



UNIVERSIDAD NACIONAL
AVENIDA DE
MEXICO

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

**PROGRAMA DE MAESTRIA Y DOCTORADO EN
INGENIERIA**

FACULTAD DE INGENIERIA

**INDEXADO DE IMÁGENES APLICANDO
LA TRANSFORMADA POLINOMIAL**

T E S I S

QUE PARA OPTAR POR EL GRADO DE:

MAESTRO EN INGENIERIA

CAMPO DE CONOCIMIENTO: INGENIERIA ELECTRICA

CAMPO DISCIPLINARIO: PROCESAMIENTO DIGITAL DE SEÑALES

P R E S E N T A :

VICENTE SANCHEZ LUNA

TUTOR:

BORIS ESCALANTE RAMIREZ

2008





Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

JURADO ASIGNADO:

Presidente: FRANCISCO GARCÍA UGALDE

Secretario: CARLOS JOEL RIVERO MORENO

Vocal: BORIS ESCALANTE RAMÍREZ

1^{er}. Suplente: JESÚS SAVAGE CARMONA

2^{do}. Suplente: FERNANDO ARÁMBULA COSIO

POSGRADO DE INGENIERIA

TUTOR DE TESIS:

BORIS ESCALANTE RAMIREZ

Agradecimientos.

A mis padres, quienes han sido mi principal e incondicional apoyo durante toda mi vida, y especialmente, en esta etapa. Este logro también es suyo. Gracias.

A mis hermanas y hermanos por el apoyo y la comprensión que me han tenido.

A todos mis familiares que confiaron y creyeron en mí. Espero seguir siendo alguien de quien puedan sentirse orgullosos.

Al Dr. Boris Escalante Ramírez por todo su apoyo y orientación durante el tiempo que trabajé en este posgrado y, en especial por haberme introducido en el interesante campo del Procesamiento Digital de Imágenes.

Al Dr. Carlos Joel Rivero Moreno, por su orientación y consejos en la elaboración de este proyecto.

A mis compañeros de trabajo y de manera muy especial al Ing. Daniel Moisés Flores Romero, Ing. Enrique del Castillo e Ing. Gerardo Díaz Díaz por su apoyo y facilidades prestadas a lo largo de este tiempo.

A la Universidad Nacional Autónoma de México y a la Facultad de Ingeniería, magníficas instituciones que me brindaron mi formación académica y profesional. Siempre les estaré agradecido.

A mis profesores por haberme transmitido sus conocimientos. Son un ejemplo a seguir.

A mis compañeros de generación por la comprensión y ayuda dada.

Vicente Sánchez Luna

Índice

Resumen.....	1
Objetivo.....	3
1 Introducción.....	4
1.1 Identificación del problema.....	4
1.2 Justificación.....	6
1.3 Antecedentes (Edo. del arte).....	7
1.3.1 Evolución de la Base de Datos.....	8
1.3.1.1 Situación actual.....	8
1.3.2 Categorías: Base de datos de imágenes.....	9
1.3.3 Clasificación: Sistemas de base de datos de imágenes.....	10
1.3.3.1 Sistemas sencillos.....	10
1.3.3.2 Sistemas de recuperación de imágenes basados en contenido.....	12
1.4 Aportación (contribución y relevancia).....	13
2 Indexado.....	14
2.1 Indexación (como surgió, niveles, tutoría técnica).....	14
2.1.1. Indexación de alto y bajo nivel.....	15
2.2 Sistemas de recuperación de imágenes basados en el contenido.....	17
2.2.1 Sistemas basados en el color.....	18
2.2.2 Sistemas basados en la textura.....	18
2.2.3 Sistemas basados en la similitud espacial.....	19
2.2.4 Sistemas basados en la forma.....	19
2.2.5 Otros sistemas.....	19
2.3 Métodos alternativos.....	20
2.3.1 Discret wavelet transform. DWT.....	20
2.3.2 Dual-tree complex wavelet transform. DT-CWT.....	21
2.3.3 Gabor.....	21
2.3.4 Hermite.....	22
2.3.5 Hermite tipo Gabor.....	23
2.3.6 Hermite direccional –rotado-.....	23
3 Transformada Hermite multiescala.....	25
3.1 Introducción.....	26
3.2 Caso 1-D: Transformada hermite multiescala.....	27
3.2.1 Descomposición multiescala.....	27
3.2.2 Análisis de estructuras.....	28
3.2.3 Definición de la THM.....	29
3.2.4 Acotamiento del rango de escalas.....	30
3.2.5 Esquema piramidal continuo.....	31
3.2.6 Predicción entre niveles.....	33
3.3 Caso 2-D: Transformada hermite multiescala-multidireccional.....	34
3.3.1 Descomposición multiescala de señales 2-D.....	34
3.3.2 Análisis de estructuras 2-D.....	34

3.3.3 Definición de la THMM.....	36
3.3.4 Filtrado direccional.....	37
3.4 Aplicaciones.....	38
4 Extracción de características con la transformada hermite.....	41
4.1 Arquitectura Bertino.....	41
4.1.1 Otras arquitecturas.....	44
4.2 Extracción de características.....	45
4.2.1 Autocorrelación.....	46
4.3 Similitud cuantitativa entre vectores de características.....	50
4.3.1 Distancia métrica.....	50
5 Resultados experimentales.....	54
5.1 Imágenes de análisis. –álbum fotográfico de Brodatz-.....	54
5.2 Vectores de características.....	57
5.3 Similitud cuantitativa ó distancia entre vectores.....	62
5.3.1 Distancia Canberra.....	62
5.4 Evaluación.....	64
Evaluación cuantitativa: eficiencia de recuperación porcentual.....	64
5.5 Análisis de resultados.....	68
Conclusiones.....	80
Apéndice A: Transformada Hermite Multiescala, Caso 1-D.....	82
Apéndice B: Transformada Hermite Multiescala Multidireccional, Caso 2-D.....	84
Apéndice C: Rotación de las DG.....	86
Anexo 1.....	87
Bibliografía.....	90

Resumen

La información que guardan las imágenes es amplia y quizá crítica para algunas aplicaciones; en este sentido podemos imaginar la gran cantidad y diversidad que tienen éstas a lo largo y ancho de la vida personal y profesional.

El avance tecnológico proporciona herramientas que impulsan la exploración de nuevos métodos y técnicas para mejorar los resultados de un proyecto técnico.

Almacenar, buscar y recuperar imágenes es una necesidad real que ha involucrado la investigación multidisciplinaria de instituciones públicas y privadas. Este trabajo no ajeno a lo anteriormente dicho propone un método por contenido de textura y que esta documentado en seis capítulos.

Capítulo 1. - Introducción al tema tratado en el presente trabajo, como también descripción breve de la necesidad detectada para una posterior justificación que acredite los beneficios, innovaciones, rentabilidad que podríamos tener al usar el método propuesto.

Capítulo 2. - En la literatura especializada la clasificación de la amplia variedad de propuestas relacionadas con el tema es evidente, ya que las técnicas y métodos documentados aplican a diferentes conceptos de implementación.

Revisar los documentos relacionados con el tema no podía pasar desapercibido y, en ese mismo sentido conocer los métodos de vanguardia desarrollados en este campo y poder hacer un comparativo para saber de ventajas y desventajas que podrían tener respecto al producto aquí desarrollado esta considerado.

Capítulo 3. - Fundamentos teóricos utilizados como herramienta son sin duda la base para el entendimiento y justificación técnica del trabajo, por lo que, en este capítulo están desarrollados los principios analíticos utilizados en los algoritmos implementados. La transformada polinomial como herramienta de análisis a fenómenos naturales y en particular a los relacionados en la propuesta, será el sustento analítico del sistema a proponer. El caso particular de la transformada polinomial hermite a mostrado a través de algunos estudios y propuestas una gran

similitud con el modelado del sistema de percepción visual humana –percepción, filtrado, selección de información importante-.

Por lo tanto, investigar y estudiar la herramienta analítica y en particular el análisis teórico de los filtros hermite como modelo de representación del sistema visual humano esta considerado.

Capítulo 4. – Haciendo uso de la transformada polinomial hermite y de sus características se procederá al diseño del algoritmo que cumpla con las expectativas planteadas. La propuesta definida con base en la investigación de fundamentos teóricos y la revisión de otras propuestas relacionadas, permitirá realizar el almacenaje, búsqueda y recuperación de imágenes del álbum fotográfico de Brodatz, que en este caso es utilizado por investigadores como un patrón base de datos de imágenes.

Capítulo 5. - La implementación del algoritmo resultante para su análisis y comparación respecto a otros jugara un papel sobresaliente en el resultado y conclusiones del proyecto, por lo tanto, para comprobar el desempeño del método propuesto, esté será sometido a una serie de pruebas que certificarán ventajas y desventajas a través de una serie de evaluaciones objetivas –cuantitativa- y subjetivas –cualitativa-. Los métodos de vanguardia considerados también están sometidos a las evaluaciones mencionadas.

Conclusiones – Con base en experimentos y resultados obtenidos de evaluaciones realizadas en el capítulo anterior apunto las conclusiones observadas al implementar el método propuesto, además de hacer énfasis en el beneficio, mejora, innovación y rentabilidad obtenida.

Objetivo.

Proponer un sistema para indexado de imágenes con base en los filtros de la transformada polinomial hermite que aproxima sus modelos al sistema de percepción humano, por lo que:

La extracción de características por contenido (textura) fundamentara el proyecto ya que serán utilizadas las sub bandas de la transformada citada.

Así como la representación de imágenes a través de vectores característicos empleando para ello la autocorrelación espacial y la discriminación de los mismos para que el sistema recupere de la base de datos las imágenes más similares dada una de entrada.

Evaluación del sistema utilizando el método al cual fueron sometidas las propuestas relacionadas con el tema documentado en la literatura especializada.

Capítulo 1

Introducción.

1.1 Identificación del problema.

Los volúmenes de información generados por las nuevas tecnologías alrededor del mundo aumentan día a día de manera sorprendente, por lo que, etapas de procesamiento, compresión, almacenamiento, Etc., estarán expuestas a retos cada vez más exigentes. En el caso particular de imágenes y video, es fácil imaginar el enorme problema para almacenar y poner a disposición de los usuarios la gran cantidad de información a consultar, por ejemplo: parte del acervo bibliográfico de una biblioteca, Internet, videoteca y/o acervo fotográfico de una institución pública ó privada. Con el conocimiento objetivo de este problema es obvio que se propongan soluciones para mejorar la eficiencia de los procesos en los sistemas de consulta actuales.

Respecto a los sistemas existentes para la recuperación, existen muy pocos productos comerciales disponibles y éstos ofrecen capacidades muy limitadas. La mayoría son prototipos que se desarrollaron en laboratorios de universidades o institutos de investigación. Los productos comerciales más conocidos son: el QBIC (de IBM), Virage Search Engine (desarrollado por Virage Corporation) y Visual Retrievalware (de Excalibur Technologies Corporation). Estos sistemas permiten la recuperación de imágenes basada en contenido: color, textura, forma y relaciones espaciales. Sin embargo dado el estado actual del análisis de imágenes y de visión por computadora, no existen todavía sistemas de propósito general suficientemente robustos que puedan aplicarse a bases de datos (BD) de diferente naturaleza.

Otros sistemas también conocidos a nivel de prototipo que han sido desarrollados en Universidades e Institutos de investigación son: el Photobook, desarrollado en el MIT, y

el IRIS desarrollado por el German Software Development Laboratory en la Universidad de Bremen.

El color, la forma, la textura y el movimiento son las características básicas que nos permiten percibir, describir y reconocer objetos y por tanto interactuar de manera correcta con el medio ambiente que nos rodea. La textura juega un papel importante en el sistema visual humano para detectar y distinguir objetos, para inferir información acerca de la orientación de superficies y su perspectiva, y para determinar la forma de objetos en las escenas tridimensionales. Hoy en día la textura debido a sus propiedades estéticas juega un papel muy importante en el diseño orientado al consumidor de productos, tales como: marketing, venta e intercambio de productos.

Razones como las anteriormente mencionadas han provocado que sistemas de acceso y búsqueda de texturas en BD de imágenes reciban día a día una mayor atención. Estos sistemas son conocidos genéricamente como Sistemas de Administración de Bases de Datos de Texturas (Texture Retrieval Systems).

Si queremos buscar información en una BD tradicional primero tenemos que localizar un elemento o registro específico, el resultado de tal proceso es juzgado por un usuario humano, por lo que, parece lógico utilizar conceptos humanos de percepción y descripción de texturas, así como medidas de similitud biológicamente basadas en el diseño de estos sistemas.

Una observación psicológica interesante corresponde al hecho de que los seres humanos no son capaces de describir textura en forma clara y objetiva, sino que más bien, lo hacen de manera subjetiva a través de una caracterización difusa de ellas.

Las tecnologías ed Soft-Computing, en contraposición a las tecnologías usuales de computación (Hard-Computing), permiten introducir conceptos como incerteza, adaptabilidad, robustez, etc. en el procesamiento de la información. Usualmente se consideran parte de este grupo de tecnologías las redes neuronales, los sistemas difusos,

la computación evolucionaría, la teoría del aprendizaje, la teoría del caos y los métodos probabilísticos de análisis (Bayes, Markov, etc.).

1.2 Justificación.

Tal como se desprende de lo expuesto en la sección anterior: almacenar, consultar y recuperar imágenes hoy en día es uno de los retos más importantes considerado por investigadores e instituciones públicas y privadas a resolver.

La aplicación dada a las imágenes en los diferentes campos del conocimiento va en aumento, lo que ha motivado una serie de trabajos y propuestas alrededor del mundo para desarrollar el mejor proceso que permita el acceso a una BD de imágenes.

De los párrafos anteriormente citados, notamos la relevancia del tema tratado en este trabajo y de la atención multidisciplinaria requerida. A pesar de lo anterior, cabe mencionar que la propuesta documentada solo esta apoyada en el procesamiento digital de imágenes con base en la transformada polinomial hermite.

Algunos investigadores atendiendo las características del sistema visual humano proponen modelos que procesan información visual a múltiples resoluciones. Las etapas primarias de visión detectan amplitud, orientación y localización de estructuras primitivas (bordes y líneas). Trabajos de investigación [1] sugieren que las células retínales y corticales detectan los parámetros antes mencionados, además de poder ser modeladas por funciones gaussianas y sus derivadas.

Martens [2] propone modelos matemáticos (transformada hermite) para la representación de imágenes. Dichos modelos son caracterizados por tomar en cuenta importantes propiedades de la visión espacial: filtrado, análisis multirresolución y selección a la dirección y frecuencia.

Los polinomios hermite realizan expansiones locales de señal, analizando imágenes en el dominio espacial.

Analizar imágenes a través de la característica básica de textura es importante [3] para su indexación y recuperación posterior [4] [5]. En la caracterización de imágenes por textura es necesaria la asignación de un vector característico obtenido por la aplicación de filtros particulares [6] [7] a cada una de las imágenes. La extracción de características de textura es realizada bajo técnicas de tratamiento de imágenes, es decir, con transformadas lineales que filtran la textura [6] [7] [8] [9] detectando cantidad de energía. El objetivo de esta etapa es propiamente la indexación de imágenes a partir de sus vectores característicos. El indexado es utilizado para la búsqueda y recuperación de imágenes con medidas de similitud.

Es importante precisar que los sistemas de indexado y búsqueda propuestos en la literatura [4] [5] [10] están fundamentados en las características extraídas de la textura. El análisis de imágenes compuesto por regiones uniformes no funcionara para la navegación y recuperación, pues la textura depende de distinciones espaciales, extracción e indexado local de las características de la textura solicitada.

Para el modelo discreto, los filtros implicados en la descomposición de la señal son los filtros Krawtchouk.

En la siguiente sección presentaremos la clasificación de los sistemas para la recuperación de imágenes, donde la clasificación por contenido es de nuestro interés

1.3 Antecedentes

Siendo la BD una etapa fundamental en un sistema recuperador de imágenes no podríamos dejar de pasar por alto las razones que ocasionaron su origen y evolución, por lo que, a continuación se exponen los distintos tipos de éstas comentando brevemente sus características además de establecer una división entre los sistemas considerados como sencillos y los sistemas de recuperación de imagen basados en contenido.

1.3.1 Evolución de la Base de Datos

El amplio campo de aplicación dado hoy en día a las imágenes ha motivado el interés cada vez más frecuente de las técnicas utilizadas para su explotación y, de forma obligada la aparición de propuestas que permiten la recuperación y procesamiento de las mismas.

Los tradicionales Sistemas Gestores de Bases de Datos (SGBD) han mostrado su potencialidad y utilidad para el acceso a datos estructurados, sin embargo, dichas características se ven disminuidas cuando los datos no son alfanuméricos o no están estructurados como en el caso de las imágenes. Esta evidente falta de Sistemas ha conducido al diseño e implementación de diversas BD de Imágenes [12] [13] [14] [15], y a que, desde hace ya algún tiempo, como señalaban Arndt y Chang [12] [17], se haya producido una gran actividad investigadora en este campo.

1.3.1.1 Situación actual

Tamura y Yokoya [11] realizaron una clasificación de las tecnologías existentes en el campo del procesamiento digital de imágenes, dividiendo el área en dos corrientes. La primera comprende aplicaciones del reconocimiento-procesamiento de imágenes y, en la segunda consideran las técnicas utilizadas para la generación de gráficos por computadora. Las BD de imágenes están enmarcadas dentro de la primera línea. En la primer corriente, el procesamiento digital permite que una imagen de entrada sea transformada en otra imagen o en una descripción simbólica derivada de ella. Estas técnicas se utilizan en BD de imágenes para ó bien obtener descriptores de la imagen atendiendo a las características de sus píxeles ó también como medio que permite describir y por lo tanto posteriormente recuperar las imágenes almacenadas en una

BD. Estas descripciones suelen utilizarse como índices que faciliten y aceleren la búsqueda.

Los investigadores que trabajan dentro de la segunda corriente se ocupan de generar imágenes a partir de sus descripciones, centrando su interés en temas alejados de las principales técnicas que utilizan las BD de Imágenes.

1.3.2 Categorías: Base de datos de imágenes

Probablemente Timothy Arndt [17] fue quien realizó la primera clasificación de BD de imágenes. Dividió los sistemas desarrollados tomando en cuenta su funcionalidad. Por lo que, estableció las dos siguientes categorías:

a) BD de imágenes que recuperan una imagen que satisfaga ciertos requisitos de entre todas las almacenadas.

Los principales campos de aplicación para este tipo de Bases de Imágenes son:

- Automatización de oficinas.
- Sistemas de información médica.
- Aplicaciones de visión artificial que requieran medidas de similitud.

En este tipo de BD de Imágenes, la respuesta a una consulta da como resultado un conjunto de imágenes que presentan características similares a las solicitadas.

b) BD de imágenes que recuperan información a partir de una imagen.

Las principales áreas que requieren este tipo de Bases de Imágenes son:

- Sistemas de Información Geográfica.
- Aplicaciones de Visión Artificial que requieran la interpretación de imágenes.

La respuesta a una consulta consiste en un conjunto de características de la imagen, que a partir de una interpretación se convierte en información útil sobre el campo de conocimiento en el que se encuentre la imagen.

En ocasiones, algunos sistemas incorporan los dos modos de funcionamiento.

1.3.3 Clasificación: Sistemas de base de datos de imágenes

1.3.3.1 Sistemas sencillos

Comúnmente la técnica utilizada para almacenar imágenes en una BD consiste en la aplicación de texto descriptivo ó palabras clave asignada por un operador humano [18]. De los sistemas que se utilizan, un gran número de ellos están basados en atributos ó texto libre que presentan muchas limitaciones al realizar la consulta [19]. A continuación se indican las características de dichos sistemas.

- **Sistemas de almacenamiento secuencial no estructurado**
Primera solución adoptada para mantener un conjunto de imágenes almacenadas y acceder a ellas [19]. El sistema puede disponer de un módulo que genera una pequeña muestra de cada imagen. Dicha muestra puede obtenerse al momento de añadir una nueva imagen a la colección o, en algunos casos, en el momento de visualizar las imágenes. Para recuperar una imagen tendremos que recorrer de manera secuencial el conjunto de imágenes ó muestras. La ventaja de generar muestras estará reflejada al ahorrar tiempo de visualización. Es evidente que la limitación de estos sistemas esta presente al no disponer de características de recuperación. Hoy en día, la mayoría de los programas de edición o tratamiento de imágenes disponen de esta opción.
- **Sistemas basados en atributos**
Formados por un SGBD convencional con capacidad para almacenar grandes objetos binarios (Binary Large Objects: BLOBs). Si es un SGBD relacional, cada fila en una tabla (relación) está formada por los atributos de la imagen y la imagen en sí,

almacenada ésta en un campo tipo BLOB y sus atributos en campos con el tipo de datos apropiados según su contenido. El acceso a una imagen (datos no estructurados) se consigue a través de una consulta sobre sus atributos estructurados. La estructuración de los atributos puede hacerse de diferentes maneras. Una de ellas consiste en realizar una clasificación taxonómica de las imágenes almacenadas. Otra consistiría en utilizar un conjunto de palabras para describir la imagen, no clasificadas de una forma jerárquica como en el anterior caso, pero siempre las mismas, de manera que la descripción de una imagen estuviera restringida a ese conjunto de palabras.

El diseño de la base de datos, el mecanismo de indexación y el método de procesamiento de consulta es el de cualquier otra base de datos.

Uno de los grandes inconvenientes de este tipo de métodos es que no permiten realizar consultas basadas en contenido que resulten cómodas y potentes.

- Sistemas basados en texto libre

Sistemas que aplican conceptos derivados de técnicas para recuperar documentos de texto, ya que éstos proporcionan funcionalidades "basadas en contenido" mediante la descripción manual de la imagen. El acceso a la imagen se hace a través de la descripción que le acompaña. Hay varios autores que comentan en sus trabajos las limitaciones de estos sistemas [10] [18] [19] [22]. Las más comúnmente señaladas son:

1. La búsqueda depende únicamente de palabras clave. Esto significa que si se realiza una consulta sobre propiedades de la imagen que no fueron descritas inicialmente probablemente se obtengan resultados erróneos. La descripción es habitualmente incompleta.
2. Hay propiedades visuales que son muy difíciles, por no decir imposibles de describir mediante texto, como algunos tipos de textura o de forma. Una reflexión interesante sobre este problema se puede encontrar en [23].

3. Incluso si todas las características de la imagen se describen utilizando texto, no existe un vocabulario comúnmente aceptado para describir propiedades de las imágenes. Esto supone que la descripción es un proceso subjetivo.
4. La falta de un vocabulario normalizado supone que un mismo usuario puede obtener distintos resultados en la misma base de datos si está descrita por especialistas diferentes.
5. Que no exista este vocabulario normalizado implica que la persona que realiza la consulta deberá de contar con un conocimiento relativo a los criterios de generación de anotaciones.
6. Realizar anotaciones mediante frases textuales para grandes bases de datos supone un gran costo de tiempo y esfuerzo.
7. Incluso si todas las características útiles han sido anotadas, hay diferencias entre los lenguajes y el vocabulario que diferentes expertos puedan utilizar, lo que supone una pérdida de precisión en la recuperación.

En la actualidad se trabaja en metodologías que analicen las consultas mediante lenguaje natural [34] como una vía que facilite la realización de búsquedas sin los inconvenientes citados anteriormente.

1.3.3.2 Sistemas de recuperación de imágenes basados en contenido

Los trabajos en esta línea de investigación son muchos y variados. El método propuesto aquí está fundamentado bajo el mismo concepto, por lo que, en el siguiente capítulo estarán descritas las técnicas y herramientas propuestas en los diferentes documentos consultados bajo esta línea de investigación, es decir, por contenido.

1.4 Aportación (contribución y relevancia)

Alcances

- Obtener características más representativas de la información visual, por ejemplo: modelos paramétricos ó estadísticos
- Combinación de las características extraídas para describir medidas visuales y calidades de información visual.
- Elaboración de distintivos que ajusten a la descripción de texturas, a fin de reducir los volúmenes de información.

Objetivos específicos.

- Proponer un sistema automático para indexado de imágenes que esta fundamentado en filtros que aproximan al sistema de percepción visual humano
- Utilizar las sub bandas de la transformada hermite para caracterizar textura.
- Extraer las características por indexación de textura de imágenes fijas sensibles a la orientación.
- Formulación de vectores característicos de textura con la autocorrelación.
- Discriminación de vectores característicos.
- Aproximación local para la caracterización, indexado y recuperación por contenido de la imagen.
- Implementar una representación discreta del banco de filtros para la información espacial.

Capítulo 2

Indexado

2.1 Indexación

Recuperar información en BD de imágenes es un amplio campo que destaca, tanto por el elevado número de aplicaciones como por su diversidad. Su aplicación en galerías de arte, museos, comunicaciones, investigación criminal, identificación de personas, textiles, logotipos, diseño industrial, así como en aplicaciones médicas y geográficas, hacen que sea uno de los campos con mayor crecimiento dentro de las tecnologías de la información. La recuperación de la información visual es un área en expansión tanto a nivel teórico como de aplicación práctica.

El análisis y procesamiento de imágenes y vídeo, reconocimiento de patrones, visión por computadora, modelado de información multimedia, indexación multidimensional, modelado psicológico del comportamiento del usuario, interacción hombre-máquina y visualización de información, son algunos de los campos de investigación más importantes que contribuyen en esta nueva área de investigación.

La clave para recuperar información visual está en la disponibilidad de una representación adecuada del contenido visual de las imágenes.

Aspectos visuales como el color, la textura, la forma, las relaciones espaciales o movimiento, directamente relacionados con el contenido de la imagen, junto con otros aspectos de alto nivel como el significado de los objetos y de las escenas, se usan como claves para la recuperación de imágenes con contenido similar en la BD.

Usar información visual para realizar una consulta surge de las propias limitaciones de los sistemas clásicos de recuperación basados en texto que son poco adecuados para modelar el concepto perceptivo que le damos a las imágenes. Las búsquedas

convencionales basadas en texto se complementan, hoy en día, con las consultas basadas en contenido visual, que tienen en cuenta los aspectos perceptivos anteriormente citados. Estas consultas implican especificar las propiedades visuales de las imágenes que se buscan.

Los sistemas de consulta que trabajan con el contenido visual deberán de tener una imagen de entrada que funcionara como patrón de búsqueda. Las imágenes obtenidas se ordenan de acuerdo a una medida de similitud que estará en concordancia con la forma en que los humanos establecen la similitud entre los objetos.

El análisis de imágenes, el reconocimiento de patrones y la visión por computador juegan un papel fundamental en los sistemas de recuperación de la información visual y proporcionan la base teórica y los métodos que permitirán extraer la información sobre la que se basará la consulta.

2.1.1. Indexación de alto y bajo nivel.

Una línea de investigación para recuperar información visual esta en el desarrollo de técnicas para la extracción automática de características a nivel bajo o intermedio de la imagen. Como características de bajo nivel ampliamente utilizadas en la literatura destacan el color, forma y textura, descritos normalmente como información numérica fácilmente guardada en una base de datos.

Otro parámetro corresponde a descriptores funcionales, es decir, aquellos en los que la característica no es un número, sino una función que expresa tamaño (directamente relacionado con la escala de representación) o tiempo (relacionado con el cambio en secuencias de imágenes).

En la tabla 1 podemos observar la clasificación de niveles que tenemos para el indexado.

Indexado de alto nivel	<p>Predefine términos de indexado para acotar imagen y video. Estos términos son la base para organizar las categorías ontológicas de alto nivel (acción, tiempo, espacio, Etc.).</p> <p>Las técnicas de indexado de alto nivel son diseños que contemplan el indexado y acotación manual, es decir, la calificación de videos que permitirán hacer anotaciones en la base de datos.</p>
Indexado de bajo nivel	<p>Esta técnica prevé el acceso a la base de datos en función de las propiedades (color, textura, Etc.) de la imagen ó video, además la clasificación podrá ser bajo el etiquetado de indexado de bajo nivel. La principal limitación es la poca semántica que relacione características.</p>
Indexado con dominio especificado	<p>Esta técnica usa estructuras de alto nivel del video para obtener características de bajo nivel de video en la extracción y procesamiento.</p> <p>La principal limitación de estas técnicas es su limitado rango de aplicación</p>

Tabla 1. Niveles de clasificación en el indexado

Tamura y Yokoya [11] clasificaron en dos corrientes las tecnologías existentes en el campo del procesamiento digital de imágenes. La primera comprende las aplicaciones para reconocimiento procesamiento de imágenes y, en la otra incluyeron las técnicas utilizadas para generación de gráficos por computadora. Las BD de Imágenes se encuentran dentro de la primera y es en esta que el procesamiento digital de imágenes utiliza una imagen de entrada que será transformada en otra imagen o en una descripción simbólica derivada de ella. Las técnicas utilizadas en BD de imágenes proporcionan descriptores de la imagen atendiendo características de sus píxeles, para que posteriormente sean utilizados en la recuperación imágenes almacenadas en una BD. Estos descriptores suelen utilizarse para que faciliten y aceleren la imagen solicitada.

2.2 Sistemas de recuperación de imágenes basados en el contenido.

Algunos autores, como es el caso de Gong [18], opinan que si la composición y el contenido de una imagen puede ser totalmente capturado por una computadora, las anotaciones objetivas, generales y comprensivas, en lugar de subjetivas y parciales también han de ser posibles.

Aunque la mayoría de los investigadores puedan estar de acuerdo con este planteamiento, no se puede olvidar que el gran inconveniente para llegar a las BD de imágenes inteligentes es que todavía no se ha solucionado de forma general el problema del reconocimiento de objetos e interpretación del contenido de una imagen, lo que supondría que las computadoras pudieran realizar una comprensión automática de las imágenes.

Los actuales sistemas de recuperación de imágenes basados en contenido ofrecen una solución para encontrar imágenes en función de sus propiedades. Emplean para ello características de bajo nivel como color, forma y textura que puede ser extraído de manera automática [10] [13]. Estos sistemas resultan útiles para recuperar imágenes cuando utilizamos otra imagen que sirve de ejemplo para encontrar las imágenes deseadas. La mayor ventaja que presentan estos sistemas está en la extracción de características, ya que al ser de bajo nivel, se obtienen mediante operaciones relativamente básicas sobre los píxeles de la imagen. El gran inconveniente se encuentra en que no permiten realizar consultas a un nivel conceptual superior.

2.2.1 Sistemas basados en el color

Los sistemas se caracterizan por utilizar una representación del color de la imagen para describir su contenido. Lo más habitual es utilizar un histograma de color de la imagen completa. Algunos sistemas utilizan varios histogramas, uno para cada región o zona de la imagen mientras que otros sistemas utilizan una representación de la señal cromática y acromática de la imagen [25]. Los modelos de color utilizados difieren según los autores, empleando unos el HVC y otros el HSI o incluso el RGB, YIQ o LAB [20]. Mientras que la medida de similitud varía entre unos autores y otros, el método de consulta suele ser utilizando una imagen de ejemplo.

2.2.2 Sistemas basados en la textura

Los sistemas basados en textura pueden clasificarse en dos grandes grupos: los que se basan en características heurísticas que se ocupan de cuantificar la percepción humana de los patrones de textura, y los que aplican técnicas del procesamiento de señal y análisis estadístico. Según esto, las características se representan bien sea mediante descriptores heurísticos [20] o mediante valores aportados por las transformaciones wavelet como es el caso de Manjunath y Ma [4] y de Huang y Dai [5]. La medida de similitud pertenecerá a las métricas [26], siendo la distancia Euclidiana una de las más utilizadas. En este caso, como en el anterior, lo más habitual es realizar una consulta mediante un ejemplo, aunque también hay sistemas que admiten una descripción textual de la textura buscada [13].

2.2.3 Sistemas basados en la similitud espacial

Los sistemas basados en la similitud espacial se centran en conocer y utilizar las relaciones espaciales entre los objetos de la imagen. El trabajo pionero en este campo es obra de Chang [27] que propuso utilizar dos cadenas bidimensionales que contuvieran las posiciones relativas de los objetos de la imagen. A partir de esas cadenas se realiza el proceso de consulta. Muchos trabajos posteriores, tanto del autor [27] como de otros investigadores aumentaron el número de relaciones y de operadores espaciales añadiéndole funcionalidad a estos sistemas. En la actualidad otros autores han propuesto sistemas que sin utilizar las cadenas 2D utilizan diferentes tipos de relaciones espaciales [28] [29].

2.2.4 Sistemas basados en la forma

Utilizar descriptores de contorno o de región de los objetos de la imagen permite a estos sistemas realizar búsquedas entre las imágenes almacenadas [30]. Otros sistemas utilizan características estructurales obtenidas a partir de un mapa de bordes [31]. Los métodos de consulta varían con respecto a los anteriores sistemas al emplear además de la consulta mediante ejemplo, la consulta mediante esbozo en la cual se realiza un sencillo dibujo de la forma de los objetos que ha de contener la imagen buscada [32]. La medida de distancia utilizada en este caso suele centrarse en aspectos relacionados con la correlación [32], con el error cuadrático medio de una firma generada [33] o con la distancia Euclidiana [20].

2.2.5 Otros sistemas

Además de estos sistemas, diferentes autores han propuesto algunos otros que, o bien utilizan otras características para describir el contenido de la imagen, o bien utilizan una combinación de las anteriores. El empleo de redes neuronales para clasificar, o de gráficos o redes semánticas para definir las relaciones entre los objetos son algunos de

los enfoques utilizados. Algunos de estos sistemas realizan las consultas con la petición de objetos de una forma más natural [16] o bien utilizan las técnicas habituales de consultas mediante ejemplo [34] o esbozo [35].

Otra dirección de la investigación es proponer modelos de similitud, no sólo basados en la evaluación de una función distancia en un espacio métrico de características, sino también basados en aspectos psicológicos más cercanos al criterio humano de similitud.

2.3 Métodos alternativos

2.3.1 Discret wavelet transform. DWT

La transformada wavelet en su análisis multiresolución es una de las principales herramientas para la caracterización de textura desde hace algunos años. Varios estudios demuestran que el comportamiento del sistema visual humano asemeja a dicho análisis y es de esta característica la que permitió en su momento la mejora de los resultados. El análisis wavelet puede ser visto como una descomposición de la imagen en una familia de ortonormales reales y compactas soportadas en una base obtenida de la traslación y dilatación de una wavelet madre. Una de las principales ventajas de la wavelet sobre el tradicional análisis de Fourier es la detección y estudio sistemático de las frecuencias bajas y altas que componen una señal ó imagen.

2.3.2 Dual-tree complex wavelet transform. DT-CWT

Métodos posteriores al anterior como el de la transformada wavelet compleja de doble árbol fueron propuestos para mejorar la eficiencia en la caracterización de las imágenes; sus filtros encuentran la parte real e imaginaria de los coeficientes wavelet. La buena selectividad direccional y eficiente carga computacional la postulan como un buen

método para obtener las características de textura. El método propuesto por Kingsbury [36] usa eficientemente las propiedades de DT-CWT al encontrar un buen análisis en el espacio frecuencia y orientación para clasificar diferentes texturas.

Con la transformada wavelet de árbol la imagen es descompuesta en 4 sub-imágenes que corresponderán a las sub-bandas paso bajo-bajo, bajo-alto, alto-bajo y alto-alto. Después de calcular la energía de cada sub-imagen en el mismo nivel, la descomposición continúa en la sub-banda con la máxima energía. El método se concentra en el canal frecuencial más significativo para el caso particular. Generalmente consideramos que la energía de la imagen esta concentrada en la sub-banda paso bajo-bajo, pero esto no necesariamente valido para algunas texturas y, es de la observación que el método propuesto por Kingsbury mejora los resultados que ofrece la transformada wavelet piramidal. Otra propiedad sobresaliente dada por el autor es la selectividad direccional que en comparación con DWT mejora el desempeño. La DT-CWT descompone la imagen en seis sub-bandas con una fuerte orientación en seis diferentes ángulos y dos sub-bandas paso bajo. Las principales características del método propuesto son la gran selectividad direccional y análisis del canal frecuencial dominante con gran precisión.

2.3.3 Gabor

Alan Conrad Bovik [37] utiliza el filtro Gabor ya que la textura es modelada como un parámetro que contiene un rango limite de frecuencias y es de está que la distinción de textura es significativa. La descomposición de la imagen está dada en múltiples canales espacio frecuencia y de orientación. Las diferentes características espacio frecuencia, orientación y fase son usadas para segmentar regiones de textura.

Los filtros Gabor 2D son usados para diferentes propósitos ya que puede ser seleccionada la orientación, ancho de banda y frecuencia central, además de tener un optimo desempeño en la resolución espacial y espacio frecuencia. Los límites entre

texturas pueden ser detectados al comparar la respuesta en amplitud del canal. Las grandes variaciones en fase pueden detectar discontinuidades en la textura.

2.3.4 Hermite

Los filtros cartesianos hermite bidimensionales permiten descomponer una señal localizada en $l_v(x-p, y-q) = v^2(x-p, y-q)l(x, y)$ por una ventana gaussiana $v(x, y)$, σ y de energía unitaria, bajo un conjunto de polinomios ortogonales de hermite $H_{n-m,m}(x/\sigma, y/\sigma)$. Los coeficientes $l_{n-m,m}(p, q)$ localizados en la posición $(p, q) \in P$ son por lo tanto obtenidos por convolución entre la señal $l(x, y)$ y los filtros de hermite. Los filtros de hermite en el dominio frecuencial son del tipo paso banda gaussiano donde su máximo esta dado por $(\omega\sigma)^2 = 2n$. Por consecuencia los filtros de orden extremo analizan sucesivamente las frecuencias superiores de la señal. El filtro de orden cero representa la componente continua (C. C.) de la señal y en el caso de la textura puede no interesar.

También es importante resaltar que los filtros son sensibles a una orientación en relación a los ejes coordenados. La selección de la frecuencia radial y la orientación de cada uno de los filtros de hermite no pueden hacerse directamente, ya que el orden de los filtros hermite determinan la frecuencia radial y orientación de estos. Esta limitación es la más importante en los filtros cartesianos de hermite.

2.3.5 Hermite tipo Gabor

Indexado de imagen y video es sugerido por Rivero Moreno en su trabajo de investigación [38]. Reporta la transformada hermite y Laguerre con excelentes propiedades en el dominio espacial y temporal respectivamente.

Como el tema de interés no aplica al video, en adelante hermite será solo mencionada.

El filtro gabor han sido usado en el análisis de textura por su buena selectividad direccional a diferentes frecuencias, sin embargo, en la base no ortogonal que la genera está su principal limitación.

Al consultar el documento de Rivero Moreno observamos un análisis teórico del filtro hermite y gabor que nos permite conocer propiedades frecuenciales de cada uno de ellos. Además una relación analítica entre ambos permite modelar un filtro que tenga la estructura del filtro hermite pero con una descomposición en frecuencia similar al filtro gabor. El filtro resultante es llamado hermite tipo gabor.

La extracción de primitivas son la base del trabajo.

2.3.6 Hermite direccional –rotada-

La transformada hermite direccional [39] genera un campo vectorial de energía asociado a la imagen en estudio. Este campo vectorial contiene por cada píxel, un vector cuya magnitud y ángulo determinan la cantidad y dirección de energía presente. Esta representación de imágenes la tenemos con los coeficientes polinomiales unidimensionales y, que consiste en un mapeo de los coeficientes 2D a coeficientes 1D logrando maximizar la energía direccional de los primeros. Lo anterior significa que el vector asignado a cada píxel está orientado en la dirección de máxima energía (máxima magnitud). Los vectores con magnitud cero son regiones planas y homogéneas de la imagen (patrones 0D), vectores con magnitudes considerables y ángulos orientados están referidos a bordes, líneas, calles, carreteras o estructuras bien orientadas (patrones 1D).

El campo vectorial de energía es particularmente importante debido a la cantidad de aplicaciones que a partir de él se pueden derivar. Algunas de las más importantes tienen que ver con la detección de bordes y estructuras orientadas en escenas, lo cual puede generar resultados considerablemente buenos en el ámbito de la segmentación y clasificación para la detección de caminos en planeación urbana, estudio de zonas costeras y sus límites, detección de cauces de ríos y extensión de vegetación en ecología, también ha sido utilizado con éxito en la detección de flujo. Por lo anterior, la correcta determinación de la magnitud y dirección de los vectores asociados a cada píxel es importante.

Capítulo 3

Transformada Hermite multiescala

Con el objetivo de fundamentar la herramienta matemática que utilizaremos, en las siguientes secciones describiremos un modelo multicanal con potencial para representar imágenes naturales en la teoría espacio-escala. Con base en una inspiración biológica los parámetros de dicho modelo pueden definirse de acuerdo a ciertas propiedades del sistema visual humano. En la primera parte estudiaremos el caso unidimensional (1-D) y posteriormente extenderemos los resultados al caso bidimensional. Los operadores obtenidos en la fase de análisis y síntesis son derivadas de Gaussiana, de tal forma que para el caso bidimensional (2-D) podremos representar la imagen en función de la respuesta de filtros direccionales. Sustituir la normalización de coordenadas por un muestreo nos permite un análisis de la señal similar al de la Transformada de Hermite Jerárquica (THJ). También definiremos la versión discreta de las transformadas multiescala y multidireccional a partir de la representación espacio-escala de señales discretas, en la cual se mantienen muchas de las propiedades del caso continuo. Finalmente mencionaremos aplicaciones prácticas dadas al modelo considerado.

3.1 Introducción

Frecuentemente los humanos evalúan los resultados en el procesamiento de imágenes, en tal caso los errores son significativos solo si se perciben. Con un profundo entendimiento del Sistema Visual Humano (SVH) podremos obtener mejor representación de imágenes tras aplicar las propiedades que lo caracterizan.

El modelo Gabor es uno de los más conocidos del SVH a nivel de la corteza visual [40], [41]. Diseñar filtros para simular el procesamiento visual ha sido de interés en esta área y, tal es caso de la transformada *Cortex* [42] cuyas funciones de análisis fueron diseñadas para aproximar el perfil de las funciones de Gabor, con la ventaja de poder ajustar los anchos de banda radial y angular. Un modelo alternativo dado a conocer por Young en 1987 esta fundamentado en la Gaussiana y sus derivadas [1]. Las funciones anteriores al igual que las funciones de Gabor son espacialmente locales y, presentan regiones de excitación e inhibición alternadamente en una envolvente que se desvanece gradualmente. La derivada Gaussiana (DG) como demostró Young, modela con mayor precisión los datos cuantificados en los campos receptivos que las funciones de Gabor, además de la ventaja adicional que representa la ortogonalidad de las funciones situadas en un mismo punto de análisis. También se ha demostrado que las DG pueden modelar los campos receptivos a nivel de las neuronas sensibles a la disparidad binocular [43].

Basados de estos operadores han derivado algunos modelos matemáticos a una sola escala [44],[45] y [2]. El último extiende al caso multiescala en [46] y en ambos casos la DG es interpretada como el producto de un polinomios Hermite con la ventana Gaussiana. La señal multiplicada por la ventana expande sobre la base de los polinomios Hermite.

3.2 Caso 1-D: Transformada Hermite Multiescala

Construir una transformada multiescala basada en DG y a la cual referiremos como transformada Hermite Multiescala (THM) será el objetivo de la presente sección. El desarrollo de las expresiones puede ser consultado en el Apéndice A.

3.2.1 Descomposición multiescala

Poder asimilar el concepto de escala se facilita si consideramos la imagen como una representación de una escena física, es decir, dicha representación estará caracterizada por cantidades físicas a medir ó estimarse. Por ejemplo, el efecto físico de la distancia de el objeto hacia el punto de observación lo podemos traducir en: un escalamiento de la distribución de intensidades de la representación. Esto significa que el efecto de alejarnos del objeto observado incrementa la escala de representación del mismo y, por lo tanto, una pérdida apreciable en el detalle; mientras que al acercarse disminuye la escala y percibimos más detalles.

Representación espacio-escala [47] implica considerar versiones escaladas de la Gaussiana de integral unitaria $G(x) = \exp(-x^2)/\sqrt{\pi}$ y mediante la operación de convolución el filtrado de la luminancia $L(x)$, es decir, $L(x,s) = L(x) \otimes G(x,s)$ es la representación de la señal a la escala s , donde

$$G(x,s) = \frac{G\left(\frac{x}{\sqrt{4s}}\right)}{\sqrt{4s}}$$

es la Gaussiana a la escala $s > 0$.

La variable normalizada $\xi = \frac{x}{\sqrt{4s}}$ es referida como coordenada natural.

Para entender la relación existente entre estructuras de la imagen que son

percibidas a diferentes escalas, primeramente tendremos que comprender la forma en que el SVH realiza dicho proceso. Resultados psicofísicos avalan la existencia de por lo menos cuatro canales en cada punto del campo visual a nivel de la retina (específicamente en las células ganglionares [48]). Dichos canales exhiben una respuesta aproximada a la diferencia de Gaussianas (o DoG del inglés *Difference of Gaussians*) cuya respuesta al impulso se escribe como $DoG(x, s_1, s_2) = G(x, s_1) - G(x, s_2)$ para $s_2 > s_1$.

La respuesta del filtro DoG contiene información que deberá agregarse a la representación de la escala s_2 para obtener la representación de la escala menor s_1 .

Así, la señal puede ser expresada como una suma de canales DoG:

$$L(x) = \sum_{k=-\infty}^{\infty} L(x) \otimes DoG(x, s_{k-1}, s_k) \quad (3.1)$$

con $s_k > s_{k-1}$.

El parámetro de espaciamiento $\tau = (s_k - s_{k-1})/s_k$ será constante y condición para todo entero k , lo cual está en acuerdo con datos psicofísicos que soportan la existencia de un número de canales en el SVH y cuyas frecuencias centrales mantienen una relación de aproximadamente una octava [48], es decir $\tau \approx 0.75$.

3.2.2 Análisis de estructuras

En un gran número de aplicaciones es condición indispensable saber qué parte de la información contenida en cada canal hablando en términos perceptivos deberá ser agregada con mayor intensidad (bordes, líneas, esquinas, etc.) y qué parte de información deberá ser atenuada al resultar molesta para la vista (texturas finas o ruido). Mencionado lo anterior resulta claro y apropiado considerar un análisis cuantitativo para discernir entre las diferentes estructuras de interés [49]. Así, la estructura local de una señal se infiere a partir de las derivadas escaladas. Por otra

parte, las derivadas de $L(x, s)$ son obtenidas con la convolución de la función $L(x)$ con el operador derivada de Gaussiana: $G_n(x, s) = \partial_x^n G(x, s)$.

Para expresar el filtro DoG en función de las DG tendremos que desarrollar la expansión en serie de Taylor de la Gaussiana a la escala s_{k-1} alrededor de la escala s_k , es decir,

$$DoG(x, s_{k-1}, s_k) = \sum_{n=1}^{\infty} \frac{(-\tau s_k)^n}{n!} G_{2n}(x, s_k) \quad (3.2)$$

Limitar el número de términos en la serie a un número pequeño es aconsejable en la práctica. Estudios neurofisiológicos resaltan que el SVH trabaja con derivadas hasta de cuarto orden [1], [50]. Sin embargo en este documento se consideran todos los términos de la serie ya que las aproximaciones podrán ser estudiadas como casos particulares.

3.2.3 Definición de la THM

De la expresión (3.2) observamos que las derivadas de orden impar no contribuyen a la descripción de la imagen como se indica en [51] y en donde se reportan experimentos de realce basado en DG de orden par. Por otro lado, es aconsejable que al menos en la etapa de análisis aparezcan todas las derivadas con el fin de tener una descripción más rica de la estructura de la señal. Distinguir entre la fase de análisis y síntesis es necesario y para lograrlo factorizamos las DG que aparecen en la serie (3.2) con la propiedad de cerradura bajo convolución. Además, si interpretamos el operador convolución como producto interno, podremos escribir la expresión para la reconstrucción de la señal como:

$$L(x) = \sum_{k=-\infty}^{\infty} \sum_{n=1}^{\infty} \tau^n D_n^{(k)}(x) \quad (3.3)$$

donde

$$D_n^{(k)}(x) = \left\langle L_n^{(k)}(\xi), G_n^{(k)}(x, \xi) \right\rangle_{\xi} \quad (3.4)$$

Y para $n = 1, 2, \dots$ tenemos las funciones que describen *detalles* de la señal a todas las escalas y que están determinadas por las derivadas escaladas,

$$L_n^{(k)}(\xi) = \left\langle L(x), G_n^{(k)}(x, \xi) \right\rangle_x \quad (3.5)$$

las cuales son determinadas proyectando la señal de entrada sobre las *funciones de análisis*

$$G_n^{(k)}(x, \xi) = \frac{1}{\sqrt{2s_k}} G_n^* \left(\frac{x}{\sqrt{2s_k}} - \xi \right) \quad (3.6)$$

para $n = 1, 2, \dots$ y $s_k = (1 - \tau)^{-k} s_0$ para todo k entero.

El proceso de análisis esta dado en la ecuación (3.5) y es lo que llamaremos THM directa, mientras que el proceso de síntesis y que llamaremos THM inversa estará descrito en la ecuación (3.3) y las *funciones de detalle* (3.4).

3.2.4 Acotamiento del rango de escalas

Si consideremos un número finito K de canales DoG y suponemos que la escala menor es suficientemente pequeña. La descomposición multicanal se reduce a

$$L(x, s_0) = L(x, s_K) + \sum_{k=1}^K \sum_{n=1}^{\infty} \tau^n D_n^{(k)}(x) \quad (3.7)$$

donde s_0 y s_K determinan las resoluciones limite del análisis. El residuo paso bajas $L(x, s_K)$ no puede ser despreciado para valores de K en un rango práctico, pero su variación es generalmente menor que la de la señal original. Por otro lado, el residuo paso altas $L(x) - L(x, s_0)$ podemos despreciarlo ya que es suficientemente pequeño. En los casos para los cuales el residuo paso altas contiene información relevante, podemos

incluir un canal definido entre las escalas 0 y s_0 ($\tau = 1$), ya que la señal representa la escala cero. Por lo tanto, si generalizamos las definiciones de las ecuaciones (3.4), (3.5) y (3.6) para $n = 0$ escribiremos la THM de la siguiente forma

$$L(x) = D_0^{(K)}(x) + \sum_{n=1}^{\infty} D_n^{(0)}(x) + \sum_{k=1}^K \sum_{n=1}^{\infty} \tau^n D_n^{(k)}(x) \quad (3.8)$$

donde los primeros dos términos del miembro derecho son el residuo paso bajas y paso altas respectivamente. Cuando $K = 0$ se conoce como transformada Hermite de una sola escala (o simplemente TH). En tal caso no aparece el tercer término de la Ecuación 3.8.

3.2.5 Esquema piramidal continuo

Cuando la imagen es descompuesta en un número de subimágenes paso bajas o paso banda las cuales son submuestreadas en proporción a su resolución, le llamaremos *esquema piramidal*. Para cada nivel de la pirámide una relación lineal con los niveles adyacentes esta presente y dicha relación es invariante con la escala. Esta definición aplica en el contexto de señales discretas, pero puede ser extrapolada para el caso continuo si reemplazamos el remuestreo por un reescalamiento de la coordenada espacial. En tal caso todos los niveles de la pirámide quedan definidos en un dominio continuo.

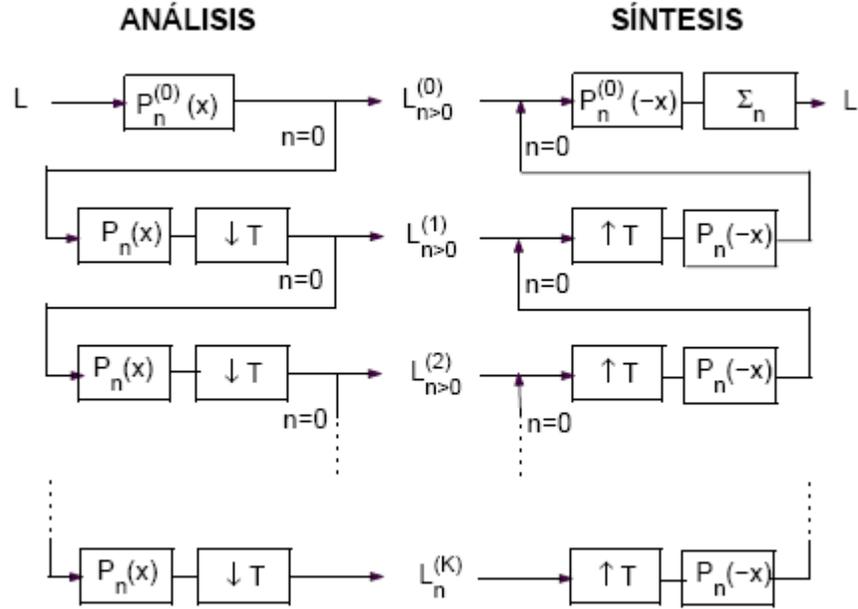


Figura 3.1: Esquema piramidal de análisis y síntesis de la THM.

Las funciones de análisis de un orden dado forman una pirámide Gaussiana conocida como relación piramidal de la THM y que esta definida por la relación lineal

$$\sqrt{\tau^n} G_n^{(k)}(x, \xi) = \int_{-\infty}^{\infty} G_0^{(k-1)}(x, \eta) P_n(T\xi - \eta) d\eta \quad (3.9)$$

Los coeficientes de la pirámide $P_n(x) = G_n^*(-x/\sqrt{4\sigma})/\sqrt{4\sigma}$ son derivadas de Gaussiana a la escala $\sigma = \tau T^2 / 4$, la cual se mantiene constante para todos los niveles. El factor $T = 1/\sqrt{1-\tau}$ indica la proporción entre las coordenadas de un nivel dado y el anterior. Para $\tau = 0.75$ se tiene $T = 2$.

Aplicando sucesivamente la propiedad (3.9) en la expresión (3.4) obtenemos la relación piramidal *ascendente*:

$$\tau^n D_n^{(k)} = \left\{ \left[\left(\sqrt{\tau^n} L_n^{(k)\uparrow T} \otimes P_n \right)^{\uparrow T} \otimes P_0 \right]^{\uparrow T} \dots \otimes P_0 \right\} \otimes P_0^{(0)} \quad (3.10)$$

$$\text{para } k = 1, \dots, K - 1 \text{ y } P_n^{(0)}(x) = \frac{G_n^* \left(\frac{-x}{\sqrt{4s_0}} \right)}{\sqrt{4s_0}}.$$

Por otra lado, si aplicamos a la fase de análisis -Ec. (2.5)- la misma propiedad obtenemos la relación piramidal *descendente*:

$$\sqrt{\tau^n} L_n^{(k)} = \left\{ \left[\left(L \otimes P_0^{(0)} \otimes P_0 \right)^{\downarrow T} \dots \otimes P_0 \right]^{\downarrow T} \otimes (-1)^n P_n \right\}^{\downarrow T} \quad (3.11)$$

Los simbolos $\uparrow T$ y $\downarrow T$ indican que la coordenada se multiplica por T y T^{-1} respectivamente. Las ecuaciones (3.10) y (3.11) son interpretadas en el esquema piramidal mostrado en la Figura 3.1 donde, por simplicidad, se escribió $L_n^{(k)}$ en lugar de $\sqrt{\tau^n} L_n^{(k)}$. Al mirar con atención el diagrama observamos que en el esquema piramidal la reconstrucción es realizada por etapas agregando sucesivamente las contribuciones de cada nivel comenzando con el residuo paso bajas.

3.2.6 Predicción entre niveles

La THM incrementa las dimensiones del espacio de representación de la señal (de posición a posición-escala-orden de derivación) que en la práctica se traduce en un incremento en la redundancia. Parte de la redundancia es debida a la dependencia lineal entre las funciones de análisis de distintas escalas. Particularmente es demostrable que las funciones del nivel $k - 1$ están relacionadas con las del nivel k mediante

$$\sqrt{\tau^n} G_n^{(k-1)}(x, \xi) = \sum_{m=n}^{\infty} c_{n,m} \int_{-\infty}^{\infty} \sqrt{\tau^m} G_m^{(k)}(x, \eta) P_{m-n}(\eta - T^{-1} \xi) d\eta \quad (3.12)$$

donde la constante $c_{n,m} = \sqrt{C_m^n T^{-m} [(1 + T^2)/2]^{m-n}}$.

La dependencia lineal de las funciones de análisis califican a la THM en un esquema de predicción del cual cada nivel es estimado a partir del siguiente.

3.3 Caso 2-D: Transformada Hermite multiescala-multidireccional

La generalización de la THM para señales bidimensionales será tema de interés en la presente sección ya que se desarrollara un esquema multidireccional basado en derivadas direccionales. Para tal caso, la transformada obtenida será referida como THM multidireccional (THMM). Demostraremos cómo la THMM puede ser obtenida mediante la THM 2-D. Para detalles consulte el Apéndice B.

3.3.1 Descomposición multiescala de señales 2-D

Cuando tenemos una señal multidimensional la representación multiescala es realizada filtrando la señal en cascada a lo largo de cada coordenada ya que la Gaussiana multidimensional es definida como un producto de Gaussianas 1-D de cada coordenada. Para el caso bidimensional se define el operador derivada escalada como $G_{m,n-m}(x, y, s) = G_m(x, s)G_{n-m}(y, s)$ para $m = 0, 1, \dots, n$ y $n = 0, 1, \dots$, donde n es el orden de derivación total. Por lo tanto, descomponer una señal bidimensional en canales DoG es análogo al caso 1-D.

3.3.2 Análisis de estructuras 2-D

Para la representación de patrones visuales, un factor físico importante a considerar es la inclinación relativa del sistema visual con respecto a un eje de referencia en la escena. La inclinación referida hará que las estructuras de la imagen sean percibidas

con una orientación determinada. Para el análisis con dicha orientación sólo podrá ser efectivo utilizando operadores sensibles a la orientación.

Asimilado lo anterior se hace evidente la aplicación de la derivada de Gaussiana. Con la función DoG la representación de estructuras orientadas se simplifica gracias a su comportamiento isotrópico. Es demostrable que el conjunto de las $n + 1$ derivadas bidimensionales de orden n permite representar patrones de hasta n orientaciones simultáneas [52].

En el análisis de orientación la conveniencia de trabajar con las versiones rotadas de estos operadores es sugerido en [53].

Sean los operadores $G_{m,n-m}(x, y, s, \theta) = G_{m,n-m}(x \cos \theta + y \sin \theta, -x \sin \theta + y \cos \theta, s)$, donde θ es el ángulo de rotación. Podemos demostrar que estas funciones están relacionadas linealmente con las funciones referidas al sistema original (ver Apéndice C).

Las derivadas direccionales de Gaussiana (DDG), denotadas por $G_{n,0}(x, y, s, \theta)$, han sido empleadas de manera eficiente en la detección y discriminación de orientaciones [52]. Dichos operadores son ideales para discriminar estructuras orientadas porque satisfacen la propiedad de direccionamiento (*steering*), a partir de la cual es posible demostrar que sólo necesitamos conocer la respuesta de la DDG de orden n en $n + 1$ direcciones para conocer la respuesta en cualquier dirección [54].

La serie de Taylor de la DoG bidimensional puede escribirse en función de las DG bidimensionales o bien en función de las DDG. La primera expansión produce la THM en dos dimensiones y que corresponde a la definida en la sección anterior pero aplicada sobre un sistema coordenado de dos dimensiones (ver Apéndice B).

Para el segundo caso apuntaremos un mayor interés, por lo que, será descrito con mayor detalle.

La expansión de la DoG para este caso es la siguiente.

$$DoG(x, y, s_{k-1}, s_k) = \sum_{n=1}^{\infty} \sum_{j=0}^n \frac{c_n (-\tau s_k)^n}{n!} G_{2n,0}(x, y, s_k, \theta_j) \quad (3.13)$$

donde $\theta_j = \theta_0 + j\Pi/(n+1)$ para $j = 0, \dots, n$ y c_n son las constantes definidas en (B.4). En la ecuación anterior observemos que el orden de derivación n determina la resolución angular y que el ángulo θ_0 es un parámetro libre que puede ser elegido según nuestro interés.

3.3.3 Definición de la THMM

A partir de la expansión (3.13) y haciendo un desarrollo similar al caso 1-D llegaremos a una expresión que resultara más apropiada para la descripción de imágenes, ya que las funciones de análisis obtenidas de dicha expresión incluirán las tres transformaciones geométricas fundamentales: traslación, rotación y escalamiento.

El proceso de análisis

$$L_n^{(k,j)}(\xi, \eta) = \langle L(x, y) G_n^{(k,j)}(x, y, \xi, \eta) \rangle_{(x,y)} \quad (3.14)$$

mediante las *funciones de análisis*

$$G_n^{(k,j)}(x, y, \xi, \eta) = \frac{1}{2s_k} G_{n,0}^* \left(\frac{x \cos \theta_j + y \sin \theta_j}{\sqrt{2s_k}} - \xi, \frac{-x \sin \theta_j + y \cos \theta_j}{\sqrt{2s_k}} - \eta \right) \quad (3.15)$$

con $\theta_j = \theta_0 + j\Pi/(n+1)$ para $j = 0, \dots, n$; $n = 1, 2, \dots$ y $s_k = (1-\tau)^{-k} s_0$ para todo k entero, es la llamada transformada Hermite multiescala-multidireccional (THMM), porque, emplea las DDG a múltiples escalas y orientaciones. La reconstrucción de la señal a partir de esta representación se le denomina THMM inversa.

En este caso, la reconstrucción de la señal

$$L(x, y) = \sum_{k=-\infty}^{\infty} \sum_{n=1}^{\infty} \sum_{j=0}^n c_n \tau^n D_n^{(k,j)}(x, y) \quad (3.16)$$

Es realizada sumando las contribuciones de las *funciones de detalle direccional*

$$D_n^{(k,j)}(x, y) = \left\langle L_n^{(k,j)}(\xi, \eta) G_n^{(k,j)}(x, y, \xi, \eta) \right\rangle_{(\xi, \eta)} \quad (3.17)$$

sobre el conjunto de escalas y orientaciones previamente definido.

3.3.4 Filtrado direccional

Representar imágenes con la THMM presenta ciertos inconvenientes ya que las funciones de análisis de la Ecuación (3.15) son no-separables y, por lo tanto, su implantación puede resultar costosa en cuanto a recursos computacionales. Sin embargo, esa desventaja puede ser reducida al calcular la transformada del caso separable y posteriormente transformar los coeficientes del mismo orden total a través de las *funciones de ángulo*

$$\alpha_{m,n-m}(\theta) = \sqrt{C_n^m} \cos^m \theta \operatorname{sen}^{n-m} \theta \quad (3.18)$$

para $m = 0, 1, \dots, n$, ya que las funciones de análisis en el punto $(\xi, \eta) = (0, 0)$ están relacionadas linealmente como

$$G_n^{(k,j)}(x, y) = \sum_{m=0}^n \alpha_{m,n-m}(\theta_j) G_m^{(k)}(x) G_{n-m}^{(k)}(y) \quad (3.19)$$

para $n = 1, 2, \dots$

Proyectar la señal sobre las funciones de análisis equivale a la convolución con los filtros direccionales, que son obtenidos al reflejar las funciones de la ecuación (3.19) respecto a x y y , más la normalización de coordenadas respecto a escala y orientación.

En la literatura el filtrado direccional es ampliamente citado como mecanismo para hacer explícita la información relevante de una imagen [54], [55], [53]. Por ejemplo, consideremos una estructura 1-D orientada un ángulo ϕ medido respecto al eje x . La THMM obtenida para esta estructura tiene la forma $L_n^{(k,j)} = \cos^n(\theta_j - \phi)H_n^{(k)}$ donde $H_n^{(k)}$ es la THM del perfil 1-D de la estructura. Los coeficientes 1-D contienen información de contraste y luminancia, mientras que los coeficientes $L_n^{(k,j)}$ presentan información de localización y orientación de la estructura a distintas escalas [46]. Trabajos anteriores sobre codificación [56] y estimación de flujo óptico en secuencias de imágenes [57] han demostrado la relación entre el filtrado direccional y el análisis de proyecciones locales.

3.4 Aplicaciones

Aplicaciones de las DG son reportadas en diferentes artículos de la literatura especializada. José Luís Silvan Cárdenas reporta en su trabajo: Compresión de imágenes basada en modelos Gaussianos de percepción visual [39] un esquema de representación de imágenes basado en una clasificación perceptual de los coeficientes de la transformada Hermite discreta rotada (THDR) sobre una latís de muestreo aproximadamente hexagonal. Donde las funciones de análisis de la THD corresponden a funciones binomiales que aproximan a las derivadas de Gaussiana con la ventaja de que pueden ser calculadas mediante un algoritmo rápido. El muestreo hexagonal muy eficiente en la transferencia de información espacial. Los coeficientes de un mismo orden son mapeados mediante una transformación unitaria especificada localmente. La transformación basada en funciones binomiales generalizadas, y como resultado se obtienen los coeficientes de la THD referidos a un sistema coordenado rotado de tal forma que la energía de los coeficientes se concentra a lo largo de una coordenada. Dicha representación permite un proceso de clasificación perceptual en el cual es

definido un umbral de error máximo tolerado para la aproximación del bloque por una constante (patrón 0-D) o por una estructura orientada (patrón 1-D). La compresión se consigue entonces eliminando los coeficientes que son visualmente irrelevantes.

Penélope López Quiroz [58] propone una clasificación de imágenes SAR (Radar de Apertura Sintética). El radar trabaja con una longitud de onda que permite en la mayoría de los casos monitorear zonas que se encuentran bajo la acción de fuertes y continuos cambios climáticos (fuertes lluvias, exceso de humedad, nubes). Las imágenes obtenidas con este dispositivo son utilizadas en el estudio de suelos, cultivos vegetación, elaboración de mapas de zonas de difícil acceso en geografía, estudio de zonas costeras y ciudades, estudio marítimo en oceanografía, Etc.

Una vez tratada la imagen para eliminar ruido en lo posible, utiliza la Transformada Hermite para conocer la configuración de patrones 0D, 1D y 2D presentes. La clasificación (0D,1D,2D) es una descripción dimensional de cada pixel, en otras palabras; aquellos pixeles que después de la aplicación de la transformada son identificados como patrones 0D, son elementos que forman parte de estructuras no orientadas, lo cual significa que no tienen una orientación definida. Este tipo de patrones se encuentra generalmente asociado con áreas planas u homogéneas de una imagen, como grandes áreas de cultivo o grandes concentraciones de agua, donde no existe estructura orientada en una dirección particular si no simplemente pixeles ó elementos de la imagen que a través de su intensidad representan dichas superficies planas u homogéneas. Los patrones 1D, se identifican en los pixeles que después de la aplicación de la transformada se reconocen como orientados, lo cual significa que tienen una orientación definida. Generalmente este tipo de patrones se encuentra en pixeles que forman parte de líneas y bordes como calles, carreteras, límites, bordes de grandes edificios y en general, toda clase de objetos y estructuras con una orientación predominante y definida dentro de la imagen. Finalmente los patrones 2D corresponden a pixeles que cuentan con más de una orientación definida. Generalmente se encuentran

presentes en esquinas, juntas, y píxeles ruidosos.

Juan Carlos Muñoz Torres propone la fusión de imágenes médicas (tomografía computarizada, resonancia magnética y tomografía por emisión de positrones) utilizando la transformada hermite rotada [59], con el propósito de obtener una mejor delineación en los contornos. El método intenta aumentar la información existente para facilitar el análisis, que en el caso de imágenes médicas resulta particularmente útil en estudios que utilizan técnicas como Detección de Bordes, Segmentación y Clasificación.

Martens en su artículo aplicaciones de la transformada hermite indica que los bordes representan en su mayoría las regiones no constantes de la imagen y su importancia ha sido reconocida ampliamente en visión computacional, donde muchos algoritmos han sido diseñados específicamente para detectar y localizar estas estructuras. La descomposición resalta las características más importantes, para que posteriormente sean seleccionadas y al fusionar las imágenes el resultado tenga la mayor información del proceso.

Visión automática y procesamiento de imágenes predice la utilidad de la THM en problemas de realce [51], detección de orientación [52], reducción de ruido [60], estimación del desenfoque para el cálculo de profundidad [61], estimación de movimiento [62], codificación [45], etc.

Capítulo 4

Extracción de características con la transformada hermite.

4.1 Arquitectura Bertino

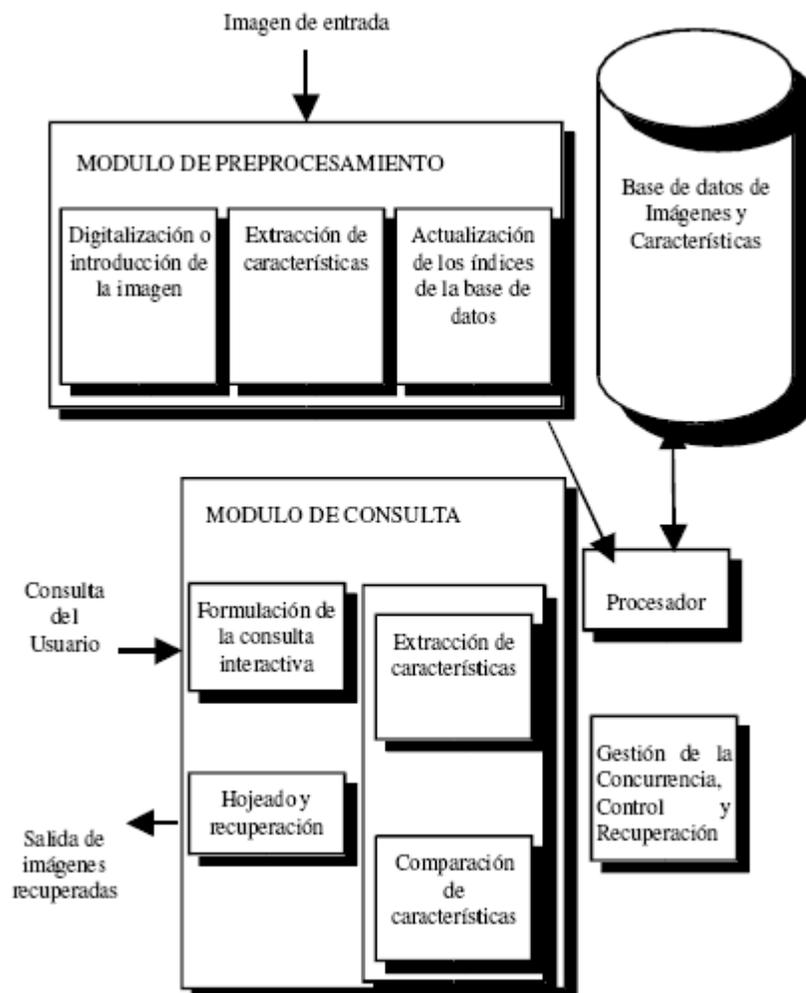


Figura 4.1 Arquitectura de un sistema para BD de imágenes

Elisa Bertino y sus colegas [21] proponen la arquitectura para una BD de Imágenes. Esta arquitectura, que se representa en la figura 4.1, muestra los módulos y funcionalidad de estos sistemas. Observando sus características, puede decirse que este modelo refleja la arquitectura de un subconjunto de posibles Bases de Datos de Imágenes: los sistemas de recuperación de imágenes basados en el contenido y, dentro de ellos, los que realizan consultas mediante ejemplo.

Según esta arquitectura, una base de datos de imágenes estaría integrada por varios subsistemas. El módulo de pre-procesamiento, de consulta, un sistema de almacenamiento secundario y un conjunto de procesos que permitan controlar la lectura-escritura a disco, o la gestión de la concurrencia si es que ésta se admite.

Antes de utilizar el sistema para recuperar imágenes es necesario poblar la BD introduciendo las imágenes que se desea almacenar. Para ello se utiliza el módulo de pre-procesamiento mediante el cual se obtiene la imagen, se extraen las características que previamente se haya decidido utilizar como descriptores y, luego, a partir de los descriptores se crean los índices para cada imagen. Estos descriptores se guardarán en algún sistema de almacenamiento secundario.

El módulo de consulta permite al usuario realizar búsquedas, utilizando para ello los distintos tipos de métodos de consulta posibles. En el modelo presentado por Bertino y otros, se asume que se realiza una extracción de características en el módulo de consulta, motivo por el que se puede afirmar que es un modelo orientado a las consultas mediante un ejemplo o un esbozo de la imagen. El proceso de consulta se puede realizar de diversas maneras:

1. Utilizando una imagen de ejemplo para solicitar al sistema imágenes similares.
2. Dibujando un esbozo de la imagen que se desea recuperar.

En el primer caso el usuario elige una imagen de ejemplo y le solicita al sistema que le devuelva todas las imágenes parecidas. El sistema extrae las características -descriptores- de la imagen utilizada como ejemplo y crea un patrón de características. Si el sistema tiene prevista alguna transformación sobre el patrón de características, transforma este vector en el índice final y, en caso contrario, lo utiliza directamente para consultar. El sub módulo de comparación de características es el encargado de comparar los índices buscando los más similares, utilizando para ello alguna técnica de reconocimiento de patrones. Según sea el diseño del sistema, se realizará una comparación exacta o una aproximada. Si la comparación es exacta la respuesta sólo devolverá aquellas imágenes cuyas descripciones coincidan con la imagen de ejemplo. Si es aproximada, el sistema devolverá varias imágenes ordenadas de mayor a menor similitud en función de una medida de distancia empleada. El segundo caso se diferencia del primero en la forma de introducir la imagen de ejemplo. En lugar de elegir una imagen determinada, el usuario realiza un dibujo o esbozo- de la imagen que desea recuperar.

Como es habitual en toda BD, son necesarios procesos que gestionen la obtención de la imagen final (asociación de los índices con sus respectivas imágenes), concurrencia (para el caso de múltiples usuarios) y control (gestión de errores en los procesos anteriores) realizando estas funciones el módulo correspondiente.

4.1.1 Otras arquitecturas

Varios autores han presentado propuestas de arquitectura para Sistemas de Recuperación de Imágenes Basados en Contenido (SRIBC), o para otro tipo de BD de Imágenes. Algunos de estos autores centraron su atención en la descripción del sistema desde un punto de vista de la obtención de características a nivel de píxeles y el posterior reconocimiento de patrones [63] o de la recuperación basada en textura [5]. Otros autores comentan las partes principales que estos sistemas han de tener [64], presentan una arquitectura general pero construida entorno a la recuperación basada en forma [65], o explican una arquitectura de tres componentes pensada para bases de datos multimedia distribuidas [66].

En las últimas revisiones del estado del arte, pueden encontrarse varios esquemas que presentan una arquitectura general para los SRIBC. En todos los casos [13], [14] se muestra un diagrama general, con poco detalle y que no contempla alguno de los métodos de indexación y consulta utilizados.

4.2 Extracción de características.

Existen muchos métodos y modelos propuestos en la literatura [3] para la caracterización de la textura. Podemos notar que la orientación, frecuencia y escalamiento forman en conjunto las características que definen la textura. La relación entre píxeles y sus vecindades caracteriza la textura por variación de luminosidad.

La figura 4.2 [38] muestra un esquema general para la caracterización de textura basada en su vector de características. La textura de una imagen T esta representada por un conjunto de vectores c_{pq} , cada uno asociado a un píxel de la imagen, donde (p, q) es la posición espacial del píxel y $[Q \times P]$ las dimensiones de la imagen. K define la dimensión del vector.

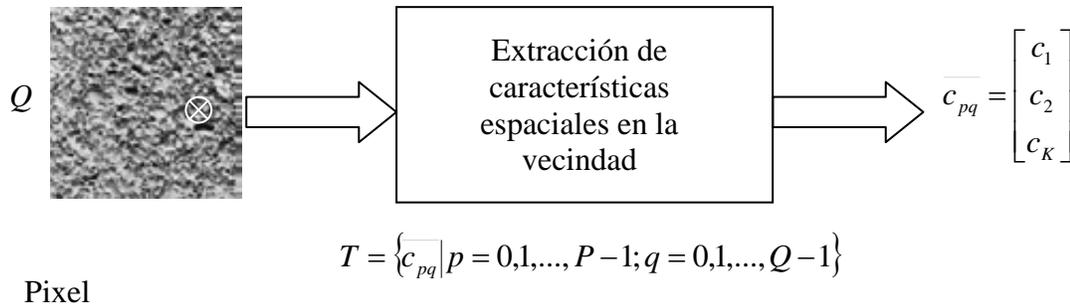


Figura 4.2 esquema general para caracterizar textura basado en el vector característico.

La figura 4.3 muestra el proceso que implica la técnica empleada para extraer características (orientación, frecuencia y escalamiento) por filtrado. Esta es realizada por una transformada lineal o filtrado de la textura [7] [8] para que posteriormente sea medida la energía ó la aplicación de un operador no lineal.

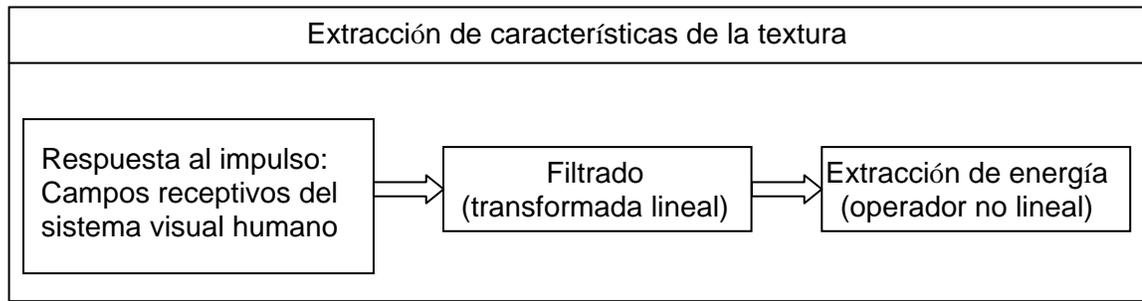


Figura 4.3 proceso para extraer características de textura por filtrado.

La textura de una imagen esta representada (vector de características) por un conjunto de componentes asociados a un operador particular. En nuestro caso, el operador que emplearemos fue utilizado en el trabajo de Portilla y Simoncelli [67] salvo que en este proyecto fue utilizado un calculo alternativo para obtener el resultado de mencionado operador –autocorrelación de imagen-.

Calcular la autocorrelación mediante la transformada de fourier como esta descrito en el trabajo de Simoncelli no fue realizo. Para ello obtuvimos los diferentes coeficientes de autocorrelación por cálculo espacial directo. Esto permite limitar los efectos de bordes relacionados al utilizar la transformada de fourier (consecuencia por periodicidad).

4.2.1 Autocorrelación

La autocorrelación es una herramienta matemática utilizada frecuentemente en el procesamiento de señales.

La función de autocorrelación se define como la correlación cruzada de la señal consigo misma. La función de autocorrelación resulta de gran utilidad para encontrar patrones repetitivos dentro de una señal, como por ejemplo, la periodicidad de una señal enmascarada bajo el ruido o para identificar la frecuencia fundamental de una señal que no contiene dicha componente, pero aparecen numerosas frecuencias armónicas de esta.

Definiciones de la función de autocorrelación

Dependiendo del campo de estudio se pueden definir diferentes tipos de autocorrelación sin que estas definiciones sean equivalentes. En algunos campos se utilizan indistintamente las funciones de autocorrelación y de autocovarianzas, dado que ambas sólo difieren entre sí en una constante de proporcionalidad que es la varianza (en este caso, la autocovarianza de orden $k=0$).

Estadística

En estadística, la autocorrelación de una serie temporal discreta de un proceso X_t es simplemente la correlación de dicho proceso con una versión desplazada en el tiempo de la propia serie temporal.

Si X_t representa un proceso estacionario de segundo orden con un valor principal de μ se define entonces:

$$R(k) = \frac{E[(X_i - \mu)(X_{i+k} - \mu)]}{\sigma^2}$$

donde E es el valor esperado y k el desplazamiento temporal considerado (normalmente denominado *desfase*). Esta función varía dentro del rango $[-1, 1]$, donde 1 indica una correlación perfecta (la señal se superpone perfectamente tras un desplazamiento temporal de k) y -1 indica no correlación. Es una práctica común en muchas disciplinas el abandonar la normalización por σ^2 y utilizar los términos *autocorrelación* y *autocovarianza* de manera intercambiable.

Procesado de señal

En el procesado de señales, dada una señal temporal $f(t)$, la autocorrelación continua $R_f(\tau)$ es la correlación continua cruzada de $f(t)$ consigo mismo tras un desfase τ , y se define como:

$$R_f(\tau) = f^*(-\tau) \otimes f(\tau) = \int_{-\infty}^{\infty} f(t+\tau) f^*(t) dt = \int_{-\infty}^{\infty} f(t) f^*(t-\tau) dt$$

Donde f^* representa el conjugado complejo y el círculo representa una convolución.

Para una función real, $f^* = f$.

Formalmente, la autocorrelación discreta R con un desfase j para una señal x_n es

$$R(j) = \sum_n (x_n - m)(x_{n-j} - m)$$

donde m es el valor esperado de x_n .

Frecuentemente la autocorrelación se calcula para señales centradas alrededor del cero, es decir con un valor principal de cero. En ese caso la definición de la autocorrelación viene dada por:

$$R(j) = \sum_n x_n x_{n-j}$$

La autocorrelación multidimensional puede definirse de manera similar. Por ejemplo, en tres dimensiones puede definirse la autocorrelación de una función como:

$$R(j, k, l) = \sum_{n, q, r} (x_{n, q, r} - m)(x_{n-j, q-k, r-l} - m)$$

Propiedades

Definiremos las propiedades de la autocorrelación unidimensional. La mayoría de sus propiedades son extensibles fácilmente a los casos multidimensionales.

- *Simetría:* $R(i) = R(-i)$,
- La función de autocorrelación alcanza un valor máximo en el origen, donde alcanza un valor real. El mismo resultado puede encontrarse en el caso discreto.
- Como la autocorrelación es un tipo específico de correlación mantiene todas las propiedades de la correlación.
- La autocorrelación de una señal de ruido blanco tendrá un fuerte pico en $\tau = 0$ y valores cercanos a cero y sin ninguna estructura para cualquier otro τ . Esto muestra que el ruido blanco carece de periodicidad.

- Según el teorema de Wiener-Khinchin, la función de autocorrelación es la transformada de Fourier de la densidad espectral:

$$R(\tau) = \int_{-\infty}^{\infty} S(f) e^{j2\pi f\tau} df$$

Igualmente, el espectro se relaciona con la función de autocorrelación:

$$S(f) = \int_{-\infty}^{\infty} R(\tau) e^{-j2\pi f\tau} d\tau$$

Por lo tanto la señal puede expresarse indistintamente en el dominio del tiempo (t) o el dominio de las frecuencias (f), al existir esta correspondencia entre ambos, y entendiendo que la señal está completamente determinada a partir del total de sus momentos o del total de sus frecuencias.

Para el caso particular de este trabajo la función de autocorrelación de una imagen está definida por la siguiente expresión:

$$R(m, n) = \frac{\sum_{x=0}^N \sum_{y=0}^N I(x, y) I(x + m, y + n)}{\sum_{x=0}^N \sum_{y=0}^N I^2(x, y)}$$

La función de autocorrelación de una imagen puede ser utilizada para cuantificar la regularidad y la aspereza (granular) de una textura.

Si la textura de una imagen es áspera, esta función de autocorrelación decrece suavemente y si la textura es regular la función de autocorrelación muestra un extremo local.

4.3 Similitud cuantitativa entre vectores de características

Para recuperar las imágenes con mayor similitud a la imagen de entrada según se describe en la sección 4.1 tendremos que comparar los vectores de características de la BD con el vector característico de la entrada. Para ello tenemos expresiones que permiten calcular la distancia entre vectores, es decir, aquellos vectores con la menor distancia corresponderán a las imágenes que presenten mayor parecido a la imagen de entrada. En los artículos especializados [68] [69] reportan los resultados obtenidos al experimentar con un gran numero de estas expresiones y de estas la capacidad para ser utilizadas para la comparación ya que presentan diferencias apreciables. A continuación aparecen aquellas que han sido utilizadas en mayor medida.

4.3.1 Distancia métrica

Sea X, Y vectores característicos de longitud n . Entonces podemos calcular la distancia entre los vectores característicos con las expresiones siguientes.

1. Distancia Minkowski (L_p métricas)

$$d(X, Y) = L_p(X, Y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}}$$

Donde $p > 0$.

2. Distancia Manhattan (L_1 métricas,)

$$d(X, Y) = L_{p=1}(X, Y) = \sum_{i=1}^n |x_i - y_i|$$

3. Distancia Euclidiana (L_2 métricas)

$$d(X, Y) = L_{p=2}(X, Y) = \|X - Y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

4. Cuadrado de la distancia euclidiana (sum square error, SSE), (mean square error, MSE)

$$d(X, Y) = L_{p=2}^2(X, Y) = SSE = \|X, Y\|^2 = \sum_{i=1}^n (x_i - y_i)^2 ;$$

$$d(X, Y) = \frac{1}{n} L_{p=2}^2(X, Y) = MSE = \frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2 ;$$

5. Distancia ángulo

$$d(X, Y) = -\cos(X, Y); \quad \cos(X, Y) = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2}}$$

6. Distancia coeficiente de correlación

$$d(X, Y) = -r(X, Y) = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{\left(n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right) \left(n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 \right)}}$$

7. Distancia Mahalanobis y distancia Mahalanobis entre vectores normales.

$$d(X, Y) = -\sum_{i=1}^n z_i x_i y_i$$

$$d(X, Y) = -\frac{1}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}} \sum_{i=1}^n z_i x_i y_i$$

Donde $z_i = \sqrt{\frac{\lambda_i}{\lambda_i + \alpha^2}}$, $\alpha = 0.25$, λ_i ó la distancia Mahalanobis simplificada

$$z_i = \sqrt{\lambda_i / (\lambda_i + \alpha^2)} \approx \sqrt{1/\lambda_i}$$

8. Distancia peso Manhattan

$$d(X, Y) = \sum_{i=1}^n z_i |x_i - y_i|, \quad z_i = \sqrt{1/\lambda_i}$$

9. Distancia peso SSE

$$d(X, Y) = \sum_{i=1}^n z_i (x_i - y_i)^2, \quad z_i = \sqrt{1/\lambda_i}$$

10. Distancia peso ángulo

$$d(X, Y) = -\frac{\sum_{i=1}^n z_i x_i y_i}{\sqrt{\sum_{i=1}^m x_i^2 \sum_{i=1}^m y_i^2}}, \quad z_i = \sqrt{1/\lambda_i}$$

11. Distancia Chi cuadrado

$$d(X, Y) = X^2 = \sum_{i=1}^n \frac{(x_i - y_i)^2}{x_i + y_i}$$

12. Distancia Canberra

$$d(X, Y) = \sum_{i=1}^n \frac{|x_i - y_i|}{|x_i| + |y_i|}$$

13. Distancia Manhattan modificada

$$d(X, Y) = \frac{\sum_{i=1}^n |x_i - y_i|}{\sum_{i=1}^n |x_i| \sum_{i=1}^n |y_i|}$$

14. Distancia SSE modificada

$$d(X, Y) = \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2}$$

15. Distancia peso Manhattan modificada

$$d(X, Y) = \frac{\sum_{i=1}^n z_i |x_i - y_i|}{\sum_{i=1}^n |x_i| \sum_{i=1}^n |y_i|}; \quad z_i = \sqrt{1/\lambda_i}$$

16. Distancia peso SSE modificada

$$d(X, Y) = \frac{\sum_{i=1}^n z_i (x_i - y_i)^2}{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i^2}; \quad z_i = \sqrt{1/\lambda_i}$$

Si el vector X es almacenado en una base de datos, algunos componentes $\sum_{i=1}^n x_i^2$, $n \sum_{i=1}^n x_i^2$, $\left(\sum_{i=1}^n x_i\right)^2$, $\sum_{i=1}^n |x_i|$ de las expresiones de distancia podrán ser calculadas por adelantado y almacenadas en la base de datos para una comparación y búsqueda en menor tiempo.

Capítulo 5

Resultados experimentales

5.1 Imágenes de análisis. –álbum fotográfico de Brodatz-

Entorno a las investigaciones relacionadas con el indexado de imágenes varios reportes están documentados en las publicaciones especializadas y de éstas aprendemos las distintas herramientas matemáticas que fundamentan los proyectos. De lo anterior es fácil entender que cada uno de los trabajos concluya en resultados distintos, sin embargo, dichos métodos buscan objetivos comunes y en este mismo sentido podemos observar la necesidad de cuantificar la calidad de las propuestas. En general los trabajos de investigación están sometidos a parámetros cuantitativos y cualitativos que serán utilizados en el proyecto para la evaluación de la propuesta dada.

La evaluación a la que han sido sometidos los trabajos relacionados con el tema, nos compromete para considerar el mismo procedimiento y saber de los alcances del proyecto. En secciones siguientes estará descrita la metodología empleada que nos permitirá obtener y reportar resultados para la comparación respectiva con otros métodos.

Por lo pronto, identificaremos el conjunto de imágenes empleadas por los diferentes trabajos de investigación y que hacen referencia al álbum de texturas de Brodatz. La figura 5.1 muestra algunas texturas de mencionado álbum.

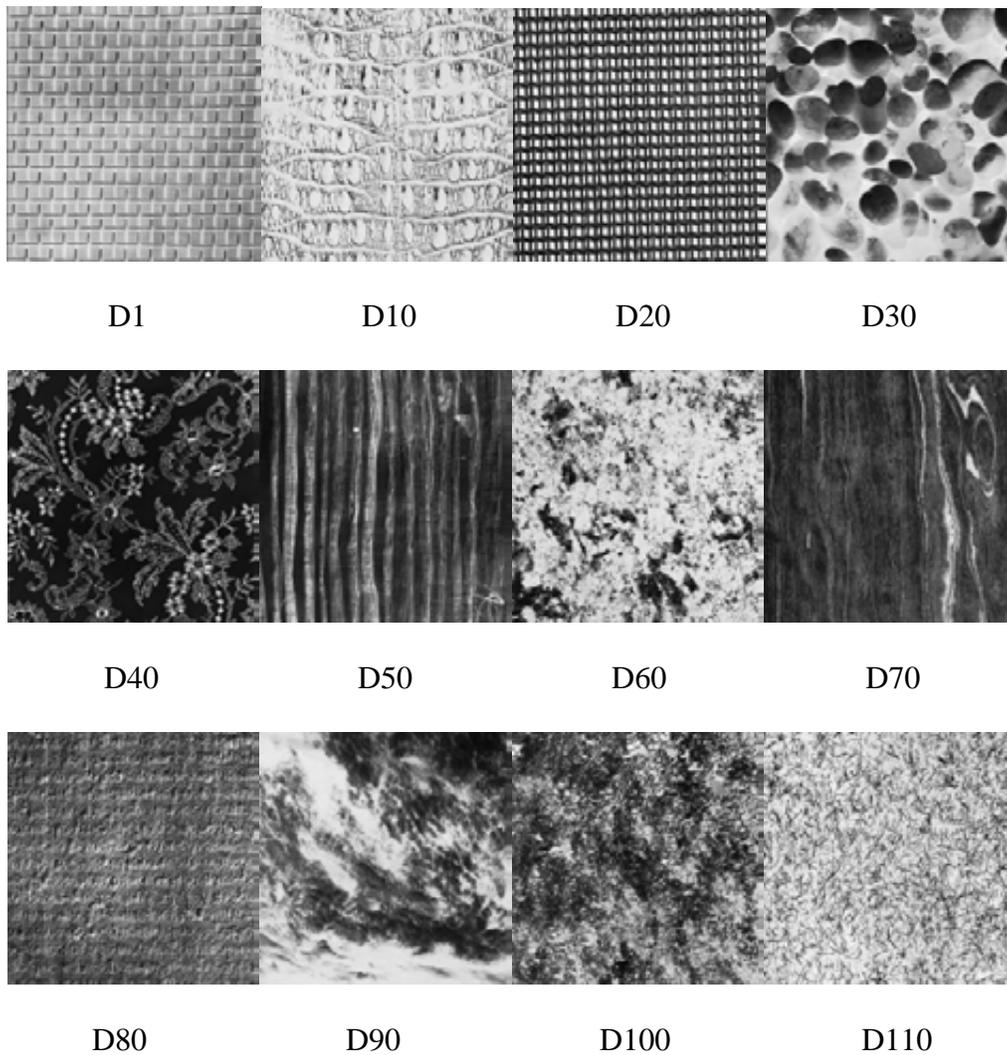


Figura 5.1 Algunas texturas del álbum de Brodatz.

Cabe la mención de que algunos autores utilizaron texturas diferentes, pero ya que compararemos resultados con métodos mencionados en el capítulo 2 tendremos que repetir los experimentos con los mismos parámetros.

Para comprobar la eficiencia del método propuesto, el álbum de texturas de Brodatz con 111 imágenes fue utilizado. Cada imagen de 640 x 640 fue dividida en 16 fracciones del mismo tamaño, así que la base de datos de textura obtenida fue de 1776. Por lo tanto, las fracciones de 160 x 160 son suficientemente grandes para ser empleadas en el sistema.

La figura 5.2 muestra la textura D108 y la figura 5.3 sus 16 fracciones.

IMAGEN SOLICITADA



Figura 5.2 textura D108

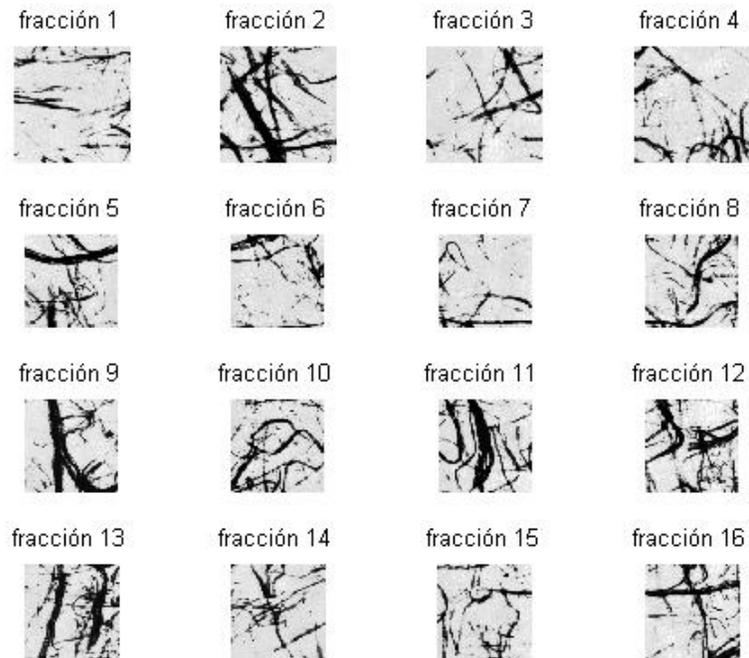


Figura 5.3 16 fracciones de la textura D108

5.2 Vectores de características.

De la sección 5.1 derivamos 1776 fracciones de imagen, que en adelante formaran la BD de textura que emplearemos para probar los algoritmos diseñados. A continuación describimos los pasos a seguir para obtener los vectores de características que formaran la BD que emplearemos para relacionar el total de texturas considerado en el experimento.

1. Descomponer la textura utilizando la transformada hermite con filtrado direccional apoyados del gradiente.

El número total de filtros depende del orden máximo D y este es calculado con la siguiente expresión $(D+1)(D+2)/2$. La figura 5.4 muestra la distribución de los filtros que son aplicados a la textura para obtener las sub-bandas correspondientes.

		m			
		0	1	2	3
n	0	00	01	02	03
	1	10	11	12	13
	2	20	21	22	23
	3	30	31	32	33

Figura 5.4 representación de los coeficientes de la transformada hermite. Los cuadrantes (m,n) representan el orden de la derivada de la ventana de análisis. El cuadrante $(0,0)$,

corresponde a la imagen original (componente continua). En general, los coeficientes dados en los cuadrantes grises son los más representativos desde el punto de vista perceptivo.

De la figura anterior notamos que el orden (0,0) corresponde a la componente continua y si esta no es considerada podemos calcular el número de sub-bandas con la expresión $D(D+3)/2$.

La figura 5.5 muestra las sub-bandas tras aplicar la transformada hermite de orden $D=3$ a una imagen.

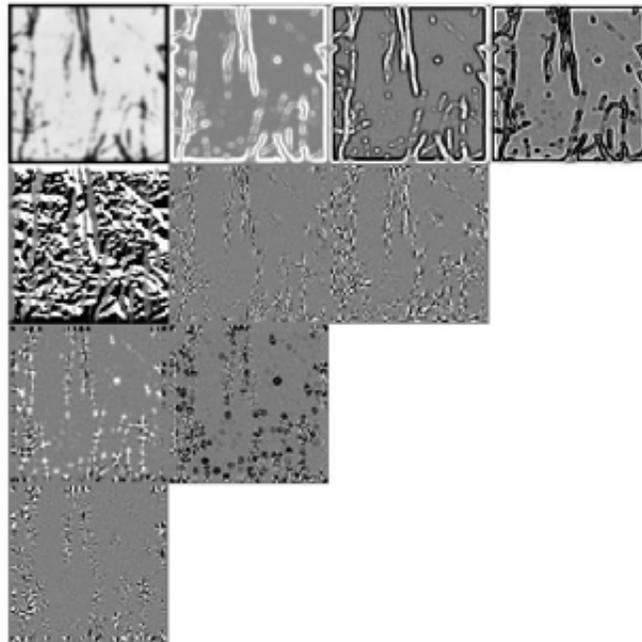


Figura 5.5 Transformada hermite de orden $D=3$ para la primera fracción de la imagen D108 -álbum de Brodatz-.

2. Calcular la autocorrelación de cada una de las sub-bandas siguiendo el procedimiento de la sección 4.2.

La autocorrelación implica desplazamientos espaciales a partir del píxel central en las direcciones x y y . Para nuestro caso el total de desplazamientos en cada dirección es $M=7$. Para la dirección x tenemos tres desplazamientos a la izquierda, tres desplazamientos a la derecha del píxel central más el origen. En la dirección y tres desplazamientos arriba, tres desplazamientos abajo del píxel central más el origen. La figura 5.6 muestra de manera esquemática los desplazamientos descritos anteriormente.

Modelo paramétrico, $M=7$ desplazamientos

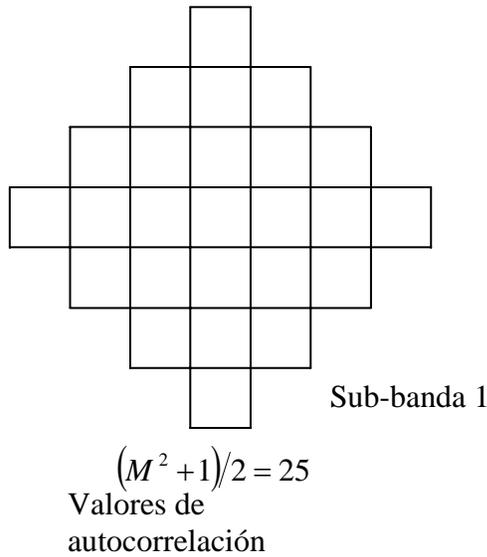


Figura 5.6 Desplazamientos de la imagen para calcular los valores de autocorrelación

Los valores de autocorrelación son simétricos, por lo que, no hay necesidad de hacer todos los cálculos. La mitad de la figura anterior cubre con los desplazamientos, más el píxel central. La figura 5.7 muestra los desplazamientos de imagen para obtener los valores de autocorrelación considerados inicialmente más otros no considerados.

El número de coeficientes para $M=7$ es $\frac{(M^2 + 1)}{2} = \frac{(7^2 + 1)}{2} = 25$. Este numero es suficiente para no hacer cálculos poco relevantes.

Modelo paramétrico, M=7 desplazamientos

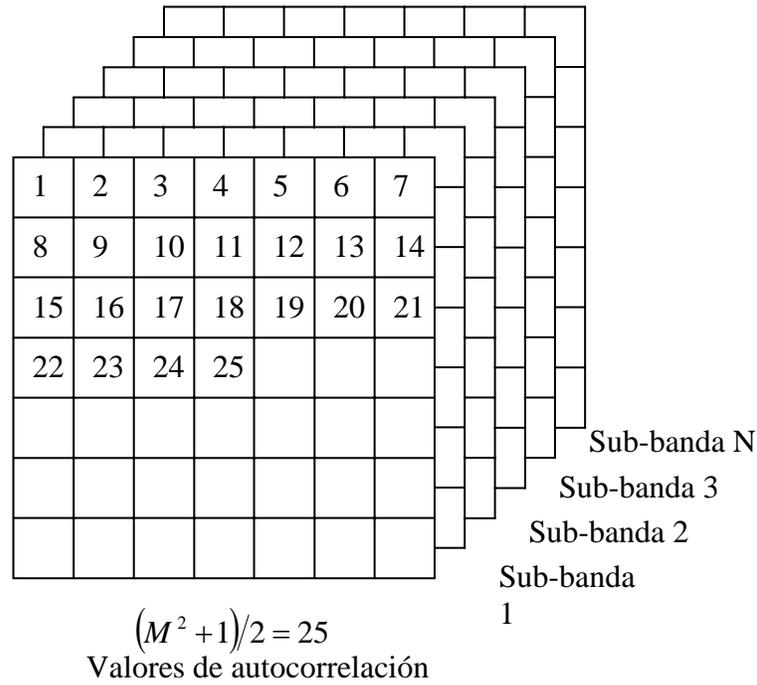


Figura 5.7 Desplazamientos de la imagen para obtener los valores de autocorrelación originales mas otros no considerados

3. Los valores obtenidos de calcular la autocorrelación, serán los índices ó componentes del vector característico.

La figura 5.8 representa el segmento de vector para una sub-banda y que implica una reducción dimensional del vector característico.

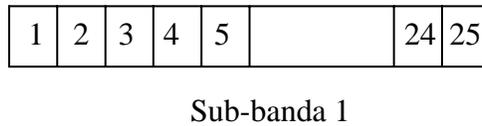


Figura 5.8 Segmento de vector que representa una sub-banda

La figura 5.9 muestra los N segmentos de vector con sus respectivos valores de autocorrelación que formar el vector característico para representar una imagen.

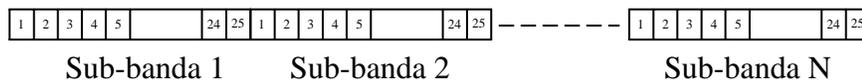


Figura 5.9 Vector característico de una imagen con N sub-bandas. En nuestro caso son fracciones de imagen.

Para nuestro caso tenemos que el orden $D=5$ será aplicado en la transformada hermite, por lo tanto, el numero de sub-bandas que tendremos en cada fracción de imagen es $N = (D + 1)(D + 2)/2 = (5 + 1)(5 + 2)/2 = 21$. Del punto anterior tenemos $M=7$ que deriva en 25 valores de autocorrelación, por lo tanto, tendremos $(25)(21)=525$ índices o componentes en el vector característico de cada fracción de imagen.

4. Para formar la BD, aplicar los puntos 1,2 y 3 a cada una de las texturas que tendrá la BD de imágenes.

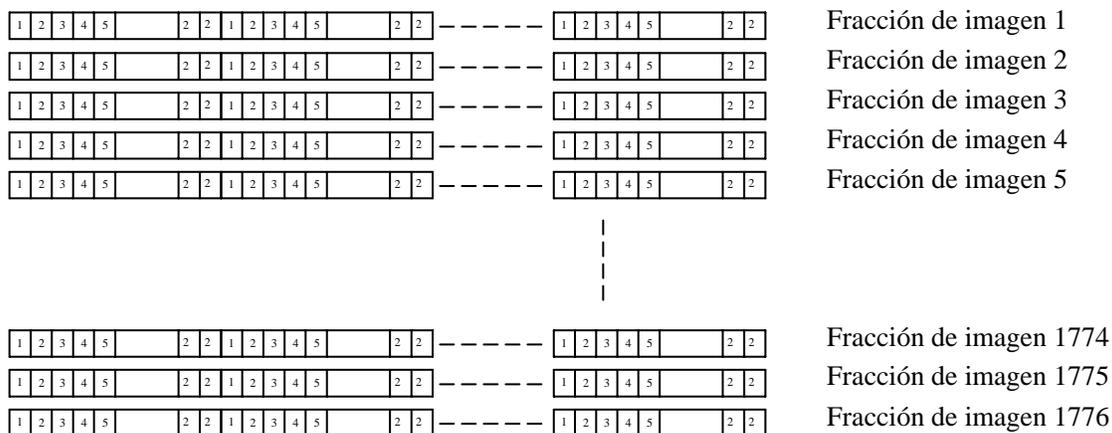


Figura 5.10 Representación de la BD para las 1776 fracciones de imagen utilizadas para someter los algoritmos propuestos.

5.3 Similitud cuantitativa ó distancia entre vectores.

De la sección 4.1 recordamos que en la arquitectura Bertino cuando hacemos una solicitud a la BD de imágenes, deberemos dar una imagen de entrada al sistema que en nuestro caso y con el propósito de evaluación, la textura de entrada será cualquiera de las 1776 fracciones derivadas de la sección 5.1. A la textura de entrada o referencia el sistema calculara el vector de característica respectivo para compararlo ó medir la similitud cuantitativa con los vectores de características de la BD de imágenes.

5.3.1 Distancia Canberra

La distancia métrica Euclidiana es una medida de similitud pero está no siempre es la mejor métrica. El hecho de que la distancia en cada dimensión este al cuadrado antes de la suma, resalta las características para una disimilaridad grande. Entonces será necesario normalizar las componentes del vector antes de encontrar la distancia entre dos texturas. La normalización a sido considerado en la distancia Canberra motivo por el cual la utilizaremos como medida de disimilaridad.

Si x y y es el vector característico de la BD y de la imagen de entrada respectivamente, entonces la distancia Canberra d esta definida por.

$$Canb(x, y) = \sum_{i=1}^d \frac{|x_i - y_i|}{|x_i| + |y_i|}$$

Las distancias son almacenadas en orden creciente, de las cuales las primeras corresponderán a las fracciones de imagen que más asemejan a la imagen de entrada ó referencia.

La eficiencia es medida con la tasa de recuperación promedio, que esta definida como el promedio porcentual de parámetros que pertenecen a la misma imagen solicitada de las 16 (9) primeras que más asemejan.

5.4 Evaluación.

Evaluación cuantitativa: eficiencia de recuperación porcentual .

Medir el desempeño del sistema fue realizado en términos de la tasa promedio de recuperación que esta definida como el promedio porcentual del número de fracciones de imagen que pertenecen a la imagen original después de haber dado una fracción de imagen como referencia y tener a la salida 16 ó 9 -según sea el caso- fracciones de imagen mas similares a la de entrada.

La figura 5.11 y 5.12 muestran la precisión del sistema donde el eje horizontal representa el número de imágenes recuperadas y el eje vertical la tasa de recuperación porcentual.

Para calcular dicha tasa de recuperación porcentual utilizamos la siguiente expresión:

$$P = (Q_1/Q_2)/N \times 100 \quad \text{donde}$$

Q_1 = número de fracciones de imágenes recuperadas que pertenecen a la original.

Q_2 = número de fracciones de imágenes que pertenecen a la original, según el caso:

$Q_2 = 9$, $Q_2 = 16$.

N = numero de imágenes recuperadas. $N = [9 \ 20 \ 30 \ \dots \ 120]$ ó $N = [16 \ 20 \ 30 \ \dots \ 120]$.

La tabla 5.1 resume la eficiencia de recuperación porcentual aplicando la *transformada hermite*. El primer valor (78.695 %) corresponde cuando la imagen original de tamaño 640x640 es dividida en 4x4 quedando fracciones de imagen de 160x160 y haber recuperado 16 fracciones de imagen. Cuando la imagen es dividida en 3x3 las fracciones de imagen quedan de 213x213 y una eficiencia del 83.78 % al haber recuperado 9 fracciones de imagen. Para 87.8212 % la imagen es dividida en 3x3 pero con fracciones de imagen de 256x256, resultando imágenes traslapadas; dicha eficiencia corresponde al recuperar 9 fracciones de imagen.

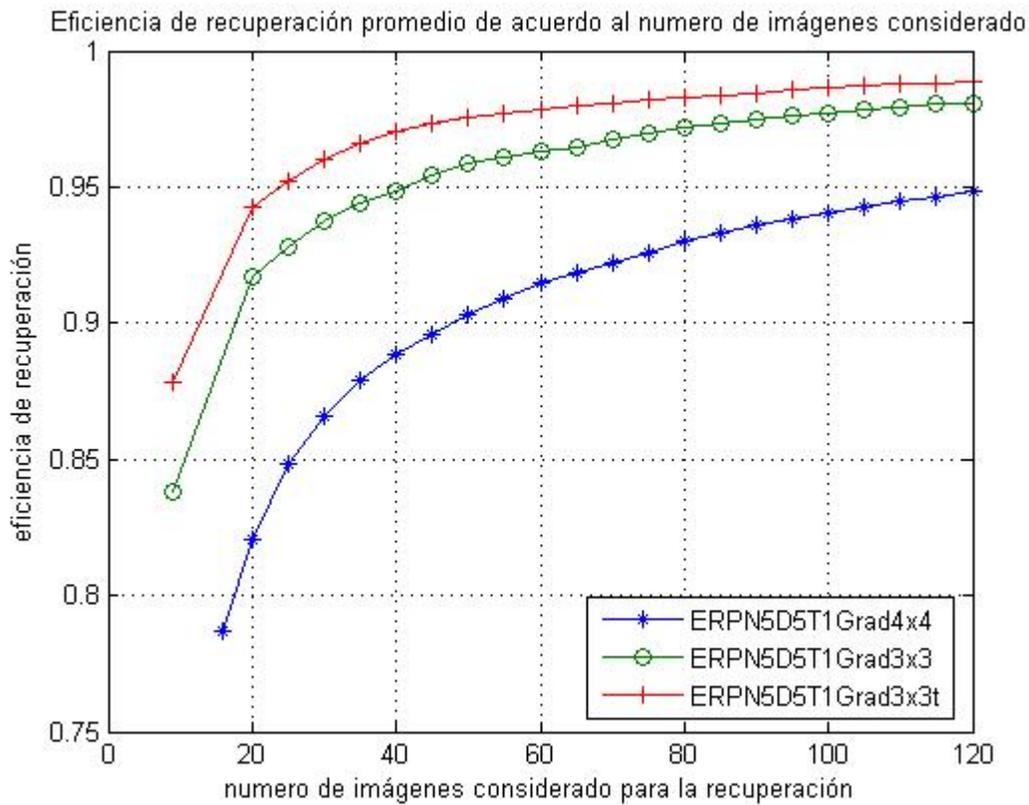


Figura 5.11 eficiencia de recuperación porcentual contra el numero de imágenes recuperado tras aplicar la *transformada hermite*. La línea --*-- corresponde para imágenes divididas en 4x4 fracciones de imagen. La línea --o-- corresponde para imágenes divididas en 3x3 fracciones de imagen. La línea --+-- corresponde para imágenes divididas en 3x3 fracciones de imagen traslapadas.

Imagen original 640x640, dividida en 4x4 quedando fracciones de 160x160	Imagen original 640x640, dividida en 3x3 quedando fracciones de 213x213	Imagen original 640x640 dividida en 3x3 traslapadas quedando fracciones de 260x260
N5D5T1Grad	N5D5T1Grad3x3	N5D5T1Grad3x3t
78.6951	83.7838	87.8212

Tabla 5.1 Tasa de recuperación porcentual aplicando *transformada hermite*.

La eficiencia de recuperación porcentual de la tabla 5.2 es resultado de considerar las mismas fracciones de imagen de los experimentos anteriores, salvo que en esta ocasión la

transformada hermite multiescala fue utilizada en los algoritmos respectivos.

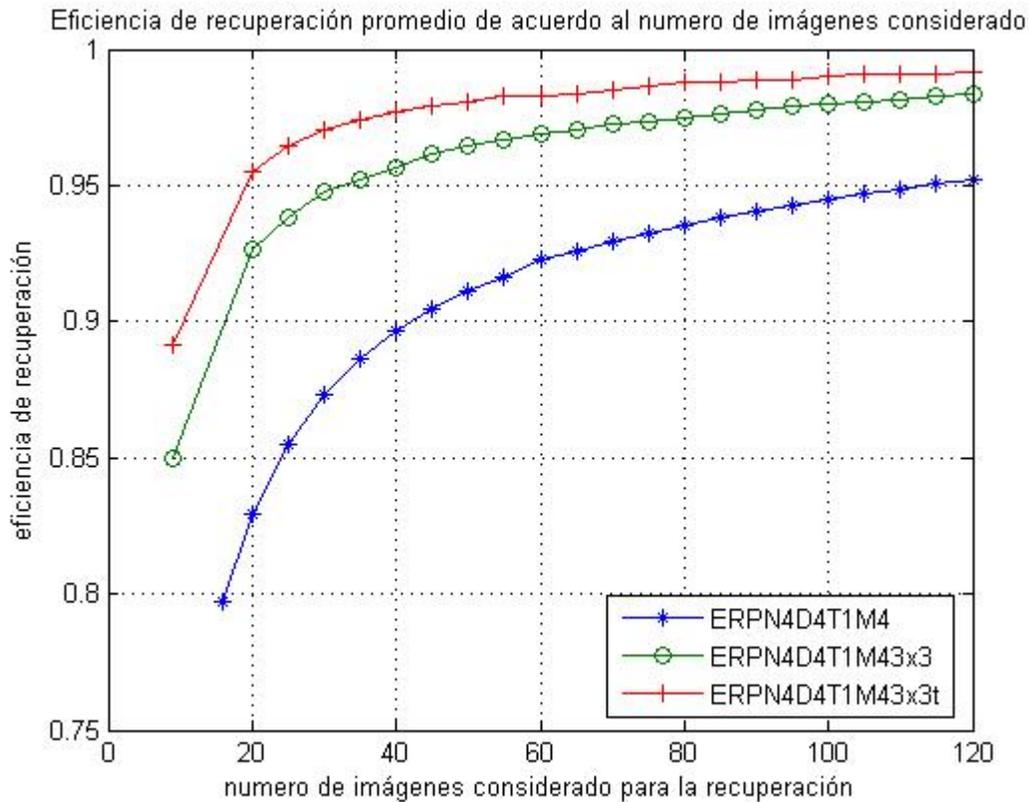


Figura 5.12 eficiencia de recuperación porcentual contra el numero de imágenes recuperado tras aplicar la *transformada hermite multiescala*. La línea --*-- corresponde para imágenes divididas en 4x4 fracciones de imagen. La línea --o-- corresponde para imágenes divididas en 3x3 fracciones de imagen. La línea --+-- corresponde para imágenes divididas en 3x3 fracciones de imagen traslapadas.

Imagen original 640x640, dividida en 4x4 quedando fracciones de 160x160	Imagen original 640x640, dividida en 3x3 quedando fracciones de 213x213	Imagen original 640x640 dividida en 3x3 traslapadas quedando fracciones de 260x260
N4D4T1M4	N4D4T1M43x3	N4D4T1M43x3t
79.7227	84.985	89.1113

Tabla 5.2 Tasa de recuperación porcentual aplicando *transformada hermite multiescala*.

De las tablas 5.1 y 5.2 podemos resaltar el incremento de un poco más del 1% al haber aplicado la TH y THM bajo las mismas condiciones de los experimentos realizados, sin embargo, este incremento trae como consecuencia un mayor calculo.

5.5 Análisis de resultados

Para conocer la eficiencia de recuperación de cada una de las imágenes utilizadas fue implementado un algoritmo que mostrara dicho porcentaje y así poder detectar las imágenes que para el sistema tiene dificultades en detectar.

La tabla 5.3 resume la eficiencia de recuperación porcentual de las imágenes consideradas bajo las siguientes condiciones: la imagen original -640x640- fue dividida en 4x4 fracciones de imagen -160x160- con la aplicación de TH.

imagen	porciento	imagen	porciento	Imagen	porciento	imagen	Porciento
45	21.8750	52	64.4530	46	87.8910	4	100.0000
43	23.0470	67	65.6250	103	87.8910	6	100.0000
44	28.5160	105	66.4060	28	89.4530	16	100.0000
90	31.6410	89	66.7970	10	90.2340	17	100.0000
31	37.8910	97	66.7970	79	90.2340	20	100.0000
42	41.4060	106	68.7500	86	90.2340	21	100.0000
58	41.7970	60	69.1410	24	90.6250	26	100.0000
91	42.1880	108	69.1410	71	90.6250	29	100.0000
7	46.0940	19	70.3130	78	90.6250	34	100.0000
38	46.0940	12	72.6560	111	91.4060	49	100.0000
73	48.4380	37	75.0000	109	91.7970	53	100.0000
88	49.2190	72	75.3910	110	92.5780	55	100.0000
99	49.2190	15	75.7810	51	92.9690	56	100.0000
27	50.3910	112	77.3440	48	94.1410	57	100.0000
13	50.7810	50	79.2970	68	94.9220	64	100.0000
69	50.7810	8	79.6880	95	94.9220	66	100.0000
59	51.1720	22	80.0780	101	95.7030	77	100.0000
54	52.7340	1	80.4690	76	96.0940	82	100.0000
62	52.7340	104	81.2500	75	96.4840	84	100.0000
61	55.4690	107	81.6410	96	96.4840	87	100.0000
98	57.0310	40	82.8130	18	96.8750	92	100.0000
100	60.1560	80	82.8130	94	97.6560		
23	61.3280	81	82.8130	102	97.6560		
2	62.5000	35	83.2030	85	98.0470		
36	62.5000	41	83.2030	83	98.4380		
63	62.5000	9	84.7660	33	98.8280		
70	63.2810	25	85.9380	47	98.8280		
30	64.0630	74	85.9380	11	99.2190		
5	64.4530	93	85.9380	32	99.6090		
39	64.4530	3	87.8910	65	99.6090		

Tabla 5.3. Eficiencia promedio de recuperación de las imágenes originales –tamaño 640x640-. Fraccionada 4x4 y tamaño de fracción 160x160.

De la tabla notamos que la eficiencia promedio de recuperación tiene un orden creciente que permite detectar cuales son las imágenes más difíciles de recuperar, así como también aquellas que presentan una menor dificultad en la recuperación. La figura 5.13 contiene las texturas respectivas de la tabla 5.3 las cuales están ordenadas de izquierda a derecha y de arriba-abajo.

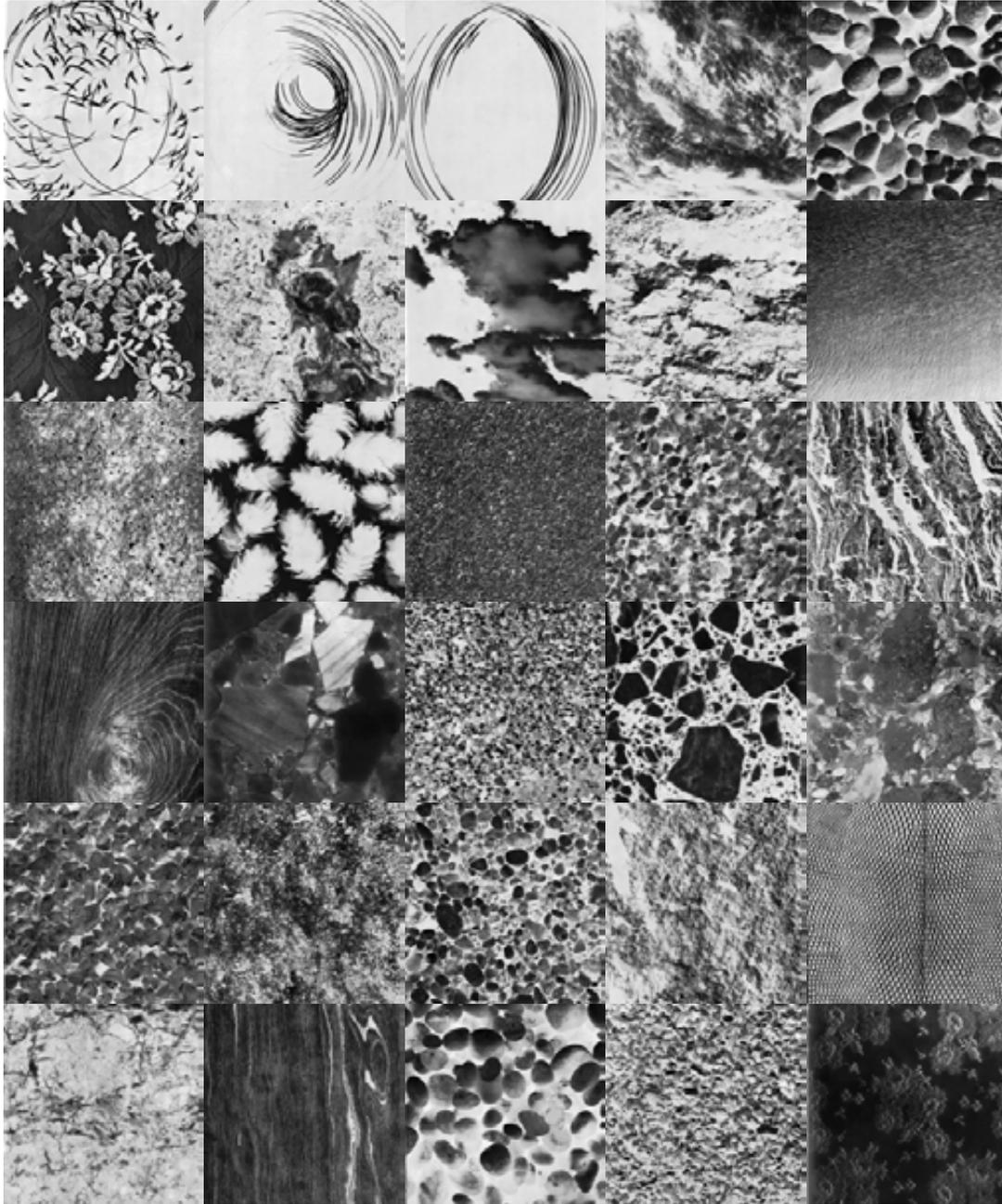


Figura 5.13. Primeras 30 imágenes que corresponden a la primera columna de la tabla 3

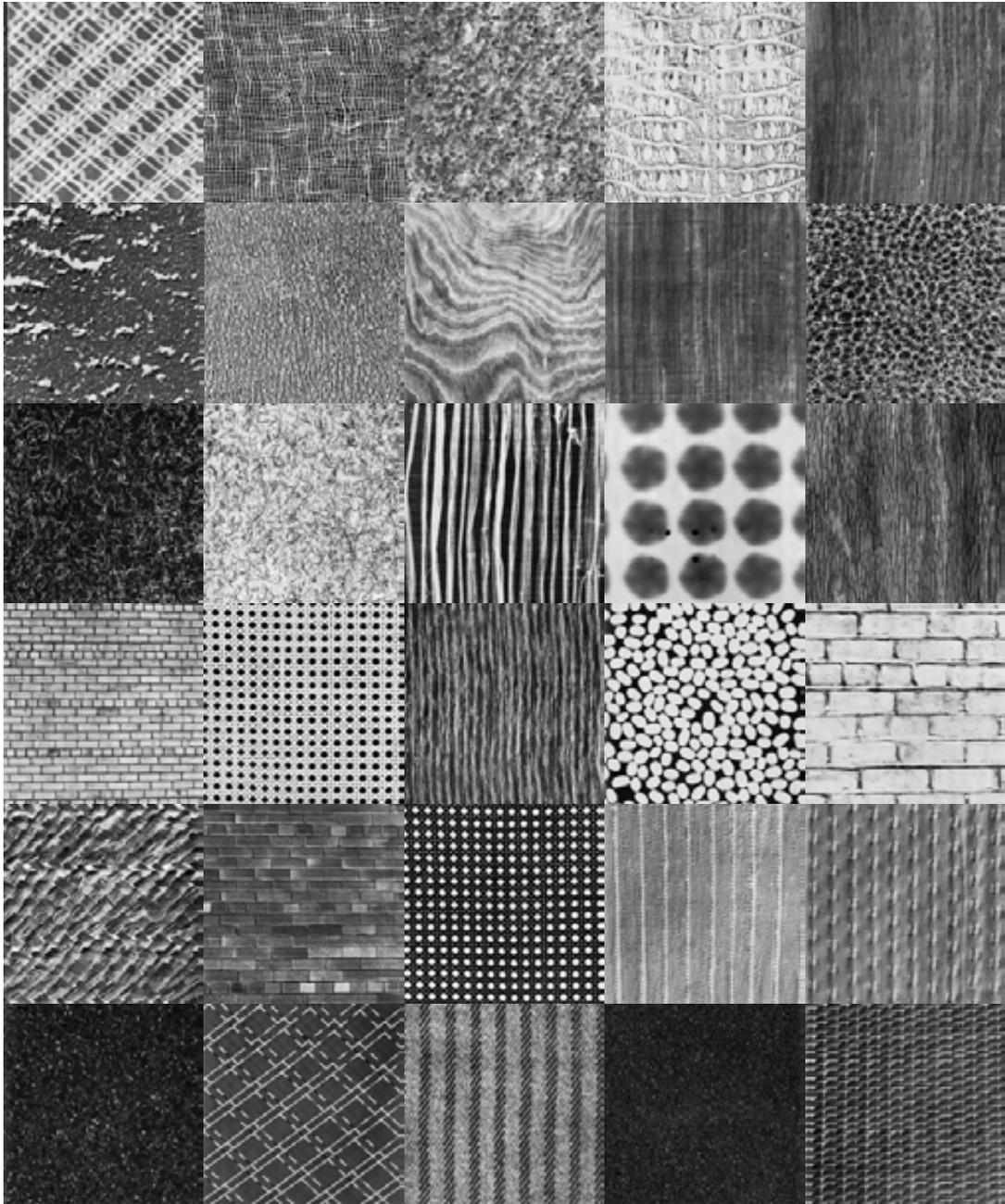


Figura 5.13. Sigüientes 30 imágenes que corresponden a la tercera columna de la tabla 3

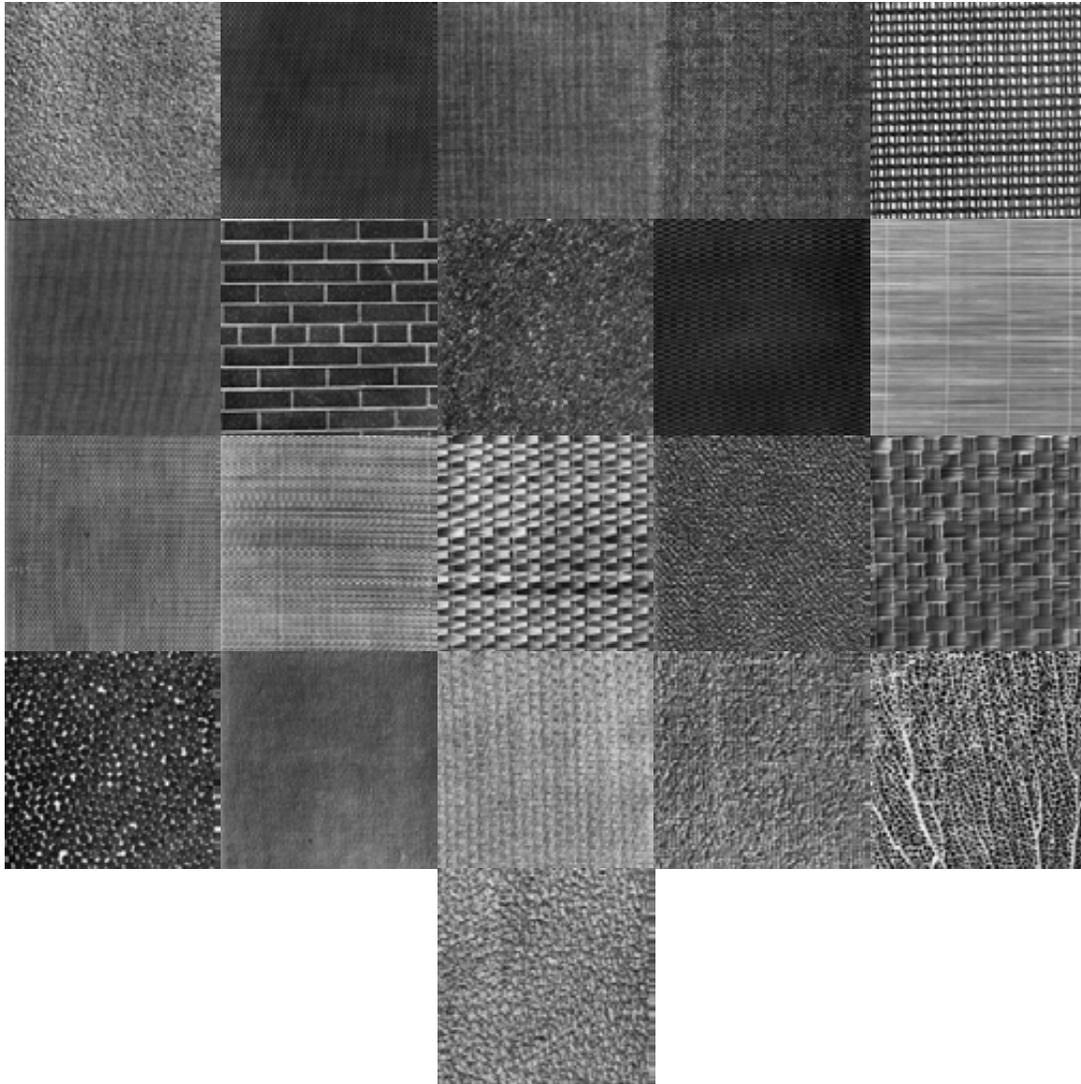


Figura 5.13. Sigüientes 21 imágenes que corresponden a la cuarta columna de la tabla 3

Al observar la figura 5.13, identificamos que las primeras imágenes tienen en común una estructura irregular y al avanzar las estructuras son cada vez más regulares ú homogéneas. El anexo 1 contiene las tablas correspondientes de eficiencia promedio de recuperación de las imágenes originales para cada uno de los experimentos.

Con el propósito de análisis en la recuperación de fracciones de imagen seleccionamos la imagen D43 con la tasa de recuperación menos eficiente según las tablas del anexo 2. La figura 5.14 muestra dicha textura con sus fracciones respectivas al ser dividida en 4x4.

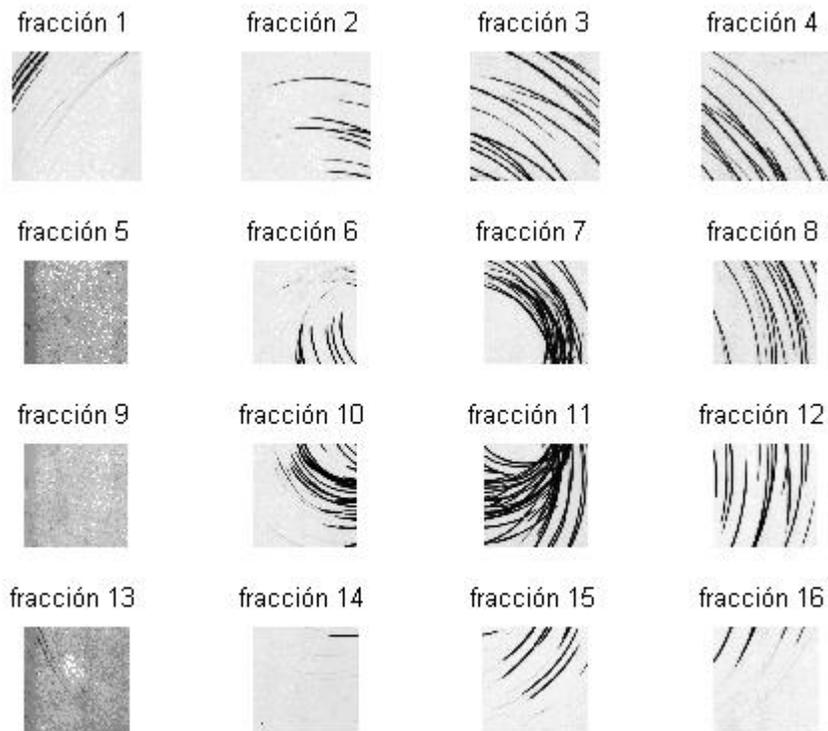


Figura 5.14. D43 de tamaño 640x640. Dividida en 4x4 quedando fracciones de imagen de 160x160.

De la figura 5.14 observamos fracciones de imagen visualmente distintas, es decir, con frecuencia y orientación diferentes. De la figura también notamos que algunas fracciones no comparten ninguna relación con sus fracciones adyacentes teniendo como consecuencia una baja de recuperación.

Por ejemplo, si solicitamos al sistema nos recupere la imagen original dando como entrada la fracción 2 de D43, tendremos como resultado las fracciones de imagen de la figura 5.15.



Figura 5.15. Resultado al dar fracción de imagen 2 de D43 como entrada de referencia al sistema.

De la figura anterior comprobamos, por que, la baja tasa de recuperación para esta imagen en particular. La tabla 5.4 muestra las fracciones de las imágenes originales mostradas en la figura 5.15. Las celdas en color gris corresponden a la imagen original.

Fracción de imagen	Imagen original
2	43
15	45
9	96
10	45
16	38
14	1
16	96
6	1
3	1
10	1
15	1
7	1
10	43
2	1
13	95
15	96

Tabla 5.4. Fracciones de las imágenes originales mostradas en la figura 5.15.

Si ahora damos como entrada la fracción 13 de D43, tendremos las fracciones de la figura 5.16.



Figura 5.16. Resultado cuando la fracción 13 de D43 es dada como referencia al sistema

De esta solicitud tenemos 6 fracciones que corresponden a la original a diferencia de la petición anterior que solo tuvo 2 fracciones. Las fracciones correspondientes están dadas en la tabla 5.5, donde las celdas grises pertenecen a la imagen original.

Este resultado podría explicarse a la poca orientación y baja presencia de frecuencias en la fracción de entrada y fracciones resultantes en este ejercicio en particular.

Fracción de imagen	Imagen original
13	43
9	43
5	43
10	44
14	44
11	44
6	44
14	43
4	91
9	91
16	43
4	44
1	43
16	48
13	48
9	48

Tabla 5.5. Fracciones de las imágenes originales mostradas en la figura 5.16.

Al repetir el experimento anterior con imágenes originales divididas en 3x3 para tener fracciones de mayor dimensión como se muestra en la figura 5.17, la eficiencia de recuperación porcentual mejora según la grafica de la figura 5.11.

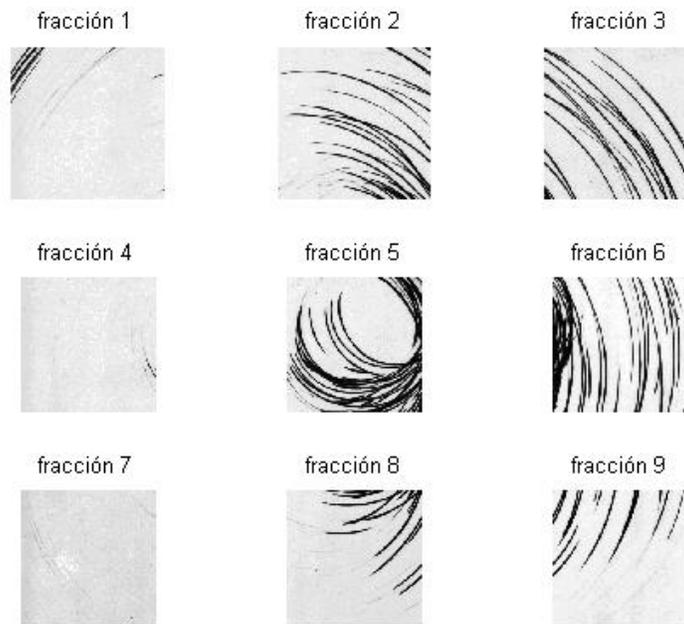


Figura 5.17. D43 -640x640-. Dividida 3x3 quedando fracciones de imagen de 213x213.

Al tener fracciones de mayor tamaño aumenta la probabilidad para que las fracciones adyacentes compartan características similares, lo que justificaría el incremento en la eficiencia de recuperación.

Atendiendo las observaciones anteriores fueron realizadas dos solicitudes extras al sistema para comprobar lo antes dicho.

En el primer experimento la imagen original es dividida en 3x3 quedando fracciones de “213x213”.

La fracción 2 de la figura 5.17 es dada como patrón de entrada.

La figura 5.18 y tabla 5.6 muestran el resultado dado por los algoritmos respectivos, donde notaremos que de esta petición solo tenemos una fracción recuperada de la imagen original y que está identificada como la fracción 5.



Figura 5.18. Fracciones resultantes cuando la fracción 2 de D43 es dada como patrón de entrada. La imagen D43 fue dividida en 3x3.

Imagen original	Fracción de imagen
43	2
45	8
45	7
43	5
44	2
45	4
75	2
1	5
45	2

Tabla 5.6. Fracciones de las imágenes originales mostradas en la figura 5.18.

En la segunda petición con el propósito de comparación se da al sistema una fracción aproximada a la fracción 2 de la figura 5.14 y esta corresponde a la fracción 2 de la figura 5.19. Dicha fracción es obtenida de dividir la imagen original en 3x3 pero con dimensiones de fracción de 256x256, es decir, fracciones traslapadas.

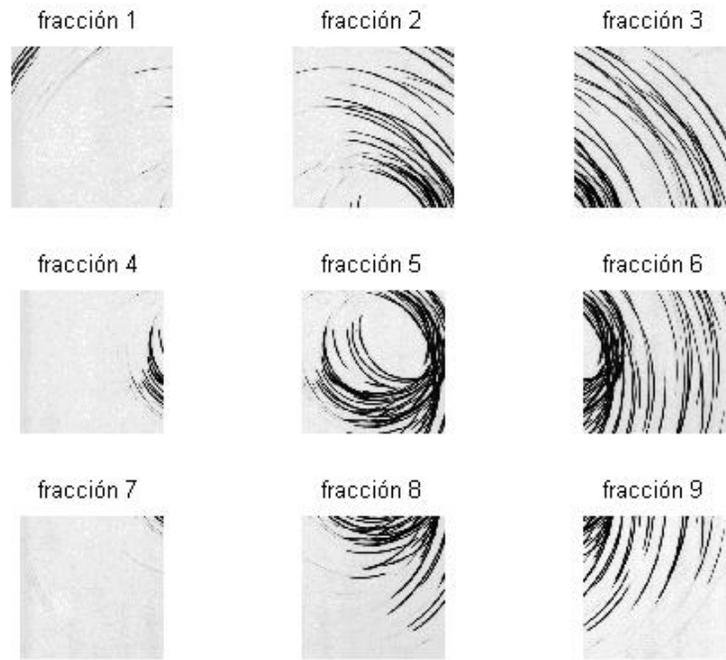


Figura 5.19. D43 de tamaño 640x640. Dividida en 3x3 quedando fracciones de imagen traslapadas de 256x256.

La tabla 5.7 muestra el resultado de la petición realizada, donde podemos resaltar que el número de fracciones que pertenecen a la imagen original fue incrementado en relación al primer experimento.



Figura 5.20. Fracciones de imagen resultante a la solicitud de búsqueda cuando fracción traslapada 2 de D43 es dada como patrón.

Imagen original	Fracción de imagen
43	2
45	7
45	4
45	8
43	8
43	3
45	2
75	1
45	5

Tabla 5.7. Fracciones de las imágenes originales mostradas en la figura 5.20.

Al realizar una comparación visual de las fracciones 2 de las figuras 5.17 y 5.19, notamos la relación que tienen estas con sus adyacentes. La segunda como es de esperar contiene información adicional y de relación que en el resultado final se ve reflejado, es decir, la eficiencia de recuperación porcentual de la figura 5.11 es incrementada de acuerdo a las consideraciones dadas en los experimentos descritos. Las tablas del anexo 2 confirman estas mejoras para cada una de las imágenes.

La figura 5.12 resulta de experimentos similares a los descritos salvo que en estos fue empleada la THM en lugar de la TH. El incremento en la eficiencia de recuperación porcentual en cada uno de los experimentos equivalentes es un poco mas del 1% pero esta se ve reflejada en la carga computacional.

Conclusiones.

La literatura especializada para indexado y búsqueda de textura es tan amplia que por su diversidad esta clasificada en alto y bajo nivel. El trabajo descrito en este documento esta clasificado en bajo nivel, es decir, extracción de características por contenido (color, forma, textura). Para obtener dichas características los algoritmos propuestos implementan transformaciones lineales que permitirán realizar un análisis en escala, frecuencia y orientación de las cuales podemos citar: wavelet discreta, wavelet compleja, wavelet Gabor.

La transformada hermite (TH) y transformada hermite multiescala (THM) fueron las transformaciones utilizadas en este proyecto por su flexibilidad en el análisis multidireccional y multiescala a partir de la representación espacio-escala.

La construcción de vectores característicos fue decidida a través de la reproducción de las experiencias reportadas en los diferentes documentos consultados que están relacionados con el tema, ya que las componentes o índices que los forman pueden ser estadísticos (media y desviación estándar), autocorrelación y correlación cruzada de las sub bandas.

Dentro de la configuración del sistema debemos resaltar la importancia que se debe tener para seleccionar la medida de similitud. En los reportes consultados fueron seleccionados aquellos que presentaban las mejores tasas de discriminación para ser considerados en los algoritmos. En esta selección fueron realizadas pruebas hasta encontrar la distancia Canberra como la mejor.

Con el propósito de comparar la propuesta desarrollada aquí, fueron implementados los algoritmos respectivos considerando las condiciones bajo las cuales fueron sometidos los otros proyectos. En la mayoría de los reportes fue utilizado el álbum fotográfico de Brodatz por lo que este fue empleado. Las texturas que contiene fueron divididas en fracciones de 4x4, 3x3 y 3x3 traslapadas para tener una base de texturas variada.

Como punto de partida iniciamos con fracciones de 160x160 pixeles al dividir la imagen original en 4x4. Tras aplicar la transformada hermite (TH), formación de vectores característicos con autocorrelación y como medida de similitud la distancia Canberra fue obtenida una eficiencia de recuperación promedio del 78.6951 %.

Al termino del anterior experimento fue implementado un algoritmo que permitiera conocer las texturas que para el sistema tuvieran una baja tasa de recuperación. El resultado puede ser visto en la tabla 5.3 y figura 5.13, donde de esta podemos observar que las primeras texturas tienen una estructura poco homogénea pero conforme la eficiencia mejora las imágenes son más regulares. De las primeras vemos que al dividir la imagen original en fracciones éstas en algunos casos no presentan relación alguna con sus fracciones adyacentes y de éstas podemos citar como ejemplo la figura 5.14.

Bajo la anterior observación fue implementado el mismo algoritmo pero con fracciones de imagen mas grande. Ahora la textura original fue dividida en 3x3 para tener fracciones de 213x213 pixeles obteniendo una mejora en la tasa de recuperación del 83.7838 %. Este porcentaje es justificado por que algunas fracciones comparten características con sus adyacentes que en el anterior experimento no presentaban de acuerdo a lo que podemos ver en la figura 5.17.

Para terminar de apuntalar la hipótesis anterior ahora la imagen original fue dividida en 3x3 pero con fracciones 256x256 para tener fracciones traslapadas que compartieran características en común con sus adyacentes. El resultado final fue incrementado al 87.8212 % como era de esperar.

Cada uno de los experimentos mencionados fueron reproducidos en una segunda ocasión salvo que ahora fue utilizada la transformada hermite multiescala. El incremento en la eficiencia de recuperación porcentual en cada uno de los experimentos equivalentes es un poco mas del 1% pero esta se ve reflejada en la carga computacional.

Apéndice A

Transformada Hermite Multiescala, Caso 1-D

De la serie de Taylor para el caso 1D tenemos:

$$G(x, s_{k-1}) = \sum_{n=0}^{\infty} \frac{(s_{k-1} - s_k)^n}{n!} \left[\frac{\partial^n}{\partial s^n} G(x, s) \right]_{s=s_k} \quad (\text{A.1})$$

y dado que la Gaussiana es solución de la ecuación de difusión:

$$\partial_s G(x, s) = \partial_x^2 G(x, s)$$

Al derivar sucesivamente respecto a s se obtiene

$$\frac{\partial^n G(x, s)}{\partial s^n} = G_{2n}(x, s) \quad (\text{A.2})$$

Para $n = 1, 2, \dots$. Sustituyendo (A.2) en (A.1) y restando el primer término de la serie en ambos miembros, llegamos a la ecuación 3.2.

Si factorizamos apoyándonos de la propiedad de cerradura bajo convolución, tenemos

$$(-1)^n G_{2n}(x, s_k) = G_n\left(-x, \frac{s_k}{2}\right) \otimes G_n\left(x, \frac{s_k}{2}\right)$$

Y aplicando la relación

$$\sqrt{(4s)^n} G_n(x, s) = \frac{1}{\sqrt{4s}} G_n\left(\frac{x}{\sqrt{4s}}\right) \quad (\text{A.3})$$

Donde, por definición

$$G_n(x) = G_n\left(x, \frac{1}{4}\right)$$

para $n = 0, 1, 2, \dots$ son las DG en coordenadas naturales. Para llegar finalmente a

$$DoG(x, s_{k-1}, s_k) = \sum_{n=1}^{\infty} \tau^n \frac{1}{\sqrt{2s_k}} G_n^*\left(\frac{-x}{\sqrt{2s_k}}\right) \otimes \frac{1}{\sqrt{2s_k}} G_n^*\left(\frac{x}{\sqrt{2s_k}}\right) \quad (\text{A.4})$$

Donde $G_n^*(x)$ denota la normalización

$$\frac{G_n\left(x, \frac{1}{4}\right)}{\sqrt{2^n n!}}$$

Respecto al producto interno

$$\int_{-\infty}^{\infty} \frac{G_n^*(\xi)G_m^*(\xi)}{G(\xi)} d\xi = \delta_{n,m} \quad (\text{A.5})$$

Donde $\delta_{n,m}$ es la delta de Kronecker.

Sustituyendo (A.4) en la ecuación (3.1), se tiene

$$L(x) = \sum_{k=-\infty}^{\infty} \sum_{n=1}^{\infty} \tau^n \left\{ L(x) \otimes \frac{1}{\sqrt{2s_k}} G_n^*\left(\frac{-x}{\sqrt{2s_k}}\right) \right\} \otimes \frac{1}{\sqrt{2s_k}} G_n^*\left(\frac{x}{\sqrt{2s_k}}\right) \quad (\text{A.6})$$

Pero si definimos

$$L_n^{(k)}(\xi) = \left[L(x) \otimes \frac{1}{\sqrt{2s_k}} G_n^*\left(\frac{-x}{\sqrt{2s_k}}\right) \right]_{x=\sqrt{2s_k}\xi} \quad (\text{A.7})$$

como la fase de análisis de la señal, entonces las definiciones

$$D_n^{(k)}(x) = L_n^{(k)}\left(\frac{x}{\sqrt{2s_k}}\right) \otimes \frac{1}{\sqrt{2s_k}} G_n^*\left(\frac{x}{\sqrt{2s_k}}\right) \quad (\text{A.8})$$

$$L(x) = \sum_{k=-\infty}^{\infty} \sum_{n=1}^{\infty} \tau^n D_n^{(k)}(x) \quad (\text{A.9})$$

corresponden a la fase de síntesis de la señal.

Apéndice B

Transformada Hermite Multiescala Multidireccional, Caso 2-D

La expansión de Taylor para el caso 2D es la siguiente:

$$DoG(x, y, s_{k-1}, s_k) = \sum_{n=1}^{\infty} \frac{(-\tau s_k)^n}{n!} \left[\frac{\partial^n}{\partial s^n} G(x, y, s) \right]_{s=s_k} \quad (B.1)$$

Donde las derivadas las obtenemos con la ecuación de difusión isotrópica en 2D:

$$\partial_s G(x, y, s) = \partial_x^2 G(x, y, s) + \partial_y^2 G(x, y, s), \text{ esto es}$$

$$\frac{\partial^n}{\partial s^n} G(x, y, s) = \sum_{m=0}^n C_n^m G_{2m, 2n-2m}(x, y, s) \quad (B.2)$$

o, alternativamente usando las DDG,

$$\frac{\partial^n}{\partial s^n} G(x, y, s) = c_n \sum_{j=0}^n G_{2n,0}(x, y, s, \theta_j) \quad (B.3)$$

Para las orientaciones $\theta_j = \theta_0 + \frac{j}{(n+1)}$, $j = 0, \dots, n$ y las constantes

$$c_n = \frac{1}{\sum_{j=1}^n \text{sen}^{2n} \left(\frac{j\pi}{n+1} \right)} = \frac{1}{n+1} \sum_{m=0}^n \frac{(C_n^m)^2}{C_{2n}^{2m}} \quad (B.4)$$

para $n = 1, 2, \dots$

Sustituyendo (B.2) en (B.1) tenemos la expresión para la DoG bidimensional:

$$DoG(x, y, s_{k-1}, s_k) = \sum_{n=1}^{\infty} \tau^n \sum_{m=0}^n \frac{1}{2s_k} G_{m, n-m}^* \left(\frac{-x}{\sqrt{2s_k}}, \frac{-y}{\sqrt{2s_k}} \right) \otimes \frac{1}{2s_k} G_{m, n-m}^* \left(\frac{x}{\sqrt{2s_k}}, \frac{y}{\sqrt{2s_k}} \right) \quad (B.5)$$

Para la expresión alternativa (B.3) llegamos:

$$DoG(r, s_{k-1}, s_k) = \sum_{n=1}^{\infty} c_n \tau^n \sum_{j=0}^n \frac{1}{2s_k} G_{n,0}^* \left(\frac{-rR_j}{\sqrt{2s_k}} \right) \otimes \frac{1}{2s_k} G_{n,0}^* \left(\frac{rR_j}{\sqrt{2s_k}} \right) \quad (B.6)$$

Con los vectores $r = (x, y)$, $\rho = (\xi, \eta)$ y la matriz de rotación

$$R_j = \begin{pmatrix} \cos \theta_j & -\operatorname{sen} \theta_j \\ \operatorname{sen} \theta_j & \cos \theta_j \end{pmatrix} \quad (\text{B.7})$$

De la ecuación (B.5) se construye la THM-2D cuyas etapas están dadas por:

Análisis:

$$L_{m,n-m}^{(k)}(\xi, \eta) = \left[L(x, y) \otimes \frac{1}{2s_k} G_{m,n-m}^* \left(\frac{-x}{\sqrt{2s_k}}, \frac{-y}{\sqrt{2s_k}} \right) \right]_{(x,y)=\sqrt{2s_k}(\xi,\eta)} \quad (\text{B.8})$$

Síntesis:

$$D_{m,n-m}^{(k)}(x, y) = L_{m,n-m}^{(k)} \left(\frac{x}{\sqrt{2s_k}}, \frac{y}{\sqrt{2s_k}} \right) \otimes \frac{1}{2s_k} G_{m,n-m}^* \left(\frac{x}{\sqrt{2s_k}}, \frac{y}{\sqrt{2s_k}} \right) \quad (\text{B.9})$$

$$L(x, y) = \sum_{k=-\infty}^{\infty} \sum_{n=1}^{\infty} \sum_{m=0}^n \tau^n D_{m,n-m}^{(k)}(x, y) \quad (\text{B.10})$$

Para la ecuación (B.6) tenemos la THMM definida como:

Análisis:

$$L_n^{(k,j)}(\rho) = \left[L(r) \otimes \frac{1}{2s_k} G_{n,0}^* \left(\frac{-rR_j}{\sqrt{2s_k}} \right) \right]_{r=\rho R_j^T \sqrt{2s_k}} \quad (\text{B.11})$$

Síntesis:

$$D_n^{(k,j)}(r) = L_n^{(k,j)} \left(\frac{rR_j}{\sqrt{2s_k}} \right) \otimes \frac{1}{2s_k} G_{n,0}^* \left(\frac{rR_j}{\sqrt{2s_k}} \right) \quad (\text{B.12})$$

$$L(r) = \sum_{k=-\infty}^{\infty} \sum_{n=1}^{\infty} \sum_{j=0}^n c_n \tau^n D_n^{(k,j)}(r) \quad (\text{B.13})$$

Apéndice D

Rotación de las DG

En el dominio de Fourier la DG es

$$g_{n-m,m}(w_x, w_y) = (jw_x)^{n-m} (jw_y)^m g(w_x, w_y)$$

donde $g(w_x, w_y)$ es la transformada de Fourier de $G(x, y)$, y sean las constantes $c = \cos \theta$ y $s = \text{sen} \theta$.

La versión rotada de las DG está dada por:

$$g_{n-m,m}(w_x, w_y, \theta) = (cjw_x + sjw_y)^{n-m} (-sjw_x + cjw_y)^m g(w_x, w_y) = \sum_{k=0}^{n-m} \sum_{i=0}^m (-1)^i C_m^i C_{n-m}^k c^{m-i+k} s^{n-m+i-k} g_{i+k, n-i-k}(w_x, w_y)$$

haciendo cambio de índices $k + i$ por k y $m - i$ por i , tenemos

$$g_{n-m,m}(w_x, w_y, \theta) = \sum_{k=0}^n g_{k, n-k}(w_x, w_y) c^{-k} s^k \sum_{i=0}^m (-1)^{m-i} C_m^i C_{n-m}^{k+i-m} c^{2k+2i-m} s^{n+m-2k-2i}$$

$$g_{n-m,m}(w_x, w_y, \theta) = \sum_{k=0}^n g_{k, n-k}(w_x, w_y) c^{-k} s^k \Delta^m \{ C_{n-m}^{k-m} c^{2k-m} s^{n+m-2k} \} = \sum_{k=0}^n A_{m,k}^{(n,\theta)} g_{k, n-k}(w_x, w_y)$$

o bien, en el dominio espacial,

$$G_{n-m,m}(x, y, \theta) = \sum_{k=0}^n A_{m,k}^{(n,\theta)} G_{k, n-k}(x, y)$$

donde

$$A_{m,k}^{(n,\theta)} = c^{-k} s^k \Delta^m \{ C_{n-m}^{k-m} c^{2k-m} s^{n+m-2k} \}$$

Son funciones de ángulo que definen una familia de funciones binomiales generalizadas en la variable discreta k .

Anexo 1

Tasa de recuperación porcentual de las imágenes originales.

4x4, 160x160						4x4, 160x160, multiescala									
imagen	porciento	imagen	porciento	imagen	porciento	imagen	porciento	imagen	porciento	imagen	porciento	imagen	porciento		
45	21.8750	52	64.4530	46	87.8910	4	100.0000	43	22.2660	67	66.7970	9	88.6720	26	100.0000
43	23.0470	67	65.6250	103	87.8910	6	100.0000	45	26.9530	89	67.9690	28	88.6720	29	100.0000
44	28.5160	105	66.4060	28	89.4530	16	100.0000	91	31.2500	112	67.9690	19	89.0630	34	100.0000
90	31.6410	89	66.7970	10	90.2340	17	100.0000	44	32.0310	108	68.3590	111	92.1880	47	100.0000
31	37.8910	97	66.7970	79	90.2340	20	100.0000	58	34.7660	63	69.9220	74	92.5780	49	100.0000
42	41.4060	106	68.7500	86	90.2340	21	100.0000	90	35.1560	52	72.6560	79	92.5780	53	100.0000
58	41.7970	60	69.1410	24	90.6250	26	100.0000	42	42.5780	36	73.8280	51	93.7500	55	100.0000
91	42.1880	108	69.1410	71	90.6250	29	100.0000	99	43.3590	12	74.2190	24	94.9220	56	100.0000
7	46.0940	19	70.3130	78	90.6250	34	100.0000	31	44.1410	25	74.6090	110	94.9220	57	100.0000
38	46.0940	12	72.6560	111	91.4060	49	100.0000	73	46.0940	107	75.3910	93	95.3130	65	100.0000
73	48.4380	37	75.0000	109	91.7970	53	100.0000	13	47.2660	104	76.5630	4	96.0940	66	100.0000
88	49.2190	72	75.3910	110	92.5780	55	100.0000	59	48.0470	105	76.9530	1	96.4840	76	100.0000
99	49.2190	15	75.7810	51	92.9690	56	100.0000	69	49.6090	8	77.3440	32	96.4840	77	100.0000
27	50.3910	112	77.3440	48	94.1410	57	100.0000	88	49.6090	5	77.7340	33	96.4840	82	100.0000
13	50.7810	50	79.2970	68	94.9220	64	100.0000	62	51.1720	40	77.7340	94	97.2660	83	100.0000
69	50.7810	8	79.6880	95	94.9220	66	100.0000	100	52.7340	103	78.5160	96	97.2660	84	100.0000
59	51.1720	22	80.0780	101	95.7030	77	100.0000	7	53.1250	50	79.6880	68	98.0470	85	100.0000
54	52.7340	1	80.4690	76	96.0940	82	100.0000	38	53.1250	72	79.6880	101	98.0470	87	100.0000
62	52.7340	104	81.2500	75	96.4840	84	100.0000	70	54.2970	80	80.4690	18	98.4380	92	100.0000
61	55.4690	107	81.6410	96	96.4840	87	100.0000	27	56.2500	15	81.2500	35	98.4380	95	100.0000
98	57.0310	40	82.8130	18	96.8750	92	100.0000	39	57.0310	71	82.8130	48	99.2190	102	100.0000
100	60.1560	80	82.8130	94	97.6560			54	57.0310	37	84.3750	75	99.2190		
23	61.3280	81	82.8130	102	97.6560			97	59.7660	10	85.1560	16	99.6090		
2	62.5000	35	83.2030	85	98.0470			61	60.9380	3	85.9380	46	99.6090		
36	62.5000	41	83.2030	83	98.4380			23	61.7190	81	85.9380	64	99.6090		
63	62.5000	9	84.7660	33	98.8280			60	64.4530	109	86.7190	6	100.0000		
70	63.2810	25	85.9380	47	98.8280			106	64.8440	41	87.1090	11	100.0000		
30	64.0630	74	85.9380	11	99.2190			30	66.0160	22	87.5000	17	100.0000		
5	64.4530	93	85.9380	32	99.6090			98	66.0160	78	87.5000	20	100.0000		
39	64.4530	3	87.8910	65	99.6090			2	66.4060	86	87.5000	21	100.0000		

3x3, 213x213

3x3, 213x213, multiescala

imagen	porciento														
43	22.2220	67	71.6050	51	95.0620	47	100.0000	43	24.6910	108	74.0740	78	96.2960	64	100.0000
45	28.3950	105	71.6050	71	95.0620	49	100.0000	45	33.3330	36	75.3090	9	96.2960	65	100.0000
44	35.8020	36	72.8400	24	96.2960	53	100.0000	58	37.0370	112	75.3090	109	96.2960	66	100.0000
91	44.4440	70	74.0740	102	96.2960	55	100.0000	44	38.2720	88	76.5430	111	96.2960	68	100.0000
90	45.6790	23	75.3090	3	96.2960	56	100.0000	91	39.5060	52	80.2470	48	97.5310	74	100.0000
38	50.6170	52	75.3090	48	96.2960	57	100.0000	59	51.8520	103	80.2470	86	97.5310	75	100.0000
59	50.6170	1	77.7780	93	96.2960	64	100.0000	73	51.8520	105	81.4810	24	98.7650	76	100.0000
73	51.8520	22	80.2470	40	97.5310	65	100.0000	90	53.0860	15	82.7160	51	98.7650	77	100.0000
54	53.0860	108	81.4810	79	97.5310	66	100.0000	69	55.5560	5	83.9510	79	98.7650	82	100.0000
58	53.0860	112	81.4810	101	97.5310	68	100.0000	54	56.7900	60	83.9510	1	100.0000	83	100.0000
69	53.0860	12	81.4810	109	97.5310	75	100.0000	38	58.0250	63	83.9510	6	100.0000	84	100.0000
7	54.3210	50	81.4810	111	97.5310	77	100.0000	62	58.0250	89	83.9510	11	100.0000	85	100.0000
98	54.3210	72	83.9510	9	98.7650	82	100.0000	99	59.2590	80	85.1850	16	100.0000	87	100.0000
99	55.5560	37	85.1850	11	98.7650	83	100.0000	97	59.2590	25	85.1850	17	100.0000	92	100.0000
30	58.0250	35	85.1850	74	98.7650	84	100.0000	7	60.4940	22	86.4200	18	100.0000	93	100.0000
31	58.0250	63	85.1850	76	98.7650	85	100.0000	70	60.4940	71	87.6540	20	100.0000	94	100.0000
62	58.0250	89	87.6540	86	98.7650	87	100.0000	98	62.9630	12	88.8890	21	100.0000	95	100.0000
42	61.7280	104	87.6540	110	98.7650	92	100.0000	42	62.9630	50	88.8890	26	100.0000	96	100.0000
13	64.1980	8	88.8890	4	100.0000	94	100.0000	61	62.9630	3	90.1230	29	100.0000	101	100.0000
27	65.4320	80	88.8890	6	100.0000	95	100.0000	39	65.4320	8	90.1230	32	100.0000	102	100.0000
100	66.6670	81	88.8890	16	100.0000	96	100.0000	13	66.6670	72	90.1230	33	100.0000	110	100.0000
5	67.9010	46	90.1230	17	100.0000			106	66.6670	37	91.3580	34	100.0000		
2	67.9010	60	91.3580	18	100.0000			2	67.9010	107	91.3580	35	100.0000		
39	67.9010	25	91.3580	20	100.0000			23	69.1360	81	92.5930	46	100.0000		
61	69.1360	41	91.3580	21	100.0000			104	70.3700	40	92.5930	47	100.0000		
88	70.3700	78	92.5930	26	100.0000			100	71.6050	28	93.8270	49	100.0000		
97	70.3700	28	92.5930	29	100.0000			27	72.8400	10	95.0620	53	100.0000		
106	70.3700	107	92.5930	32	100.0000			30	72.8400	4	96.2960	55	100.0000		
15	71.6050	103	93.8270	33	100.0000			31	72.8400	19	96.2960	56	100.0000		
19	71.6050	10	93.8270	34	100.0000			67	74.0740	41	96.2960	57	100.0000		

3x3, 256x256, traslapada

imagen	porciento														
43	28.3950	97	80.2470	72	98.7650	56	100.0000	43	37.0370	31	83.9510	41	98.7650	64	100.0000
45	44.4440	67	80.2470	74	98.7650	57	100.0000	45	40.7410	108	83.9510	72	98.7650	65	100.0000
44	46.9140	23	81.4810	78	98.7650	64	100.0000	91	46.9140	15	85.1850	74	98.7650	66	100.0000
38	55.5560	88	81.4810	93	98.7650	65	100.0000	44	48.1480	88	86.4200	110	98.7650	68	100.0000
90	55.5560	112	81.4810	107	98.7650	66	100.0000	58	49.3830	5	87.6540	1	100.0000	75	100.0000
59	55.5560	30	82.7160	109	98.7650	68	100.0000	69	50.6170	105	87.6540	6	100.0000	76	100.0000
73	55.5560	22	83.9510	111	98.7650	75	100.0000	59	56.7900	36	90.1230	10	100.0000	77	100.0000
54	58.0250	52	85.1850	4	100.0000	76	100.0000	73	58.0250	50	90.1230	11	100.0000	79	100.0000
69	58.0250	36	86.4200	6	100.0000	77	100.0000	27	62.9630	22	90.1230	16	100.0000	82	100.0000
91	58.0250	37	87.6540	9	100.0000	82	100.0000	54	62.9630	63	90.1230	17	100.0000	83	100.0000
7	59.2590	63	87.6540	10	100.0000	83	100.0000	90	64.1980	60	91.3580	18	100.0000	84	100.0000
58	60.4940	108	87.6540	11	100.0000	84	100.0000	70	66.6670	30	91.3580	20	100.0000	85	100.0000
98	61.7280	50	88.8890	16	100.0000	85	100.0000	99	66.6670	52	92.5930	21	100.0000	86	100.0000
27	64.1980	8	91.3580	17	100.0000	86	100.0000	62	69.1360	71	92.5930	24	100.0000	87	100.0000
31	64.1980	81	93.8270	18	100.0000	87	100.0000	98	69.1360	89	92.5930	25	100.0000	92	100.0000
13	67.9010	60	93.8270	20	100.0000	92	100.0000	7	70.3700	3	93.8270	26	100.0000	93	100.0000
2	69.1360	89	95.0620	21	100.0000	94	100.0000	103	72.8400	8	93.8270	29	100.0000	94	100.0000
61	69.1360	101	95.0620	25	100.0000	95	100.0000	38	74.0740	80	95.0620	32	100.0000	95	100.0000
99	69.1360	104	95.0620	26	100.0000	96	100.0000	97	74.0740	81	95.0620	33	100.0000	96	100.0000
62	70.3700	103	95.0620	29	100.0000	102	100.0000	2	75.3090	78	96.2960	34	100.0000	101	100.0000
106	71.6050	35	96.2960	32	100.0000	110	100.0000	39	76.5430	107	96.2960	35	100.0000	102	100.0000
42	72.8400	80	96.2960	33	100.0000			42	76.5430	19	97.5310	46	100.0000		
70	74.0740	41	96.2960	34	100.0000			104	77.7780	28	97.5310	47	100.0000		
5	75.3090	12	97.5310	40	100.0000			13	77.7780	37	97.5310	48	100.0000		
19	75.3090	24	97.5310	47	100.0000			106	77.7780	109	97.5310	49	100.0000		
15	76.5430	46	97.5310	48	100.0000			112	79.0120	111	97.5310	51	100.0000		
100	76.5430	79	97.5310	49	100.0000			100	80.2470	4	98.7650	53	100.0000		
105	76.5430	3	98.7650	51	100.0000			61	80.2470	9	98.7650	55	100.0000		
39	77.7780	28	98.7650	53	100.0000			67	81.4810	12	98.7650	56	100.0000		
1	79.0120	71	98.7650	55	100.0000			23	83.9510	40	98.7650	57	100.0000		

Bibliografia

- [1] Young, R. (1987). The Gaussian derivative theory of spatial vision: I. Retinal mechanisms. *Spatial Vision* 2, 273-293.
- [2] Martens, J.-B. (1990). The Hermite Transform - Theory. *IEEE Transactions on Acoustics, Speech and Signal Processing ASSP-38*, 1595-1606.
- [3] J. Zhang and T. Tan, "Brief review of invariant texture analysis methods", *Pattern Recognition*, vol. 35, pp. 735-747, 2002.
- [4] P. Wu, B.S. Manjunath, S. Newsam, and H.D. Shin, "A texture descriptor for browsing and similarity retrieval", *Signal Processing: Image Communication*, vol. 16, no. 1,2, pp. 33-43, 2000.
- [5] P.W. Huang and S.K. Dai, "Image retrieval by texture similarity", *Pattern Recognition*, vol. 36, pp. 665-679, 2003.
- [6] A.C. Bovik, M. Clark, and W.S. Geisler, "Multichannel texture analysis using localized spatial filters", *IEEE Trans. Pattern Analysis Mach. Intell.*, vol. 12, pp. 55-73, 1990.
- [7] T. Randen and J.H. Husøy, "Filtering for texture classification: A comparative study", *IEEE Trans. Pattern Analysis Mach. Intell.*, vol. 21, pp. 291-310, 1999.
- [8] C.C. Chen and C.C. Chen, "Filtering methods for texture discrimination", *Pattern Recognition Letters*, vol. 20, pp. 783-790, 1999.
- [9] P. Kruizinga and N. Petkov, "Nonlinear operator for oriented texture", *IEEE Trans. Image Processing*, vol. 8, pp. 1395-1407, 1999.
- [10] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by image and video content: The QBIC system", *IEEE Computer*, vol. 28, no. 9, pp. 23-32, 1995.
- [11] Tamura, H. and Yokoya, Y. "Image database Systems: A survey..Pattern Recognition", vol. 17, n. 1, pp. 29-43, 1984.
- [12] Chang, S. K. and Hu, A., "Image Information Systems: Where Do We Go From Here?..", *IEEE Transactions on Knowledge and Data Engineering*, vol. 5, n. 4, pp. 431-442, 1992.
- [13] Veltkamp, R.C. and Tanase, M., "A survey in content-based image retrieval systems"., in Marques, O and Furht, B (eds). *Content-Based Image and Video Retrieval*, pp. 47-101, 2002. Boston. Kluwer Academic Publishers.

[14] Rui, Y., Huang, T.S. and Chang, S.F., "Image retrieval: current techniques, promising directions and open issues"., *Journal of Visual Communication and Image Representation*, vol. 10, n. 1, pp. 39-62, 1999.

[15] Johansson, B., "A survey on: Contents Based Search in Image Databases", 2000 Technical report Linköping University [Online]. Available: <http://www.isy.liu.se/cvl/Projects/VISIT-bjojo/>

[16] Swanberg, D., Weymouth, T. and Jain, R., "Content-based modeling in multimedia information systems", *Proc. of 1st International Conference on Information and Knowledge Management*, Baltimore, November, 1992.

[17] Arndt T., "A Survey of Recent Research in Image Database Management", *Proceedings of the 1990 IEEE Workshop on Visual Languages*. October 4-6, pp. 92-97, 1990.

[18] Gong, Y., "Intelligent image databases". *Towards Advanced Image Retrieval*. Kluwer Academic Publishers, 1998.

[19] Hermes, Th., Klauck, Ch., Kreyb, J. and Zhang, J., "Image Retrieval for information systems". *Storage and Retrieval for Image and Video Databases III.*, Wayne Niblack, Ramesh C. Jain, Editors, *Proc. SPIE 2420*, pp. 394-405, 1995.

[20] Niblack, W., Barber, R., Equitz, W., Flickner, M., Glasman, E., Petrovic, D. and Yanker, P., "The QBIC project: Querying images by content using color, texture and shape", Wayne Niblack editor, *Storage and Retrieval for Image and Video Databases*, pp. 173-187, 1993.

[21] Bertino, E., Ooi, B.Ch., Sacks-Davis, R., Zobel, J. and Catania, B., "Indexing Techniques for advanced database systems", Kluwer Academic Publishers, 1997.

[22] Liang, K.Ch. and Kuo, C.C. J., "Implementation and Performance Evaluation of A Progressive Image Retrieval System", *Storage and Retrieval for Image and Video Databases VI.*, Ishwar K. Sethi, Ramesh C. Jain, Editors, *Proc. SPIE 3312*, pp. 371-382, 1998.

[23] Dyson, M. C., "How do you describe a symbol? the problems involved in retrieving symbols from a database", *Information Services and Use*, n° 76, pp. 12-65, 1992.

[24] Mittendorfer, M. and Winiwarter, W., "Exploiting syntactic analysis of queries for information retrieval", *Data & Knowledge Engineering*, n. 42, pp. 315-325, 2002.

[25] Qiu, C., "Indexing chromatic and achromatic patterns for content-based colour image retrieval", *Pattern Recognition*, n. 35, pp. 1675-1686, 2002.

[26] Korfhage, R. R., "Information Storage and Retrieval", Wiley Computer Publishing, 1997.

- [27] Chang, S.K., Shi, Q.Y., and Yan, C. W., “Iconic indexing by 2D string”, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol 9, n. 3, pp. 413-428, 1987.
- [28] Sciascio, E., Donini, F.M. and Mongiello, M., “Spatial layout representation for query-by-sketch content-based image retrieval”, Pattern Recognition Letters, n. 23, pp. 1599-1612, 2002.
- [29] Nascimento, M., Sridhar, V. and Li, X., “Effective and efficient region based image retrieval”, Journal of Visual Languages and Computing, n. 14, pp. 151-179, 2003.
- [30] Tao, Y. and Grosky, W.I., “Delaunay triangulation for image object indexing: a novel method for shape representation”, Storage and Retrieval for Image and Video Databases VII, Minerva M. Yeung, Boon-Lock Yeo, Charles A. Bourman, Editors, Proceedings of SPIE, vol. 3656, pp. 631-642, 1999.
- [31] Sean, X. and Huang, T. S., “Edge-based structural features for content-based image retrieval”, Pattern Recognition Letters, n. 22, pp. 457-468, 2001.
- [32] Kato, T., Kurita, T. and Shimogaki, H., “Intelligent visual interaction with image database systems . toward the multimedia personal interface”, Journal of Information Processing (Japan), vol 14, n° 2, pp. 134-143, 1991.
- [33] Wilkin, D.J., Debure, K.R. and Roberts, Z. W., “Query by Sketch in DARWIN . Digital Analysis to Recognize Whale Images on Network”, Storage and Retrieval for Image and Video Databases VII, Minerva M. Yeung, Boon-Lock Yeo, Charles A. Bourman, Editors, Proceedings of SPIE, vol. 3656, pp. 41-48, 1999.
- [34] Ricard, R.W. and Minka, T.P., “Vision texture for annotation”, Multimedia Systems, Springer Verlag, n. 3, 1995.
- [35] Mikunoki, M., Minoh, M. and Ikeda, K., “A retrieval method of outdoor scenes using object sketch and an automatic index generation method”, Journal of IEICE of Japan, D-II, J79-DII, n. 6, 1996.
- [36] Nick Kingsbury, Serkan H., Sanjit K., “Texture classification using dual-tree complex wavelet transform”, Image Processing and Applications, Conference Publication No. 465, 1999.
- [37] Alan Conrad Bovik, Marianna Clark, Wilson S. Geisler, “Multichannel Texture Analysis Using Localized Spatial Filters”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, No I , January 1990.
- [38] Carlos Joel Rivero Moreno, “Contribution à la caractérisation des images par transformée polynomiale : Application à l’indexation des images et des videos”, L’Institut National des Sciences Appliquées de Lyon, 2005.

- [39] José Luís Silván Cárdenas, “Compresión de Imágenes Basada en Modelos Gaussianos de Percepción Visual”, UNAM, 2002.
- [40] M. Bastiaans, “Gabor's signal expansion and degrees of freedom of a signal”, *Optic Acta* 29 (1982), 1223{1229.
- [41] M. Porat and Zeevi, “The generalized gabor scheme of image representation in biological and machine vision”, *IEEE Transactions on Pattern Analysis Machine Intelligence* 10 (1988), 452{467.
- [42] A. B. Watson, “The cortex transform: Rapid computation of simulated neural Images”, *Computer Vision, Graphics and Image Processing* 39 (1987), 311{327.
- [43] R. E. Frye and R. S. Ledley, “Derivative of gaussian functions as receptive field models for disparity sensitive neurons of the visual cortex”, *Biomedical Engineering Conference (1996), 270-273, Proceedings of the 1996 Fifteenth Southern , 1996.*
- [44] J. A. Bloom and T. R. Reed, “A gaussian derivative-based transform”, *IEEE Transactions on Image Processing* 5 (1996), no. 3, 551-553.
- [45] R. A. Haddad and A. N. Akansu, “A new orthogonal transform for signal coding”, *IEEE Transactions on Acoustics, Speech and Signal Processing* 36 (1988), no. 9, 1404-1411.
- [46] J. B. Martens, “The hermite transform-applications”, *IEEE Transactions on Acoustics, Speech and Signal Processing* 38 (1990), no. 9, 1607-1618.
- [47] A. Witkin, “Scale-space filtering: A new approach to multiscale description”, *Image Understanding* 3 (1984), 79-95.
- [48] H. R. Wilson and J. R. Bergen, “A four mechanism model for spatial vision”, *Vision Research* 19 (1979), 19-32.
- [49] W. R. Hendee and P. N. T. Wells, “The perception of visual information”, Springer-Verlag New York Inc., NY, 1993.
- [50] “Simulation of human retinal function with the gaussian derivative model”, *Proceedings IEEE CCVP (Miami, FL.)*, 1986, pp. 564-569.
- [51] M. Basu and L. M. Kennedy, “An experiment with gaussian derivatives for image enhancement”, *Systems, Man and Cybernetics* 4 (1995), 3778-3783.
- [52] M. Shizawa and T. Iso, “Direct representation and detecting of multi-scale, multi-orientation fields using local differentiation filters”, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (1993), 508-514, Proceedings CVPR '93, 1993.*

- [53] “Local orientation analysis in images by means of the hermite transform”, IEEE Transactions on Image Processing 6 (1997), no. 8, 1103-1116.
- [54] W. T. Freeman and E. H. Adelson, “The design and use of steerable filters”, IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (1991), no. 9, 891-906.
- [55] J. J. Koenderink and A. J. Van Doorn, “Generic neighborhood operators”, IEEE Transactions on Pattern Analysis and Machine Intelligence 14 (1992), no. 6, 597-605.
- [56] J. L. Silván Cárdenas, M. B. Escalante Ramírez, “Radon-Hermite analysis applied to image coding”, Proceedings of the IEEE International Symposium on Industrial Electronics (Slovenia), vol. 3, ISIE, 1999, pp. 1204-1207.
- [57] B. Escalante Ramírez and J. L. Silván Cárdenas, “Optical-flow estimation by means of local projection analysis with the Radón-Hermite transform”, Mathematical Modeling, Bayesian Estimation, and Inverse Problems (Denver, CO) (F. Petreux, A. Mohammad-Djafari, and E. Dougherty, eds.), vol. 3816, SPIE, 1999, pp. 121-129.
- [58] Penélope López Quiroz, “Clasificación de imágenes SAR por medio de la Transformada Orientada de Hermite y Campos Aleatorios de Harkov”, UNAM, 2004.
- [59] Juan Carlos Muñoz Torres, “Fusión de imágenes médicas a través de la transformada de hermite”, UNAM, 2005.
- [60] B. Escalante Ramírez and J. B. Martens, “Noise reduction in computerized tomography images by means of polynomial transforms”, Visual Communication and Image Representation 3 (1992), 272-285.
- [61] D. Zion, S. Wang, and J. Vaillancourt, “Depth from defocus using the hermite transform”, International Conference on Image Processing (1998), 958-962, Proceedings ICIP98, vol. 2.
- [62] J. L. Silván Cárdenas and B. Escalante Ramírez, “Application of the polynomial transforms to motion estimation”, Mathematical Modeling and Estimation Techniques in Computer Vision (San Diego, CA) (F. Petreux, J. L. Davison, and E. Dougherty, eds.), vol. 3457, SPIE, 1998, pp. 216-227.
- [63] Antani, S., Kasturi R. and Jain R. (2002) .A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video., Pattern Recognition, n. 35, pp. 945-965.
- [64] Vasconcelos, N. (2002), .Decisiontheoretic Image Retrieval with Embedded Multi-resolution Mixtures., Technical Report Cambridge Research Laboratory, CRL 2002/04.
- [65] Han, J. W. and Guo, L. (2003), .A shape-based image retrieval method using salient edges., Signal Processing: Image Communication, n. 18, pp. 141-156.

[66] Chang, W., Murthy, D., Zhang, A. and Syeda-Mahmood, T. (1998), .Global Integration of Visual Databases., IEEE International Conference on Data Engineering.

[67] Javier Portilla and Eero P. Simoncelli, A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients, International Journal of Computer Vision

[68] Humberto Sossa, Roberto A. Vázquez, Reconocimiento y localización de instrumental médico usando análisis automatizado de imágenes, Revista mexicana de ingeniería biomédica, Vol. XXVI, Núm. 2 Septiembre 2005 pp 75 - 85

[69] Vytautas Perlibakas, Distance measures for PCA-based face recognition