text{if } \bigvee_s G_x > 0 \\ 0 & \text{if } \bigvee_s G_x = 0 \end{cases} \quad (4.3)$$

Para este fin se consultan las imágenes contenidas en cada s -escenario, haciendo uso de la prueba de resubstitución. Cada consulta de imagen en la base de datos, proporciona el escenario al que corresponde, con esta información se llena la matriz de confusión que se utiliza para calcular el porcentaje de clasificación. El proceso de consulta de imagen se describe en el siguiente párrafo. Más tarde, en la Sección 4.1.3 la *función de aptitud global* se toma en cuenta para el proceso de evolución de la población en el proceso del AG.

Creación de una consulta a partir de una sola imagen

Para clasificar una imagen de un s -escenario considerando las relaciones especificadas en un individuo G_x , se debe construir una matriz de relación causal MC_{si} . El primer paso consiste en crear un conjunto de M imágenes sintéticas, L_1, L_2, \dots, L_M , a partir de una sola imagen L . Esto se produce mediante la manipulación de la primera lectura de la imagen, haciendo un desplazamiento circular de d posiciones para cada nueva imagen sintética, a fin de crear varias muestras de la misma imagen como se muestra en la Figura 4.3. De esta manera se tiene, la matriz de consulta respectiva MC_{si} de tamaño $|k \times w \times M|$. Donde $k :=$ número de texturas en el diccionario, $w :=$ número de ventanas, y $M :=$ número de imágenes sintéticas, la cual se genera para alimentar el análisis de CWG, con la finalidad de obtener la matriz de relación causal normalizada resultante η_L^M de tamaño $|k \times k|$. Estos pasos los llevan a cabo las ecuaciones (3.1) y (3.2).

Entonces, la manipulación de la matriz η_L^M se realiza como se muestra en la etapa presentada en la Figura 4.1 para obtener la representación lineal de la matriz de relaciones causales. El último paso de la consulta consiste en aplicar el clasificador k -NN, con $k := 1$, para determinar qué escenario (fila) de la matriz τ tiene las relaciones más cercanas a la representación de la relación lineal de la imagen i -ésima, considerando solo la relación indicada con los valores G_x distintos de cero.

4.1.3. Implementación del AG

Se aplica un algoritmo genético para cada línea de τ con la finalidad de seleccionar automáticamente las relaciones causales más representativas de cada escenario. La Figura 4.4(a) muestra el diagrama de flujo del algoritmo general de este enfoque. Se genera aleatoriamente una población inicial, PG , de $sizeP$ individuos, donde $sizeP$ es un número impar y cada individuo tiene un tamaño $k^2 - k$, este es el tamaño en columnas de la matriz τ .

Luego, los individuos PG se evalúan con la función de aptitud, Ecuación (4.3), para un s -escenario particular, considerando el conjunto total de imágenes que lo conforman, como se muestra en la Figura 4.4(b). La aptitud del individuo se almacena dentro de una matriz de aptitud $\{\bullet\}$, como se aprecia en la Ecuación (4.4). Por consiguiente, la matriz $\{\bullet\}$ se ordena, de más alto a más bajo, para encontrar al mejor individuo con la mayor aptitud.

$$\{\bullet\} = \{f_1(G_x), f_2(G_x), \dots, f_{sizeP}(G_x)\}, \quad (4.4)$$

tal que $f_p(G_x) \geq 0$ para $1 \leq p \leq sizeP$. Para esta propuesta, el tamaño de la población $sizeP = 21$, la longitud del genoma = 12, y el número de iteraciones fue $maxGen = 100$ generaciones.

Para generar la nueva población, $(sizeP - 1)/2$ tripletas de números aleatorios son creadas, por ejemplo, $\{1, 5, 1\}$ o $\{2, 4, 0\}$, donde los dos primeros números son los números individuales seleccionados que generan los nuevos individuos, el tercer elemento de la tripleta es uno de las dos posibles operaciones genéticas a ejecutar; ya sea cruza “1” o mutación “0”.

A partir de cada tripleta, generamos dos nuevos individuos sea cual sea la operación genética ya sea por cruza, o por mutación, los dos individuos seleccionados se alteran por separado, para generar dos elementos nuevos para la nueva generación.

La operación de cruza se aplica en un punto aleatorio uniforme de los dos cromosomas participantes, y el operador de mutación se realiza sobre 10 % de los elementos de un cromosoma, como se muestra en la Figura 4.5.

Las operaciones genéticas de mutación y cruza se aplican a 30 % y 70 % de la población para nuestra implementación, respectivamente, lo que favorece la selección de los individuos con mejor función de aptitud para su reproducción. El individuo con el mejor función de aptitud pasa a la siguiente generación por medio del elitismo. De esta manera, la población evolucionará hacia una selección de relaciones causales relevantes con el objetivo de caracterizar cada escenario de la base de datos.

El criterio de paro del AG se da cuando el porcentaje de clasificación llega al 100 % o se alcanza un número máximo de generaciones.

Después de aplicar el AG Cs -veces, se encuentran los individuos que contienen las relaciones causales más relevantes para cada escenario. Luego, la matriz τ se actualiza y sus entradas se reemplazan

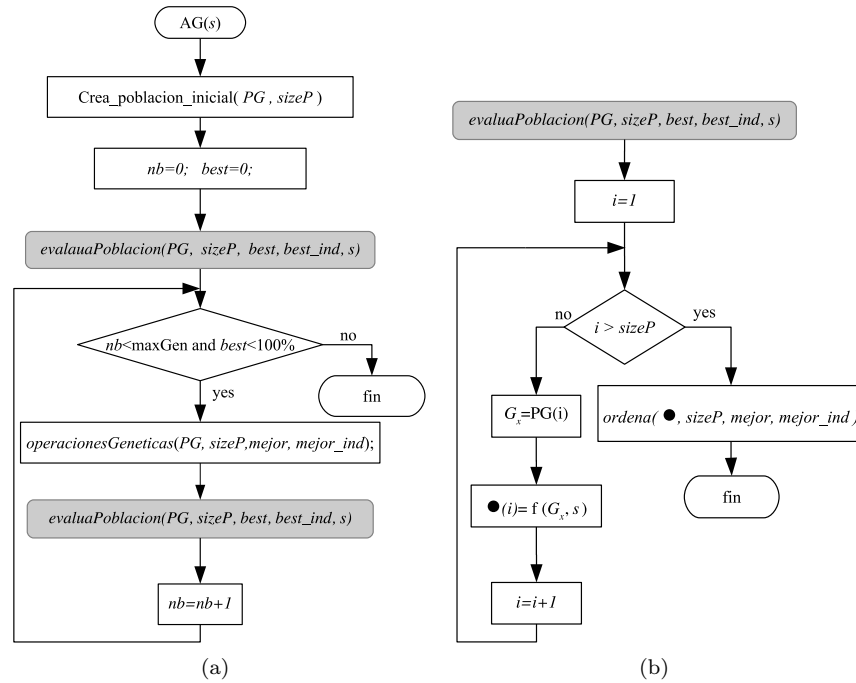


Figura 4.4: Diagrama general de flujo de la implementación del AG.

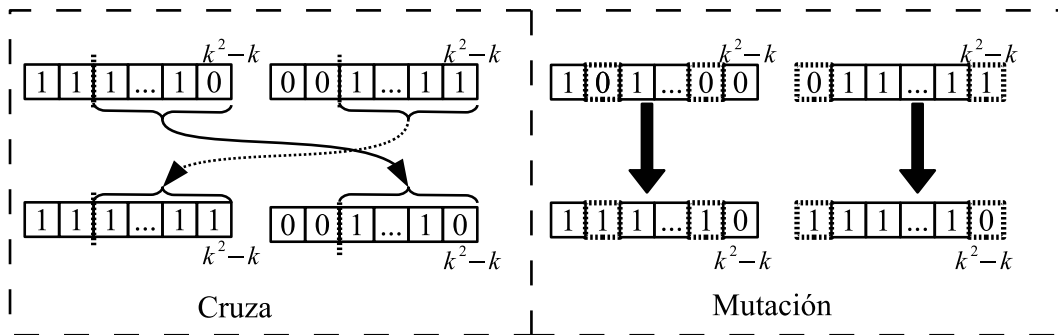


Figura 4.5: Aplicación de los operadores genéticos.

por un valor cero siempre que las entradas individuales correspondientes tengan un cero y mantengan su valor en otro caso.

4.2. Paralelización del algoritmo genético

En esta sección, se presenta un algoritmo paralelo para acelerar el rendimiento de la metodología CWG propuesta. El enfoque paralelo funciona en una arquitectura de memoria distribuida utilizando la biblioteca MPI; es decir, hay un conjunto de procesos sin memoria compartida, y estos procesos funcionan en paralelo, y la comunicación pasa por el intercambio de mensajes para determinar las relaciones causales relevantes de todo el escenario. Cada proceso puede acceder a la BDEN para extraer y trabajar con el conjunto de imágenes correspondiente. La complejidad del algoritmo en esta propuesta se da para la ecuación (4.5).

$$O(N_{class} \times k \times CIP \times r \times c \times nImg \times t_{comp} \times WGC_{tb}) \quad (4.5)$$

dónde, N_{class} := número de clases, k := número de texturas, CIP := constante para cada procesamiento de imagen, r := número de filas en la cuadrícula, c := número de columnas en la cuadrícula, $nImg$:= número total de imágenes, t_{comp} := tiempo de comparación con las texturas base, WGC_{tb} := tiempo de análisis de causalidad, (por ejemplo, para una imagen de tamaño 640×480 píxeles, $r = (640/20)$, $c = (480/30)$ y p es el tamaño del de la vecindad dentro de la ventana, de modo que $10 \text{ veces } 10$ implica $p = 10$) Eso significa que si el número de filas r , el número cols c en la cuadrícula, el número de imágenes $nImg$ en BDEN, y el número de texturas base k en el diccionario, y el número de clases N_{class} aumentan, estas incrementan el costo computacional. De esta manera, es necesario concebir una arquitectura paralela para resolver este problema en una gran cantidad de imágenes relacionadas con problemas de datos masivos o Big Data.

El Algoritmo 1 muestra el procedimiento que se ejecuta simultáneamente por cada proceso, y el proceso general de esta propuesta paralela se representa en la Figura 4.6.

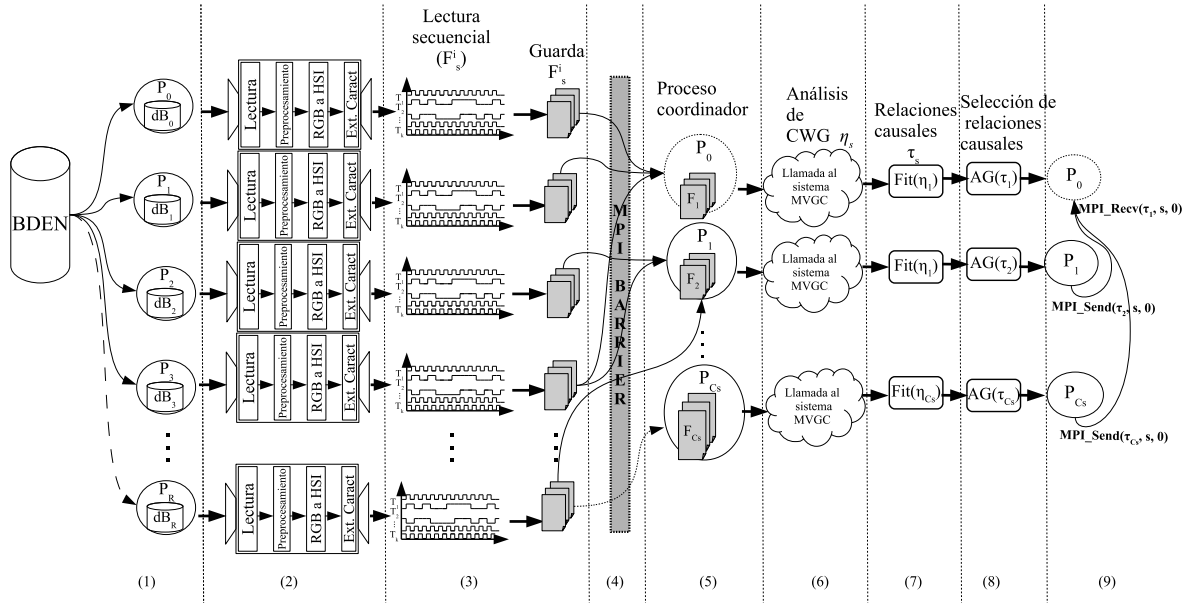


Figura 4.6: Arquitectura paralela diseñada de la búsqueda de relaciones causales para cada escenario natural.

Algoritmo 1 Algoritmo paralelo para la construcción de la matriz de relaciones causales.

```

1: procedure CONSTRUCCIÓN DE LA MATRIZ DE CAUSALIDAD(identificador)
2:   inicialización(bloqueImágenes, imágenesTotales, procesosTotales, identificador);
3:   for cada i en bloqueImágenes do
4:     leerImagen_RGB(imagen, s, i);
5:     preproesar_Imagen(imagen);
6:     RGB_a_HSI(imagen);
7:      $M_s^i = \text{extracción\_Características}(\text{imagen});$ 
8:      $F_s^i = \text{construcción\_LecturaSecuencial}(M_s^i, \text{diccionario\_Texturas});$ 
9:     almacena_LecturaSecuencial( $F_s^i$ );
10:  end for
11:  Barrera_Sincronización();
12:  if identificador in {identificadores_escenario_coordinador} then
13:     $F_s = \text{CargarLecturasSecuenciales}(s);$ 
14:     $\eta_s = \text{llamada\_Al\_Sistema}(MVGC(F_s));$ 
15:     $\tau_s = \text{adecuarMatriz}(\eta_s);$ 
16:    algoritmoGenetico( $\tau_s$ );
17:    Enviar( $\tau_s, s, \text{identificador\_coordinador\_general}$ );
18:  end if
19:  if (identificador == identificador_coordinador_general) then
20:    for (cada id en {identificadores_escenario_coordinador}) do
21:      Recibe( $\tau_s, s, id$ );
22:       $\tau(s) = \tau_s;$ 
23:    end for
24:  end if
25: end procedure
    
```

Al principio (línea 2 del Algoritmo 1, Figura 4.6, etiqueta (1)), cada proceso determina la cantidad de conjunto de imágenes a leer (*bloqueImágenes*), teniendo en cuenta el número total de imágenes (*imagenesTotales*), el número total de procesos (*procesosTotales*) y el identificador de proceso (*identificador*). Un solo proceso puede funcionar con imágenes que pertenecen a diferentes escenarios (por ejemplo, $imagenesTotales = 700$, $procesosTotales = 70$, $bloqueImágenes = 700/10 = 10$, esto es calculado para la *BDEN*).

Cada proceso funciona simultáneamente con la sección de *BDEN*, *bloqueImágenes*, que se le asignó, realizando los siguientes pasos (líneas 3 a 10). La lectura de la imagen i en el espacio RGB es la primera acción que se ejecuta, a continuación, el escenario, s , de modo que también se obtiene i -imagen $\in s$ (línea 4). Luego se realiza el preprocesamiento de la imagen y la conversión de los espacios de color RGB a HSI (líneas 5 y 6, respectivamente). En la línea 7, se calculan las características estadísticas, incluida la construcción de la cuadrícula de imágenes y vecindarios, entonces las características CBIR por cada vecindario generan la matriz M_s^i ; La Figura 4.6, etiqueta (2), representa la ejecución de las líneas 4 a 7 del Algoritmo 1. Luego, M_s^i y el diccionario de texturas se utilizan para construir las respectivas lecturas secuenciales, F_s^i , que se almacenan en el archivo (líneas 8,9 del algoritmo, Figura 4.6, etiqueta (3)).

Hasta este punto, todos los procesos funcionan de manera independiente; sin embargo, para garantizar que cada proceso haya cumplido completamente su tarea, se debe introducir una sincronización de barrera paralela (línea 11 del algoritmo, Figura 4.6, etiqueta(4)), antes de continuar con el siguiente paso. Aquí (línea 12), solo los procesos identificados como *escenario coordinador* (un proceso por escenario) continúan con la construcción de la matriz F_s correspondiente (línea 13, Figura 4.6, etiqueta (5)), al cargar el conjunto respectivo de matrices F_s^i (una por imagen de escenario), generadas previamente. Luego, se realiza una llamada al sistema (línea 14, Figura 4.6, etiqueta (6)) para ejecutar el toolbox MVGC y obtener la matriz de relaciones causales, η_s , a partir del análisis CWG.

La función $Fit(\eta_s)$ en la línea 15 (Figura 4.6, etiqueta (7)) está a cargo de la normalización y la representación vectorial de la matriz de relación de causalidad, η_s . De este modo se genera el τ_s respectivo, correspondiente a la fila s -ésimo (escenario) de la matriz τ . La línea 16 del Algoritmo 1 (Figura 4.6, etiqueta (8)) muestra la llamada AG que se ejecuta por cada uno de los procesos del coordinador de escenarios Cs , siendo Cs el Número de escenarios. Después de identificar las relaciones causales más relevantes por medio de la AG, τ_s se actualiza y se envía al coordinador general (línea 17) a través de un mensaje.

Finalmente, en las líneas 19–24, el proceso coordinador general recibe, por medio de varios mensajes, los resultados generados por los procesos del coordinador de escenarios (Figura 4.6, etiqueta (9)). Cuando se logran todas las recepciones de mensajes, la matriz τ se construye con éxito.

Capítulo 5

Resultados

Se seleccionaron cuatro texturas de imagen, $k = 4$, para conformar el diccionario base, como se muestra en la Tabla 5.1. Para cada textura, los valores generados se obtuvieron manualmente dentro de las imágenes de la base de datos, se tomaron un conjunto de 20 muestras de textura de un conjunto de 5 imágenes por clase, véase Figura 5.1, de cada textura se extrajo el promedio en la capa H más dos veces la desviación estándar, con esto se generaron los valores de umbral máximos y mínimos para cada textura.

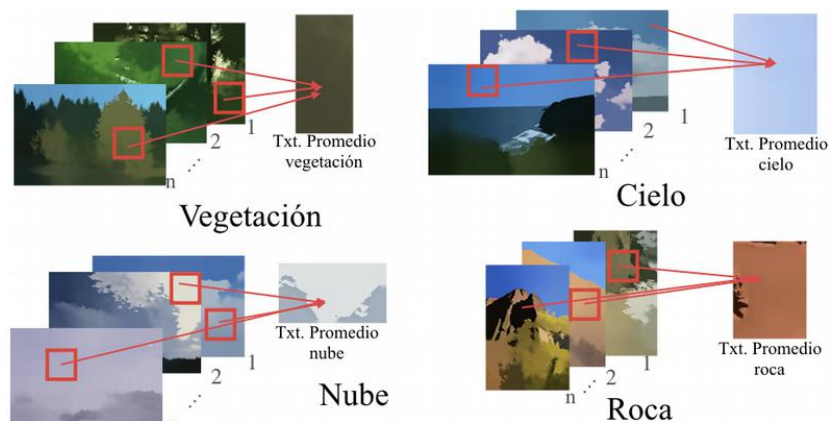


Figura 5.1: Generación del diccionario de texturas.

La cantidad de elementos contenidos en la BDEN se muestra en la Tabla 5.2, como se puede observar no es una base de datos completamente balanceada. Las imágenes se adaptaron para que se

Tabla 5.1: Rangos HSI para las texturas base de prueba.

Textura	H-max	H-min	S-max	S-min	I-max	I-min
Nube (N)	180	0	25	0	255	61
Cielo (C)	113	93	255	25	61	255
Roca (R)	24	6	255	20	190	30
Vegetación (V)	102	28	255	10	229	3

Tabla 5.2: Descripción de la BDEN utilizada para el trabajo de investigación.

Clase	Bosque	Cielo	Costa	Montaña	Pradera	Río
Imágenes	120	100	120	120	120	120

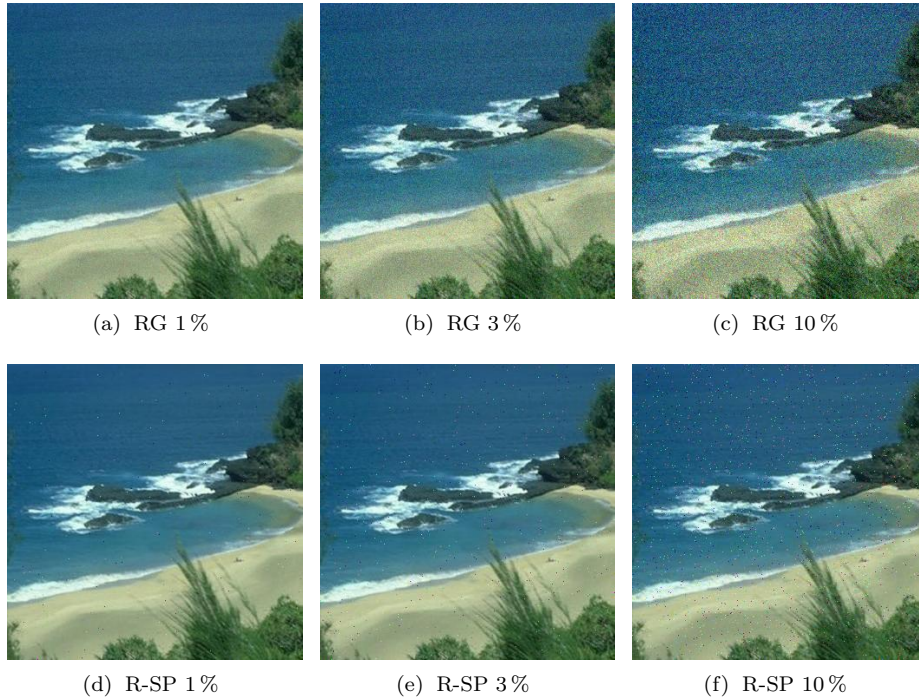


Figura 5.2: Ejemplo de imágenes contaminadas con ruido Gaussiano RG, y ruido sal y pimienta R-SP, al 1%, 3%, y 10%, respectivamente.

consideraran algunos desafíos de clasificación típicos. El conjunto completo de imágenes se probó en un estado normal e introdujo ruido gaussiano (RG), ruido de sal y pimienta (R-S&P) de 1%, 3% y 10% niveles respectivamente, como se muestra en la Figura 5.2. También se introdujo una transformación de rotación en cada imagen considerando 0° , 45° , 90° , 135° y 180° , como se muestra en la Figura 5.3. Se realizó una consulta de consulta de imagen siguiendo el mismo procedimiento descrito en la Sección 4.1.2.

Los resultados en esta sección están organizados de la siguiente manera. Primero se presenta el rendimiento de clasificación de imágenes obtenido al aplicar la teoría CWG. Luego se muestran los tiempos de ejecución de la metodología paralela propuesta.

5.1. Resultados de clasificación

Para demostrar que la implementación propuesta de AG fue una buena solución para seleccionar algunas relaciones de textura relevantes que describan un escenario, comparamos nuestra propuesta (versión AG) con la estrategia manual (versión Manual) presentada en la Referencia [20]; Bajo la estrategia manual solo se seleccionaron los valores de relación causal entre las texturas que tienen los

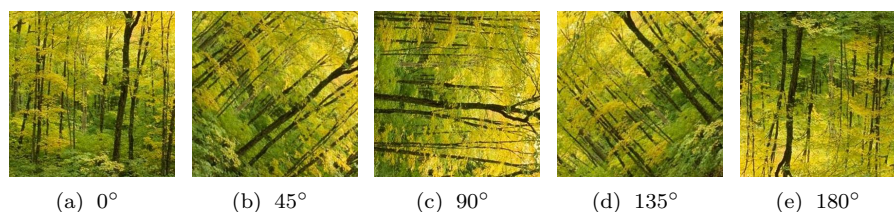


Figura 5.3: Ejemplo de imágenes transformadas para la prueba de rotación.

valores más altos, estableciendo un umbral específico. En ambas versiones, se ejecutó la metodología presentada en la Sección 3 para la construcción de la matriz τ . La tabla 5.3 muestra los valores resultantes de τ .

Debido a que no había criterios *a priori* para determinar un valor umbral, en la versión Manual 25% de las relaciones causales menos significativas por escenario se eliminaron. La tabla 5.4 muestra la matriz τ actualizada después de la selección manual.

Tabla 5.3: Los valores obtenidos de la matriz τ .

Escenario/ F	$F_{T1,T2}$	$F_{T1,T3}$	$F_{T1,T4}$	$F_{T2,T1}$	$F_{T2,T3}$	$F_{T2,T4}$	$F_{T3,T1}$	$F_{T3,T2}$	$F_{T3,T4}$	$F_{T4,T1}$	$F_{T2,T4}$	$F_{T3,T4}$
Bosque	0.370	0.289	0.0051	0.0269	0.037	0.0013	0.108	0.075	0.0849	0.00332	1.611e-05	0.00062
Cielo	0.0441	0.0481	0.0321	0.0269	0.343	0.197	0.0044	0.0385	0.211	0.011	0.0070	0.0371
Costa	0.0099	0.233	0.188	0.021	0.0363	0.0503	0.0542	0.0014	0.0646	0.1558	0.0772	0.109
Montaña	0.085	0.405	0.0132	0.0497	0.0799	0.0698	0.1115	0.0102	0.0162	0.1076	0.0392	0.0133
Pradera	0.2401	0.1143	0.2400	0.0045	0.0140	0.0268	0.2151	0.1146	0.0006	0.0161	0.0140	0.0002
Río	0.1619	0.4053	0.1112	0.0061	0.0046	0.0035	0.0377	0.0322	0.1794	0.0066	0.0084	0.0432

Tabla 5.4: La matriz τ resultante de la selección manual de las relaciones causales más altas.

Escenario/ F	$F_{T1,T2}$	$F_{T1,T3}$	$F_{T1,T4}$	$F_{T2,T1}$	$F_{T2,T3}$	$F_{T2,T4}$	$F_{T3,T1}$	$F_{T3,T2}$	$F_{T3,T4}$	$F_{T4,T1}$	$F_{T2,T4}$	$F_{T3,T4}$
Bosque	0.370	0.289	0.0051	0.0269	0.037	0	0.108	0.075	0.0849	0.00332	0	0
Cielo	0.0441	0.0481	0.0321	0.0269	0.343	0.197	0	0.0385	0.211	0	0	0.0371
Costa	0	0.233	0.188	0	0.0363	0.0503	0.0542	0	0.0646	0.1558	0.0772	0.109
Montaña	0.085	0.405	0	0.0497	0.0799	0.0698	0.1115	0	0	0.1076	0.0392	0
Pradera	0.2401	0.1143	0.2400	0	0.0140	0.0268	0.2151	0.1146	0	0.0161	0.0140	0
Río	0.1619	0.4053	0.1112	0	0	0	0.0377	0.0322	0.1794	0	0	0.0432

Al ejecutar la versión AG para la selección de relaciones causales de la matriz τ , se exploró un espacio mayor de posibles soluciones tratando de buscar las relaciones causales que mejor representan un escenario. Los individuos obtenidos se presentan en la Tabla 5.5, y la matriz τ actualizada se muestra en la Tabla 5.6.

Tanto las versiones AG como las manuales se probaron con 300 imágenes, 50 por escenario. La versión manual solo obtuvo la obtención de un porcentaje de clasificación general de 12.53%. La matriz de confusión que muestra la asociación de imágenes por escenario se puede ver en la Tabla 5.7; observamos que la mayoría de las imágenes estaban asociadas al paisaje costero y, como resultado, la prueba de selección manual dio un porcentaje de clasificación pobre.

Con la información en la Tabla 5.6, las relaciones más representativas de cada escenario natural se generan como una representación visual, los gráficos representan la intensidad de las relaciones causales entre las texturas base $k = 4$ del diccionario. Estos gráficos mostrarán cómo se relacionan las texturas

Tabla 5.5: Mejores individuos resultantes de la evaluación del AG para cada escenario.

Escenario/ F	1	2	3	4	5	6	7	8	9	10	11	12
Bosque	1	0	0	0	0	0	0	0	1	0	0	0
Cielo	0	1	1	0	1	1	1	1	0	0	0	1
Costa	0	0	1	1	0	0	0	1	0	0	1	1
Montaña	1	0	0	1	0	0	1	1	0	0	0	1
Pradera	1	0	1	0	1	0	1	1	1	0	1	0
Río	0	0	1	1	0	0	1	1	0	0	0	1

Tabla 5.6: Valores finales de la matriz τ cuando se aplican los mejores individuos del AG.

Escenario/ F	$F_{T1,T2}$	$F_{T1,T3}$	$F_{T1,T4}$	$F_{T2,T1}$	$F_{T2,T3}$	$F_{T2,T4}$	$F_{T3,T1}$	$F_{T3,T2}$	$F_{T3,T4}$	$F_{T4,T1}$	$F_{T4,T2}$	$F_{T4,T3}$
Bosque	0.370	0	0	0	0	0	0	0	0.0849	0	0	0
Cielo	0	0.0481	0.0321	0	0.343	0.197	0.0044	0.0385	0	0	0	0.0371
Costa	0	0	0.188	0.021	0	0	0	0.0014	0	0	0.0772	0.109
Montaña	0.085	0	0	0.0497	0	0	0.1115	0.0102	0	0	0	0.0133
Pradera	0.2401	0	0.2400	0	0.0140	0	0.2151	0.1146	0.0006	0	0.0140	0
Ríos	0	0	0.1112	0.0061	0	0	0.0377	0.0322	0	0	0	0.0432

Tabla 5.7: Matriz de confusión para una prueba con 50 imágenes por escenario, usando la selección manual de relaciones causales.

$Escenario_i/Escenario_j$	Bosque	Cielo	Costa	Montaña	Pradera	Río
Bosque	0	0	31	19	0	0
Cielo	1	0	39	10	0	0
Costa	3	0	36	11	0	0
Montaña	2	0	47	1	0	0
Pradera	5	0	40	5	0	0
Río	6	0	32	11	1	0

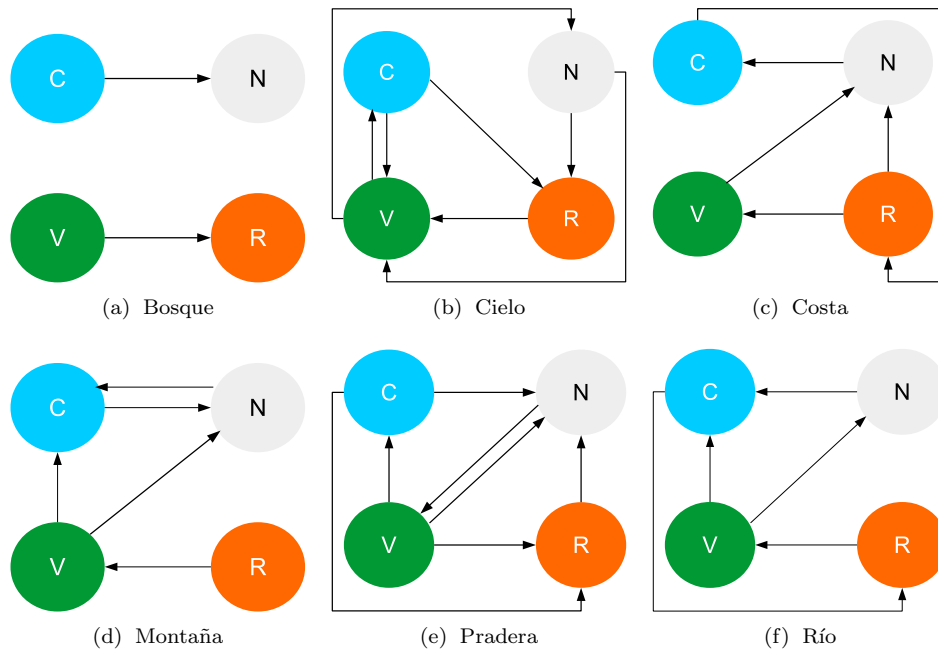


Figura 5.4: Representación gráfica de las relaciones causales para cada escenario natural.

dentro del escenario correspondiente, obteniendo el patrón que representa a cada uno de ellos, como se puede apreciar en la Figura 5.4.

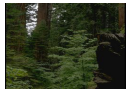
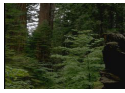






















Dados estos primeros resultados, se puede observar que no necesariamente las relaciones con valores más altos fueron las mejores para ser seleccionadas. Para medir la eficiencia técnica de nuestra propuesta utilizando un AG, se estimaron las medidas de *Recuperación* (administrado como porcentaje de clasificación), *Precisión*, *Exactitud* y *medida-F1*, (*Precision*, *Accuracy* and *F1-Score*), a partir de las matrices de confusión de cada prueba.

Los resultados de clasificación de la Figura 5.6 muestran que al rotar las imágenes en 45°, 90° y 315°, el rendimiento de la clasificación disminuye significativamente. Además, el RG y R-SP alteran significativamente la clasificación que se espera en las imágenes de paisajes naturales, ya que la textura es representativa del tipo de imagen, y las alteraciones con ruido degeneran en otro significado posible. En condiciones normales, evitando ruidos y rotaciones, el rendimiento de clasificación alcanza el 100%.













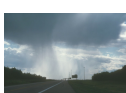











Además, la Figura 5.7 muestra las estimaciones promedio de la *precisión* (Figura 5.7(a)), la *Recuperación* (Figura 5.7(b)), la *exactitud* (Figura 5.7(c)), y finalmente la *medida-F1* (Figura 5.7(d)), para las 6 clases contenidas en la BDEN. En general, la clasificación de imágenes ideales (0°) sin rotaciones y sin ruido obtiene una clasificación del 100%. Sin embargo, cuando se agregan rotaciones y ruido, este porcentaje disminuye, particularmente la clase de cielo es la más afectada.

La prueba de 50 imágenes desconocidas en el sistema se representa por las Figuras 5.8(a), 5.8(b), 5.8(c), 5.8(d), 5.8(e) y 5.8(f), contenidas en la Figura 5.8. De esta manera se puede observar la correcta recuperación del sistema para las imágenes de consulta.



La Figura 5.9 muestra los porcentajes de clasificación y desempeño del sistema con imágenes desconocidas. La Figura 5.9(e) muestra el porcentaje de clasificación, la Figura 5.9(a) muestra la *precisión*, la

Test	1-NN	2-NN	3-NN	4-NN	5-NN
					
					
					
					

























(a) Elementos recuperados para la clase bosque

Test	1-NN	2-NN	3-NN	4-NN	5-NN
					
					
					
					

(b) Elementos recuperados para la clase cielo.

Test	1-NN	2-NN	3-NN	4-NN	5-NN
					
					
					
					

(c) Elementos recuperados para la clase costa.

Test	1-NN	2-NN	3-NN	4-NN	5-NN
					
					
					
					

(d) Elementos recuperados para la clase montaña.

Test	1-NN	2-NN	3-NN	4-NN	5-NN

(e) Elementos recuperados para la clase pradera.

Test	1-NN	2-NN	3-NN	4-NN	5-NN

(f) Elementos recuperados para la clase río.

Figura 5.5: Elementos recuperados en la prueba de resubstitución para la BDEN.

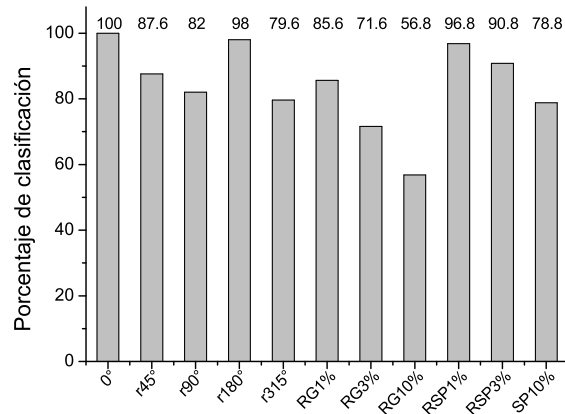


Figura 5.6: Resultados de clasificación para el AG, considerando las pruebas de ruido y rotación.

Recuperación se observa en la Figura 5.9(b), la *exactitud* en la Figura 5.9(c), y finalmente la *medida-F1* se muestra en la Figura 5.9(d).

El resultado promedio de la evaluación AG se muestra en la Figura 5.10. La Figura 5.10(a) muestra la evolución de la *función de aptitud* dentro de una ejecución con 100 generaciones. La Figura 5.10(b) muestra que, si bien en algunas clases se logra la aptitud física más alta en las primeras iteraciones, en otras 100 generaciones no son suficientes para lograr la mejor aptitud. La Figura 5.10(c) muestra la mejor aptitud obtenida a través de 200 repeticiones considerando 100 generaciones por repetición y el tamaño de la población establecido en 21 individuos; toda la aptitud converge cerca del valor esperado de 100 % de clasificación.

5.2. Rendimiento de la metodología paralela

El tiempo de ejecución tomado por la metodología de causalidad paralela propuesta aplicada a la identificación y clasificación de paisajes naturales, para un total de 700 imágenes, 6 escenarios diferentes y variando el número de procesos, se muestra en la Figura 5.11. Estos valores fueron el tiempo promedio necesario para 200 ejecuciones. Podemos observar que el tiempo de ejecución disminuyó rápidamente al tiempo que aumentó el número de procesos, obteniendo el mejor tiempo de ejecución cuando se definieron procesos concurrentes de 125. Con esta configuración, cada proceso funcionó con 5 o 6 imágenes de de la base de datos, favoreciendo la programación interna utilizando de manera más eficiente los recursos de la computadora en comparación con el tiempo de ejecución de la versión secuencial disminuyéndolo en 88.9 % como se muestra en la Figura 5.11.

5.3. Discusión de resultados

Para encontrar las relaciones de causalidad de textura buscando caracterizar un escenario natural, encontramos la necesidad de la implementación de un AG. Con esta propuesta, se logró con éxito el

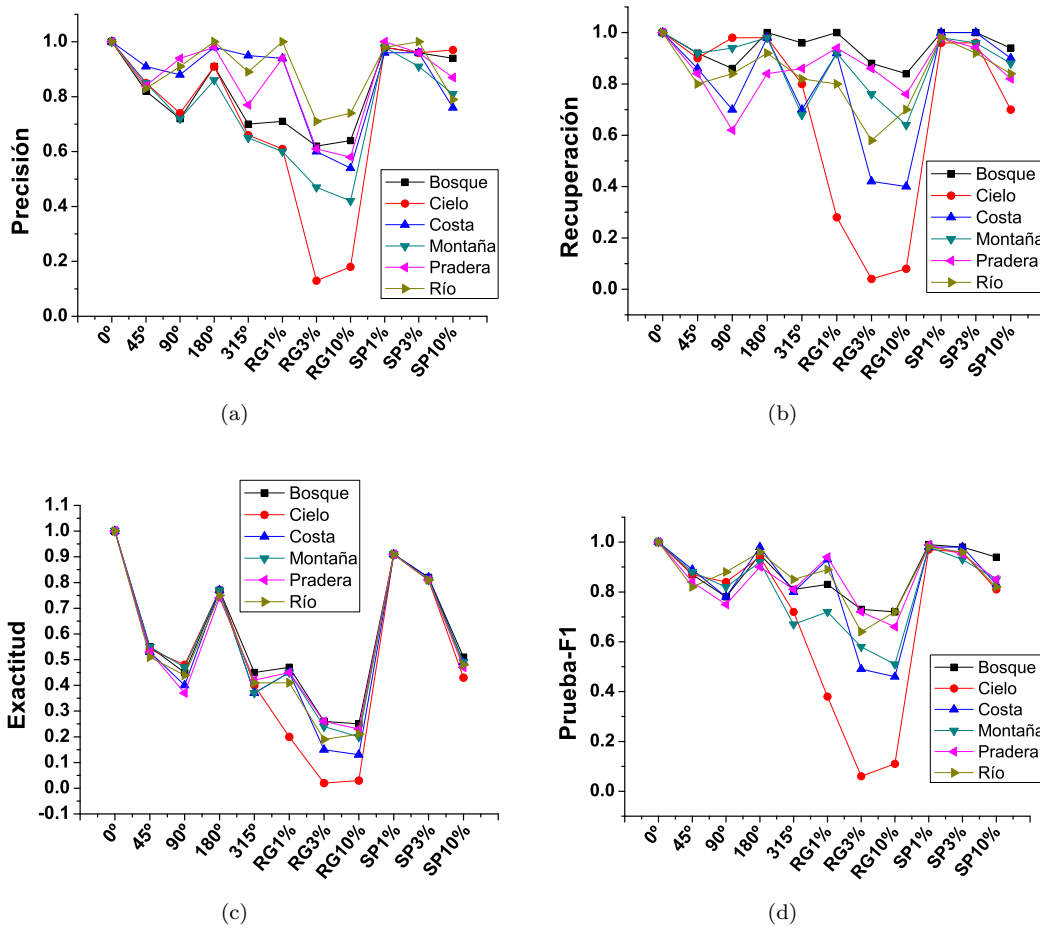














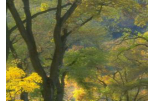




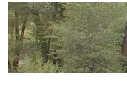






Figura 5.7: Medidas de eficiencia para los mejores individuos del AG para cada escenario natural.

proceso de discriminación automática para seleccionar las relaciones causales que son más importantes o relevantes para la clasificación del escenario propuesto. En comparación con algunos de los artículos en la literatura, la posibilidad de nuestro enfoque para realizar la selección de forma automática permite el estudio del problema de clasificación de escenarios considerando más parámetros. De esta manera, se podría tener en cuenta un mayor número de escenarios o texturas base para futuras implementaciones de varios propósitos de clasificación de textura considerando la metodología eficiente y el algoritmo evolutivo propuesto en este trabajo.

























Esta propuesta podría ayudar a reconocer escenarios naturales en la navegación de un automóvil autónomo o posiblemente un dron, siendo un elemento importante en la seguridad de la navegación de vehículos autónomos. Como podemos ver en la Figura 5.6, el porcentaje de clasificación obtenido por medio de las características seleccionadas para el proceso evolutivo supera las obtenidas por la versión de selección manual. Este resultado es importante porque las relaciones causales representativas del escenario se seleccionan de tal manera que escapen numéricamente de la perspectiva manual; es decir, en la estrategia de selección manual se especifica un valor umbral no significativo, de modo que cualquier

valor inferior al umbral se establece en cero, pero la estrategia evolutiva resulta que algunas de estas relaciones causales son relevantes para clasificar el escenario, marcando diferencias con otro escenario similar.

























A partir de la teoría de la causalidad aplicada a las secuencias de lectura de imágenes, estamos tratando de inferir el orden de aparición de las texturas tipificadas en el diccionario base buscando representarlas como lectura visual temporal que vemos como un tipo de escenario natural.

Test	1-NN	2-NN	3-NN	4-NN	5-NN
					
					
					
					


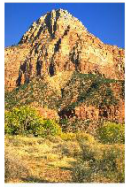






















(a) Elementos recuperados para la clase bosque con imágenes no conocidas

Test	1-NN	2-NN	3-NN	4-NN	5-NN
					
					
					
					











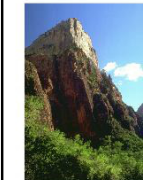













(b) Elementos recuperados para la clase cielo con imágenes no conocidas.

Test	1-NN	2-NN	3-NN	4-NN	5-NN
					
					
					
					













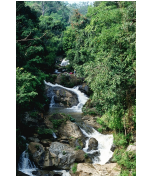

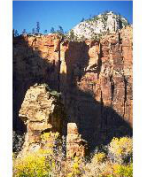

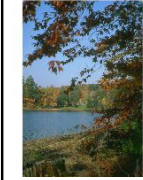







(c) Elementos recuperados para la clase costa con imágenes no conocidas.

Test	1-NN	2-NN	3-NN	4-NN	5-NN
					
					
					
					

(d) Elementos recuperados para la clase montaña con imágenes no conocidas.

Test	1-NN	2-NN	3-NN	4-NN	5-NN
					
					
					
					

(e) Elementos recuperados para la clase pradera con imágenes no conocidas.

Test	1-NN	2-NN	3-NN	4-NN	5-NN
					
					
					
					

(f) Elementos recuperados para la clase río con imágenes no conocidas.

Figura 5.8: Imágenes recuperadas para elementos desconocidos en la base de conocimiento.

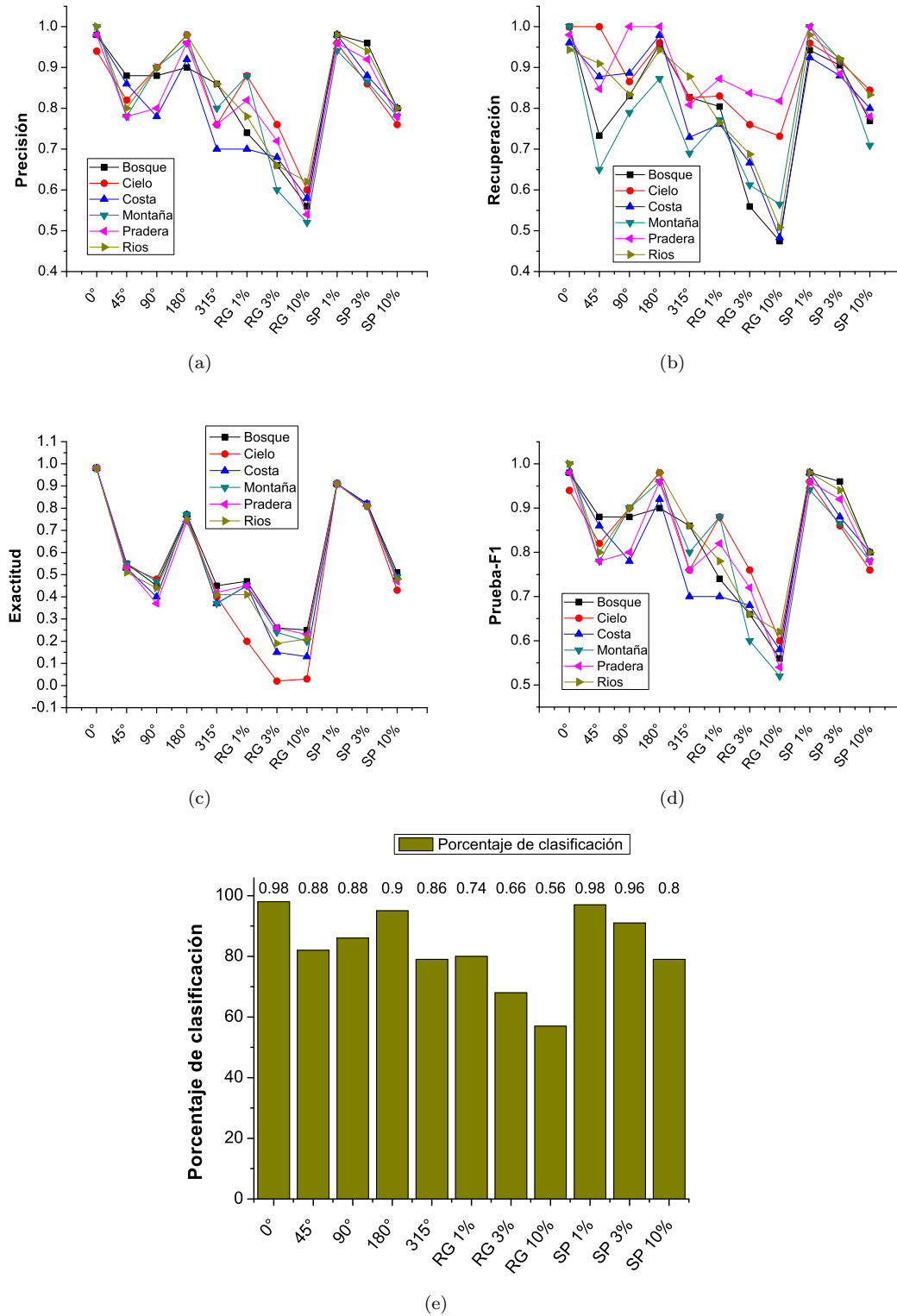
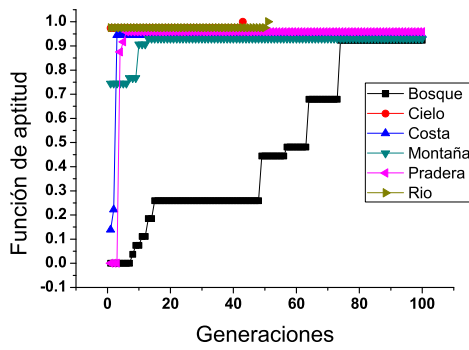
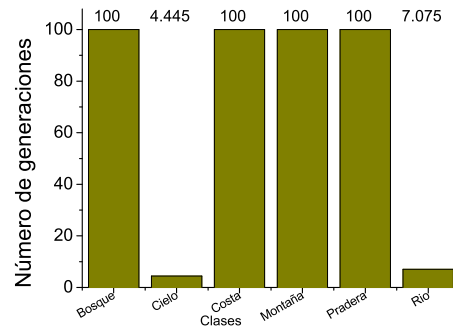


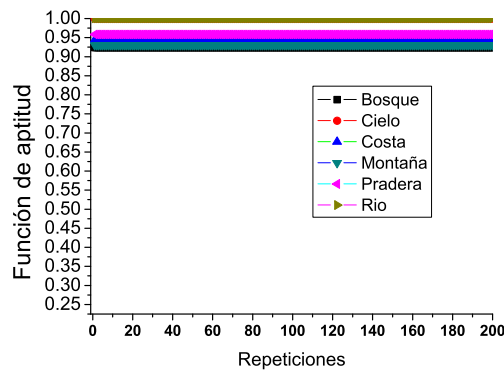
Figura 5.9: Porcentaje de clasificación y medidas de eficiencia para 50 imágenes desconocidas de cada escenario natural.



(a) Representación de una sola ejecución del AG.



(b) Convergencia de la ejecución del AG.



(c) Evaluación del AG con 200 repeticiones y 100 generaciones por repetición.

Figura 5.10: Desempeño del AG para los escenarios naturales contenidos en la BDEN.

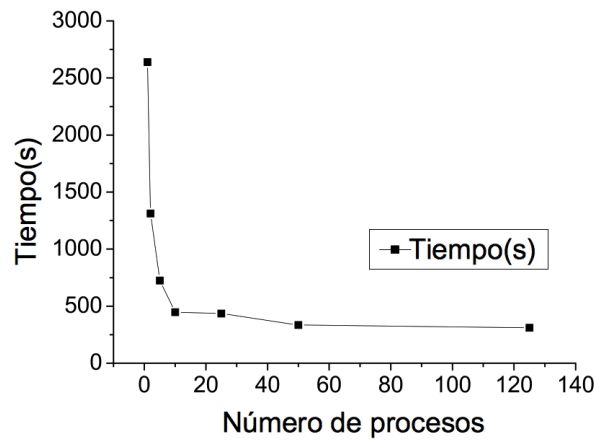


Figura 5.11: Rendimiento del algoritmo paralelo.

Capítulo 6

Conclusiones

En este trabajo se presentó una nueva propuesta para el uso de la teoría de la Causalidad de Wiener-Granger (CWG) respaldada por un algoritmo genético (AG), junto con el análisis de CBIR, aplicado para la identificación y clasificación de 6 paisajes naturales tales como: costa, bosque, montaña, pradera, río/lago y cielo/nube. Teniendo en cuenta la nueva formulación desde la perspectiva de proponer una nueva forma de leer a las imágenes para comprender o tratar de proporcionar un descriptor del contenido de las mismas.

Fue posible encontrar un conjunto de descriptores de las matrices de causalidad para representar una clase de escenario, a partir de un conjunto base de texturas de referencia, proponiendo una caracterización de imágenes basadas en la aparición continua de texturas dentro de ellas; el diccionario base en este enfoque incluía las texturas: Nube, Cielo, Roca y Vegetación. A diferencia de otros enfoques, nuestra metodología aborda las consideraciones de rotación y ruido de imagen, y los resultados muestran porcentajes de clasificación excelentes.

En este enfoque tenemos una clasificación de imágenes del 100% para todo el conjunto de datos, y la metodología proporcionó una tasa de clasificación buena para la prueba de rotación a 180°, por otro lado una sensibilidad para niveles de rotación intermedios (45°, 90°), y tuvo buenos resultados para el ruido de la imagen de sal y pimienta.

El uso del algoritmo evolutivo ayudo con la optimización de resultados, y como se mostró, no necesariamente las relaciones causales con menor fuerza son sujetas a eliminación, además de la necesidad de encontrar los mejores patrones que describan un escenario natural, fue necesaria una búsqueda de soluciones rápida para manejar una gran cantidad de operaciones en tiempos cortos de procesamiento sin perder precisión en las operaciones de punto flotante.

En relación con el rendimiento de la propuesta, se desarrolló el diseño de un algoritmo de computación paralela. Se logró una reducción en los tiempos de ejecución utilizando un servidor de clúster de doble núcleo de 19 procesadores y la herramienta MPI, alcanzando una disminución del 88.9% del tiempo de ejecución de la versión secuencial cuando se iniciaron 125 procesos.

El trabajo futuro incluye el estudio del desempeño de esta propuesta utilizando otras arquitecturas paralelas; por ejemplo, la tecnología de GPU podría funcionar de manera eficiente para la etapa de extracción de características de las imágenes, así como la implementación de tipo de modelos bioinspirados, como por ejemplo la Programación Genética, para analizar todas las texturas de imágenes

que buscan caracterizar todo el escenario natural y sus asociaciones con el paradigma de comprensión visual, mejorar la cantidad de texturas base y automatizar la generación de las mismas por medio de la clusterización.

Bibliografía

- [1] Altaf Alam and Zainul Jaffery. *A Vision-Based System for Traffic Light Detection: SIGMA 2018, Volume 1*, pages 333–343. 01 2019.
- [2] T. Baltrusaitis, P. Robinson, and L. Morency. Openface: An open source facial behavior analysis toolkit. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–10, March 2016.
- [3] Lionel Barnett and Anil K. Seth. The {MVGC} multivariate granger causality toolbox: A new approach to granger-causal inference. *Journal of Neuroscience Methods*, 223:50 – 68, 2014.
- [4] Rafiya Begum and S.V. Halse. The smart car parking system using gsm and labview. *Journal of Computer and Mathematical Sciences*, 9:135–142, 02 2018.
- [5] Md Bhuiyan. Content -based image retrieval for image indexing. *International Journal of Advanced Computer Science and Applications*, 6, 06 2015.
- [6] Sid-Ali Blaifi, S Moulahoum, Ilhami Colak, and Walid Merrouche. Monitoring and enhanced dynamic modeling of battery by genetic algorithm using labview applied in photovoltaic system. *Electrical Engineering*, 100:1–18, 05 2017.
- [7] Sing-Tze Bow. *Pattern Recognition and Image Preprocessing*. Number 77 in Electrical Engineering and Electronics. Dekker, 1992.
- [8] Steven L. Bressler and Anil K. Seth. Wiener-granger causality: A well established methodology. *NeuroImage*, 58(2):323 – 329, 2011.
- [9] C.Benavides, J.Villegas-Cortez, G.Roman, and C.Aviles-Cruz. Reconocimiento de rostros a partir de la propia imagen usando técnica cbir. *X Congreso Español sobre Metaheurísticas, Algoritmos Evolutivos y Bioinspirados (MAEB 2015)*, pages 733–740, 2015.
- [10] C.Benavides, J.Villegas-Cortez, G.Roman, and C.Aviles-Cruz. Face classification by local texture analysis through cbir and surf points. *IEEE Latin America Transactions*, 14:2418–2424, 2016.
- [11] Lalitha Dabbiru, James V. Aanstoos, John E. Ball, and Nicolas H. Younan. Screening mississippi river levees using texture-based and polarimetric-based features from synthetic aperture radar data. *Electronics*, 6(2), 2017.

- [12] Charles Darwin. *On the Origin of Species by Means of Natural Selection*. Murray, London, 1859. or the Preservation of Favored Races in the Struggle for Life.
- [13] J.P.M. de Sa. *Pattern Recognition: Concepts, Methods, and Applications*. Springer, 2001.
- [14] Kalyanmoy Deb and Deb Kalyanmoy. *Multi-Objective Optimization Using Evolutionary Algorithms*. John Wiley & Sons, Inc., New York, NY, USA, 2001.
- [15] Rupali Desai and Bhakti Sonawane. Gist, hog, and dwt-based content-based image retrieval for facial images. 468:297–307, 08 2017.
- [16] M. Ding, Y Chen, and S. L. Bressler. Granger causality: Basic theory and application to neuroscience. Handbook of Time Series Analysis. Björn Schelter, Matthias Winterhalder, Jens Timmer, WILEY-VCH Verlag GmbH Co., 2006.
- [17] E.Cuevas and D.Zaldívar. *Procesamiento digital de imágenes usando MatLAB y Simulink*. Alfaomega, 1ra edition, Septiembre 2010.
- [18] R Fablet, P. Bouthemy, and P. Pérez. Nonparametric motion characterization using causal probabilistic models for video indexing and retrieval. *IEEE TRANSACTIONS ON IMAGE PROCESSING*, 11(4):393–407, 2002.
- [19] Y. Fan, H. Yang, S. Zheng, H. Su, and S. Wu. Video sensor-based complex scene analysis with granger causality. *Sensor*, 13(1):13685–13707, 2013.
- [20] Yawen Fan, Hua Yang, Shibao Zheng, Hang Su, and Shuang Wu. Video sensor-based complex scene analysis with granger causality. *Sensors*, 13(10):13685, 2013.
- [21] Karl Friston. Causal modelling and brain connectivity in functional magnetic resonance imaging. *PLOS Biology*, 7(2):1–6, 02 2009.
- [22] Keinosuke Fukunaga. *Introduction to statistical pattern recognition (2nd ed.)*. Academic Press Professional, Inc., San Diego, CA, USA, 1990.
- [23] J. Gao, J. Yang, J. Zhang, and M. Li. Natural scene recognition based on convolutional neural networks and deep boltzmann machines. In *2015 IEEE International Conference on Mechatronics and Automation (ICMA)*, pages 2369–2374, Aug 2015.
- [24] John Geweke. Measurement of linear dependence and feedback between multiple time series. *Journal of the American Statistical Association*, 77(378):304–313, 1982.
- [25] David E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition, 1989.
- [26] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1992.
- [27] Clive Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–38, 1969.

- [28] Olague Gustavo. *Evolutionary Computer Vision: The First Footprints*. Natural Computing Series. Springer-Verlag Berlin Heidelberg, 1st edition, 2016.
- [29] J-Vogel and B-Schiele. Semantic modeling of natural scenes for content-based image retrieval. *International Journal of Computer Vision*, 72(2):133–157, 2007.
- [30] J.F.Serrano-Talamantes, C.Aviles-Cruz, J.Villegas-Cortez, and J.H.Sossa-Azuela. Self organizing natural scene image retrieval. *Expert Systems with Applications: An International Journal*, 40(7):2398–2409, 2012.
- [31] Peter Karpov, Giovanni Squillero, and Alberto Tonda. Valis: an evolutionary classification algorithm. *Genetic Programming and Evolvable Machines*, 19(3):453–471, Sep 2018.
- [32] Toshikazu Kato. Database architecture for content-based image retrieval. In Albert A. Jamberdino and Carlton Wayne Niblack, editors, *Image Storage and Retrieval Systems*, volume 1662, pages 112 – 123. International Society for Optics and Photonics, SPIE, 1992.
- [33] Amandeep Khokher and Rajneesh Talwar. Content-based image retrieval: State-of-the-art and challenges. 08 2011.
- [34] Eunwoo Kim, Dae-Shik Kim, Fayyaz Ahmad, and Hyunwook Park. Pattern-based granger causality mapping in fmri. *Brain connectivity*, 3, 09 2013.
- [35] John R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA, USA, 1992.
- [36] Dalwinder Kular and Eraldo Ribeiro. *Analyzing Activities in Videos Using Latent Dirichlet Allocation and Granger Causality*, pages 647–656. Springer International Publishing, Cham, 2015.
- [37] Ying Liu, Dengsheng Zhang, Guojun Lu, and Wei-Ying Ma. A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, 40(1):262 – 282, 2007.
- [38] Springer Nature Singapore Pte Ltd. *Content-Based Image Retrieval: Ideas, Influences, and Current Trends*. 2017.
- [39] David G. Luenberger. *Programación lineal y no lineal*. Addison - Wesley Iberoamericana, Buenos Aires, 1a edición edition, 1989.
- [40] Gonzalo Pajares M. and Jesús M. de la Cruz G. *Visión por computador imágenes digitales y aplicaciones*. Alfaomega, 2 edition, Marzo 2008.
- [41] M-Gavrilova M-Sultana. A content based feature combination method for face recognition. 226:197–206, 2013.
- [42] D Madhavi and Ramesh Patnaik. *Genetic Algorithm-Based Optimized Gabor Filters for Content-Based Image Retrieval*, pages 157–164. 01 2018.
- [43] Michael Mannino and Steven L. Bressler. Foundational perspectives on causality in large-scale brain networks. *Physics of Life Reviews*, 15:107 – 123, 2015.

- [44] Alianna J. Maren, Craig T. Harston, and Robert M. Pap. *Handbook of Neural Computing Applications*. Academic Press, Inc., Orlando, FL, USA, 1990.
- [45] Yuliana Martínez, Leonardo Trujillo, Pierrick Legrand, and Edgar Galvan-Lopez. Prediction of expected performance for a genetic programming classifier. *Genetic Programming and Evolvable Machines*, 17, 02 2016.
- [46] Fernanda S. Matias, Leonardo L. Gollo, Pedro V. Carelli, Steven L. Bressler, Mauro Copelli, and Claudio R. Mirasso. Modeling positive granger causality and negative phase lag between cortical areas. *NeuroImage*, 99:411 – 418, 2014.
- [47] Fanjie Meng, Xinqing Wang, Faming Shao, Dong Wang, and Xia Hua. Energy-efficient gabor kernels in neural networks with genetic algorithm training method. *Electronics*, 8(1), 2019.
- [48] Sayan Nag. Vector quantization using the improved differential evolution algorithm for image compression. *Genetic Programming and Evolvable Machines*, Jan 2019.
- [49] N.Wiener. *The theory of prediction*. Modern mathematics for engineers Ed. McGraw-Hill, 1956.
- [50] Aude Oliva and Antonio Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int. J. Comput. Vision*, 42(3):145–175, May 2001.
- [51] Y. Pérez-Pimentel, I. Osuna-Galan, J. Villegas-Cortez, and C. Avilés-Cruz. A genetic algorithm applied to content-based image retrieval for natural scenes classification. In *2014 13th Mexican International Conference on Artificial Intelligence*, pages 155–161, Nov 2014.
- [52] Y. Pérez-Pimentel, I. Osuna-Galan, J. Villegas-Cortez, and C. Avilés-Cruz. A genetic algorithm applied to content-based image retrieval for natural scenes classification. In *2014 13th Mexican International Conference on Artificial Intelligence*, pages 155–161, Nov 2014.
- [53] K. Prabhakar, S. Oh, P. Wang, G. D. Abowd, and J. M. Rehg. Temporal causality for the analysis of visual events. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1967–1974, June 2010.
- [54] Azam Shirali, Javidan Kazemi Kordestani, and Mohammad Reza Meybodi. Self-adaptive multi-population genetic algorithms for dynamic resource allocation in shared hosting platforms. *Genetic Programming and Evolvable Machines*, 19(4):505–534, Dec 2018.
- [55] W. Tao, H. Jin, and Y. Zhang. Color image segmentation based on mean shift and normalized cuts. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 37(5):1382–1389, Oct 2007.
- [56] A. J. M. Traina, A. G. R. Balan, L. M. Bortolotti, and C. Traina. Content-based image retrieval using approximate shape of objects. In *Computer-Based Medical Systems, 2004. CBMS 2004. Proceedings. 17th IEEE Symposium on*, pages 91–96, June 2004.
- [57] V-S.Nalwa. *A Guided Tour of Computer Vision*, volume 1 of *TA1632*. Addison Wesley, 1993.

- [58] Julia Vogel and Bernt Schiele. Performance evaluation and optimization for content-based image retrieval. *Pattern Recognition*, 39:897–909, 2006.
- [59] Robert A. Yaffee and Monnie McGee. *Introduction to Time Series Analysis and Forecasting: With Applications of SAS and SPSS*. Academic Press, Inc., USA, 1st edition, 2000.
- [60] Lotfi A. Zadeh and John R. Ragazzini. An extension of wiener’s theory of prediction. *Journal of Applied Physics*, 21(7):645–655, 1950.
- [61] Peilong Zeng, Zhixin Li, and Canlong Zhang. Scene Classification Using Spatial and Color Features. In Zhongzhi Shi, Zhaohui Wu, David Leake, and Uli Sattler, editors, *8th International Conference on Intelligent Information Processing (IIP)*, volume AICT-432 of *Intelligent Information Processing VII*, pages 259–268, Hangzhou, China, October 2014. Springer. Part 7: Multimedia.
- [62] Cong Zhang, Xiaokang Yang, Weiyao Lin, and Jun Zhu. Recognizing human group behaviors with multi-group causalities. In *Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology - Volume 03, WI-IAT ’12*, pages 44–48, Washington, DC, USA, 2012. IEEE Computer Society.
- [63] Hui Zhang and Xiaoting Li. Effective connectivity of facial expression network by using granger causality analysis. *Parallel Processing of Images and Optimization and Medical Imaging Processing*, 8920, 2013.
- [64] Mengjie Zhang and Phillip Wong. Genetic programming for medical classification: a program simplification approach. *Genetic Programming and Evolvable Machines*, 9(3):229–255, Sep 2008.

Anexo 1

Glosario de términos

A continuación se muestran los conceptos básicos para entender la manipulación de imágenes digitales, iniciando por el análisis de la disposición de los píxeles que la conforman, su métrica, región, etc.:

Métrica

Expresa la proximidad entre dos objetos:

Distancia entre dos puntos (distancia Euclidiana), distancia de un punto a una recta, distancia de un punto a un plano.

Ruta (Path). Se define como el camino entre dos puntos (pasos) moviendo entre vecindades.

Región conectada. $b(x, y) \in R$ es un conjunto de puntos $p(x, y)$, tales que existe un camino que comienza en p_1 y termina en p_2 para todo $p_1, p_2 \in R$.

Vecindades de un Píxel

Un píxel p de coordenadas (x, y) tiene dos vecinos horizontales y dos vecinos verticales cuyas coordenadas están dadas por:

$$(x + 1, y), (x - 1, y), (x, y + 1), (x, y - 1)$$

Este conjunto se llama: vecindad-4 de p y se denota como $N_4(p)$, véase la Figura 1.1.

Un píxel p de coordenadas (x, y) tiene 4 vecinos diagonales cuyas coordenadas están dadas por:

$$(x + 1, y + 1), (x + 1, y - 1), (x - 1, y + 1), (x - 1, y - 1)$$

Se denota como $N_8(p)$. A estos puntos junto con los de vecindad-4, se les llama vecindad-8 de p , como se muestra en la Figura 1.2.

		2		
	2	1	2	
2	1	0	1	2
	2	1	2	
		2		

Figura 1.1: Vecindad $N_4(p, q) = |x - s| + |y - t|$.

2	2	2	2	2
2	1	1	1	2
2	1	0	1	2
2	1	1	1	2
2	2	2	2	2

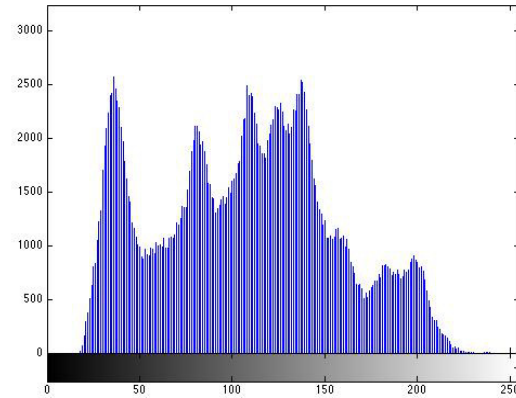
Figura 1.2: Vecindad $N_8(p, q) = \max(|x - s|, |y - t|)$.

Figura 1.3: Histograma de un rostro Pre-procesado.

1.0.1. Información estadística de una imagen.

Entendemos por información estadística de una imagen digital, al cúmulo de valores extraídos a partir de la información de la conformación de la propia imagen (histograma, media, varianza, desviación estándar) [26][22]. A continuación se mencionan las definiciones de cómo se definen los valores estadísticos más usados en el campo del procesamiento digital de imágenes.

1. Histograma.

El histograma de una imagen digital con niveles de gris en el rango $0, L - 1$ es una función discreta $h(r_k) = n_k/L$ donde r_k es el nivel de gris k -ésimo y n_k es el número de píxeles en la imagen con nivel de gris r_k , como se ejemplifica en la Figura 1.3. Con un simple vistazo, puede proporcionarnos una idea muy aproximada de la distribución de niveles de gris de la imagen, el contraste y el brillo que presenta [17].

2. Media.

Consiste en generar una nueva imagen $g(x, y)$ cuya intensidad para cada píxel se obtiene promediando los valores de intensidad de los píxeles $f(x, y)$ incluidos en un entorno de vecindad predefinido.

$$\bar{f}(x, y) = \frac{1}{n} \sum_{k=1}^n f_k(x, y) \quad (1.1)$$

3. Varianza.

Una medida simple del contraste de la imagen. Es una medida de la diferencia existente entre todos los valores de la muestra y la media, se define matemáticamente como el promedio del cuadrado de esas diferencias.

$$\alpha^2(x, y) = \frac{1}{n-1} \sum_{k=1}^n (f_p(x, y) - f_k(x, y))^2 \quad (1.2)$$

4. Desviación estándar.

$$\alpha(x, y) = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (f_p(x, y) - f_k(x, y))^2} \quad (1.3)$$

1.0.2. Acondicionado de la imagen.

Entendemos al acondicionamiento de la imagen como la adecuación de la misma, su preprocesamiento, acorde a un fin específico.

■ Ruido en imágenes.

El ruido en una imagen es cualquier entidad o resultado intermedio no interesante para los propósitos del cómputo principal. El ruido puede resultar de varias causas:

- La óptica y electrónica asociadas al sistema de adquisición.
- Las condiciones de iluminación.
- Las sombras creadas entre objetos.
- Las condiciones atmosféricas.

■ Algunos tipos de ruido y su ejemplo en la Figura 1.4.

- **Ruido aditivo.** $\tilde{f}(x, y)$ es obtenida a partir de $f(x, y)$, en este caso, al adicionar al azar a cada píxel de $f(x, y)$ un real c , tal que:

$$\tilde{f}(x, y) = f(x, y) + c \quad (1.4)$$

- **Ruido substractivo.** $\tilde{f}(x, y)$ es obtenida a partir de $f(x, y)$, en este caso, al restar al azar a cada píxel de $f(x, y)$ un real c , tal que:

$$\tilde{f}(x, y) = f(x, y) - c \quad (1.5)$$

- **Ruido mezclado.** $\tilde{f}(x, y)$ es obtenida a partir de $f(x, y)$, en este caso, al sumar o restar al azar a cada píxel de $f(x, y)$ un real c , tal que se tenga uno de los casos anteriores.

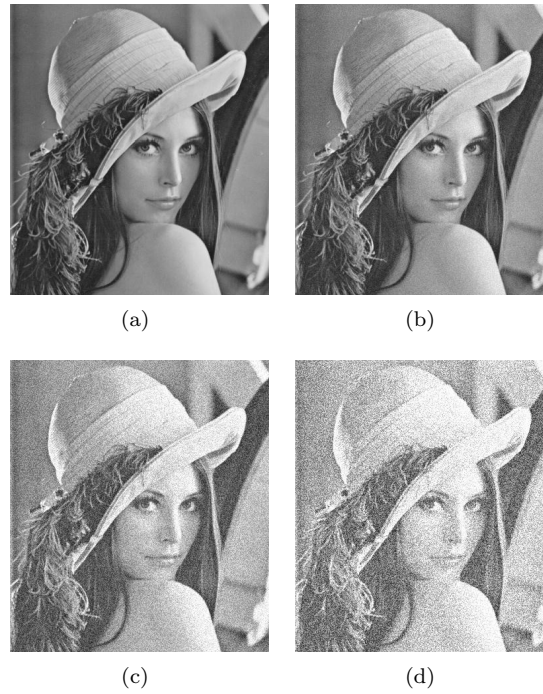


Figura 1.4: Ejemplo de una imagen (a) sin ruido, (b) con ruido gaussiano al 10%, (c) con ruido gaussiano al 30%, (d) con ruido gaussiano al 50%.

- **Ruido gaussiano.** Este tipo de ruido cumple con una distribución gaussiana, en otras palabras, la variable c toma valores como se observa en la ecuación 1.6 [26] véase la Figura 1.4:

$$c = (L - 1) * e^{\frac{-z^2}{2\sigma^2}} \quad (1.6)$$

con $z \geq 0$ real, y σ es el parámetro de amplitud de la Gaussiana.

- **Ruido sal.** Se caracteriza por saturar el valor de uno o más píxeles de la imagen hacia el límite superior[26] Figura 1.3 b).

$$\tilde{f}(x, y) = f(x, y) + b * c \quad (1.7)$$

con $b = 1$ y $c = L - 1$

- **Ruido pimienta.** Se caracteriza por saturar el valor de uno o más píxeles de la imagen hacia el límite inferior[26] Figura 1.5 c).



Figura 1.5: ejemplo de una imagen digital: (a) sin ruido, (b) contaminada con ruido sal, (c) contaminada con ruido pimienta.

$$\tilde{f}(x, y) = f(x, y) + b * c \quad (1.8)$$

$$\text{con } b = 1 \text{ y } c = 0$$

■ El filtrado de una imagen.

Las imágenes al ser adquiridas a través de un medio óptico o electrónico es susceptible a ser degradada por el ambiente de captación.

El objetivo de las técnicas de filtrado consiste en reducir o eliminar, en el mejor de los casos, el ruido presente en una imagen.

Debido a que el modelado del ruido es algo muy complicado, en el mejor de los casos lo que se logra es una reducción del ruido presente en la señal.

Este puede llevarse a cabo en el *dominio de la frecuencia*, como en el *dominio espacial*.

● Filtrado en el dominio de la frecuencia

1. Se calcula la transformada de Fourier de la imagen a ser filtrada,
2. luego el resultado es multiplicado por una función de transferencia de filtrado adecuada.
3. Finalmente, al resultado obtenido se le calcula la transformada inversa de Fourier para obtener la imagen mejorada.

● Filtrado en el dominio espacial.

- En este caso se trabaja en el mismo dominio de los píxeles al cambiar el valor de un píxel tomando en cuenta el valor de dicho píxel y el de alguno de sus vecinos.

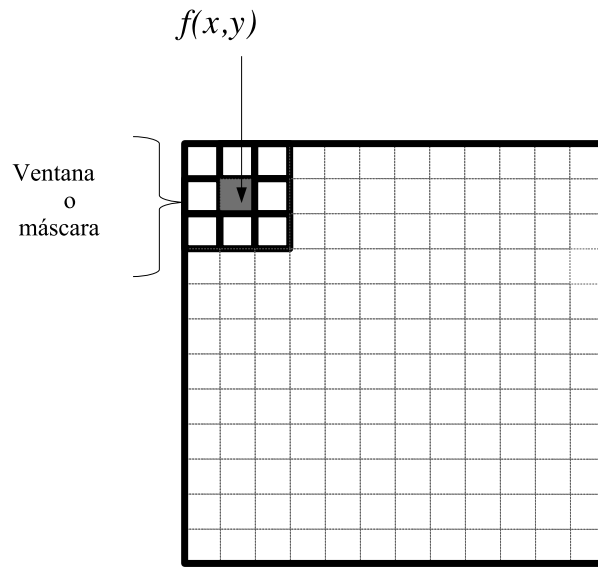


Figura 1.6: Ejemplo de una ventana o máscara cuadrada de 3x3 aplicada a una imagen.

○ → fácil y rápido de implementar.

■ **Filtros lineales.**

La operación de un filtro en el dominio espacial puede expresarse como:

$$s(x, y) = T[f(x, y)] \quad (1.9)$$

con f la imagen dada, s la imagen transformada y T es el operador sobre f , definido en una vecindad de (x, y) .

La vecindad puede ser cualquiera, las más usadas son la cuadrada Figura 1.6 y la circular, ambas con centro en (x, y) .

El efecto de aplicar una operación T dentro de una ventana deslizante e.g 3x3, no es más que situar la ventana dentro de la imagen e ir la recorriendo:

1. Se sitúa la ventana dentro de la imagen y se aplica el operador T al píxel del centro como se mostró en la Figura 1.6,
2. se desliza la venta una columna hacia la derecha y se ejecuta el operador de nuevo,
3. una vez que el borde de la ventana llega al final de la imagen, esta se pasa al siguiente renglón Figura 1.7,
4. cuando los bordes de la ventana llega al último renglón y a la última columna, se termina de aplicar el operador T y se tiene la imagen procesada.

Los principales objetivos de la aplicación de los *filtros* son:

- Suavizar la imagen.

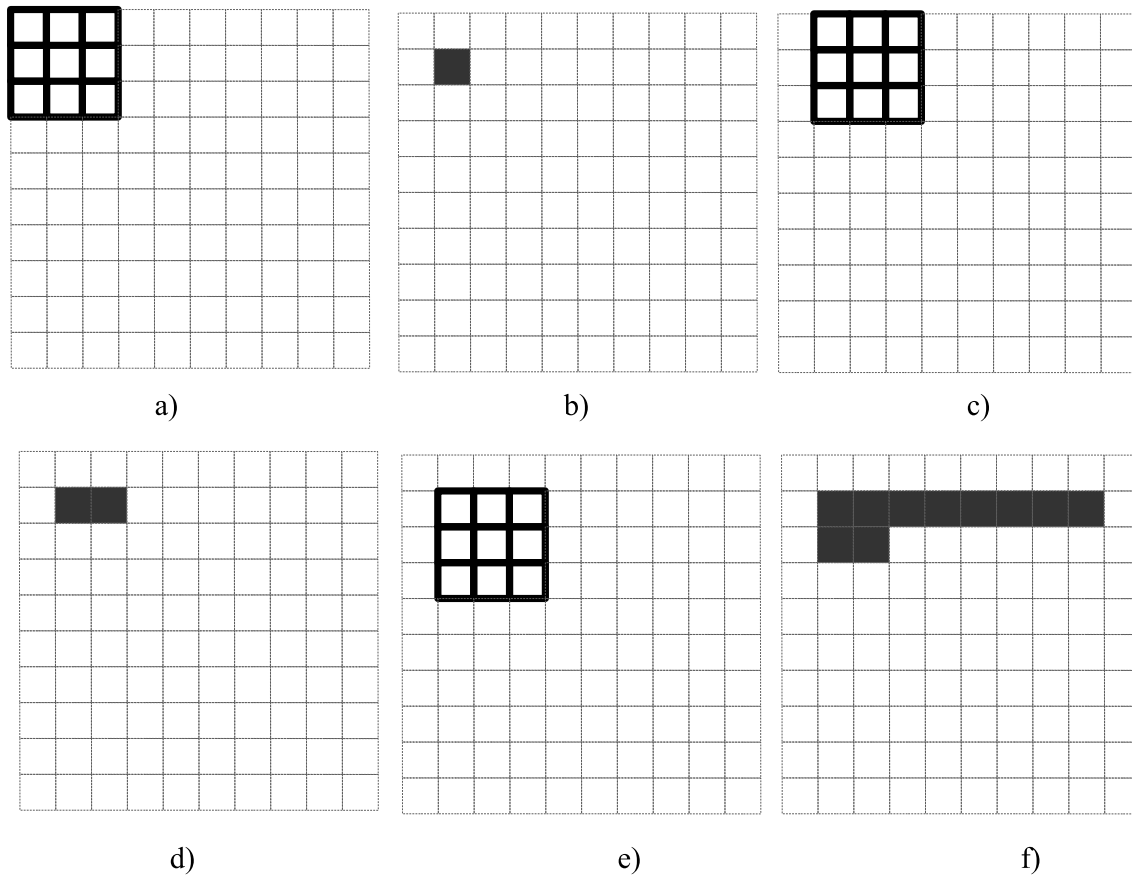


Figura 1.7: Efecto de la aplicación de *T-Ventana deslizante*.

- Eliminar ruido.
- Detectar bordes.

■ Tipos de filtros

- **Filtro promedio aritmético.** Estos tipo de filtros encuentran algún tipo de promedio sobre una ventana de $n \times n$ píxeles. El *filtro promedio aritmético* es el filtro más simple de ellos, halla el valor promedio de los píxeles dentro de la ventana ec. (1.10):

$$f_{pa} = \frac{1}{n^2} \sum_{(x,y) \in V} f(x,y) \quad (1.10)$$

El resultado de aplicar un filtro promedio aritmético, es el de “suavizar” variaciones locales en la señal \rightarrow filtro “pasa-bajos” Figura 1.8. Tiende a desenfocar la imagen \rightarrow mitiga los efectos de ruido.

Este es u filtro muy fácil de entender, nos ayuda a comprender la forma de aplicación de los filtros, y la formación y aplicación de la ventana o máscara. Pero no es un filtro eficiente:

- Las señales con frecuencias comunes a las del ruido son atenuadas.



Figura 1.8: (a) imagen original, (b) imagen con ruido gaussiano al 20 %, (c) imagen filtrada con ventana 3x3 filtro promedio.

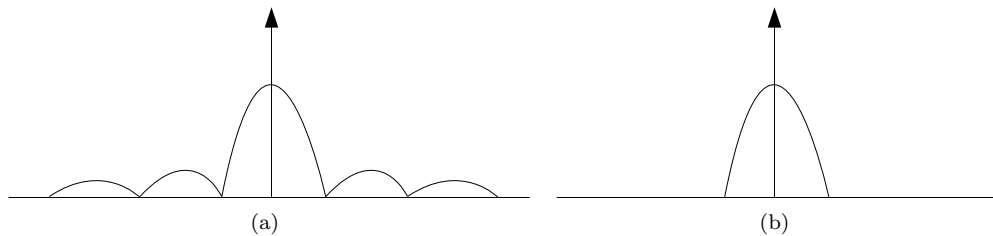


Figura 1.9: (a) Transformada de Fourier del núcleo promedio aritmético $1 - D$, (b) Transformada de Fourier del núcleo gaussiano $1 - D$.

- El ruido de tipo impulsivo es solo atenuado más no removido e.g. Ruido sal o ruido pimienta.
- **Filtro gaussiano.** Es un tipo especial de filtro promedio, cuyos pesos siguen el comportamiento de una gaussiana.

En el dominio de la frecuencia, la transformada de Fourier de un filtro promedio aritmético uno-dimensional de ancho $2W$ Figura 1.8 esta dada por:

$$\text{sinc}(\omega) = \frac{2\text{sen}(\omega W)}{\omega} \quad (1.11)$$

- **Observaciones**

Esto significa que convolucionar una imagen f con un núcleo gaussiano G es equivalente a convolucionar primero todas las filas de f , y después todas las columnas con un núcleo gaussiano uno-dimensional con la misma σ .

Una de las ventajas obvias de usar un núcleo uno-dimensional en lugar de su equivalente núcleo bidimensional es la reducción de tiempo obtenida.

- **Construcción de núcleos gaussianos**

Existen varias formas de calcular los coeficientes o pesos de un núcleo gaussiano. Una buena aproximación de los coeficientes de un núcleo gaussiano uno-dimensional se puede obtener a través de los coeficientes de la expansión binomial:

$$(1 + x)^n = \binom{n}{0} + \binom{n}{1}x + \binom{n}{2}x^2 + \dots + \binom{n}{n}x^n \tag{1.12}$$

→ equivalente a tomar la n-ésima fila del triángulo de pascal.

$$\begin{array}{ccccccc}
 & & & & 1 & & & \\
 & & & & 1 & & 1 & \\
 & & & 1 & 2 & & 1 & \\
 & & 1 & 3 & 3 & & 1 & \\
 & 1 & 4 & 6 & 4 & & 1 & \\
 1 & 5 & 10 & 10 & 5 & & 1 & \\
 1 & 6 & 15 & 20 & 15 & 6 & 1 & \\
 & & & & \cdot & & & \\
 & & & & \cdot & & & \\
 & & & & \cdot & & &
 \end{array}$$

Así los coeficientes de un filtro uno-dimensional de tamaño cinco son:

$$1 \quad 4 \quad 6 \quad 4 \quad 1$$

Esta máscara puede ser usada para alisar primero las filas de la imagen de entrada y luego las columnas.

Desventaja: no se tiene control sobre el parámetro σ .

- **Alternativa** Otra forma de diseñar un núcleo gaussiano consiste en calcular los pesos de la ventana al usar directamente la ecuación gaussiana (1.13).

$$G(x, y) = ce^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1.13}$$

Al escoger el valor de σ^2 se puede evaluar el núcleo para un tamaño de ventana n dado. e.g., para $\sigma = 2$ y $n = 7$ se tiene:

(x,y)	-3	-2	-1	0	1	2	3
-3	0.011	0.039	0.082	0.105	0.082	0.039	0.011
-2	0.039	0.135	0.287	0.368	0.287	0.135	0.039
-1	0.082	0.287	0.606	0.779	0.606	0.287	0.082
0	0.105	0.368	0.779	1.0	0.779	0.368	0.105
1	0.082	0.287	0.606	0.779	0.606	0.287	0.082
2	0.039	0.135	0.387	0.368	0.387	0.135	0.039
3	0.011	0.039	0.082	0.105	0.082	0.039	0.011

Es necesario re-escalar para que el valor más pequeño sea 1.0. para esto se calcula el valor: $k = \frac{1}{0.011} = 91$



Figura 1.10: (a) Imagen original, (b) imagen contaminada con ruido gaussiano al 20 %, (c) Imagen procesada con filtro gaussiano ventana 7x7.

1	4	7	10	7	4	1
4	12	26	33	26	12	4
7	26	55	71	55	26	7
10	33	71	91	71	33	10
7	26	55	71	55	26	7
4	12	26	33	26	12	4
1	4	7	10	7	4	1

Para evitar la saturación en los valores de la imagen resultante al aplicar el núcleo obtenido, se agrega un factor de normalización, que se obtiene de la suma de todos los

$$\text{pesos del núcleo: } c = \sum_{h=-\frac{n}{2}}^{\frac{n}{2}} \sum_{k=-\frac{n}{2}}^{\frac{n}{2}} G(h, k) = 1115$$

Tabla 1.1: Filtro gaussiano con ventana de 7x7.

	1	4	7	10	7	4	1
	4	12	26	33	26	12	4
	7	26	55	71	55	26	7
$\frac{1}{1115}$	10	33	71	91	71	33	10
	7	26	55	71	55	26	7
	4	12	26	33	26	12	4
	1	4	7	10	7	4	1

- **Filtro promedio contra-armónico**

Funciona muy bien en presencia de ruido sal o pimienta dependiendo del orden del filtro usado.

si \mathbf{R} es el orden del filtro entonces en la ec (1.14) tenemos:

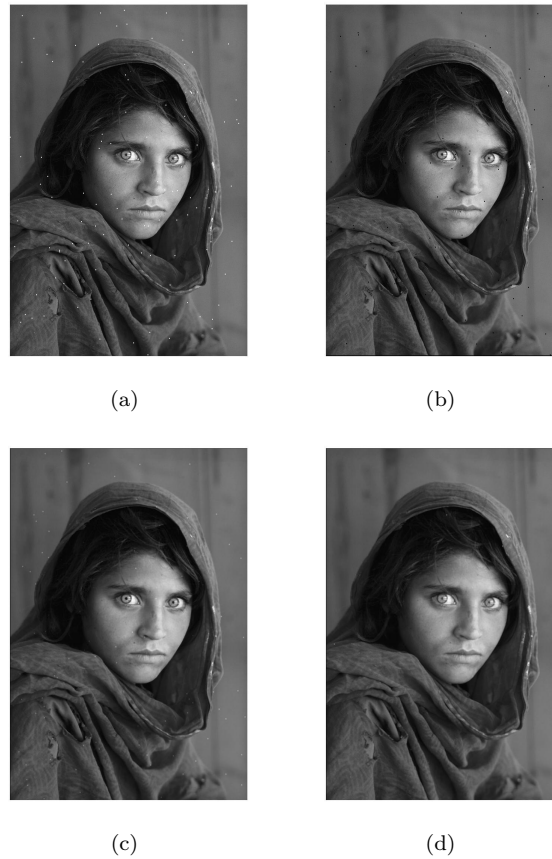


Figura 1.11: (a)Imagen con ruido sal, (b)imagen con ruido pimienta, (c)Filtro contra-harmónico con $R = -1$ para la Figura a, (d)filtro corta-harmónico con $R = 1$ para la Figura b.

$$f_{ca} = \frac{\sum_{(x,y) \in V} f(x,y)^{R+1}}{\sum_{(x,y) \in V} f(x,y)^R} \quad (1.14)$$

si $R < 0$, el filtro elimina el ruido tipo sal de la Figura 1.11 a) y se obtiene como resultado la Figura 1.11 c),

si $R > 0$, el filtro elimina el ruido tipo pimienta de la Figura 1.11 b) y se obtiene como resultado la Figura 1.11 d),

si $R=0$, realiza la media aritmética (*filtro promedio*).

Funciona de una u otra forma, pero no de forma simultánea.

- **Filtro promedio geométrico**

Funciona muy bien en presencia de ruido gaussiano, por encima del filtro promedio aritmético Figura 1.12.



Figura 1.12: (a)Imagen original, (b)Imagen con ruido gaussiano al 5%, (c)Imagen con filtro promedio geométrico ventana 3x3.



Figura 1.13: (a)Imagen original, (b)imagen contaminada con ruido SAL, (c)Imagen procesada con filtro Promedio Harmónico ventana 3x3.

$$f_{pg} = \prod_{(x,y) \in V} [f(x,y)]^{\frac{1}{n^2}} \quad (1.15)$$

- **Filtro promedio armónico** Tiene un buen desempeño ante presencia de ruido tipo sal Figura 1.13, pero no así con el ruido tipo pimienta.

$$f_{ph} = \frac{n^2}{\sum_{(x,y) \in V} \frac{1}{f(x,y)}} \quad (1.16)$$

Al igual que el filtro promedio geométrico, éste se comporta bien ante la presencia de ruido gaussiano, respetando también mejor el detalle de la imagen original que el filtro promedio aritmético Figura 1.13.

- **Filtro Y_p**

$$f_{Y_p} = \left[\sum_{(x,y) \in V} \frac{f(x,y)^p}{n^2} \right]^{1/p} \quad (1.17)$$

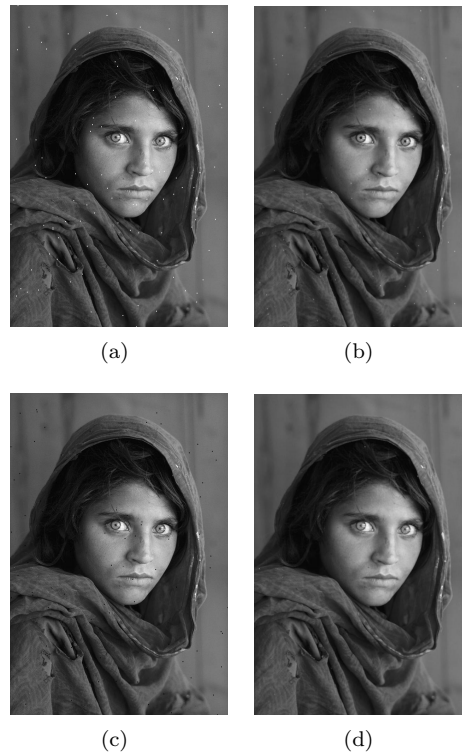


Figura 1.14: (a) Imagen contaminada con ruido SAL, (b) Imagen procesada con filtro Y_p ventana 7×7 y $p = -3$, (c) Imagen contaminada con ruido PIMIENTA, (d) Imagen procesada con filtro Y_p ventana 7×7 y $p = 3$.

- Remueve ruido tipo SAL para $p < 0$.
- Remueve ruido tipo PIMIENTA para $p > 0$
- **Filtros laplacianos** Se usan comúnmente para resaltar los detalles en una imagen. Este tipo de filtros tiene a mejorar los detalles en todas las direcciones por igual. La Figura 1.15 muestra dos máscaras de 3×3 para un filtro del tipo Laplaciano que se puede aplicar a una imagen.
- **Filtros diferenciadores.** Al contrario de los filtros Laplacianos, mejoran los detalles de la imagen en la dirección especificada de acuerdo a la máscara seleccionada Figura 1.16
- **Filtros enfatizadores del altas frecuencias “High-Boost”.** Una imagen filtrada en sus altas frecuencias puede ser obtenida como la diferencia entre la imagen original y su

0	-1	0
-1	5	-1
0	-1	0

1	2	1
-2	5	-2
1	-2	1

Figura 1.15: Máscaras de 3×3 para resaltar detalles filtro Laplaciano.

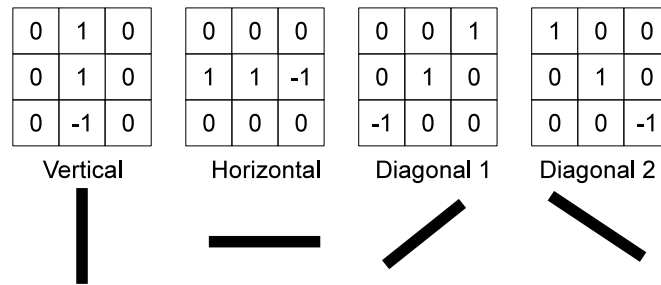


Figura 1.16: Máscaras de 3x3 para resaltar líneas filtro Diferenciador.

versión pasa bajo.

$$f_a = f - f_b \quad (1.18)$$

Si se multiplica por un factor de amplificación A , tenemos un *filtro elevador* o *filtro enfatizador de altas frecuencias*.

$$f_e = A \cdot f - f_b \quad (1.19)$$

La ecuación (1.19) puede descomponerse como:

$$\begin{aligned} f_e &= A \cdot f - f_b \\ &= (A - 1)f + f - f_b \\ f_e &= (A - 1)f + f_a \end{aligned} \quad (1.20)$$

Si $A = 1 \rightarrow$ da como resultado la versión pasa-alto estándar.

Si $A > 1 \rightarrow$ parte de la información de la imagen original es adicionada a la imagen filtrada en sus altas frecuencias \rightarrow permite restaurar parte de las componentes de bajas frecuencias perdidas durante el filtrado pasa-alto de la imagen \rightarrow mejora los bordes de los objetos dependiendo el valor de A . El filtro elevador puede ser implementado por medio de la máscara ponderada que se muestra en la Figura 1.17 cuyo peso central puede ser obtenido como:

$$P = 9 \cdot A - 1 \text{ para } A > 0$$

El valor de A determina la naturaleza del filtro.

1.0.3. Modelos de color

El propósito de un modelo de color (también llamado espacio de color o sistema de colores) es facilitar la especificación de colores en alguna forma estándar de aceptación general. En esencia, un

$$\frac{1}{9} \begin{array}{|c|c|c|} \hline -1 & -1 & -1 \\ \hline -1 & P & -1 \\ \hline -1 & -1 & -1 \\ \hline \end{array}$$

Figura 1.17: Máscara de 3x3 para el filtro enfatizador.

modelo de color es una especificación de un sistema coordinado y un subespacio dentro de ese sistema donde cada color se representa por un punto.

Los modelos de color que existen son: RGB, NTSC, YCbCr, HSV, CMY, CMYK y HSI (ver [13, 16]).

La mayoría de los modelos de color hoy en día están orientados ya sea al hardware (como monitores a color o impresoras) o a alguna aplicación donde el objetivo es la manipulación del color (como la creación de gráficas a color para animaciones). En términos de procesamiento digital de imágenes, los modelos orientados al hardware más comúnmente utilizados en la práctica son el modelo RGB (rojo, verde, azul) modelo para monitores a color y para una gran variedad de cámaras de video en color; el CMY (cyan, magenta, amarillo) y el CMYK (cyan, magenta, amarillo, negro), modelos para impresoras en color; y el HSI (brillo (hue), saturación, intensidad) que corresponde al modelo más cercano a la manera en que los humanos percibimos el color. A continuación se describen los espacios de color RGB y HSI.

Una imagen de color RGB es una matriz de píxeles de tamaño $M \times N \times 3$, donde cada pixel es una combinación de colores rojo, verde y azul. Por convención, cada matriz se referencia como componente rojo, verde y azul de la imagen RGB. Una imagen RGB es de 8 bits (uint8) si su rango de valores es de $[0,255]$, o de 16 bits (uint16) para $[0,65535]$, o double para $[0,1]$. El número de bits empleados para representar el pixel de las componentes de la imagen determinan la profundidad de bit (*bit depth*). Por ejemplo, si cada una de las componentes de la imagen es de 8 bits, entonces la imagen RGB correspondiente se dice que es de 24 bits de fondo. El número total de colores en una imagen RGB de 24 bits es de $(2^8)^3 = 16,777,216$. El espacio de color RGB, generalmente se muestra como un cubo de colores como se muestra en la Figura 1.18 Los vértices del cubo son los colores primarios (rojo, verde y azul) y secundarios (cyan, magenta y amarillo) de la luz.

El termino nivel de gris (gray level) se emplea para referirse a la intensidad de las imágenes monocromáticas.

El modelo RGB esta relacionado con el hardware, sin embargo no es útil para describir los colores en términos prácticos para la interpretación humana. Por ejemplo, uno no se refiere al color de un auto dando porcentajes del contenido de cada uno de los colores primarios. Cuando los humanos vemos un color, lo describimos en términos de su tono (H), saturación (S) y su brillo o intensidad (I). El modelo HSI debe su utilidad a dos hechos:

1. La componente de intensidad (I) esta desacoplada de la información cromática.
2. Las componentes de tono (H) y saturación (S) están íntimamente relacionadas con la forma en que los humanos percibimos el color.

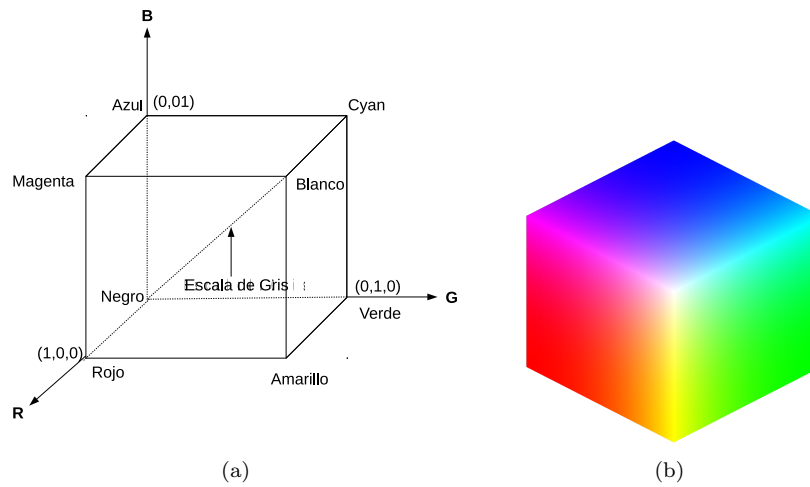


Figura 1.18: Cubo 24 bits RGB. (a)Esquema normalizado a valores entre [0,1], (b)Cubo en colores.

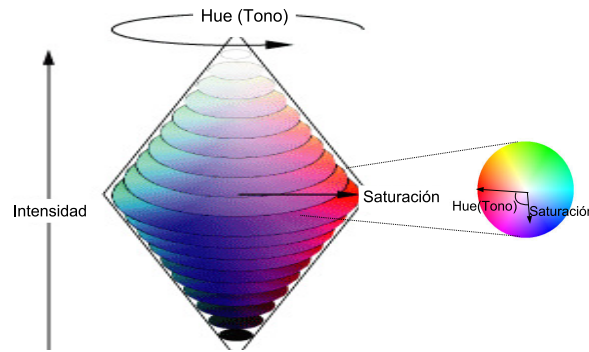


Figura 1.19: Ejemplo de una representación del espacio de color HSI.

El modelo de color HSI es una herramienta ideal para el procesamiento de imágenes en color. Los componentes importantes del espacio de color HSI, son el eje vertical de intensidad, el tamaño del vector al punto de color y el ángulo de ese vector hace respecto a la referencia (eje rojo). Por lo tanto no es inusual el plano HSI está definido en términos de los cortes en círculos vistos en Figura 1.19.

1.0.4. Teorema de Parseval

El valor medio de una señal se define como la media de todos los valores que definen y componen la misma cuya suma representa el área bajo la curva entre un periodo de tiempo que matemáticamente se representa por a la Ecuación 1.21. Gráficamente corresponde a un triángulo que contiene el área equivalente a la que tiene la señal bajo su curva como se muestra en la Figura 1.20.

$$Area = \int_0^T f(t) \quad (1.21)$$

En los sistemas de comunicaciones es importante conocer la potencia promedio de las señales lo que es equivalente al valor cuadrático medio en un periodo definido (T) de la señal como lo muestra la

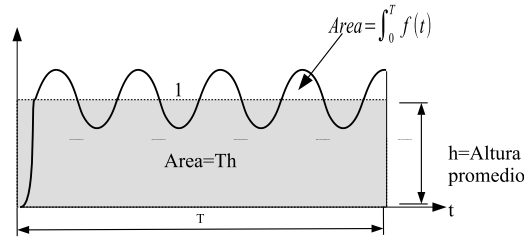


Figura 1.20: Área bajo la curva de una señal.

Ecuación 1.22, si $f(t)$ corresponde a una señal de corriente o voltaje representa la potencia promedio entregada por la misma a una resistencia de 1Ω . Los límites de integración para una señal periódica corresponden a un periodo de la señal ya que si se toma n periodos, se aumenta el tiempo n veces y de igual manera pasaría con el área, por lo tanto, se obtendría el mismo resultado.

$$P = \int_{-T/2}^{T/2} \{f(t)\}^2 dt \quad (1.22)$$

El teorema de Parseval define que la potencia de las señales es equivalente a la suma de la potencia de sus componentes espectrales y se toma dependiendo de si la señal es periódica o no ya que para su análisis se implementa la serie y la transformada de Fourier respectivamente.

Teorema de Parseval para señales periódicas:

Si por un lado, la serie de Fourier corresponde a la serie trigonométrica o, por el otro, a la exponencial compleja el Teorema de Parseval corresponde a las Ecuaciones 1.23 y 1.24 respectivamente. En donde se define que la potencia de la señal es equivalente a la suma de la potencia de los componentes espectrales representados por los coeficientes a_0 , a_n y b_n para el primer caso y C_n para el segundo. El valor cuadrático medio es correspondiente al valor cuadrático medio de los componentes espectrales como lo muestra la ecuación 1.25, donde C_0 es el nivel de offset de la señal y C_n la amplitud de la n -ésimo armónico.

$$\frac{1}{T} \int_{-T/2}^{T/2} \{f(t)\}^2 dt = \frac{1}{4} a_0^2 + \frac{1}{2} \sum_{n=1}^{\infty} (a_n^2 + b_n^2) \quad (1.23)$$

$$\frac{1}{T} \int_{-T/2}^{T/2} \{f(t)\}^2 dt = \sum_{n=-\infty}^{\infty} |C_n|^2 \quad (1.24)$$

$$\frac{1}{T} \int_{-T/2}^{T/2} \{f(t)\}^2 dt = C_0^2 + \sum_{n=1}^{\infty} \left| \frac{C_n}{\sqrt{2}} \right|^2 \quad (1.25)$$

Anexo 2

Cómputo paralelo

2.1. Introducción

Durante varios años el modo habitual del desarrollo computacional se ha hecho de forma secuencial. De esta manera, una computadora con un único procesador sólo cuenta con la capacidad de ejecutar una instrucción en un momento determinado. Esto es debido a la adopción del modelo que propuso Von Neuman, donde la estabilidad es la base para el desarrollo de diversas aplicaciones. Sin embargo, desde el comienzo de la computación secuencial se vislumbraba que, tarde o temprano, se necesitaría recurrir a la computación paralela para resolver problemas que demandaban muchos recursos de cómputo por lo que su implementación en máquinas secuenciales no sería lo más óptimo o en su caso posible. Por ejemplo:

- Sistema de tamaño N
- Atracción de $N - 1$ cuerpos $\rightarrow N * (N - 1) \approx N^2$
- Predicción del clima.

A la ejecución de dos o más procesos al mismo tiempo usando más de un procesador se le llama cómputo paralelo. Procesar un programa en paralelo hace que se ejecute de manera más rápida ya que existen más procesadores trabajando en forma simultánea y coordinada en diferentes partes del problema. El objetivo es reducir al mínimo el tiempo total de cómputo, esto se hace mediante la distribución de la carga de trabajo entre los procesadores disponibles en ese momento.

Las aplicaciones paralelas se pueden encontrar en una amplia gama de disciplinas, que van desde la predicción del clima, el modelado de la biósfera, el procesamiento de imágenes, el modelado de océanos, etc. La velocidad de procesamiento no es la única razón para utilizar el paralelismo. La construcción de aplicaciones más complejas ha requerido computadoras más rápidas, y las limitaciones en el desarrollo de computadoras seriales han llegado a ser más y más evidentes. Anteriormente, el procesamiento paralelo se empleaba para resolver problemas que requerían una gran escala simulación, un gran número de cálculos y procesamiento de datos (por ejemplo, el cómputo de los datos de un censo). En la actualidad, el costo del hardware disminuye rápidamente, por lo que el procesamiento paralelo se está utilizando cada vez más en tareas comunes. No obstante, además de las facilidades de

hardware que permiten llevar a cabo este tipo de procesamiento, es necesario contar con software que soporte la ejecución y la coordinación de procesos en forma paralela.

En 1966 Michael Flynn propuso un mecanismo de clasificación de las computadoras paralelas. Aunque Flynn no cubre todas las posibles arquitecturas, su taxonomía facilita de manera significativa la comprensión del funcionamiento, basándose en el número de instrucciones y en la secuencia de datos simultáneos que utiliza una computadora para procesar información. Los tipos más sobresalientes dentro de esta clasificación son SISD (Single Instruction Single Data), SIMD (Single Instruction Multiple Data), y MIMD (Multiple Instruction Multiple Data).

La Figura 2.1 muestra una clasificación general de las arquitecturas paralelas, las cuales se pueden dividir en dos grandes grupos, los multiprocesadores y las multicomputadoras.

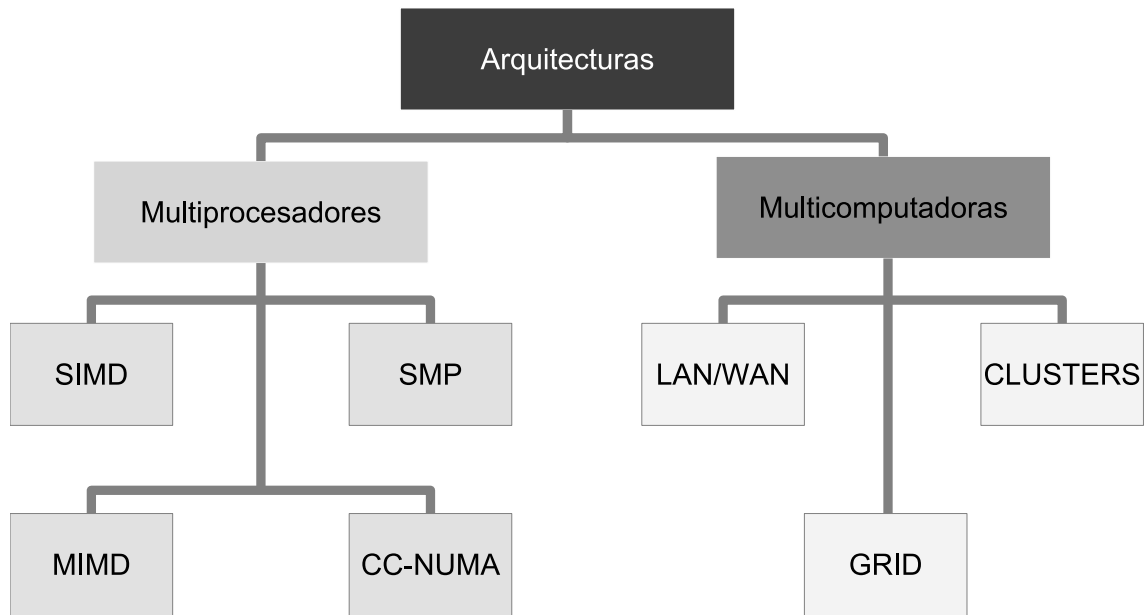


Figura 2.1: Arquitecturas paralelas.

2.1.1. SISD(Single instruction-Single data)

Este es el modelo tradicional de computación secuencial donde una unidad de procesamiento recibe una sola secuencia de instrucciones que opera en una secuencia de datos Figura2.2.

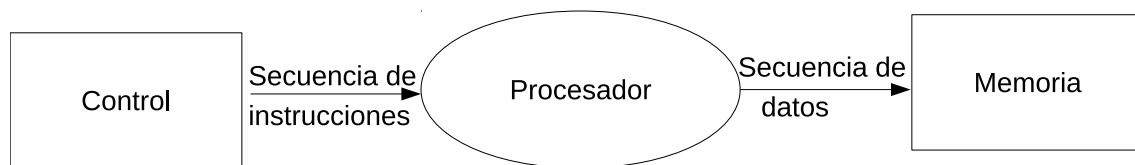


Figura 2.2: Arquitectura SISD.

2.1.2. Multiprocesadores

Estas arquitecturas consisten en una computadora que se compone de un arreglo de procesadores y un banco de memoria común, conectados a través de un bus rápido o uno de los varios tipos de red de interconexión dinámica multietapas. Existen dos tipos de sistemas multiprocesador:

- **Simétrico.** En estos los microprocesadores del sistema comparten los recursos del sistema, como la entrada/salida de memoria y disco. Se distribuye uniformemente la carga de trabajo a los procesadores disponibles.
- **Asimétrico.** En estos, diferentes procesadores gobiernan las distintas tareas y recursos del sistema, El enfoque principal que adoptan, es incrementar el rendimiento del sistema para múltiples usuarios de red, en múltiples segmentos de red de área local. Cada CPU se dedica a una función específica entrada/salida de la tarjeta de interfaz de red u operaciones con manejo de archivos.

La mejor manera de programar dentro de los multiprocesadores es a través de los hilos ya que estos se comunican por medio de la memoria compartida ya que estando dentro de un mismo proceso de ejecución comparten el mismo espacio de memoria, en este tipo de programación hay que tener cuidado a la hora de manejar las variables globales ya que se pueden generar errores no deseados a la hora de leer/escribir en ellas.

2.1.3. Multicomputadoras

Estas arquitecturas se componen de una arquitectura de memoria distribuida, es decir, no compartida. En esta arquitectura se tiene un arreglo de microprocesadores, cada uno con su propia memoria local. Son sistemas donde cada unidad de proceso es autónoma e independiente de las demás pero trabajan de manera conjunta. La interconexión de unidades de proceso se hace a través de redes de computadora de preferencia de altas velocidades.

La mejor manera de programar a través de las multicomputadoras es por medio del paso de mensajes, ya que cada proceso se ejecuta en una máquina diferente, sí estos deben de enviar los resultados de algún cálculo a otro proceso que lo necesite, tiene que ser a través de un mensaje. La directiva más conocida para paso de mensajes es la de MPI, la cual, ha hecho posible el desarrollo de un grado aceptable de portabilidad del software paralelo. MPI proporciona una interfaz de programación basada en envío y recepción de mensajes que es independiente de cualquier arquitectura paralela particular, además de que puede enlazarse con C++ o Fortran.