

01170



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MEXICO

PROGRAMA DE POSGRADO EN INGENIERIA
CAMPO DE CONOCIMIENTO: INGENIERIA ELECTRICA

"CLASIFICACION DE IMAGENES SAR POR MEDIO DE LA
TRANSFORMADA ORIENTADA DE HERMITE Y CAMPOS
ALEATORIOS DE MARKOV"

T E S I S

QUE PARA OBTENER EL TITULO DE:
MAESTRA EN INGENIERIA (ELECTRICA)

P R E S E N T A :

PENELOPE LOPEZ QUIROZ

DIRECTOR DE TESIS:
DR. BORIS ESCALANTE RAMIREZ

ENERO DE 2004



Universidad Nacional
Autónoma de México



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.

ESTA TESIS NO SALE
DE LA BIBLIOTECA

Índice general

1. Introducción	5
1.1. Definición del problema	7
1.2. Objetivo	7
1.3. Hipótesis	7
1.4. Método	8
2. Transformada de Hermite o Hermitiana, [4]	9
2.1. Descripción	9
2.2. Introducción	9
2.3. Transformada polinomial 1D	10
2.4. Transformada de Hermite 1D	13
2.5. Transformada Polinomial 2D	14
2.5.1. Representación de imágenes con coeficientes polino- miales unidimensionales	15
2.6. Transformada de Hermite 2D	17
2.7. Transformadas Polinomiales Discretas	18
2.8. Transformada discreta de Hermite 2D	19
2.9. Aplicaciones de la Transformada de Hermite	21
2.10. Caso 1-D: Transformada de Hermite Multiescala, [6]	22
2.10.1. Descomposición multiescala	22
2.10.2. Análisis estructural	23
2.10.3. Definición de la THM	24
2.10.4. Acotamiento del rango de escalas	25
2.10.5. Esquema piramidal continuo	25
2.10.6. Predicción entre niveles	26
2.11. Caso 2-D: Transformada de Hermite multiescala-multidireccional	27
2.11.1. Descomposición multiescala de señales 2-D	27
2.11.2. Análisis estructural de señales 2-D	28
2.11.3. Definición de la THMM	29

Autorizo a la Dirección General de Bibliotecas de la UNAM a difundir en formato electrónico e impreso el contenido de mi trabajo recepcional.

NOMBRE: Penélope López

Quiroz

FECHA: 20 Enero 2004

FIRMA: 

2.11.4. Filtrado direccional	29
2.12. Caso discreto: THDM y THDMM	30
2.12.1. Espacio-escala discreto	30
2.12.2. Derivadas discretas	31
2.12.3. Definición de la THDM	31
2.12.4. Definición de la THDMM	32
2.13. Clasificación de estructuras en el dominio de la THDR	33
2.13.1. Descomposición de la imagen	33
2.13.2. Rotación local	34
2.13.3. Clasificación	35
3. Campo de vectores de energía y clasificación de estructuras de una imagen	38
3.1. Descripción	38
3.2. Campo de vectores de energía, [5]	38
3.3. Patrones 0D, 1D y 2D, [6]	48
4. Campos Aleatorios de Markov, MRF, [8]	57
4.1. Descripción	57
4.2. Introducción	57
4.3. Etiquetado visual	58
4.4. MRF y Distribuciones de Gibbs	61
4.5. MRF	63
4.6. Campos Aleatorios de Gibbs (GRF, Gibbs Random Fields)	64
4.7. Equivalencia Markov-Gibbs	66
4.8. Modelos de Markov útiles	67
4.9. Vision basada en optimización	68
4.10. Etiquetado de MRF's por medio de Bayes	68
4.11. Estimación de Bayes	69
4.12. Etiquetado MAP-MRF	70
4.13. Resumen de la aproximación MAP-MRF	71
4.14. Modelos de observación	71
4.15. Restauración y reconstrucción de imágenes	73
4.16. Obtención de la energía posterior	74
4.17. Minimización de la energía	74
4.18. Algoritmo ICM	75
5. Regularización del campo de vectores de energía mediante Campos Aleatorios de Markov	77
5.1. Descripción	77

5.2. Introducción	77
5.3. Regularización	78
6. Conclusiones	92

Resumen

A partir de la posesión de dos imágenes SAR-PRI previamente restauradas se las clasificó en patrones 0D (estructuras sin una dirección definida), 1D (estructuras con una sola dirección definida), 2D (estructuras con más de una dirección definida) y se obtuvo el campo de vectores de energía correspondiente (estructuras 1D-bordes) utilizando la Transformada Polinomial Hermitiana. Posteriormente, se regularizó dicho campo de vectores mediante Campos Aleatorios de Markov. El objetivo principal consistió en conocer la configuración de patrones 0D, 1D y 2D de una imagen, así como el campo de vectores de energía correspondiente. Lo anterior se logró mediante el empleo de la versión 2D discreta, multidireccional-multiescala de la Transformada Polinomial Hermitiana, la cual resulta útil para explicitar la información relevante de una imagen. Una vez que se obtuvo el campo de vectores de energía, se le regularizó mediante el empleo de campos aleatorios de markov, herramienta que permite analizar, restaurar o modificar imágenes localmente utilizando su información contextual; en particular, se utilizaron la estimación Bayesiana maximum a posteriori (MAP) y la técnica de optimización Iterated Conditional Modes (ICM). Los resultados obtenidos fueron satisfactorios aunque se trate de resultados cualitativos. La relevancia de este tipo de trabajos radica en la amplia variedad de aplicaciones que pueden beneficiarse de éste, como la planeación de infraestructura urbana y monitoreo de recursos, por ejemplo, ya que los patrones 1D como líneas o bordes juegan un rol importante en análisis de imágenes de Percepción Remota.

Capítulo 1

Introducción

Dentro del área del Procesamiento Digital de Imágenes existen numerosos tipos de imágenes, obtenidas mediante diversos dispositivos de adquisición. Uno de los dispositivos que ha cobrado relevancia por sus características y ventajas es el Radar de Apertura Sintética (SAR). Este tipo de radar se considera activo ya que es capaz de generar su propia fuente de energía. Lo anterior, resulta deseable debido a su flexibilidad para adquirir datos (escenas, imágenes) 24 horas al día. Además, la longitud de onda con que trabajan permite, en la mayoría de los casos, sensar escenas correspondientes a zonas que se encuentran bajo la acción de fuertes y continuos cambios climáticos (fuertes lluvias, exceso de humedad, nubes). Por lo anterior, el uso de este tipo de dispositivos y las imágenes que producen se ha generalizado y extendido a numerosas aplicaciones como el estudio de suelos, cultivos y vegetación en ecología y agricultura, la elaboración de mapas de zonas de difícil acceso en geografía, el estudio de las características de zonas costeras y ciudades y el estudio de hielos marítimos en oceanografía, por ejemplo, [1]. En este trabajo, se ha elegido utilizar imágenes SAR-PRI, y como paso previo a su manejo se les ha aplicado un método para eliminar el ruido que contienen. El método considera la naturaleza multiplicativa y el comportamiento del ruido que afecta a este tipo de imágenes y que se conoce como ‘speckle’. A partir de ello, propone una novedosa técnica para eliminar este tipo de ruido en imágenes SAR de uno o varios looks, documentada ampliamente en [2]. Una vez que se ha eliminado, en la medida de lo posible, el ruido en las imágenes SAR, se ha utilizado la Transformada Polinomial Hermitiana para conocer la configuración de patrones 0D, 1D y 2D presentes en la imagen. Esta clasificación (0D,1D,2D), es una descripción de la dimensionalidad de cada uno de los pixeles presentes en la imagen. 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 existen estructuras orientadas en una dirección particular sino simplemente pixeles o 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 pixeles ruidosos. Utilizando también la Transformada Polinomial Hermitiana, se ha generado el campo de vectores de energía correspondiente a las imágenes en estudio. Este campo de vectores contiene, por cada pixel de la imagen, un vector cuya magnitud y ángulo determinan la cantidad y dirección en que la energía se presenta. Esto se logra por medio de la representación de imágenes con coeficientes polinomiales unidimensionales, que consiste en un mapeo de los coeficientes 2D a coeficientes 1D el cual se logra maximizando la energía direccional en los primeros. Lo anterior significa que el vector asignado a cada pixel está orientado en la dirección de máxima energía (máxima magnitud), lo cual genera vectores con magnitud cero para regiones planas y homogéneas de la imagen (patrones 0D), vectores con magnitudes considerables y ángulos en la dirección de los bordes, líneas, calles, carreteras o estructuras bien orientadas (patrones 1D). Este campo de vectores 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 captadas por imágenes, lo cual puede generar resultados considerablemente buenos en el ámbito de la segmentación y clasificación de imágenes para la detección de caminos en planeación urbana, el estudio de zonas costeras y sus límites, la detección de cauces de ríos y de extensión de vegetación en ecología y administración de recursos, por ejemplo. También ha sido utilizado con éxito en la detección de flujo óptico en imágenes como lo demuestra [3]. Por lo anterior, la correcta determinación de la magnitud y dirección de los vectores asociados a cada pixel es importante. Desafortunadamente debido a la naturaleza del ruido

speckle presente en imágenes SAR, eliminarlo del todo se convierte en una tarea difícil. De tal forma que en ocasiones no todos los vectores del campo de vectores de energía se encuentran apropiadamente definidos. Ello como resultado de la aplicación de la Transformada Polinomial Hermitiana sobre imágenes con presencia de píxeles corrompidos por remanente de ruido, lo que implica que el cálculo del vector de energía correspondiente no se efectúe de manera adecuada. Por esta razón y para solucionar este problema, se ha utilizado la teoría de Campos Aleatorios de Markov, que se basa en la consideración de las características predominantes de la región de estudio para efectuar cambios y mejoras. Así, para cada vector del campo de vectores de energía se define una vecindad o área de interacción entre vectores, de la cual el mismo vector forma parte, y a partir de la que es posible estudiar el comportamiento que predomina entre los vectores de dicha área o vecindad, y de esta manera estimar el verdadero comportamiento (orientación) del vector en cuestión. Lo anterior genera entonces una técnica que es capaz de estimar la verdadera dirección de los vectores de energía del campo, con base en el análisis de los vectores vecinos.

1.1. Definición del problema

A partir de la posesión de dos imágenes SAR-PRI previamente restauradas [2], se desea clasificarlas en patrones 0D, 1D, 2D y obtener el campo de vectores de energía correspondiente, utilizando la Transformada Polinomial Hermitiana. Posteriormente, se desea regularizar dicho campo de vectores mediante Campos Aleatorios de Markov.

1.2. Objetivo

Conocer la configuración de patrones 0D, 1D y 2D de una imagen, así como el campo de vectores de energía correspondiente.

1.3. Hipótesis

- 1) Las técnicas para conocer la configuración de patrones 0D, 1D y 2D, así como el campo de vectores de energía de una imagen digital mediante la Transformada Polinomial Hermitiana, son válidas para imágenes SAR-PRI, que han sido sometidas previamente a un proceso de eliminación de ruido.
- 2) El campo de vectores de energía puede ser regularizado mediante Campos Aleatorios de Markov.

1.4. Método

Las técnicas empleadas en el desarrollo de este trabajo se basan en la teoría existente en torno a los temas: Transformada Polinomial Hermitiana y Campos Aleatorios de Markov, de los cuales se abunda en los capítulos siguientes.

Capítulo 2

Transformada de Hermite o Hermitiana, [4]

2.1. Descripción

En este capítulo se abordarán los fundamentos de la teoría relacionada con la Transformada de Hermite, primero para señales continuas y 1D y a continuación para señales discretas y 2D.

2.2. Introducción

Una imagen digital se encuentra especificada por un arreglo de datos que corresponden a valores de intensidad. Para una interpretación inteligente de la imagen se necesita hacer explícita la información importante. Esto usualmente implica determinar las relaciones espacio-temporales entre estas intensidades para proponer la manera en que se procesarán localmente los datos. Este tipo de procesamiento involucra dos decisiones importantes. Primero, para efectuar el procesamiento local, usualmente la imagen se multiplica por una función ventana. Este procesamiento local debe repetirse un número suficiente de veces, para describir a la imagen completamente. Algunas características de la ventana que deben ser elegidas son: el tamaño, la forma y el espaciamiento. El tamaño establece el conjunto de píxeles de la imagen que contribuyen en un paso específico de procesamiento. La forma determina el peso relativo de cada píxel de la imagen que está contribuyendo. Más adelante se hablará del espaciamiento entre ventanas. Segundo, para cada posición de la ventana, deben proponerse pasos específicos de procesamiento para ser aplicados. Estos pasos están determinados por el tipo de patrones que se es-

tén buscando y que deben definirse a priori. En general, es muy difícil escoger la función ventana y los pasos de procesamiento basándose en argumentos puramente teóricos. Es por ello que un número considerable de aplicaciones utilizan como referencia al Sistema de Visión Humano, HVS (Human Visual System). En cuanto a la función ventana, cabe mencionar que el problema de la determinación de su tamaño puede resolverse limitando la complejidad del análisis que se realice en cada ventana y determinando subsecuentemente el tamaño de ventana necesario para describir localmente a la imagen con precisión suficiente. De esta manera no se restringe el procesamiento a una sola escala (tamaño de ventana), sino que se repite el mismo procesamiento a diferentes escalas y subsecuentemente se utilizan las salidas en esta etapa del procesamiento para seleccionar la escala óptima para cada posición. Este mismo principio es utilizado por el HVS. Ahora bien, si se utilizan ventanas de diferentes tamaños, el espaciamiento entre ellas es generalmente proporcional a su tamaño, de hecho para estructuras piramidales, el espaciamiento y tamaño se incrementa en pasos de 2. La transformada de Hermite involucra estas consideraciones en su definición y desempeño. Por otro lado, en cuanto a las técnicas de procesamiento local, está demostrado que el uso de técnicas de preprocesamiento, esto es, describir a la imagen por medio de patrones seleccionados a priori, combinado con el uso de métodos estadísticos, provee mejores resultados que el uso directo de métodos de aproximación estadística. Además, el uso de patrones con diferentes orientaciones está de acuerdo con el HVS. Generalmente la búsqueda de patrones importantes en imágenes, tales como líneas y bordes, involucra el uso de derivadas de primer y segundo orden en combinación con un filtro paso bajas. La transformada de Hermite utiliza derivadas de Gaussianas para el procesamiento local, las cuales pueden modelar operaciones de filtrado en Visión Humana con la misma precisión y con menos parámetros que los ampliamente usados filtros de Gabor.

2.3. Transformada polinomial 1D

Se trata de una técnica de descomposición de una señal, en la que la señal es aproximada localmente por polinomios. El análisis en una Transformada Polinomial involucra dos pasos. Primero, la señal original $L(x)$ se localiza, multiplicándola por una función ventana $V(x)$. Como se dijo antes, una descripción completa de la señal implica que el proceso de localización se repita un número suficiente de veces para diferentes posiciones de la ventana. Se considera el caso de espaciamiento equidistante entre ventanas. A partir

de la función ventana $V(x)$, se puede construir la función peso

$$W(x) = \sum_k V(x - kT) \quad (2.1)$$

por repetición periódica. La función peso es periódica con período T . Si $W(x)$ es diferente de cero para todo x , se tiene

$$L(x) = \frac{1}{W(x)} \sum_k L(x - kT) \cdot V(x - kT) \quad (2.2)$$

de esta manera se garantiza que las señales localizadas $L(x) \cdot V(x - kT)$ en las diferentes posiciones kT , contengan información suficiente de la señal original. Segundo, se aproxima el pedazo de señal dentro de la ventana $V(x - kT)$ mediante polinomios. Para llevar a cabo esta expansión Polinomial se utilizan como funciones base los polinomios $G_n(x)$, de grado n , que son ortonormales con respecto a $V^2(x)$, esto es,

$$\int_{-\infty}^{+\infty} V^2(x) G_m(x) G_n(x) dx = \delta_{mn} \quad (2.3)$$

Estos polinomios están únicamente determinados por $V^2(x)$. Los polinomios ortonormales para una función ventana $V^2(x)$ arbitraria están dados por

$$G_n(x) = \frac{1}{\sqrt{M_{n-1} M_n}} \begin{vmatrix} c_0 & c_1 & \dots & c_n \\ c_1 & c_2 & \dots & c_{n+1} \\ \vdots & \vdots & & \vdots \\ c_{n-1} & c_n & \dots & c_{2n-1} \\ 1 & x & \dots & x^n \end{vmatrix} \quad (2.4)$$

donde el determinante M_n está definido como

$$M_n = |c_{i+j}|_{i,j=0,\dots,n}, \quad M_{-1} = 1 \quad (2.5)$$

y

$$c_n = \int_{-\infty}^{+\infty} x^n V^2(x) dx \quad (2.6)$$

es el momento de orden n . Bajo condiciones generales, se tiene que para la señal original $L(x)$,

$$V(x - kT) \left[L(x) - \sum_{n=0}^{\infty} L_n(kT) \cdot G_n(x - kT) \right] = 0 \quad (2.7)$$

con

$$L_n(kT) = \int_{-\infty}^{+\infty} L(x) \cdot G_n(x - kT) V^2(x - kT) dx \quad (2.8)$$

El hecho de que $L(x)$ sea analítica y finita para toda x es suficiente para garantizar la convergencia de la expansión en series en 2.7 para la mayoría de funciones ventana. De esta manera, el error de aproximación entre la señal y la expansión puede hacerse arbitrariamente pequeño eligiendo un orden para la expansión Polinomial suficientemente grande, como en el caso de las expansiones de Taylor. Esto implica que la descripción de las señales localizadas $L(x) \cdot V(x - kT)$, puede reducirse especificando un conjunto finito de coeficientes polinomiales $L_n(kT)$. La energía de la señal dentro de la ventana puede expresarse en términos de los coeficientes de la expansión de acuerdo con la generalización del teorema de Parseval para polinomios ortonormales,

$$\int_{-\infty}^{+\infty} L^2(x) V^2(x - kT) dx = \sum_{n=0}^{\infty} L_n^2(kT) \quad (2.9)$$

Combinando 2.2 y 2.7, se tiene la expansión para la señal completa

$$L(x) = \sum_{n=0}^{\infty} \sum_k L_n(kT) \cdot P_n(x - kT) \quad (2.10)$$

donde

$$P_n(x) = G_n(x) V(x) / W(x) \quad (2.11)$$

La ecuación 2.8 implica que los coeficientes $L_n(kT)$, pueden derivarse a partir de la convolución entre la señal $L(x)$, y los filtros

$$D_n(x) = G_n(-x) V^2(-x) \quad (2.12)$$

seguida de un muestreo en múltiplos de T . Este mapeo de la señal original $L(x)$, a los coeficientes polinomiales $L_n(kT)$, se conoce como Transformada Polinomial Directa. La reconstrucción de la señal a partir de estos coeficientes, puede lograrse de acuerdo con 2.10 que se conoce como Transformada Polinomial Inversa y consiste en interpolar los coeficientes $L_n(kT)$, para k entero, con la función patrón $P_n(x)$, sumando los n órdenes. La Transformada Polinomial Directa e Inversa se ilustra en la figura 2.1

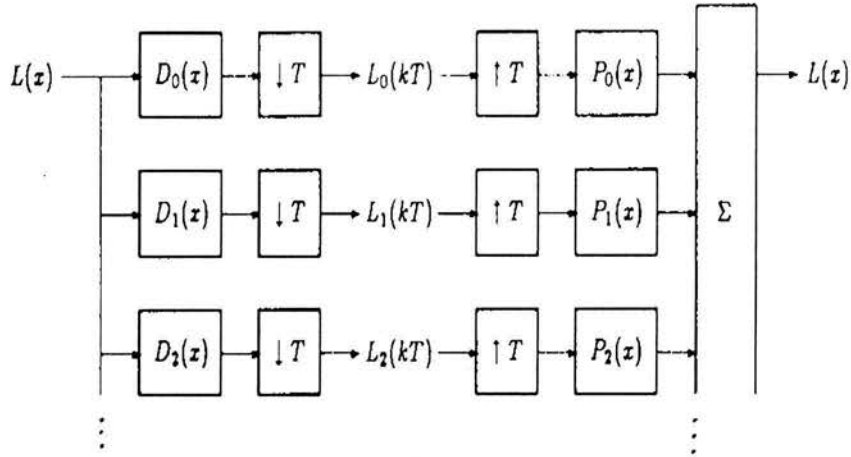


Figura 2.1: Transformada Polinomial

2.4. Transformada de Hermite 1D

Cuando la función ventana es gaussiana

$$V(x) = \frac{1}{\sqrt{\sqrt{\pi}\sigma}} e^{-x^2/2\sigma^2} \quad (2.13)$$

los polinomios ortogonales asociados con $V^2(x)$ se conocen como Polinomios de Hermite, y nos referimos a la técnica de descomposición local resultante como Transformada de Hermite. La función peso $W(x)$, es periódica con período T y puede expandirse en una serie de Fourier

$$W(x) = \frac{\sqrt{2\sqrt{\pi}\sigma}}{T} w(x) \quad (2.14)$$

con

$$w(x) = 1 + 2 \sum_{k=1}^{\infty} \exp\left[-\frac{1}{2} \left(k \frac{2\pi\sigma}{T}\right)^2\right] \cdot \cos\left(k \frac{2\pi x}{T}\right) \quad (2.15)$$

Las funciones filtro determinan que información se hará explícita en los coeficientes de la Transformada de Hermite. Las principales propiedades de la Transformada de Hermite están determinadas por estas funciones filtro. De la expresión general 2.12 se tiene que

$$D_n(x) = \frac{(-1)^n}{\sqrt{2^n n!}} \cdot \frac{1}{\sigma\sqrt{\pi}} H_n\left(\frac{x}{\sigma}\right) e^{x^2/\sigma^2} \quad (2.16)$$

ya que los polinomios de Hermite $\{H_n(x/\sigma); n = 0, 1, \dots\}$, son ortogonales sobre la ventana Gaussiana $V^2(x)$. Está demostrado que las funciones filtro $D_n(x)$, son iguales a la derivada de orden n de una Gaussiana,

$$D_n(x) = \frac{1}{\sqrt{2^n n!}} \cdot \frac{d^n}{d\left(\frac{x}{\sigma}\right)^n} \left[\frac{1}{\sigma\sqrt{\pi}} e^{-x^2/\sigma^2} \right] \quad (2.17)$$

Las funciones patron $P_n(x)$, se requieren para resintetizar la señal original a partir de los coeficientes de la Transformada de Hermite y están dados por las siguientes expresiones:

$$P_n(x) = \frac{T}{\sqrt{2^n n!}} \cdot \frac{1}{\sigma\sqrt{2\pi}} H_n\left(\frac{x}{\sigma}\right) e^{-x^2/2\sigma^2}/w(x) \quad (2.18)$$

donde $w(x)$ es la función peso de 2.15.

2.5. Transformada Polinomial 2D

Dada una función ventana $V(x, y)$, los polinomios ortonormales $G_{m,n-m}(x, y)$, donde m y $n-m$, son los grados con respecto a x y y , respectivamente, están determinados por

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} V^2(x, y) G_{m,n-m}(x, y) G_{j,i-j}(x, y) dx dy = \delta_{ni} \delta_{mj} \quad (2.19)$$

para $n, i = 0, 1, \dots, \infty$; $m = 0, \dots, n$ y $j = 0, \dots, i$. La descomposición de señales 2D en polinomiales localizadas se convierte en

$$L(x, y) = \sum_{n=0}^{\infty} \sum_{m=0}^n \sum_{(p,q) \in S} L_{m,n-m}(p, q) \cdot P_{m,n-m}(x-p, y-q) \quad (2.20)$$

donde (p, q) son todas las coordenadas en una 'Lattice' S de muestreo 2D. La única condición para la 'Lattice' de muestreo es que la función peso

$$W(x, y) = \sum_{(p,q) \in S} V(x-p, y-q) \quad (2.21)$$

sea diferente de cero para todas las coordenadas (x, y) . Los coeficientes polinomiales $L_{m,n-m}(p, q)$ se derivan convolucionando la imagen con los filtros

$$D_{m,n-m}(x, y) = G_{m,n-m}(-x, -y) V^2(-x, -y) \quad (2.22)$$

y muestreando la salida en $(p, q) \in S$. Las funciones patrón que se utilizan para interpolar los coeficientes polinomiales están definidas como

$$P_{m,n-m}(x, y) = G_{m,n-m}(x, y) V(x, y) / W(x, y) \quad (2.23)$$

para $n = 0, 1, \dots, \infty$; $m = 0, \dots, n$.

2.5.1. Representación de imágenes con coeficientes polinomiales unidimensionales

En la introducción se discutió que un análisis de imágenes debe tratar de descomponer la señal en patrones que sean perceptualmente importantes. Se sabe que en el marco de la visión computacional y la percepción visual, los patrones 1D como líneas o bordes juegan un rol importante en visión temprana. Ahora, se establecerá como encontrar el mejor patrón 1D local que encaje con una imagen con ayuda de las transformadas polinomiales. Utilizando un criterio cuadrado ponderado, se minimiza

$$E^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [K(x\cos\theta + y\sin\theta) - L(x, y)]^2 V^2(x, y) dx dy \quad (2.24)$$

sobre todos los patrones 1D K y ángulos θ . Se define la función ventana 1D

$$V_\theta^2(u) = \int_{-\infty}^{+\infty} V^2(u\cos\theta - v\sin\theta, u\sin\theta + v\cos\theta) dv \quad (2.25)$$

proyectando la función 2D $V^2(x, y)$ en un eje que haga un ángulo θ con el eje x . Esta función ventana es independiente de la orientación si $V^2(x, y)$ es rotacionalmente simétrica. Es posible expandir el patrón 1D $K(u)$ en la base $\{F_{n,\theta}(u); n = 0, 1, \dots\}$ de polinomios ortonormales sobre $V_\theta^2(u)$, esto es,

$$V_\theta(u) [K(u) - \sum_{n=0}^{\infty} K_{n,\theta} \cdot F_{n,\theta}(u)] = 0 \quad (2.26)$$

Sustituyendo las expansiones polinomiales 2D y 1D para $L(x, y)$ y $K(u)$, respectivamente, en la ecuación 2.24, y tomando la derivada parcial con respecto a $K_{n,\theta}$, se obtiene la siguiente solución óptima

$$K_{n,\theta} = \sum_{k=0}^n \sum_{l=0}^k L_{l,k-l} \cdot h_{n,\theta}(l, k-l) \quad (2.27)$$

para los coeficientes de patrones 1D, donde

$$h_{n,\theta}(l, k-l) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F_{n,\theta}(x\cos\theta + y\sin\theta) \cdot G_{l,k-l}(x, y) V^2(x, y) dx dy \quad (2.28)$$

es una función ángulo que está completamente determinada por $V^2(x, y)$. Los polinomios ortogonales $F_{n,\theta}(u)$ y la función ángulo pueden determinarse

sin necesidad de conocer explícitamente a $V_\theta(u)$. De hecho, la ecuación 2.4 implica que solo los momentos

$$c_{n,\theta} = \int_{-\infty}^{+\infty} u^n V_\theta^2(u) du \quad (2.29)$$

se necesitan para especificar completamente a los polinomios ortogonales. Sin embargo, el cálculo de estos momentos puede basarse directamente en $V^2(x, y)$, ya que

$$c_{n,\theta} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x \cos \theta + y \sin \theta)^n V^2(x, y) dx dy \quad (2.30)$$

De la ortogonalidad de los polinomios $F_{n,\theta}(u)$, es posible derivar las siguientes propiedades,

$$h_{n,\theta}(l, k-l) = 0 \quad \text{si } k > n \quad (2.31)$$

y

$$\sum_{k=0}^{\infty} \sum_{l=0}^k h_{m,\theta}(l, k-l) h_{n,\theta}(l, k-l) = \delta_{mn} \quad (2.32)$$

para la función ángulo. El error de aproximación 1D

$$E^2 = \sum_{k=0}^{\infty} \sum_{l=0}^k L_{l,k-l}^2 - \sum_{n=0}^{\infty} K_{n,\theta}^2 \quad (2.33)$$

puede minimizarse sobre el ángulo θ maximizando la energía direccional

$$\sum_{n=0}^{\infty} K_{n,\theta}^2 \quad (2.34)$$

donde $K_{n,\theta}$ se determina por medio de los coeficientes polinomiales 2D a través de la ecuación 2.27. En la práctica, sólo algunos de los primeros términos de esta energía direccional son suficientes para calcular un estimado de la dirección óptima. Cabe mencionar que la energía direccional se encuentra a través de una combinación simple de coeficientes polinomiales 2D. Esto es computacionalmente más eficiente que el hecho de usar distintos filtros para calcular la energía en cada dirección, especialmente si quieren verificarse un gran número de direcciones. Si la imagen original $L(x, y)$, es 1D localmente, entonces el error de estimación debe ser cero para el ángulo óptimo θ , por lo cual los coeficientes polinomiales 2D deben satisfacer

$$L_{l,k-l} = \sum_{n=k}^{\infty} K_{n,\theta} \cdot h_{n,\theta}(l, k-l) \quad (2.35)$$

Si la imagen $L(x, y)$, no es 1D localmente, entonces la ecuación anterior puede utilizarse como una aproximación óptima 1D, para coeficientes polinomiales 2D.

2.6. Transformada de Hermite 2D

La Transformada de Hermite se obtiene cuando en la transformada Polinomial 2D, la función ventana es una función Gaussiana. Una ventaja importante de las funciones ventana Gaussianas, en dos dimensiones, es la propiedad que tienen de ser espacialmente separables y rotacionalmente simétricas. Ahora bien, el hecho de que la función ventana sea separable permite que las funciones filtro y patrón sean también separables y, como consecuencia, que puedan ser implementadas de manera sencilla. Por ejemplo, los coeficientes polinomiales se encuentran mediante la convolución de la imagen con funciones filtro $D_m(x)D_{n-m}(y)$ donde $D_m(x)$ es la función filtro 1D para la ventana $V(x)$, seguida de un muestreo de la salida en las direcciones vertical y horizontal en múltiplos del espacio de muestreo T . Las propiedades de las funciones filtro incluyen el hecho de que son separables espacial y polarmente. La transformada de Fourier de $D_m(x)D_{n-m}(y)$, expresada en coordenadas polares $\omega_x = \omega \cos\theta$ y $\omega_y = \omega \sin\theta$ es

$$d_m(\omega_x)d_{n-m}(\omega_y) = g_{m,n-m}(\theta) \cdot d_n(\omega) \quad (2.36)$$

donde $d_n(\omega)$ es la transformada de Fourier de la función filtro de Hermite 1D $D_n(r)$, considerando a r como coordenada radial y

$$g_{m,n-m}(\theta) = \sqrt{\frac{n!}{m!(n-m)!}} \cos^m\theta \cdot \sin^{n-m}\theta \quad (2.37)$$

expresa la selectividad direccional del filtro. De esta manera los filtros de orden mayor a n analizan sucesivamente frecuencias radiales mayores, esto es, resoluciones espaciales más grandes como en el caso 1D. Los filtros del mismo orden n e índice m (direccional) diferente distinguen entre diferentes orientaciones en la imagen. La relación con la función de ángulo general presentada anteriormente es

$$h_{n,\theta}(l, k-l) = \delta_{nk} \cdot g_{l,k-l}(\theta) \quad (2.38)$$

lo cual resulta en una simplificación sustancial del caso general. Daugman demostró la importancia de filtros polarmente separables, esto es, filtros que pueden expresarse como el producto de una función de frecuencia espacial y

una función de orientación, ambas sintonizadoras. De manera más específica, sólo los filtros separables tienen la propiedad de arrojar curvas sintonizadoras de orientación idénticas para diferentes patrones 1D, como líneas y bordes. Esto implica que los filtros pueden detectar la orientación de un patrón 1D independientemente de su estructura interna.

2.7. Transformadas Polinomiales Discretas

Hasta ahora las señales y filtros se asumieron como continuos, pero las aplicaciones prácticas de las transformaciones polinomiales requieren de una formulación para señales discretas. Para ello, es posible definir a las transformadas polinomiales directamente para funciones discretas, esto es, sin la necesidad de una liga explícita entre señales digitales y analógicas. En el caso de las transformadas polinomiales 1D, las expresiones para las funciones peso, filtros y patrones siguen siendo válidas, mientras se reemplace la variable continua x por una discreta y todas las expresiones con integrales tendrán que utilizar en su lugar sumatorias. Por ejemplo, la ecuación 2.6 para calcular el momento de orden n tendrá que cambiar por

$$c_n = \sum_x x^n V^2(x) \quad (2.39)$$

para $n = 0, \dots, N$. Si la ventana discreta es finita, esto es, $V(x) = 0$ para $x < N_1$ y $x > N_2$ entonces la transformada Polinomial tendrá un orden finito $N = N_2 - N_1$. Los coeficientes polinomiales de hasta orden N , serán suficientes para reconstruir cualquier señal discreta. La razón es que la señal discreta dentro de la ventana $V(x)$ tiene únicamente $N + 1$ grados de libertad. Para el caso de las transformadas polinomiales 2D, la mayoría de los resultados para señales analógicas pueden ajustarse de manera directa a imágenes discretas. Sin embargo, existe una complicación que se debe a que cuando se proyecta la función $V^2(x, y)$ sobre un eje que hace un ángulo θ con el eje x , esto es,

$$V_\theta^2(u) = \sum_v V^2(ucos\theta - vsen\theta, usen\theta + vcos\theta) \quad (2.40)$$

sólo los puntos muestreados se deben considerar en la proyección. De esta forma, u y v asumen únicamente los valores

$$u = xcos\theta + ysen\theta, \quad v = -xsen\theta + ycos\theta \quad (2.41)$$

donde x y y toman todos los valores (enteros) para los cuales $V(x, y)$ es diferente de cero. Las funciones suaves se obtienen usualmente seleccionando

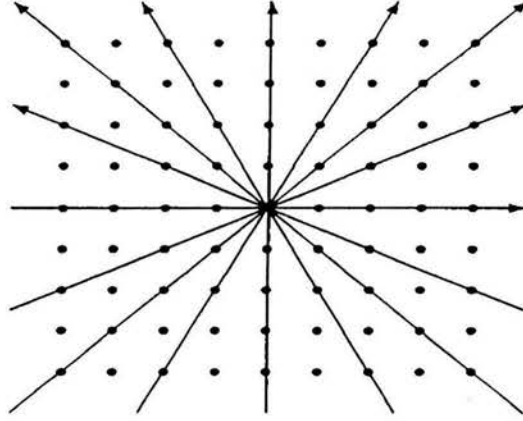


Figura 2.2: Ángulos discretos en una rejilla de muestreo cuadrada 2D

el ángulo θ para el cual las líneas de proyección tocan el número mayor de puntos posibles como se muestra en la figura 2.2. La aplicación de la técnica de aproximación 1D, requiere conocer la función ángulo

$$h_{n,\theta}(l, k-l) = \sum_{x,y} F_{n,\theta}(xcos\theta + ysen\theta) \cdot G_{l,k-l}(x, y) V^2(x, y) \quad (2.42)$$

donde $F_{n,\theta}(u)$ es el polinomio ortonormal de orden n sobre la ventana 1D $V_\theta^2(u)$.

2.8. Transformada discreta de Hermite 2D

El equivalente discreto de la ventana Gaussiana es una ventana binomial

$$V^2(x) = \frac{1}{2^M} C_M^x \quad (2.43)$$

para $x = 0, \dots, M$. Los polinomios ortonormales discretos que están asociados con esta ventana, se conocen como polinomios de Krawtchouk

$$G_n(x) = \frac{1}{\sqrt{C_M^n}} \sum_{k=0}^n (-1)^{n-k} C_{M-x}^{n-k} \cdot C_x^k \quad (2.44)$$

para $x, n = 0, \dots, M$. Para valores grandes de M , la ventana binomial se reduce a una ventana Gaussiana; de manera específica

$$\lim_{M \rightarrow \infty} \frac{1}{2^M} C_M^{x+(M/2)} = \frac{1}{\sqrt{\pi} \sqrt{\frac{M}{2}}} \exp \left[- \left(\frac{x}{\sqrt{\frac{M}{2}}} \right)^2 \right] \quad (2.45)$$

para $x = -(M/2), \dots, M/2$. También puede demostrarse que el mismo límite convierte a un polinomio de Krawtchouk en un polinomio de Hermite,

$$\lim_{M \rightarrow \infty} G_n \left(x + \frac{M}{2} \right) = \frac{1}{\sqrt{2^n n!}} H_n \left(\frac{x}{\sqrt{\frac{M}{2}}} \right) \quad (2.46)$$

De esta manera, la transformada discreta de Hermite de longitud M aproxima a la transformada analógica de Hermite de extensión $\sigma = \sqrt{M/2}$. Las propiedades de la transformada discreta de Hermite pueden predecirse adecuadamente a partir de las correspondientes propiedades del caso analógico. Si nos concentramos en el caso en el que M es par, las funciones filtro y patrón pueden centrarse en el origen mediante el movimiento de la ventana binomial sobre $M/2$. Esto permite la definición de las funciones filtro para la transformada discreta de Hermite

$$D_n(x) = G_n \left(\frac{M}{2} - x \right) \cdot V^2 \left(\frac{M}{2} - x \right) \quad (2.47)$$

para $x = -(M/2), \dots, M/2$. Estas funciones también pueden expresarse como

$$D_n \left(\frac{M}{2} - x \right) = \frac{(-1)^n}{2^M \sqrt{C_M^n}} \Delta^n [C_M^x \cdot C_x^n] \quad (2.48)$$

donde

$$(-1)^n \Delta^n L(x) = \sum_{k=0}^n (-1)^k C_n^k L(x+k) \quad (2.49)$$

es el operador diferencia de n -ésimo orden. Calculando las transformadas z de esta función filtro se tiene

$$d_n(z) = \sum_{x=-M/2}^{M/2} D_n(x) z^{-x} = z^{-M/2} \sqrt{C_M^n} \left(\frac{1-z}{2} \right)^n \left(\frac{1+z}{2} \right)^{M-n} \quad (2.50)$$

ó

$$d_n(e^{-j\omega}) = \sqrt{C_M^n} \left(j \operatorname{sen} \frac{\omega}{2} \right)^n \left(\cos \frac{\omega}{2} \right)^{M-n} \quad (2.51)$$

si se expresa en frecuencias angulares para $n = 0, \dots, M$. Es obvio que para ω 's pequeñas este filtro se reduce a una operación derivativa de orden n , como en el caso analógico. Los filtros anteriores tienen la importante ventaja práctica de que pueden implementarse como una cascada de los filtros $z^{-1}(1+z)^2$, $z^{-1}(1-z)(1+z)$, $z^{-1}(1-z)^2$, cuyos kernels son $[1 \ 2 \ 1]$, $[-1 \ 0 \ 1]$ y $[1 \ -2 \ 1]$, respectivamente. De esta manera, con excepción del factor de amplificación $\sqrt{C_M^n}$, estos filtros pueden realizarse sin la necesidad de multiplicaciones generales. El cálculo de la función ángulo $h_{n,\theta}$ es una aplicación directa de la ecuación 2.41, aunque los cálculos pueden ser más largos para n 's mayores. Algunas desviaciones sustanciales, con respecto a la función ángulo de la Transformada analógica de Hermite dada en la ecuación 2.37, ocurrirán sólo si M es pequeño. En la mayoría de las aplicaciones se estará interesado únicamente en las funciones ángulo $h_{n,\theta}$ para n pequeños. Las expresiones explícitas para $n = 0, 1, 2$ se muestran a continuación

$$\begin{aligned}
h_{0,\theta}(0,0) &= 1 \\
h_{1,\theta}(0,0) &= 0 \\
h_{1,\theta}(1,0) &= \cos\theta \\
h_{1,\theta}(0,1) &= \sin\theta \\
h_{2,\theta}(0,0) &= 0 \\
h_{2,\theta}(1,0) &= 0 \\
h_{2,\theta}(0,1) &= 0 \\
h_{2,\theta}(2,0) &= \cos^2\theta \cdot \sqrt{1 - \frac{1}{M}} \sqrt{1 - \frac{1}{M}(\cos^4\theta + \sin^4\theta)} \\
h_{2,\theta}(1,1) &= \sqrt{2}\cos\theta\sin\theta \sqrt{1 - \frac{1}{M}(\cos^4\theta + \sin^4\theta)} \\
h_{2,\theta}(2,0) &= \sin^2\theta \cdot \sqrt{1 - \frac{1}{M}} \sqrt{1 - \frac{1}{M}(\cos^4\theta + \sin^4\theta)}
\end{aligned} \tag{2.52}$$

Cabe mencionar que $h_{0,\theta}$ y $h_{1,\theta}$ son los mismos que para el caso analógico y que la diferencia máxima para $h_{2,\theta}$, entre el caso analógico y el discreto es usualmente muy pequeña, por ejemplo, para $M = 6$ es menor al 5%.

2.9. Aplicaciones de la Transformada de Hermite

En el artículo [5] de Martens, se propone la codificación jerárquica como una de las aplicaciones de la Transformada de Hermite. Un codificador

jerárquico compara las descripciones de la imagen en niveles sucesivos de detalle. Las representaciones de imágenes propuestas por Martens en el artículo [4] eran descripciones polinomiales locales y finitas en 2D. En este artículo propone la utilización de polinomiales locales 1D, motivado por el hecho de que en las imágenes naturales la mayoría de las descripciones locales 2D pueden ser aproximadas correctamente por descripciones 1D; de esta manera el poder descriptivo de la aproximación 1D es mucho más parecido o adecuado que en el caso de las descripciones 2D. Esto refleja el hecho de que las estructuras locales 1D, como los bordes, representan una mayoría de las regiones no-constantes. La importancia de las estructuras 1D ha sido reconocida ampliamente en visión computacional, donde muchos algoritmos han sido diseñados específicamente para detectar y localizar estas estructuras. En la codificación por transformada polinomial 1D, el análisis que se efectúa en cada nivel del codificador no se restringe a la transformación polinomial directa 2D en coeficientes $L_{m,n-m}$, sino que estos coeficientes son mapeados subsecuentemente en un ángulo óptimo θ y los coeficientes polinomiales 1D

$$K_{k,\theta} = \sum_{n=0}^k \sum_{m=0}^n L_{m,n-m} \cdot h_{k,\theta}(m, n-m) \quad (2.53)$$

para $k = 0, 1, \dots, N$, utilizando el procedimiento descrito en la sección ‘Representación de imágenes con coeficientes polinomiales unidimensionales’, donde también se define la función ángulo $h_{k,\theta}$. El ángulo óptimo se encuentra maximizando la energía direccional $\sum_{k=1}^N K_{k,\theta}^2$ o, equivalentemente, el coeficiente de contraste

$$C_\theta = \log \left[1 + \sum_{k=1}^N K_{k,\theta}^2 \right] \quad (2.54)$$

sobre todos los ángulos. Como resultado se tiene que para cada posición de la ventana en el nivel i , la imagen original se mapea en coeficientes 1D $K_{k,\theta}$, y un ángulo óptimo θ .

2.10. Caso 1-D: Transformada de Hermite Multiescala, [6]

2.10.1. Descomposición multiescala

Para comprender el concepto de escala se puede ver a la imagen como la representación de una escena física donde, por ejemplo, el efecto de la distancia de observación se puede traducir en un escalamiento de la distribución

de intensidades de la representación; esto significa que al alejarse del objeto observado se incrementa la escala de la representación y por lo tanto se pierden detalles, mientras que al acercarse se disminuye la escala y se perciben más detalles. La representación espacio-escala emplea las versiones escaladas de la Gaussiana, de integral unitaria, $G(x) = \exp(-x^2)/\sqrt{\pi}$, para eliminar detalles de la señal de luminancia $L(x)$ mediante la operación de convolución; es decir que $L(x, s) = L(x) * G(x, s)$ es la representación de la señal a la escala s , donde $G(x, s) = G(x/\sqrt{4s})/\sqrt{4s}$ es la Gaussiana a la escala $s > 0$. La variable normalizada $\xi = x/\sqrt{4s}$ es referida como coordenada natural. Para poder arribar a un resultado que ponga en evidencia la relación existente entre las estructuras de la imagen percibidas a diferentes escalas, es necesario tener una idea sobre la forma en la que el sistema de visión humana podría llevar a cabo dicho proceso. Los resultados psicofísicos evidencian la existencia de por lo menos cuatro canales en cada punto del campo visual a nivel de la retina (específicamente en las células ganglionares). Dichos canales exhiben una respuesta aproximada a una diferencia de Gaussianas (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 la información que debe ser agregada a la representación a la escala s_2 para obtener la representación a la escala menor s_1 . Así, la señal puede ser expresada como una suma de canales DoG

$$L(x) = \sum_{k=-\infty}^{\infty} L(x) * DoG(x, s_{k-1}, s_k) \quad (2.55)$$

con $s_k > s_{k-1}$. Se requiere que el parámetro de espaciamiento $\tau = (s_k - s_{k-1})/s_k$ sea constante para todo entero k , lo cual está en acuerdo con datos psicofísicos que soportan la existencia en el HVS (Sistema de Visión Humano) de un número de canales cuyas frecuencias centrales mantienen una relación de aproximadamente una octava, es decir $\tau \approx 0,75$.

2.10.2. Análisis estructural

En la mayoría de las aplicaciones es deseable saber que parte de la información contenida en cada canal, hablando en términos perceptivos, debe agregarse con mayor intensidad (por ejemplo, bordes, líneas, esquinas, etc.) y que parte de información debe ser atenuada porque resulta molesta a la vista (por ejemplo, texturas finas o ruido). En tal caso, resulta apropiado llevar a cabo un análisis derivativo para discernir entre las diferentes estructuras de interés. Así, la estructura local de una señal se infiere a partir de

las derivadas de las representaciones escaladas. Por otra parte, las derivadas de $L(x, s)$ se pueden obtener convolucionando la función $L(x)$ con los operadores derivadas de Gaussiana (DG), $G_n(x, s) = \partial_x^n G(x, s)$. Por tanto, el siguiente paso es expresar al filtro DoG en función de las DG, lo cual se consigue tomando la expansión en serie de Taylor de la Gaussiana a la escala s_{k-1} alrededor de la escala s_k , es decir,

$$\text{DoG}(x, s_{k-1}, s_k) = \sum_{n=1}^{\infty} \frac{(-\tau s_k)^n}{n!} G_{2n}(x, s_k) \quad (2.56)$$

En la práctica se debe limitar el número de términos de esta serie a un número pequeño. Estudios neurofisiológicos indican que el sistema visual trabaja con derivadas hasta de cuarto orden. En este estudio matemático, sin embargo, se consideran todos los términos de la serie ya que las aproximaciones pueden ser estudiadas como casos particulares.

2.10.3. Definición de la THM

La ecuación 2.56 muestra explícitamente que las derivadas de orden impar no contribuyen en la descripción de la imagen. Por otro lado, es deseable que, al menos en la etapa de análisis del modelo, aparezcan todas las derivadas con el fin de tener una descripción más rica de la estructura de la señal. Por lo tanto, es necesario distinguir la fase de análisis de la fase de síntesis en el modelo. Esto se logra factorizando las DG que aparecen en la serie 2.56 usando la propiedad de cerradura bajo convolución. Más aún, si se interpretan las operaciones de convolución como producto interno, es posible escribir las expresiones para la reconstrucción de la señal como

$$L(x) = \sum_{k=-\infty}^{\infty} \sum_{n=1}^{\infty} \tau^n D_n^{(k)}(x) \quad (2.57)$$

donde las funciones

$$D_n^{(k)}(x) = \langle L_n^{(k)}(\xi), G_n^{(k)}(x, \xi) \rangle_{\xi} \quad (2.58)$$

para $n = 1, 2, \dots$ aportan los detalles de la señal a todas las escalas y están determinadas por las derivadas escaladas,

$$L_n^{(k)}(\xi) = \langle L(x), G_n^{(k)}(x, \xi) \rangle_x \quad (2.59)$$

las cuales se determinan proyectando la señal de entrada sobre las funciones de análisis

$$G_n^{(k)}(x, \xi) = \frac{1}{\sqrt{2s_k}} G_n^* \left(\frac{x}{\sqrt{2s_k}} - \xi \right) \quad (2.60)$$

para $n = 1, 2, \dots$ y $s_k = (1 - \tau)^{-k} s_0$ para todo k entero. Al análisis de la señal expresado en la ecuación 2.59 se le conoce como THM directa, mientras que la reconstrucción de la señal mediante la ecuación 2.57 y las funciones de detalle en la ecuación 2.58 se conocen como THM inversa.

2.10.4. Acotamiento del rango de escalas

Si se considera un número finito K de canales DoG y se supone que la escala menor es suficientemente pequeña, la descomposición multicanal se reduce a

$$L(x, s_0) = L(x, s_K) + \sum_{k=1}^K \sum_{n=1}^{\infty} \tau^n D_n^{(k)}(x) \quad (2.61)$$

donde s_0 y s_K determinan las resoluciones límite del análisis. El residuo pasobajas $L(x, s_K)$ no puede ser despreciado para valores de K en un rango práctico, pero su variación es, en general, menor que la de la señal original. En cambio, el residuo pasoaltas $L(x) - L(x, s_0)$ es despreciado en virtud de que s_0 es suficientemente pequeño. En los casos en los cuales el residuo pasoaltas contiene información relevante, se puede agregar un canal definido entre las escalas 0 y $s_0(\tau = 1)$, ya que la señal es la representación a la escala cero. Por lo tanto, generalizando las definiciones de las ecuaciones 2.58, 2.59 y 2.60 para $n = 0$, es posible escribir la THM como

$$L(x) = D_0^{(K)}(x) + \sum_{n=1}^{\infty} D_n^{(0)}(x) + \sum_{k=1}^K \sum_{n=1}^{\infty} \tau^n D_n^{(k)}(x) \quad (2.62)$$

donde los dos primeros términos del miembro derecho son justamente el residuo pasobajas y el residuo pasoaltas respectivamente. El caso cuando $K = 0$ se conoce como transformada de Hermite de una sola escala (o simplemente TH). En tal caso no aparece el tercer término de la ecuación 2.62.

2.10.5. Esquema piramidal contínuo

Al esquema de procesamiento en el cual la imagen se descompone en un número de subimágenes pasobajas o pasobanda, las cuales son submuestreadas en proporción a su resolución, se le conoce como esquema piramidal. Cada nivel de la pirámide guarda una relación lineal con los niveles adyacentes y dicha relación es invariante con la escala. Esta definición está enclavada en el contexto de señales discretas, pero puede ser extrapolada al caso continuo si se reemplaza el remuestreo por un reescalamiento de la coordenada espacial. En tal caso, todos los niveles de la pirámide quedan definidos en un dominio

continuo. La relación piramidal de la THM se deriva del hecho de que las funciones de análisis de un orden dado forman una pirámide Gaussiana dada por la relación lineal

$$\sqrt{\tau^n} G_n^{(k)}(x, \xi) = \int_{-\infty}^{\infty} G_0^{(k-1)}(x, \eta) P_n(T\xi - \eta) d\eta \quad (2.63)$$

donde 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 $T = 1/\sqrt{1-\tau}$ determina la proporción entre las coordenadas de un nivel dado y del nivel anterior. Para $\tau = 0,75$ se tiene $T = 2$. Aplicando sucesivamente la propiedad 2.63 en la expresión 2.58 se obtiene la relación piramidal ascendente

$$\tau^n D_n^{(k)} = \{[(\sqrt{\tau^n} L_n^{(k)\uparrow T} * P_n)^{\uparrow T} * \underbrace{P_0]^{\uparrow T} \dots * P_0}_{k-1 \text{ factores}}\} * P_0^{(0)} \quad (2.64)$$

para $k = 1, \dots, K-1$ y $P_n^{(0)}(x) = G_n^*(-x/\sqrt{4s_0})/\sqrt{4s_0}$. Por otra parte, si se aplica la misma propiedad a la fase de análisis de la ecuación 2.59, se obtiene la relación piramidal descendente

$$\sqrt{\tau^n} L_n^{(k)} = \{[(L * P_0^{(0)} * \underbrace{P_0]^{\downarrow T} \dots * P_0}_{k-1 \text{ factores}}] * (-1)^n P_n\}^{\downarrow T} \quad (2.65)$$

Los símbolos $\uparrow T$ y $\downarrow T$ significan que la coordenada se multiplica por T y T^{-1} respectivamente. De las ecuaciones 2.64 y 2.65 se obtiene el esquema piramidal mostrado en la figura 2.3 donde, por simplicidad, se escribió $L_n^{(k)}$ en lugar de $\sqrt{\tau^n} L_n^{(k)}$. Como se puede observar en este diagrama, el esquema piramidal tiene la característica de que la reconstrucción se realiza por etapas agregando sucesivamente las contribuciones de cada nivel de la pirámide comenzando con el residuo pasobajas.

2.10.6. Predicción entre niveles

Una desventaja de la THM es que 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 esa redundancia se debe a la dependencia lineal entre las funciones de análisis de distintas escalas. En particular, se puede demostrar que las funciones del nivel $k-1$ se relacionan con las del nivel k mediante

$$\sqrt{\tau^n} G_n^{(k-1)}(x, \xi) = \sum_{m=n}^{\infty} c_{n,m} \int_{-\infty}^{\infty} \sqrt{\tau^m} G_m^{(k)}(x, \eta) P_{m-n}(\eta - T^{-1}\xi) d\eta \quad (2.66)$$

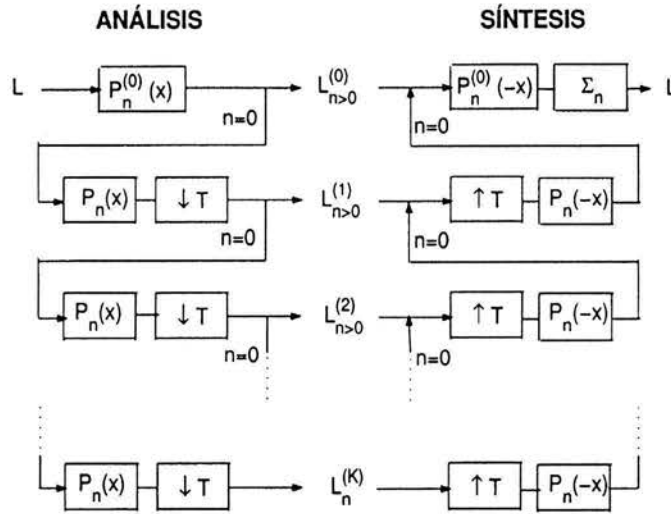


Figura 2.3: Esquema piramidal de análisis y síntesis de la THM

con las constantes $c_{n,m} = \sqrt{C_m^n} T^{-m} [(1+T^2)/2]^{m-n}$. Esta dependencia lineal entre las funciones de análisis permite insertar la THM en un esquema de predicción en el cual cada nivel se estima a partir del siguiente nivel.

2.11. Caso 2-D: Transformada de Hermite multiescala-multidireccional

2.11.1. Descomposición multiescala de señales 2-D

Para el caso de señales multidimensionales, la representación multiescala se lleva a cabo filtrando la señal en cascada a lo largo de cada coordenada ya que la Gaussiana multidimensional se expresa como un producto de Gaussianas 1-D de cada coordenada. En particular, para el caso bidimensional se definen los operadores de derivadas escaladas 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, la descomposición de señales bidimensionales en canales DoG es análoga al caso de señales 1-D.

2.11.2. Análisis estructural de señales 2-D

Otro factor físico que influye en la representación de los patrones visuales es la inclinación relativa del sistema de visión con respecto a un eje de referencia en la escena. Dicha inclinación hará que las estructuras de la imagen se perciban con una orientación determinada, la cual sólo puede ser vista por operadores sensibles a la orientación. La ventaja de usar derivadas de Gaussiana se hace evidente cuando se requiere representar estructuras orientadas ya que la función DoG es isotrópica. Se puede demostrar que el conjunto formado por las $n + 1$ derivadas bidimensionales de orden n permite representar patrones de hasta n orientaciones simultáneas. Para propósitos de análisis de orientación es conveniente trabajar con las versiones rotadas de estos operadores. 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, se puede demostrar que estas funciones están relacionadas linealmente con las funciones referidas al sistema original. Las derivadas direccionales de Gaussiana (DDG), denotadas por $G_{n,0}(x, y, s, \theta)$, se han empleado eficazmente en la detección y discriminación de orientaciones. Estos operadores son ideales para la descripción de estructuras orientadas porque satisfacen la llamada propiedad de direccionamiento (steering), a partir de la cual se demuestra que sólo se requiere conocer la respuesta de la DDG de orden n en $n + 1$ direcciones para poder conocer la respuesta en cualquier dirección. La expansión en serie de Taylor de la DoG bidimensional puede escribirse ya sea en función de las DG bidimensionales o bien en función de las DDG. La primera expansión produce la THM en dos dimensiones, que no es otra cosa que la transformación definida en la sección anterior pero aplicada sobre ambas coordenadas espaciales. El segundo caso es más interesante y vale la pena describirlo con mayor detalle. En este caso la expansión de la DoG se escribe como

$$DoG(x, y, s_{k-1}, s_k) = \sum_{n=1}^{\infty} \sum_{j=0}^n \frac{c_n (-\tau s_k)^n}{n!} G_{2n,0}(x, y, s_k, \theta_j) \quad (2.67)$$

donde $\theta_j = \theta_0 + j\pi/(n + 1)$ para $j = 0, \dots, n$ y

$$c_n = \frac{1}{\sum_{j=1}^n \text{sen}^{2n}(\frac{j\pi}{n+1})} = \frac{1}{n+1} \sum_{m=0}^n \frac{(C_n^m)^2}{C_{2n}^{2m}} \quad (2.68)$$

para $n = 1, 2, \dots$. Note 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 convenga.

2.11.3. Definición de la THMM

Partiendo de la expansión 2.67 y procediendo como en el caso 1-D, se construye una transformada que resulta más apropiada para la descripción de imágenes, ya que las funciones de análisis que se obtienen incluyen las tres transformaciones geométricas fundamentales: traslación, rotación y escalamiento. El proceso de análisis

$$L_n^{(k,j)}(\xi, \eta) = \langle L(x, y) G_n^{(k,j)}(x, y, \xi, \eta) \rangle_{(x,y)} \quad (2.69)$$

mediante las funciones de análisis

$$G_n^{(k,j)}(x, y, \xi, \eta) = \frac{1}{2s_k} G_{n,0}^* \left(\frac{x \cos \theta_j + y \sin \theta_j}{\sqrt{2s_k}} - \xi, \frac{-x \sin \theta_j + y \cos \theta_j}{\sqrt{2s_k}} - \eta \right) \quad (2.70)$$

con $\theta_j = \theta_0 + j/(n+1)$ para $j = 0, \dots, n$; $n = 1, 2, \dots$ y $s_k = (1 - \tau)^{-k}$ para todo k entero, es referido como transformada de Hermite multiescala-multidireccional (THMM) porque emplea las DDG a múltiples escalas y orientaciones. Y por tanto, al proceso de reconstrucción de la señal a partir de esta representación se le denomina THMM inversa. En este caso, la reconstrucción de la señal

$$L(x, y) = \sum_{k=-\infty}^{\infty} \sum_{n=1}^{\infty} \sum_{j=0}^n c_n \tau^n D_n^{(k,j)}(x, y) \quad (2.71)$$

se lleva a cabo sumando las contribuciones de las funciones de detalle direccional

$$D_n^{(k,j)}(x, y) = \langle L_n^{(k,j)}(\xi, \eta) G_n^{(k,j)}(x, y, \xi, \eta) \rangle_{(\xi, \eta)} \quad (2.72)$$

sobre el conjunto de escalas y orientaciones previamente definido.

2.11.4. Filtrado direccional

La representación de imágenes mediante la THMM tiene la desventaja de que las funciones de análisis de la ecuación 2.70 son no-separables y, por lo tanto, su implantación puede resultar costosa en cuanto a recursos computacionales. Sin embargo, ese costo puede ser reducido si se calcula la transformada del caso separable y posteriormente se transforman los coeficientes del mismo orden total a través de las funciones de ángulo

$$\alpha_{m,n-m}(\theta) = \sqrt{C_n^m} \cos^m \theta \sin^{n-m} \theta \quad (2.73)$$

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

$$G_n^{(k,j)}(x, y) = \sum_{m=0}^n \alpha_{m,n-m}(\theta_j) G_m^{(k)}(x) G_{n-m}^{(k)}(y) \quad (2.74)$$

para $n = 1, 2, \dots$. La proyección de la señal sobre las funciones de análisis equivale a una convolución con los filtros direccionales que se obtienen reflejando las funciones de la ecuación 2.74 respecto a x y y , más la normalización de coordenadas respecto a escala y orientación. El filtrado direccional es ampliamente citado en la literatura como mecanismo para hacer explícita la información relevante de una imagen. Por ejemplo, considérese una estructura 1-D orientada con un ángulo ϕ medido respecto al eje x . La THMM que se obtiene 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. Estos coeficientes 1-D contienen información sobre contraste y nivel de luminancia, mientras que los coeficientes $L_n^{(k,j)}$ contienen además información sobre la localización y orientación de la estructura a distintas escalas.

2.12. Caso discreto: THDM y THDMM

2.12.1. Espacio-escala discreto

Dado que en la práctica las imágenes obtenidas son representadas digitalmente para su manipulación y almacenamiento, es necesaria una formulación discreta del modelo. Un problema concerniente a la representación espacio-escala de señales discretas es cómo discretizar la teoría espacio-escala y al mismo tiempo mantener las propiedades del caso continuo. Para señales unidimensionales se puede formular una teoría discreta completa siguiendo el mismo axioma de causalidad del caso continuo; $H(\cdot, s)$ es un kernel discreto de la representación espacio-escala si, y solo si, el número de extremos locales de $L * H$ no excede el número de extremos de L . El único kernel de soporte compacto que satisface este axioma es la función binomial generalizada. Y si la definición se combina con el requerimiento de que la familia de transformaciones de suavizado debe obedecer la propiedad de semigrupo, $H(\cdot, s_2) * H(\cdot, s_1) = H(\cdot, s_2 + s_1)$, entonces la única función que puede ser candidata para kernel del espacio-escala discreta es la función binomial simétrica $B[x, N] = 2^{-2N} C_{2N}^{x+N}$ para $N \in \mathbb{Z}^+$, $x = -N, \dots, N$ y $B[x, N] = 0$ para cualquier otro valor de x . Este kernel satisface varias propiedades en el dominio discreto que son similares a las del kernel Gaussiano en el dominio

continuo; por ejemplo, tiende a la función delta discreta cuando $N \rightarrow 0$, mientras que para valores grandes de N se aproxima a la Gaussiana. A pesar de que esta formulación de la teoría espacio-escala es completa y análoga al caso continuo, no puede ser extrapolada a espacios de más de una dimensión. Se ha demostrado que usando argumentos similares al caso continuo no existe ningún kernel trivial en dos o más dimensiones que garantice no introducir nuevos extremos locales. Sin embargo, la causalidad de los extremos locales puede ser expresada alternativamente: Si para un nivel de escala s_0 , un punto x_0 es un máximo (mínimo) local a ese nivel, entonces su valor no debe incrementar (decrementar) al incrementar la escala. Intuitivamente, este requerimiento previene que los extremos locales sean realzados o aparezcan de la nada. Basado en este axioma, se puede generalizar el kernel binomial a dos dimensiones como $B[x, y, N] = B[x, N]B[y, N]$. En este caso, aunque la escala es la misma en las orientaciones vertical y horizontal, no es así para las orientaciones oblicuas; lo cual supone una limitación para el análisis de orientaciones.

2.12.2. Derivadas discretas

El equivalente discreto de las derivadas escaladas se escribe en términos de diferencias finitas del kernel binomial, es decir, $B_n[x, N] = \delta_x^n B[x - N, N - n/2]$ para $n = 0, \dots, 2N$. En lo sucesivo se hará referencia a estas funciones como Diferencias de Binomial (DB). Se puede demostrar que las DB tienden asintóticamente a las DG al crecer la escala, específicamente, $B_n[x, N] \sim G_n(x, N/4), N \rightarrow \infty$. De ahí que la mayoría de sus propiedades puedan predecirse con bastante precisión a partir de las propiedades correspondientes del caso continuo haciendo $N = 4s$. La generalización de estos operadores a 2-D es inmediata por la separabilidad del kernel binomial, es decir, $B_{n,m}[x, y, N] = B_n[x, N]B_m[y, N]$. Por lo que el filtro de derivada direccional se puede definir convenientemente como

$$B_n[x, y, N, \theta_j] = \sum_{m=0}^n C_m^n \cos^m(\theta_j) \sin^{n-m}(\theta_j) B_m[x, N] B_{n-m}[y, N] \quad (2.75)$$

donde el parámetro de orientación está discretizado como en el caso continuo: $\theta_j = \theta_0 + j\pi/(n + 1)$, para $j = 0, \dots, n$ y un valor θ_0 arbitrario.

2.12.3. Definición de la THDM

El filtro DoG puede ser aproximado por una Diferencia de Binomiales (DoB) y dado que las funciones Binomiales son de soporte compacto, su

representación mediante las DB se expresa como una suma finita. A partir de esta aproximación se construye la versión discreta de la THM a la que denominamos THDM. La descomposición de una señal discreta se representa como una suma finita de las funciones de detalle discretas $D_n^{(k)} = (L_n^{(k)})^\dagger T_k * T_k B_n^{(k)}$, esto es,

$$L[x] = D_0^{(k)}[x] + \sum_{n=1}^{N_0} D_n^{(0)}[x] + \sum_{k=1}^K \sum_{n=1}^{\tau N_k} \Gamma_n^{(k)}(\tau) D_n^{(k)}[x] \quad (2.76)$$

donde, como en el caso continuo, la fase de análisis $L_n^{(k)} = (L * (-1)^n B_n^{(k)}) \downarrow T_k$ puede verse como la proyección de la señal de entrada sobre las funciones de análisis

$$B_n^{(k)}[x, \xi] = B_n^*[x - T_k \xi, N_k/2] \quad (2.77)$$

Note la similitud de estas expresiones con las obtenidas para el caso continuo con número finito de canales DoG. De hecho, los coeficientes $\Gamma_n^{(k)}$ tienden a τ^n a medida que el índice de escala k crece ($N_k \rightarrow \infty$). El primer término de la ecuación 2.76 corresponde al residuo pasobajas; el segundo, al canal DoB entre las escalas cero y N_0 ; y el tercero, a los canales espaciados logarítmicamente con parámetro de espaciamiento $\tau = (N_k - N_{k-1})$. El factor de normalización $T_k = \sqrt{N_k/2}$ debe ser entero. Evidentemente, el valor $T_k = 2^k$ genera una secuencia de escalas $N_k = 2^{2k+1}$ cuya constante de espaciamiento $\tau = 0,75$.

2.12.4. Definición de la THDMM

La THDM se generaliza directamente a dos o más dimensiones en forma similar al caso continuo, mientras que la versión discreta de la THMM se calcula a partir de ésta mediante la definición de la ecuación 2.75 . Es decir, para un punto genérico la expresión

$$L_n^{(k,j)} = \sum_{m=0}^n \alpha_{m,n-m}(\theta_j) L_{m,n-m}^{(k)} \quad (2.78)$$

sirve para el análisis de la señal, mientras que para la síntesis de la señal se efectúa la THDM inversa con los coeficientes

$$H_{m,n-m}^{(k)} = c_n \sum_{j=0}^n \alpha_{m,n-m}(\theta_j) L_n^{(k,j)} \quad (2.79)$$

Es de esperarse que, debido al submuestreo, los coeficientes $L_{m,n-m}^{(k)}$ y $H_{m,n-m}^{(k)}$ reproduzcan imágenes ligeramente diferentes. Por lo general suele elegirse $\theta_0 = \arctan(L_{0,1}^{(k)}/L_{1,0}^{(k)})$ en todos los niveles, lo que ocasiona que $L_1^{(k,0)} = 0$. La ventaja de la THDMM es que tiende a compactar más la energía a lo largo de una coordenada.

2.13. Clasificación de estructuras en el dominio de la THDR

En esta sección se presenta un esquema de representación de imágenes basado en una clasificación perceptual de los coeficientes de la transformada de Hermite discreta rotada (THDR) sobre una latice de muestreo aproximadamente hexagonal. 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. Los coeficientes de un mismo orden se mapean mediante una transformación unitaria especificada localmente. La transformación se basa en funciones binomiales generalizadas, y como resultado se obtienen los coeficientes de la THD referidos a un sistema de coordenadas 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 se define un umbral de error máximo tolerado para la aproximación del bloque por una constante (patrón O-D) o por una estructura orientada (patrón l-D).

2.13.1. Descomposición de la imagen

El primer paso consiste en el mapeo de los datos de la imagen (mediante la THD a una sola escala) en un conjunto de coeficientes espectrales a partir de los cuales es posible obtener una reconstrucción ‘perfecta’ de la imagen. La THD se basa en los filtros binomiales

$$B_{n,k} = (-1)^n 2^{-N} \sqrt{C_N^n} \Delta_k^n \{C_{N-n}^{k-n}\} \quad (2.80)$$

para $n, k = 0, \dots, N$ donde n es el orden de derivación y $N + 1$ es la longitud del filtro. Para las pruebas se emplearon filtros con N igual a una potencia de 2. Por otro lado, se definen las funciones binomiales

$$A_{n,k} = \sqrt{2^{-N}} \Delta_k^n \{C_{N-n}^{k-n}\} \quad (2.81)$$

para $n, k = 0, \dots, N$, las cuales están relacionadas con los filtros binomiales mediante $B_{n,k} = W_k A_{n,N-k}$, donde

$$W_k = \sqrt{2^{-N} C_N^k} \quad (2.82)$$

es la ventana binomial. Las funciones binomiales satisfacen un número de propiedades importantes las cuales pueden ser estudiadas usando la representación matricial $A = [A_{n,k}]_{n,k=0,\dots,N}$. Por ejemplo, para $N = 8$ la matriz se escribe explícitamente como

$$\frac{1}{\sqrt{2^8}} \begin{pmatrix} 1 & 8 & 28 & 56 & 70 & 56 & 28 & 8 & 1 \\ 1 & 6 & 14 & 14 & 0 & -14 & -14 & -6 & -1 \\ 1 & 4 & 4 & -4 & -10 & -4 & 4 & 4 & 1 \\ 1 & 2 & -2 & -6 & 0 & 6 & 2 & -2 & -1 \\ 1 & 0 & -4 & 0 & 6 & 0 & -4 & 0 & 1 \\ 1 & -2 & -2 & 6 & 0 & -6 & 2 & 2 & -1 \\ 1 & -4 & 4 & 4 & -10 & 4 & 4 & -4 & 1 \\ 1 & -6 & 14 & -14 & 0 & 14 & -14 & 6 & 1 \\ 1 & -8 & 28 & -56 & 70 & -56 & 28 & -8 & 1 \end{pmatrix} \quad (2.83)$$

donde cada renglón representa una máscara de análisis. Las funciones binomiales satisfacen la propiedad de ortogonalidad $AA = I$, donde I es la matriz identidad. Por lo tanto, la transformación definida por la matriz A lleva a cabo las mismas operaciones tanto para el análisis como para la síntesis. La salida de los filtros binomiales en una posición genérica se puede obtener mediante la transformación definida por la matriz $B = WA$, donde $W = \text{diag}(W_k)_{k=0,\dots,N}$ es una matriz diagonal que representa la ventana binomial. De la propiedad de ortogonalidad de la matriz A se puede demostrar que $BB^T = W^2$. Si F es una matriz cuadrada que contiene los datos de un bloque de $(N + 1) \times (N + 1)$ de la imagen de entrada, entonces $G = BFB^T$ son coeficientes de la THD. Se dice que la THD es una expansión local de los datos de la imagen ya que para cada punto sobre una latice de muestreo se puede obtener la representación local $W^2FW^2 = B^TGB$.

2.13.2. Rotación local

La rotación local en el dominio de la transformada puede verse como un mapeo de los coeficientes espectrales a través de una transformación unitaria. Esta se lleva a cabo sobre los coeficientes de un mismo orden y puede ser representada matricialmente como se describe a continuación. Sean g_n para $n = 0, \dots, 2N$, los vectores columna formados por las antidiagonales de

la matriz de coeficientes G y sean las funciones binomiales generalizadas definidas como

$$A_{n,k}^{(\theta)} = c^{-k} s^k \Delta^n \{ C_{N-n}^{k-n} c^{2k-n} s^{N-2k+n} \} \quad (2.84)$$

con $c = \cos\theta$ y $s = \sin\theta$ para $0 \leq \theta < \pi$. Se puede ver que si el parámetro de sesgo θ se hace igual a $\pi/4$ se obtiene el caso simétrico, es decir, $A_{n,k} = A_{n,k}^{(\pi/4)}$. Usando la notación matricial se denotará la matriz de funciones binomiales generalizadas por $A_{(\theta)}$. Esta matriz satisface la misma propiedad de ortogonalidad del caso simétrico. Las primeras tres matrices (excepto el caso $N = 0$, el cual se reduce a la unidad) tienen la siguiente expresión explícita

$$\begin{pmatrix} s & c \\ c & -s \end{pmatrix}, \begin{pmatrix} s^2 & 2sc & c^2 \\ sc & c^2 - s^2 & -sc \\ c^2 & -2sc & s^2 \end{pmatrix}, \begin{pmatrix} s^3 & 3s^2c & 3sc^2 & c^3 \\ s^2c & 2sc^2 - s^3 & c^3 - 2s^2c & -s^2c \\ s^3 & -3s^2c & 3sc^2 & -c^3 \end{pmatrix} \quad (2.85)$$

La rotación de los vectores de coeficientes para un ángulo θ dado se obtienen a partir de la transformación definida por la matriz $R^{(\theta)} = W A^{(\theta)} W^{-1}$, es decir que el vector $g_n^{(\theta)} = R^{(\theta)} g_n$ representa los coeficientes de orden n referido a los ejes coordenados rotados por un ángulo θ medido en sentido inverso a las manecillas del reloj, mientras que la rotación inversa se consigue aplicando la matriz de rotación transpuesta a los vectores de coeficientes rotados. En lo subsecuente se hará referencia a los coeficientes rotados a través de la matriz $G^{(\theta)}$ cuyas antidiagonales son justamente los vectores $g_n^{(\theta)}$. En las pruebas presentadas aquí, se eligió el ángulo del gradiente $\theta = \arctan(G_{0,1}/G_{1,0})$

2.13.3. Clasificación

La idea de clasificar los bloques de coeficientes previamente rotados tiene la finalidad de reducir la redundancia aprovechando el hecho de que en muchos casos el valor de la imagen puede ser localmente aproximado por estructuras que son O-D o 1-D. Para bloques localizados en regiones homogéneas, todos los coeficientes, excepto $G_{0,0}$ (el coeficiente de DC), tienen poca relevancia en la reconstrucción de la señal. Mientras que los coeficientes para bloques localizados sobre los contornos presentan una concentración de energía importante sobre la primera columna de la matriz $G^{(\theta)}$ (por simplicidad se omitirá el superíndice de aquí en adelante) y, en tal caso, los coeficientes $G_{i,j}$ para $j > 0$ pueden ser despreciados. En el resto de los casos deberán codificarse todos los coeficientes. Si las clases se etiquetan como O-D, 1-D y 2-D respectivamente se tienen las siguientes hipótesis: H0 = el bloque es O-D, H1 = el bloque es 1-D, H2 = el bloque es 2-D. La partición del espacio

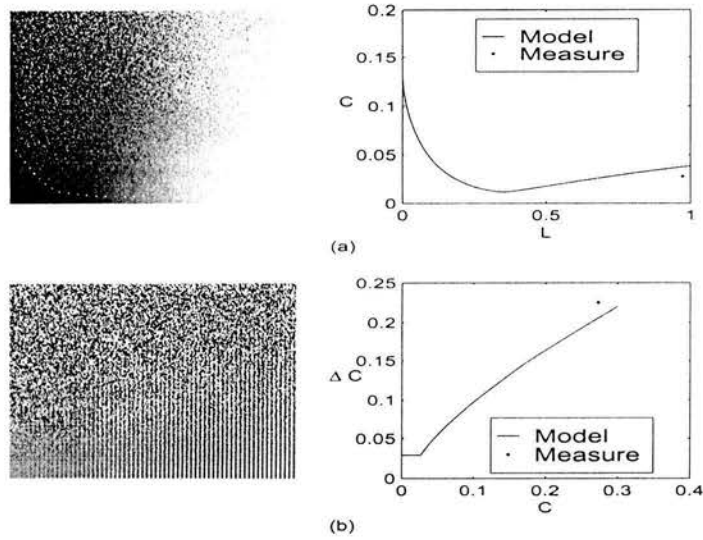


Figura 2.4: a) Umbral de detección de contraste y b) Umbral de enmascaramiento de contraste

en las clases definidas arriba puede llevarse a cabo más fácilmente asumiendo que cualquiera de ellas puede ser perfectamente descrita por las siguientes variables:

$$L = G_{0,0}, c = \left[\sum_{i=0}^N \sum_{j=0}^N G_{i,j}^2 - L^2 \right]^{1/2}, \Delta C = \left[C^2 - \sum_{i=1}^N G_{i,0}^2 \right]^{1/2} \quad (2.86)$$

La clasificación se lleva a cabo en dos pasos. En primer lugar separamos los bloques O-D mediante la comparación

$$C \leq_{H_0}^{H_1} k_0 C_{thr}(L) \quad (2.87)$$

donde la curva $C_{thr}(L)$ representa el umbral de detección de contraste. Esta curva se infiere mediante la imagen de estímulo de la figura 2.4 a) (izquierda). La imagen presenta ruido blanco gaussiano cuya media varía linealmente en la dirección horizontal y cuya desviación estándar varía linealmente en la dirección vertical. La curva trazada sobre la imagen es una curva de detección típica que señala el umbral a partir del cual el contraste local es visualmente

relevante. El umbral se ajustó por el siguiente modelo

$$C_{thr} = C_{min} + \left| \frac{L^\alpha - L_{min}^\alpha}{L^\alpha + L_{min}^\alpha} \right|^{1/\alpha} \quad (2.88)$$

donde el exponente α está entre 0 y 1 (típicamente 0.6) y C_{min} es el contraste mínimo al nivel de intensidad L_{min} el cual define a su vez el nivel de máxima sensibilidad al contraste. En la figura 2.4 a) (derecha) se muestran la curva teórica de la ecuación 2.88 y las mediciones sobre el perfil mostrado en la imagen. Se observa que el modelo ajusta bastante bien los datos subjetivos. En el segundo paso, separamos la clase 1-D del conjunto de bloques no clasificados mediante la comparación

$$\Delta C \lesssim_{H_2}^{H_1} k_1 \Delta C_{thr}(L, C) \quad (2.89)$$

donde ΔC_{thr} es el umbral de enmascaramiento de contraste. Idealmente, si un bloque contiene una estructura orientada, toda la energía queda contenida completamente en la dirección del gradiente y el contraste residual es cero. En la práctica, sin embargo, lo más común es que exista un contraste residual diferente de cero aún en los casos en que visualmente se perciba una estructura orientada. Esto se debe al fenómeno visual conocido como enmascaramiento de contraste, el cual se refiere justamente a la reducción en la visibilidad de una componente de la imagen debido a la presencia de otra. El modelo de enmascaramiento visual que se emplea está basado en los trabajos de Legge y Foley donde

$$\Delta C_{thr} = \max(C_{thr}, C^\beta C_{thr}^{1-\beta}) \quad (2.90)$$

y el exponente β está entre 0 y 1. Note que cuando $\beta = 0$ no existe enmascaramiento y el umbral es constante igual a C_{thr} . Cuando $\beta = 1$, se tiene lo que comúnmente se denomina ley de Weber. Un valor típico es $\beta = 0,7$. La figura 2.4 b) muestra la curva teórica (derecha) y las mediciones obtenidas a lo largo de la curva mostrada en la imagen de estímulo (izquierda). La curva define un umbral subjetivo que separa las estructuras 1-D de las estructuras 2-D. En este caso la imagen de estímulo está compuesta de barras senoidales cuya amplitud varía linealmente en dirección horizontal más ruido blanco con desviación estándar variando linealmente en dirección vertical.

Capítulo 3

Campo de vectores de energía y clasificación de estructuras de una imagen

3.1. Descripción

En este capítulo se presentan los resultados de haber aplicado la representación con coeficientes polinomiales unidimensionales y la clasificación de estructuras, como se hizo en [6] pero esta vez a una imagen SAR del valle de México, captada por el radar transportado por el satélite ERS-2 en el año de 1995. Cabe mencionar que dicha imagen, en formato raw, fue provista por la European Space Agency ESA, bajo el contrato: A03-318. Los resultados de estas pruebas se presentaron en [7].

3.2. Campo de vectores de energía, [5]

Las representaciones de imágenes propuestas por Martens en el artículo [4] e incluidas al inicio del capítulo 2, eran descripciones polinomiales locales y finitas en 2D; más adelante, el mismo Martens propone la utilización de polinomiales locales 1D motivado por el hecho de que en las imágenes naturales la mayoría de las descripciones locales 2D pueden ser aproximadas correctamente por descripciones 1D, y de esta manera el poder descriptivo de la aproximación 1D es mucho más parecido o adecuado que en el caso de las descripciones 2D. Esto refleja el hecho de que las estructuras locales 1D como los bordes representan una mayoría de las regiones no-constantes. La importancia de las estructuras 1D ha sido reconocida ampliamente en

visión computacional, donde muchos algoritmos han sido diseñados específicamente para detectar y localizar estas estructuras. A partir de la aplicación de la representación de coeficientes polinomiales unidimensionales a una imagen, es posible obtener el campo de vectores de energía correspondiente. En este campo de vectores, existe un vector con magnitud y dirección, para cada píxel, que representa la cantidad y dirección de energía presente en el mismo. En la representación de imágenes con coeficientes polinomiales unidimensionales, el análisis que se efectúa no se restringe a la transformación polinomial directa 2D en coeficientes $L_{m,n-m}$, sino que estos coeficientes son mapeados subsecuentemente en un ángulo óptimo θ y los coeficientes polinomiales 1D

$$K_{k,\theta} = \sum_{n=0}^k \sum_{m=0}^n L_{m,n-m} \cdot h_{k,\theta}(m, n-m) \quad (3.1)$$

para $k = 0, 1, \dots, N$, utilizando el procedimiento descrito en la sección "Transformada Polinomial 2D", donde también se define la función ángulo $h_{k,\theta}$. El ángulo óptimo se encuentra maximizando la energía direccional $\sum_{k=1}^N K_{k,\theta}^2$, o equivalentemente, el coeficiente de contraste

$$C_\theta = \log \left[1 + \sum_{k=1}^N K_{k,\theta}^2 \right] \quad (3.2)$$

sobre todos los ángulos que se muestran en esa misma sección. Como resultado se tiene que para cada posición de la ventana en el nivel i , la imagen original se mapea en coeficientes 1D $K_{k,\theta}$, y un ángulo óptimo θ . De esta manera, se obtiene un conjunto de vectores orientados cada uno en la dirección de máxima energía, esto es, en la dirección de orientación de la estructura o patrón 1D. A continuación se presentan los campos de vectores de energía correspondientes a varias secciones o recortes de la imagen.

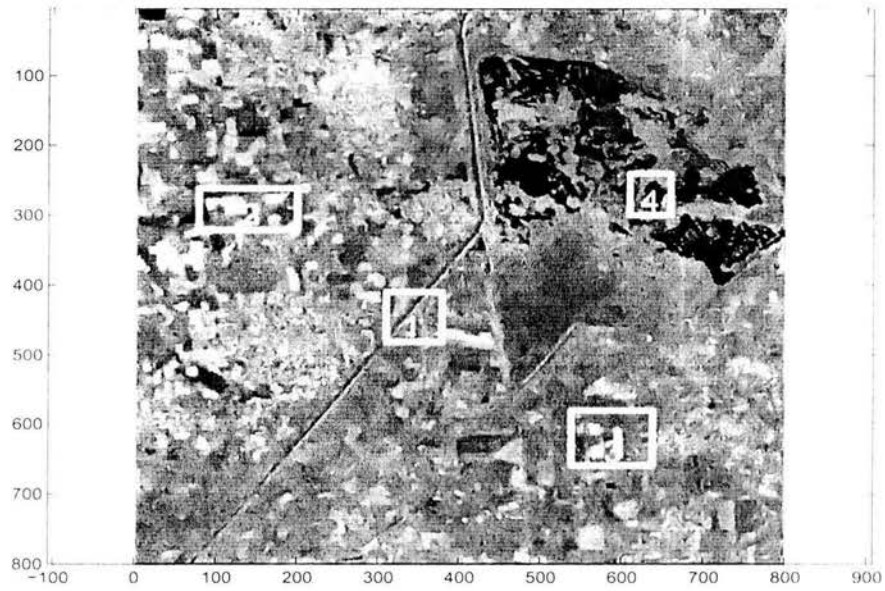


Figura 3.1: Recorte de la imagen original SAR, y señalamiento de la ubicación de secciones para pruebas sucesivas

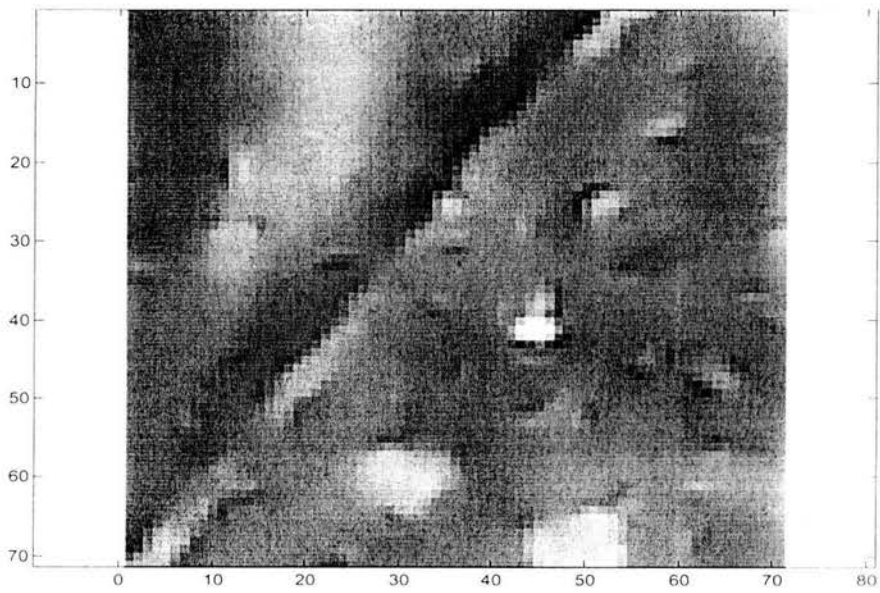


Figura 3.2: Sección de prueba no. 1

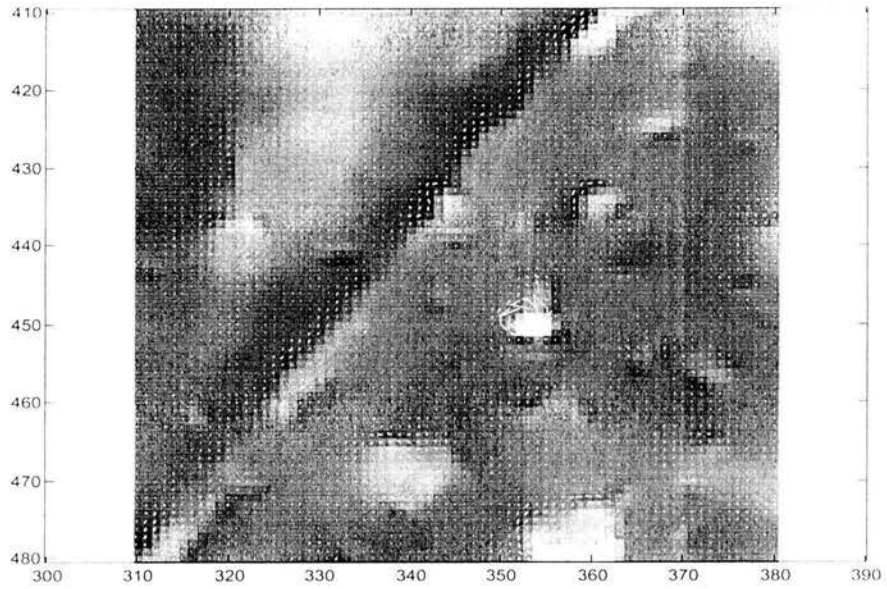


Figura 3.3: Campo de vectores de energía correspondiente a la sección 1, escala menor $s=0$

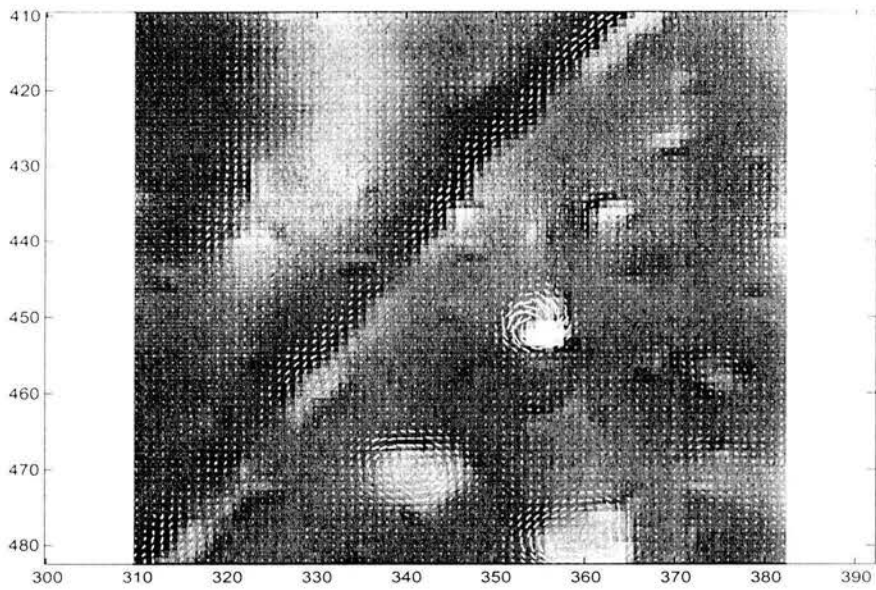


Figura 3.4: Campo de vectores de energía correspondiente a la sección 1, escala media $s=1$

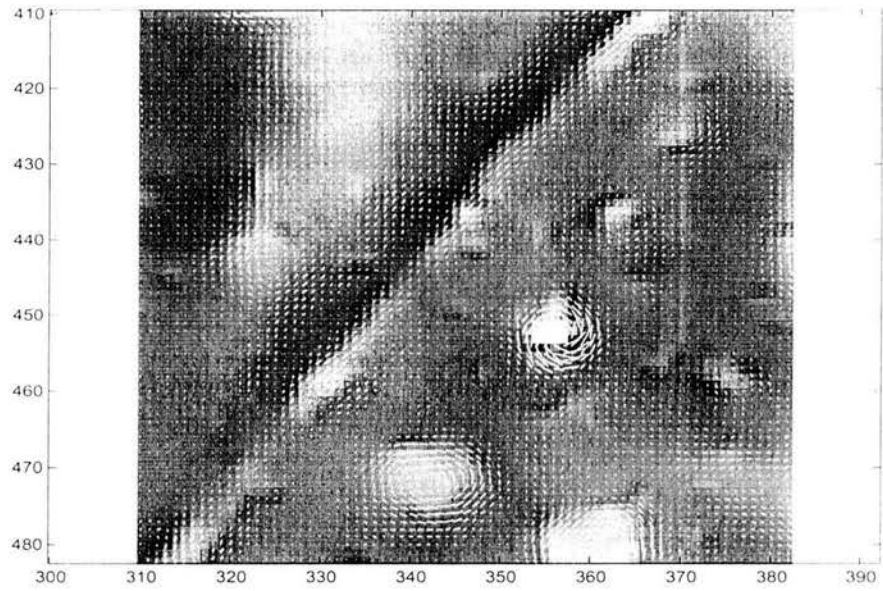


Figura 3.5: Campo de vectores de energía correspondiente a la sección 1, escala mayor s^{-2}

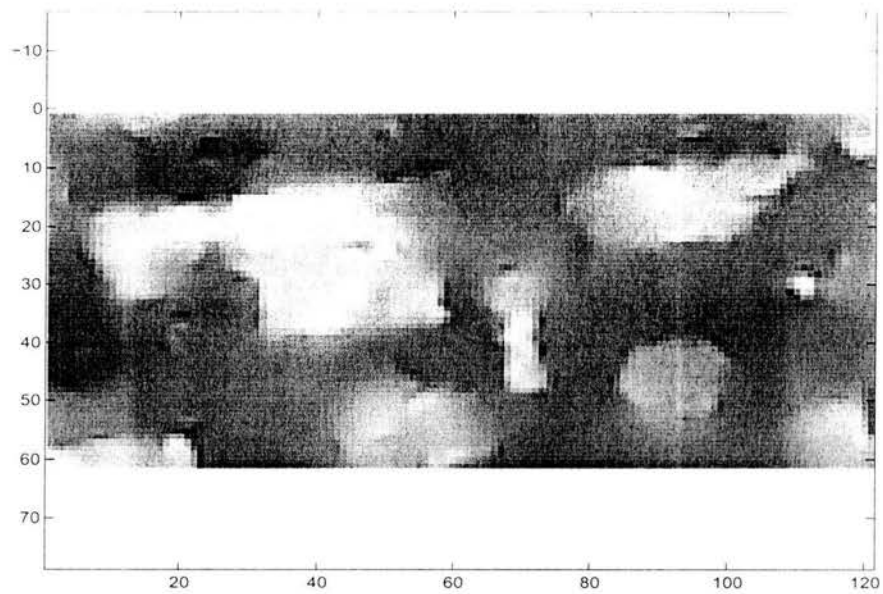


Figura 3.6: Sección de prueba no. 2

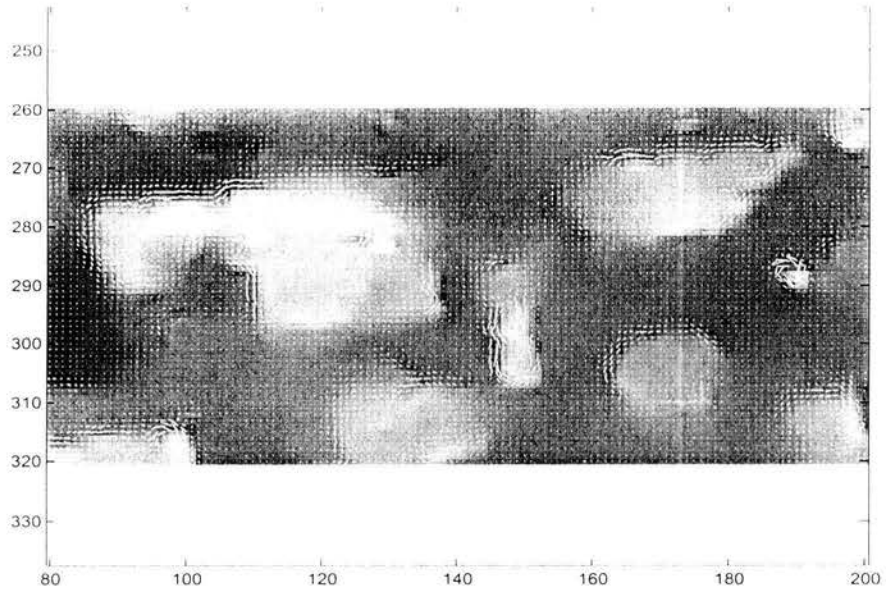


Figura 3.7: Campo de vectores de energía correspondiente a la sección 2, escala menor $s=0$

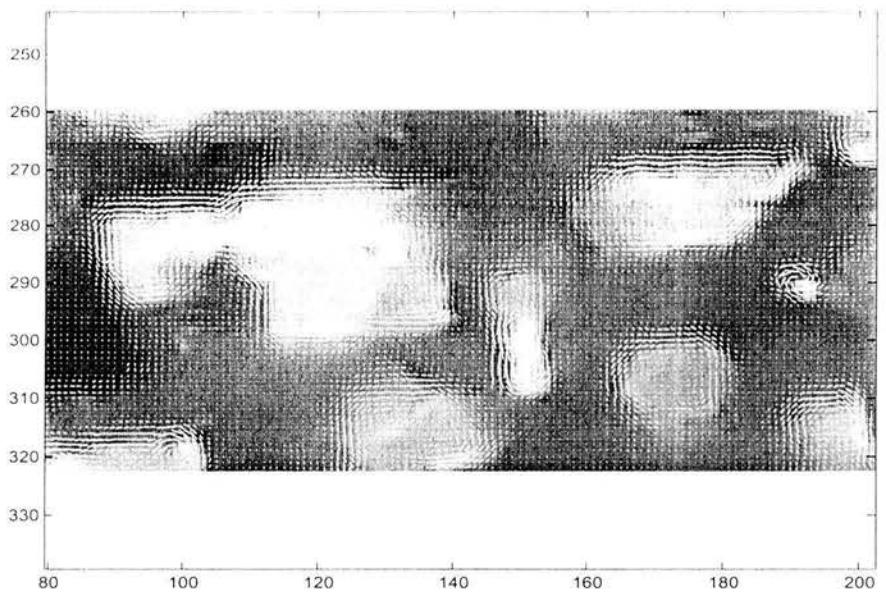


Figura 3.8: Campo de vectores de energía correspondiente a la sección 2, escala media $s=1$

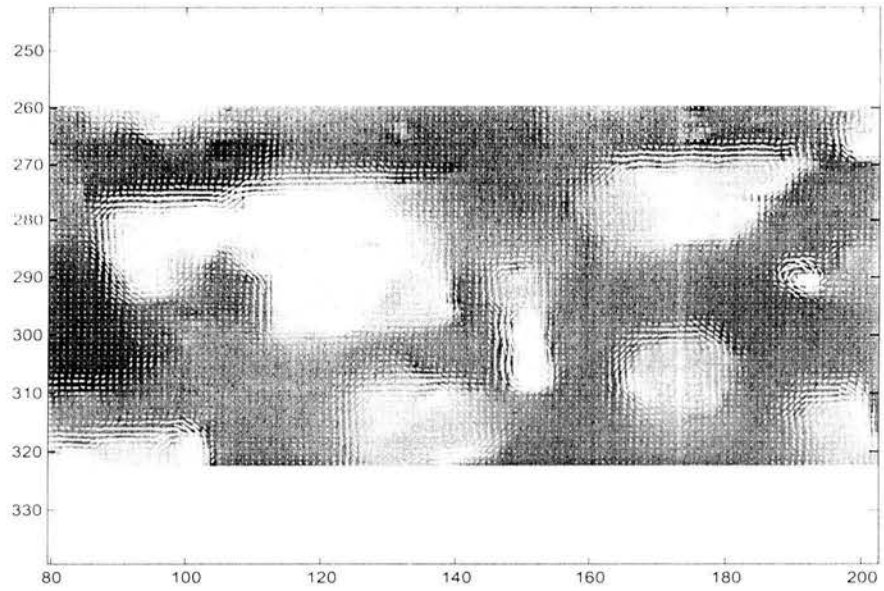


Figura 3.9: Campo de vectores de energía correspondiente a la sección 2, escala mayor $s=2$

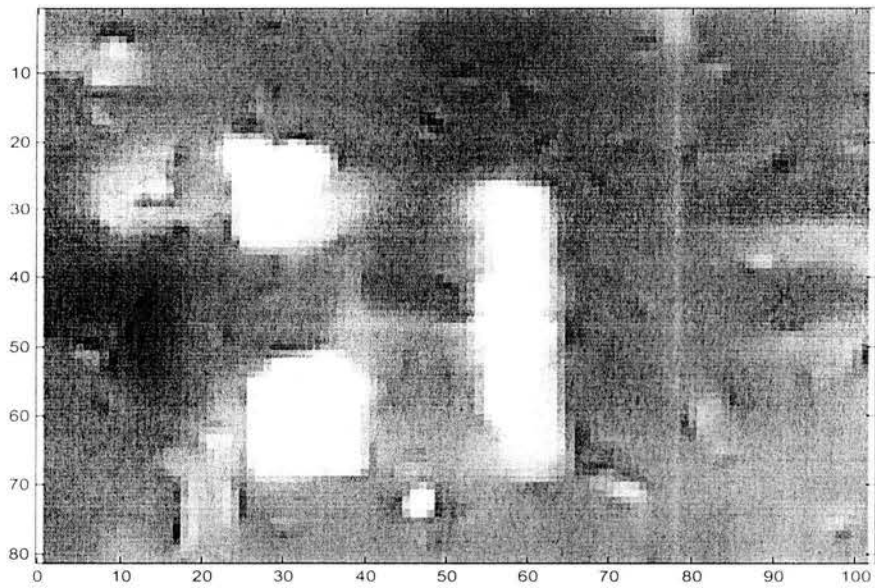


Figura 3.10: Sección de prueba no. 3

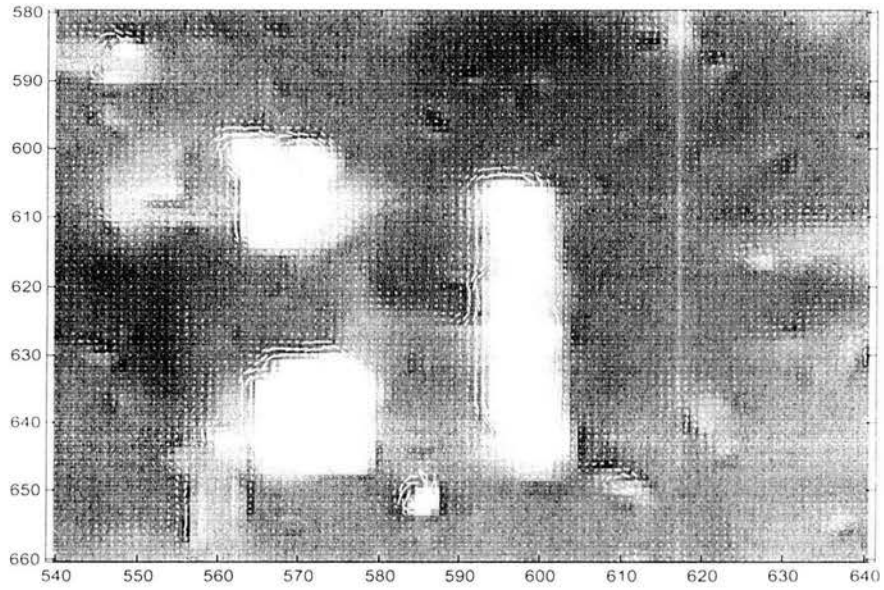


Figura 3.11: Campo de vectores de energía correspondiente a la sección 3, escala menor $s=0$

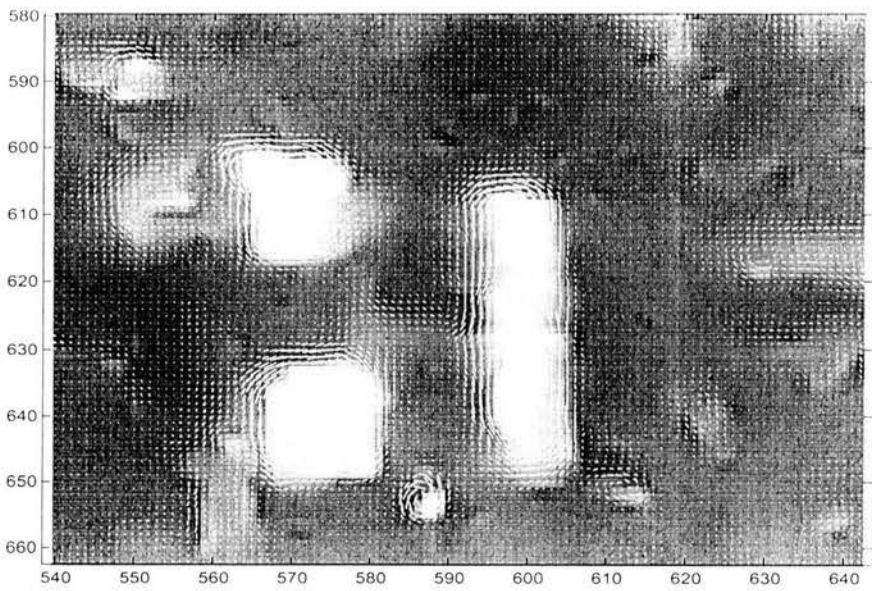


Figura 3.12: Campo de vectores de energía correspondiente a la sección 3, escala media $s=-1$

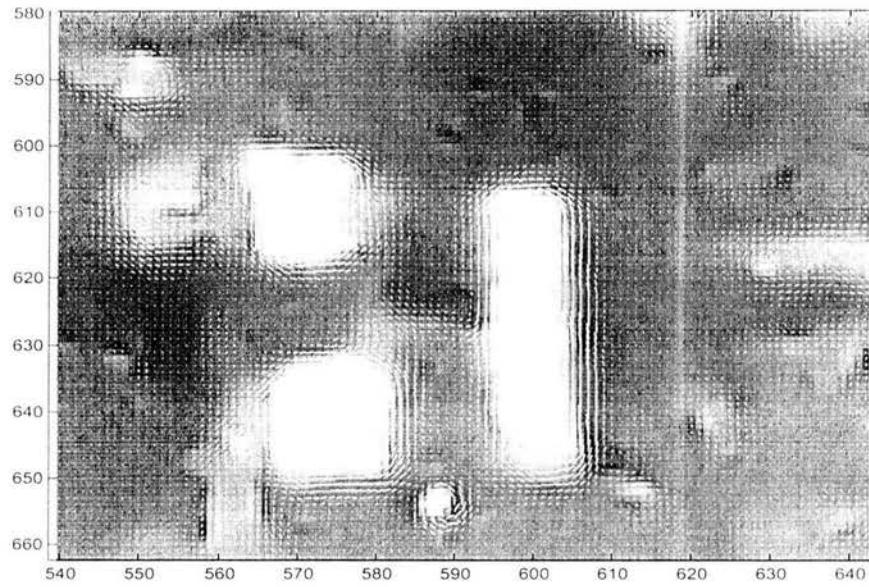


Figura 3.13: Campo de vectores de energía correspondiente a la sección 3, escala mayor $s=2$

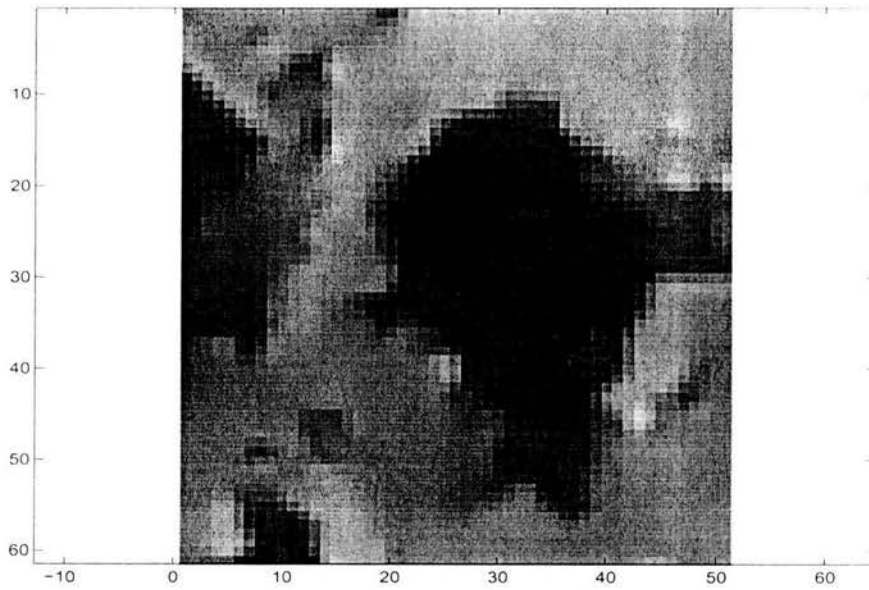


Figura 3.14: Sección de prueba no. 4

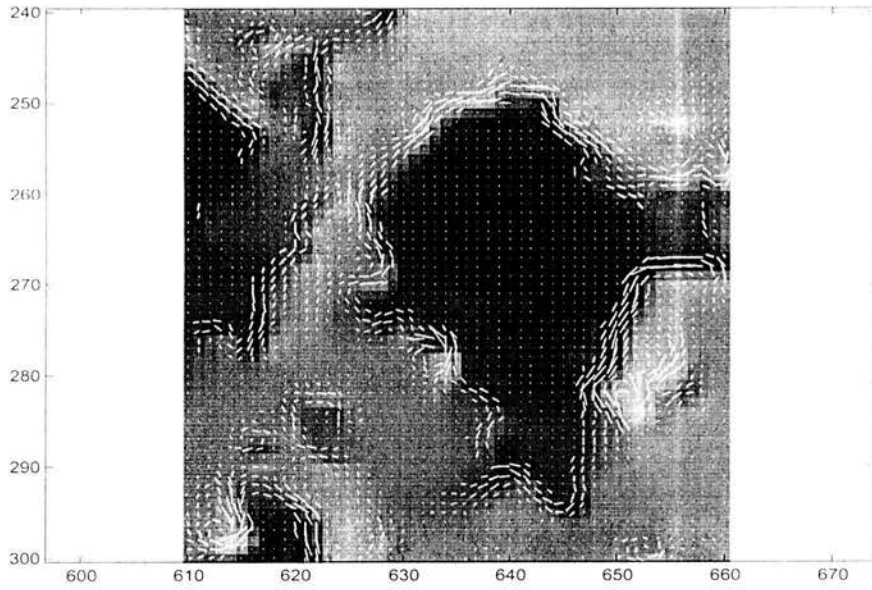


Figura 3.15: Campo de vectores de energía correspondiente a la sección 4, escala menor $s=0$

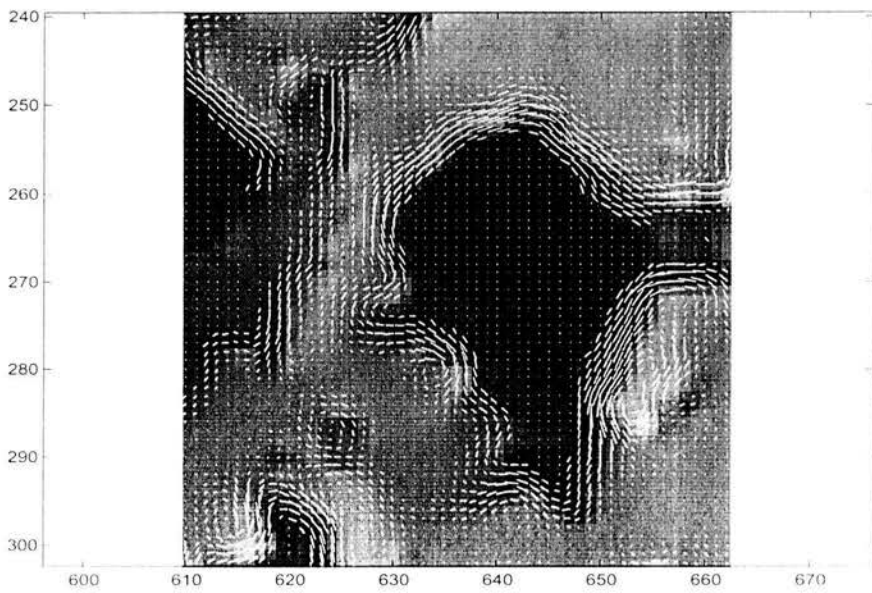


Figura 3.16: Campo de vectores de energía correspondiente a la sección 4, escala media $s=1$

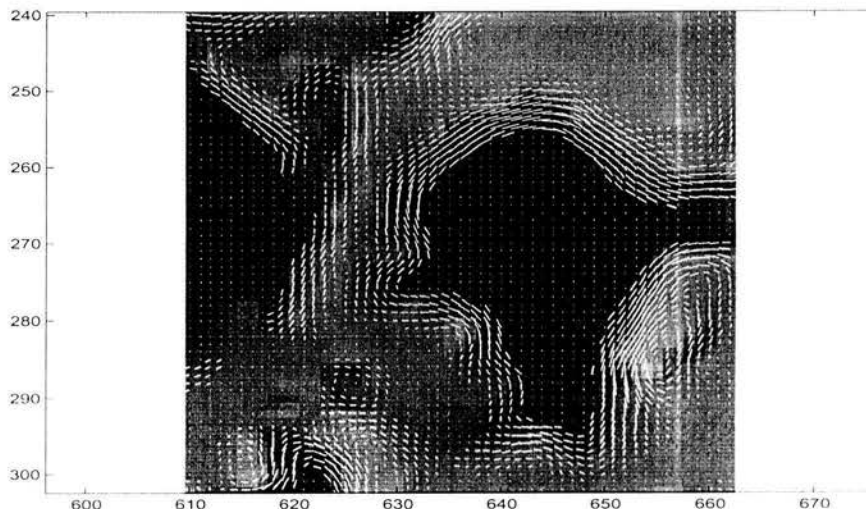


Figura 3.17: Campo de vectores de energía correspondiente a la sección 4, escala mayor $s=2$

3.3. Patrones 0D, 1D y 2D, [6]

Nuevamente bajo la premisa de representar a la señal, en este caso la imagen, en patrones que sean perceptualmente importantes, se propone la clasificación de todas las estructuras existentes en patrones denominados: 0D, 1D y 2D. Los patrones 0D, representan aquellas estructuras en la imagen que no cuentan con ninguna orientación, tal es el caso de superficies homogéneas que en imágenes SAR pueden representar grandes zonas de cultivo, explanadas, techos de grandes edificios, etc. Por otro lado los patrones 1D, corresponden a estructuras que cuentan con una sola orientación definida, como en el caso de bordes y líneas, que en imágenes SAR pueden representar grandes avenidas, cauces de ríos o bordes de grandes edificios. Finalmente, los patrones 2D representan estructuras donde está presente más de una dirección, como en esquinas o puntos con remanente de ruido en imágenes. A partir del método presentado en la sección ‘Clasificación de estructuras en el dominio de la THDR’ del capítulo 2, aplicado a una imagen SAR, es posible obtener una matriz, del mismo tamaño de la imagen original, donde a cada píxel de la imagen se le asocia una clase, 0D, 1D ó 2D, lograndose a partir de esto la clasificación. A continuación se presentan las clasificaciones realizadas para varias secciones o recortes de la imagen.

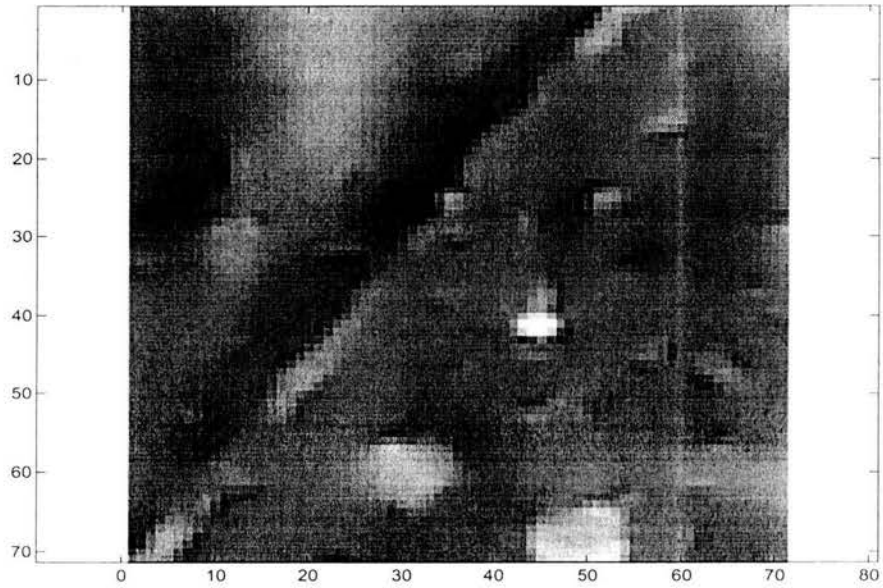


Figura 3.18: Sección de prueba no. 1

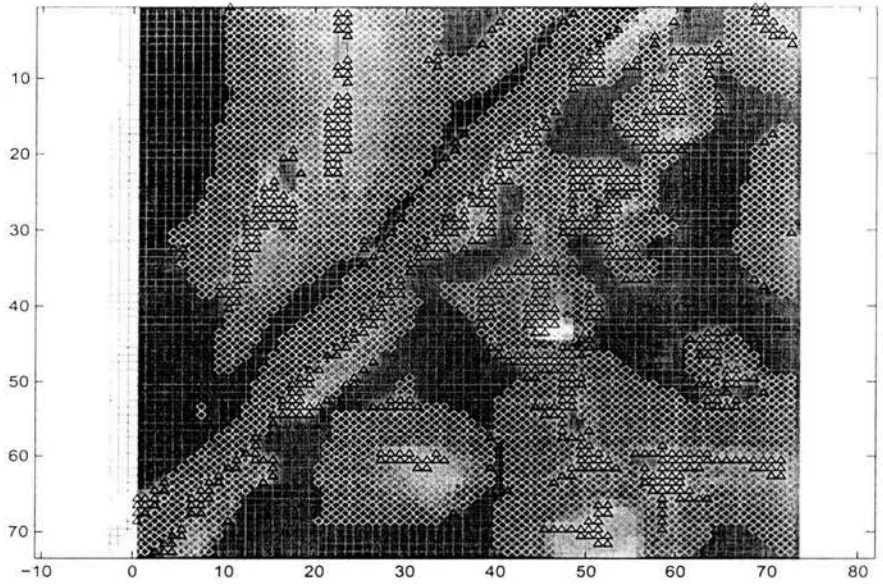


Figura 3.19: Patrones 0D +, 1D \diamond y 2D \triangle , para la sección 1, escala menor $s=0$

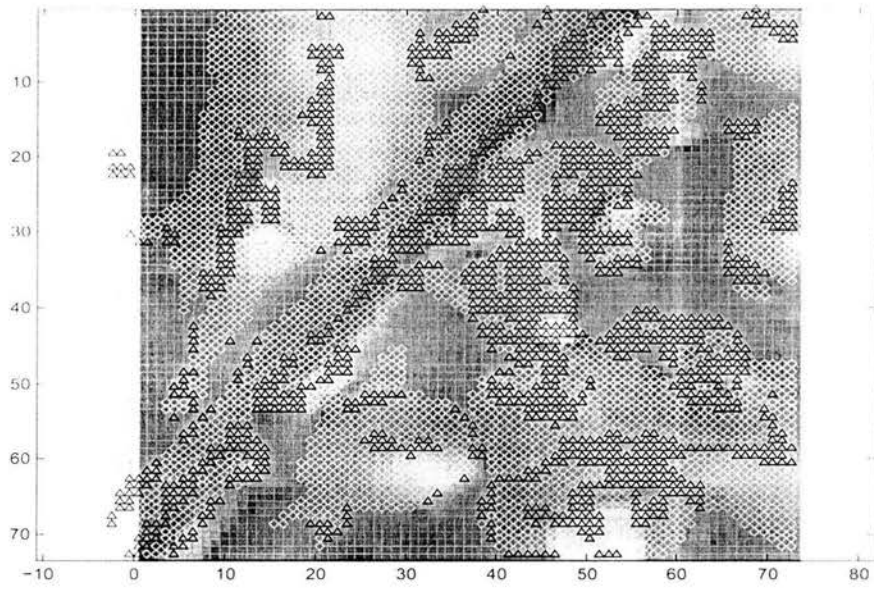


Figura 3.20: Patrones 0D +, 1D \diamond y 2D \triangle , para la sección 1, escala media $s=1$

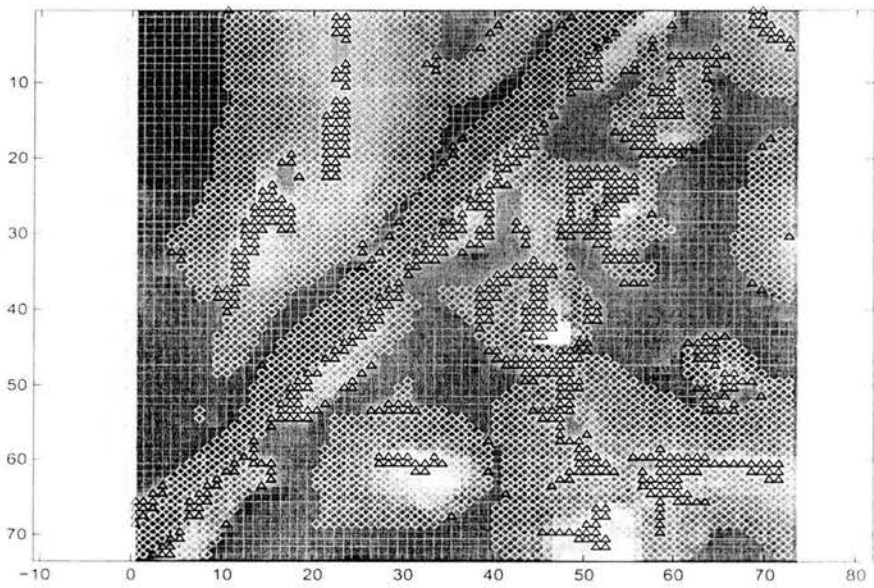


Figura 3.21: Patrones 0D +, 1D \diamond y 2D \triangle , para la sección 1, escala mayor $s=1$

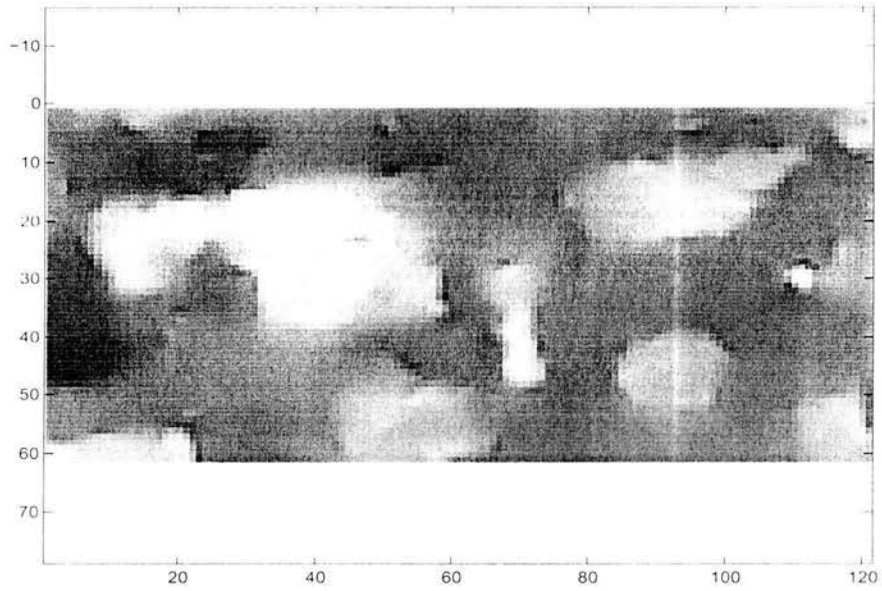


Figura 3.22: Sección de prueba no. 2

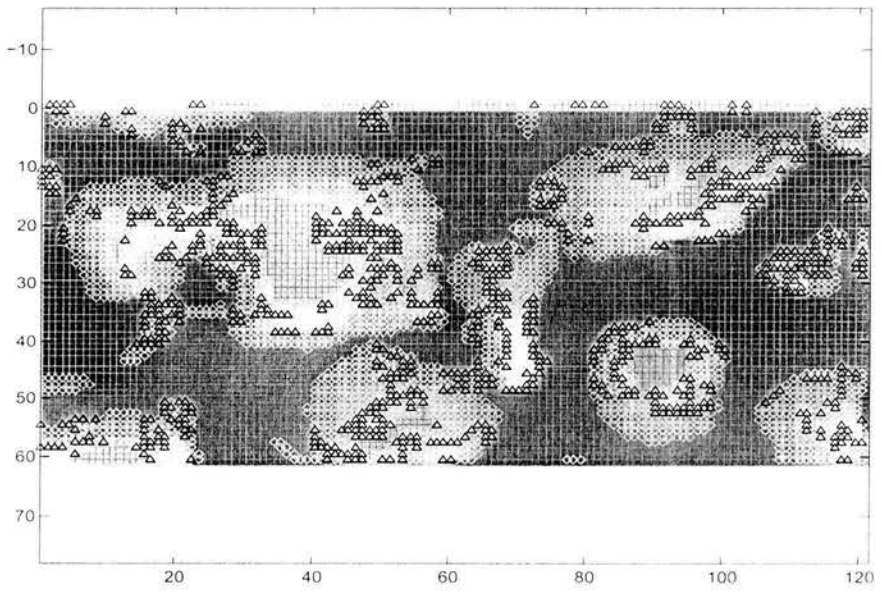


Figura 3.23: Patrones 0D +, 1D ◇ y 2D △, para la sección 2, escala menor $s=0$

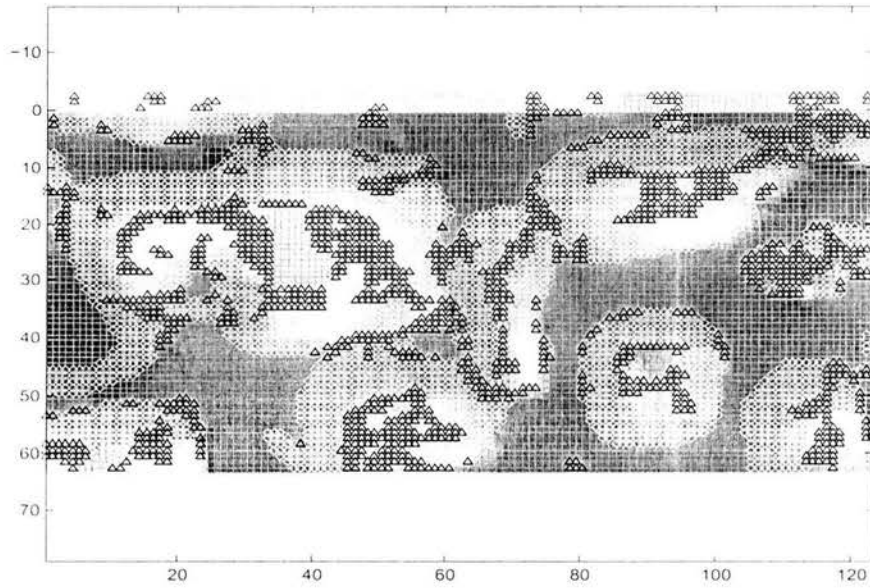


Figura 3.24: Patrones 0D $+$, 1D \diamond y 2D \triangle , para la sección 2, escala media $s=1$

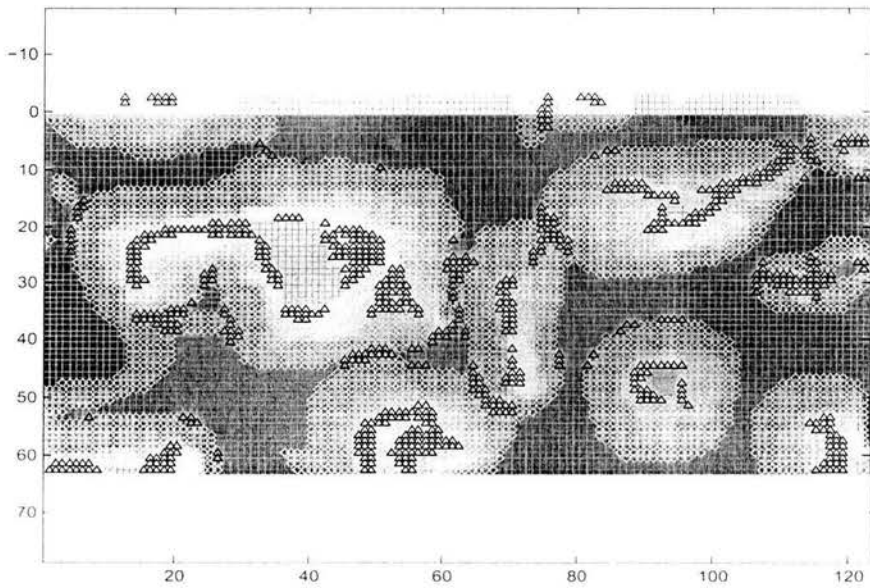


Figura 3.25: Patrones 0D $+$, 1D \diamond y 2D \triangle , para la sección 2, escala mayor $s=2$

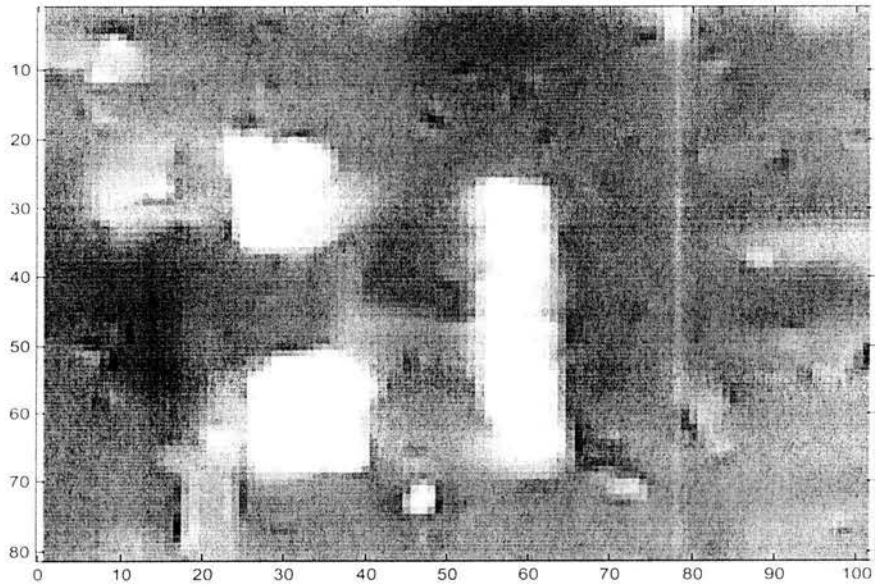


Figura 3.26: Sección de prueba no. 3

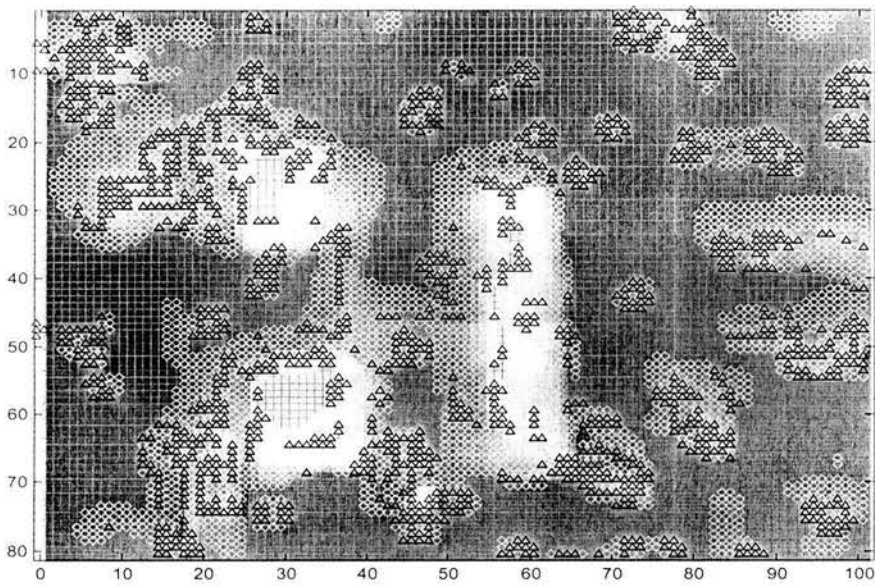


Figura 3.27: Patrones 0D +, 1D \diamond y 2D \triangle , para la sección 3, escala menor $s=0$

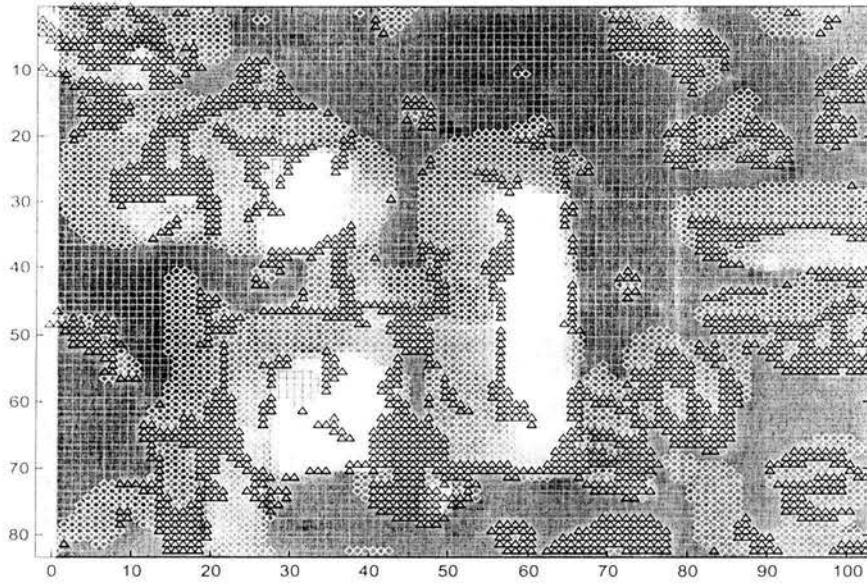


Figura 3.28: Patrones 0D $+$, 1D \diamond y 2D \triangle , para la sección 3, escala media $s=1$

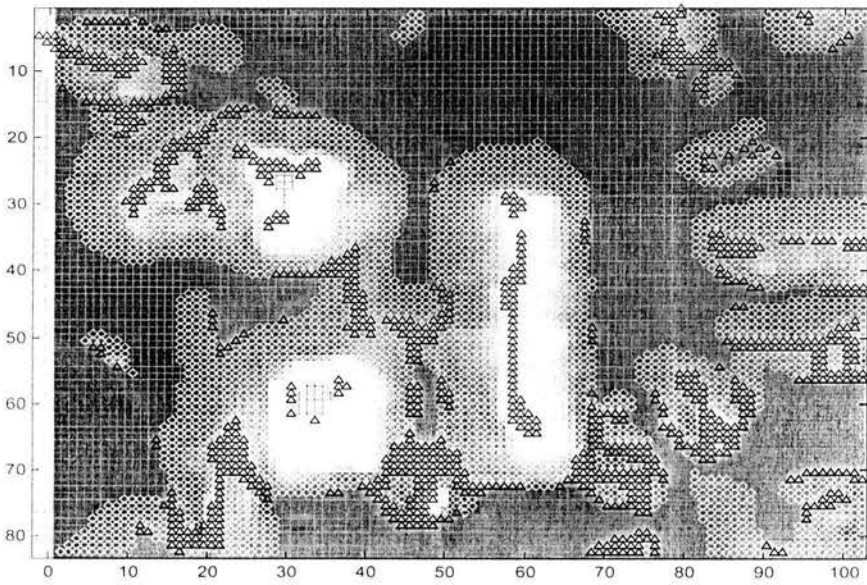


Figura 3.29: Patrones 0D $+$, 1D \diamond y 2D \triangle , para la sección 3, escala mayor $s=2$

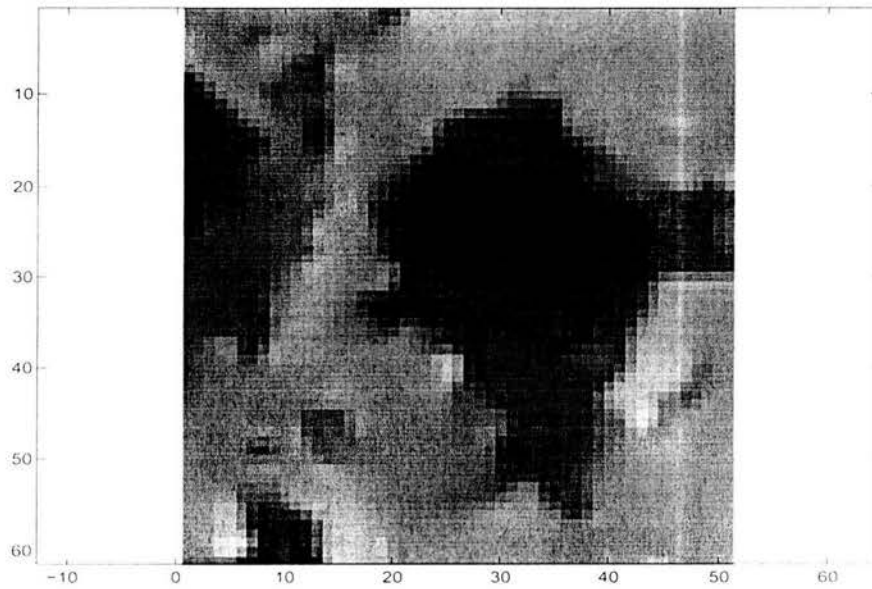


Figura 3.30: Sección de prueba no. 4

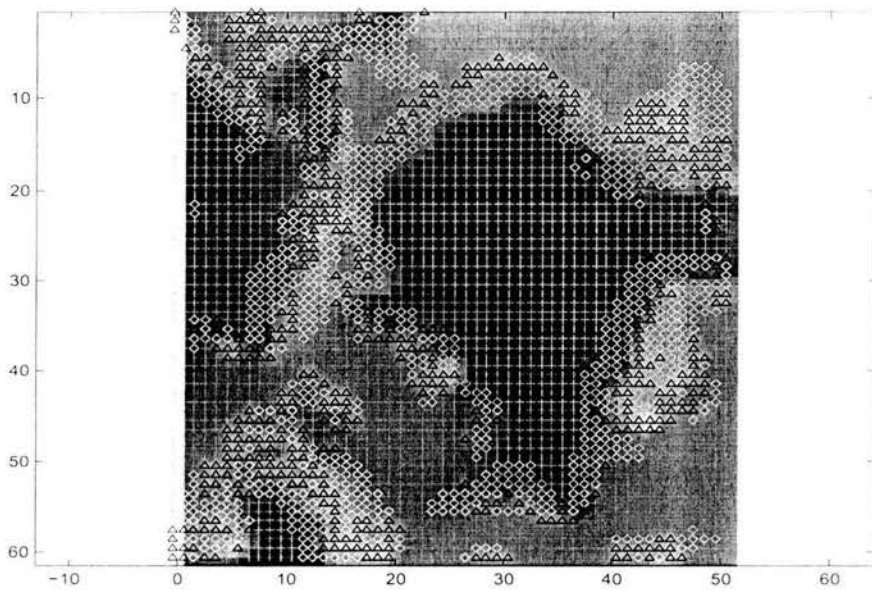


Figura 3.31: Patrones 0D $+$, 1D \diamond y 2D \triangle , para la sección 4, escala menor $s = 0$

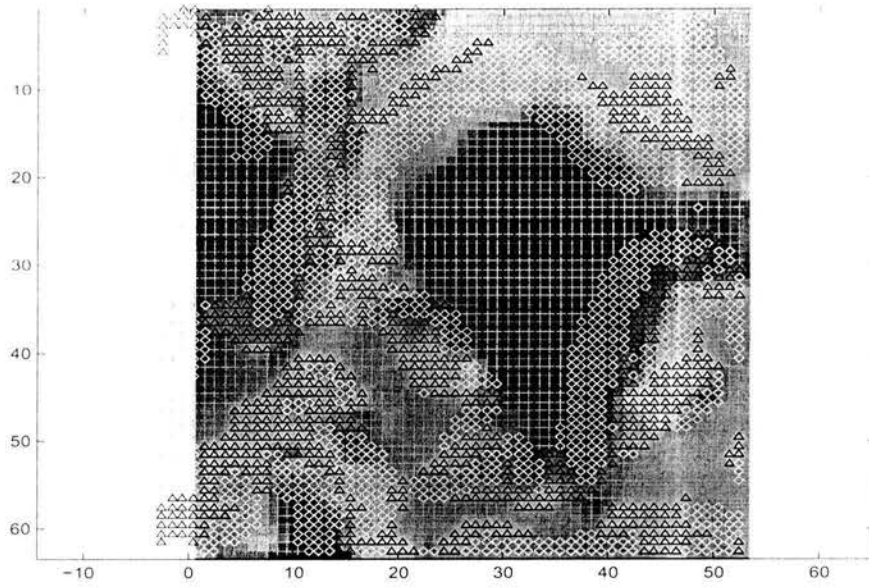


Figura 3.32: Patrones 0D +, 1D \diamond y 2D \triangle , para la sección 4, escala media $s=1$

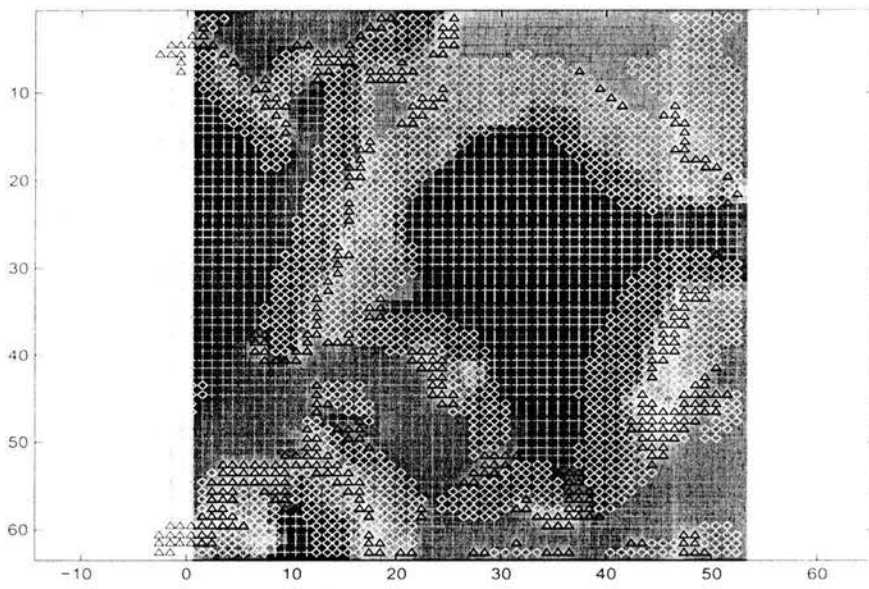


Figura 3.33: Patrones 0D +, 1D \diamond y 2D \triangle , para la sección 4, escala mayor $s=2$

Capítulo 4

Campos Aleatorios de Markov, MRF, [8]

4.1. Descripción

En el presente capítulo se abordarán los antecedentes teóricos necesarios para comprender la teoría de los Campos Aleatorios de Markov, la cual será empleada en el capítulo 5 para la regularización de los campos de vectores de energía obtenidos en el capítulo 3.

4.2. Introducción

La teoría de Campos Aleatorios de Markov, MRF (Markov Random Fields), provee una manera conveniente y consistente de modelar entidades dependientes del contexto como lo son píxeles de imágenes y otras características espacialmente correlacionadas. Lo anterior, se logra a través de la caracterización de influencias mutuas entre dichas entidades, utilizando probabilidades de los MRF. La teoría de MRF, expone cómo modelar la probabilidad a priori de patrones dependientes del contexto, como clases de texturas y disposición de características de objetos. Un modelo de MRF particular, favorece a su propia clase de patrones asociándolos con probabilidades más grandes que las de otras clases de patrones. La teoría de MRF suele utilizarse en conjunción con teorías de decisión estadística y estimación, para formular funciones objetivas en términos de principios de optimización establecidos.

El uso práctico de los MRF, se debe principalmente a la equivalencia, establecida por Hammersley y Clifford y desarrollada más tarde por Besag,

entre los MRF y las distribuciones de Gibbs. La equivalencia MRF-Gibbs, propone una fórmula explícita para las distribuciones conjuntas de los MRF. Esto permite modelar matemáticamente problemas de visión que pueden ser abordados por el análisis de imágenes, dentro del marco de la teoría Bayesiana. Desde la perspectiva computacional, la propiedad local de los MRF, permite que los algoritmos puedan implementarse de manera local y masivamente paralela. Además, la teoría de MRF, provee fundamentos para los cálculos a nivel multiresolución. Es por ello que los MRF han sido ampliamente empleados para resolver problemas de visión a varios niveles. La mayor parte de los modelos de MRF se utilizan en procesamiento de bajo nivel. Esto incluye restauración de imágenes, segmentación, reconstrucción de superficies, detección de bordes, análisis de texturas, flujo óptico y fusión de datos, entre otros. En años recientes, el uso de MRF en procesamiento de alto nivel ha emergido como en el caso del reconocimiento de objetos. El interés por el modelado de MRF en visión por computadora todavía está creciendo, como lo reflejan artículos de libros y revistas publicados en años recientes.

4.3. Etiquetado visual

Muchos problemas de visión pueden plantearse como problemas de etiquetado en los que la solución es un conjunto de etiquetas asignadas a pixeles o características de las imágenes. El etiquetado es una representación natural para el estudio de los MRF. Un problema de etiquetado, se especifica en términos de un conjunto de 'sites' y un conjunto de etiquetas. Sea S un conjunto discreto de m 'sites',

$$S = \{1, \dots, m\} \quad (4.1)$$

donde $1, \dots, m$ son índices. Un 'site' representa un punto o una región en el espacio Euclideo como un pixel o una característica de una imagen, esto es, un punto representando una esquina, un segmento de línea o el área de una superficie. Un conjunto de 'sites' puede ser caracterizado en términos de su regularidad. Los 'sites' de una 'lattice' se consideran espacialmente regulares. Una 'lattice' rectangular para una imagen 2D de tamaño $n \times n$, puede denotarse como

$$S = \{(i, j) | 1 \leq i, j \leq n\} \quad (4.2)$$

Sus elementos corresponden a los sitios en los que la imagen es muestreada. Los 'sites' que no presentan regularidad espacial se consideran como irregulares.

Los 'sites' en MRF, suelen considerarse no ordenados. Para una imagen de $n \times n$, el pixel (i, j) puede ser reindexado convenientemente por un número único k , donde k toma valores en $\{1, 2, \dots, m\}$ con $m = n \times n$. La interrelación entre 'sites' se basa en sistemas de vecindad. Una etiqueta es un evento que puede ocurrir a un 'site'. Sea \mathcal{L} el conjunto de etiquetas. Un conjunto de etiquetas puede ser caracterizado como continuo o discreto. En el caso continuo, el conjunto de etiquetas puede corresponder a la línea real \mathbb{R} o a un intervalo compacto de ella,

$$\mathcal{L}_c = [X_l, X_h] \subset \mathbb{R} \quad (4.3)$$

En el caso discreto, una etiqueta asume un valor discreto de un conjunto de M etiquetas

$$\mathcal{L}_d = \{l_1, \dots, l_M\} \quad (4.4)$$

o simplemente

$$\mathcal{L}_d = \{1, \dots, M\} \quad (4.5)$$

En detección de bordes, por ejemplo, el conjunto de etiquetas es $\mathcal{L} = \{\text{borde}, \text{noborde}\}$. Otra propiedad esencial del conjunto de etiquetas es el orden de éstas. Por ejemplo, los elementos del conjunto continuo de etiquetas \mathbb{R} , puede ordenarse por la relación "más chico que". Cuando un conjunto discreto como $\{0, \dots, 255\}$, representa el valor de intensidades cuantizadas, entonces se trata de un conjunto ordenado por que tenemos que para los valores de intensidad se cumple $0 < 1 < 2 < \dots < 255$. Cuando un conjunto representa 256 símbolos diferentes como tipos de texturas, entonces se considera como desordenado a menos que un orden artificial le sea impuesto. En un conjunto de etiquetas ordenado, puede definirse una medida numérica (cuantitativa) de similitud entre cualquier par de etiquetas, en cambio en un conjunto desordenado, la medida de similitud es simbólica (cualitativa) y generalmente toma los valores de 'igual' o 'desigual'. El problema de etiquetado consiste en asignar una etiqueta, de un conjunto de etiquetas \mathcal{L} , a cada uno de los 'sites' en \mathcal{S} . La detección de bordes en una imagen por ejemplo, consiste en asignar una etiqueta f_i , del conjunto de etiquetas $\mathcal{L} = \{\text{borde}, \text{noborde}\}$, al 'site' $i \in \mathcal{S}$, donde los elementos en \mathcal{S} indexan píxeles de una imagen. El conjunto

$$f = \{f_1, \dots, f_m\} \quad (4.6)$$

se conoce como el etiquetado de los 'sites' en \mathcal{S} en términos de las etiquetas en \mathcal{L} . Cuando a cada 'site' se asigna una etiqueta única, entonces $f_i = f(i)$ puede verse como una función con dominio \mathcal{S} e imagen \mathcal{L} . Como el soporte de la función es todo el dominio \mathcal{S} , se trata de un mapeo de \mathcal{S} a \mathcal{L} , esto es

$$f : \mathcal{S} \longrightarrow \mathcal{L} \quad (4.7)$$

En términos de MRF, un etiquetado se conoce como una configuración. En visión, una configuración o etiquetado puede corresponder a una imagen, un mapa de bordes, un interpretación de características de una imagen en términos de características de sus objetos, o una transformación. Cuando todos los 'sites' tienen el mismo conjunto de etiquetas \mathcal{L} , el conjunto de todos los posibles etiquetados, esto es, el espacio de configuración, es el siguiente producto Cartesiano

$$\mathbb{F} = \underbrace{\mathcal{L} \times \mathcal{L} \dots \times \mathcal{L}}_{m \text{ times}} = \mathcal{L}^m \quad (4.8)$$

donde m es el tamaño de \mathcal{S} . En restauración de imágenes, por ejemplo, \mathcal{L} contiene valores permitidos para los pixels, los cuales son comunes a todos los 'sites' en \mathcal{S} y \mathbb{F} define todas las imágenes permitidas. Cuando $\mathcal{L} = \mathbb{R}$ es la línea de los reales, entonces $\mathbb{F} = \mathbb{R}^m$ es el espacio real m dimensional. Cuando \mathcal{L} es un conjunto discreto, el tamaño de \mathbb{F} es combinatorio. Para un problema con m 'sites' y M etiquetas, por ejemplo, existe un total de M^m configuraciones posibles en \mathbb{F} . En términos de regularidad y continuidad, es posible clasificar un problema de etiquetado en visión dentro de una de las siguientes categorías.

LP1: 'sites' regulares con etiquetas continuas (Restauración de imágenes con valores continuos en píxeles)

LP2: 'sites' regulares con etiquetas discretas (Segmentación de regiones)

LP3: 'sites' irregulares con etiquetas discretas (Reconocimiento de objetos basado en características)

LP4: 'sites' irregulares con etiquetas continuas (Estimación de posición a partir de un conjunto de correspondencias entre puntos)

El uso de información contextual, indispensable en el entendimiento de imágenes y utilizado en análisis y reconocimiento de patrones, se debe en sus inicios a Chow y Abend. Más tarde Fu, Yu y Rosenfeld desarrollaron modelos para clasificar imágenes basados en características contextuales. En términos de probabilidad, las restricciones contextuales pueden expresarse localmente en términos de probabilidades condicionales $P(f_i | \{f_{i'}\})$, donde $\{f_{i'}\}$

denota el conjunto de etiquetas en los ‘sites’ $i' \neq i$, o globalmente como la probabilidad conjunta $P(f)$. Como la información local puede ser observada directamente, la inferencia global se hace generalmente basándose en propiedades locales. En situaciones donde las etiquetas son independientes unas de otras (no existe contexto), la probabilidad conjunta es el producto de las probabilidades locales

$$P(f) = \prod_{i \in S} P(f_i) \quad (4.9)$$

Lo anterior implica independencia condicional

$$P(f_i | \{f_{i'}\}) = P(f_i) \quad i' \neq i \quad (4.10)$$

Por lo tanto, el etiquetado global f , puede calcularse considerando localmente cada etiqueta f_i . Esta es una ventaja en la resolución de problemas. Cuando existe contexto, las etiquetas son mutuamente dependientes y las relaciones expresadas en las ecuaciones 4.9 y 4.10 ya no se cumplen. De esta manera, efectuar una inferencia global utilizando información local se convierte en una tarea no trivial. La teoría de Campos Aleatorios de Markov provee los fundamentos matemáticos para resolver este problema.

4.4. MRF y Distribuciones de Gibbs

La teoría de MRF es una rama de la teoría de la probabilidad utilizada para analizar las dependencias espaciales o contextuales de un fenómeno físico. Se emplea en etiquetado visual para establecer las distribuciones probabilísticas de las etiquetas que interactúan. Los ‘sites’ en S están relacionados unos con otros a través de un sistema de vecindad. Un sistema de vecindad en S se define como

$$\mathcal{N} = \{\mathcal{N}_i \mid \forall i \in S\} \quad (4.11)$$

donde \mathcal{N}_i es el conjunto de ‘sites’ vecinos de i . Las relaciones de vecindad tienen las siguientes propiedades:

- (1) Un ‘site’ no es vecino de él mismo: $i \notin \mathcal{N}_i$
- (2) la relación de vecindad es mutua: $i \in \mathcal{N}_{i'} \iff i' \in \mathcal{N}_i$

Para una ‘lattice’ regular S , el conjunto de vecinos de i se define como el conjunto de ‘sites’ cercanos dentro de un radio r

$$\mathcal{N}_i = \{i' \in S \mid [\text{dist}(\text{pixel}_{i'}, \text{pixel}_i)]^2 \leq r, i' \neq i\} \quad (4.12)$$

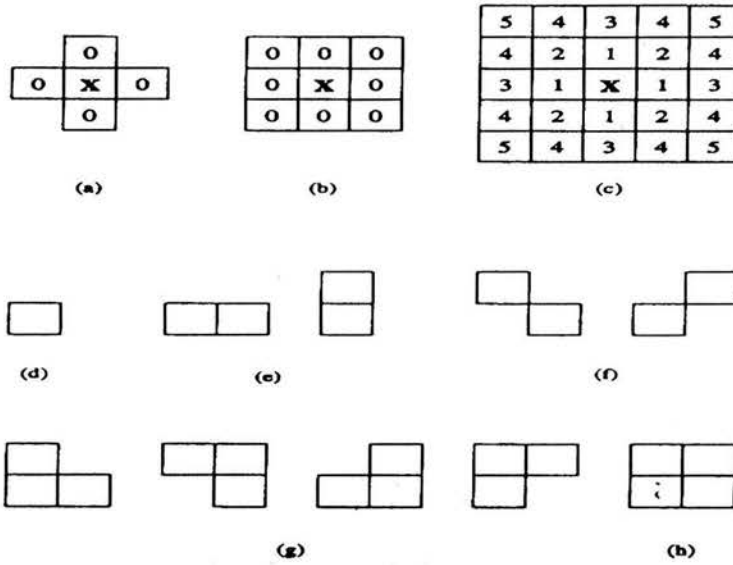


Figura 4.1: Vecindades y cliques en una lattice de sites regulares

donde $\text{dist}(A,B)$ denota la distancia Euclidiana entre A y B y r toma valores enteros. En el sistema de vecindad de primer orden, también conocido como sistema 4 conexidad, cada 'site' (interior) cuenta con 4 vecinos, como se muestra en la figura 4.1 a) donde x denota el 'site' considerado y 0 's sus vecinos. En el sistema de vecindad segundo orden, también conocido como sistema 8 conexidad, existen 8 vecinos por cada 'site' (interior), tal como se muestra en la figura 4.1 b). Los números $n = 1, \dots, 5$ mostrados en la figura 4.1 c), indican los 'sites' más exteriores de un sistema de vecindad de n -ésimo orden. Cuando el orden de los elementos en S está especificado, el conjunto de vecinos puede determinarse de manera más explícita. Por ejemplo, cuando $S = \{1, \dots, m\}$ es un conjunto ordenado de 'sites', y sus elementos indexan los pixeles de un arreglo 1D, un 'site' interior $i \in \{2, \dots, m-1\}$ tiene dos vecinos más cercanos, $\mathcal{N}_i = \{i-1, i+1\}$, y un 'site' en la frontera (primero y último, para este caso) tiene uno, $\mathcal{N}_1 = \{2\}$ y $\mathcal{N}_m = \{m-1\}$. Cuando los 'sites' en una 'lattice' regular y rectangular $S = \{(i,j) | 1 \leq i, j \leq n\}$, corresponden a los pixeles de una imagen de $n \times n$ en el plano 2D, un 'site' interno (i,j) tiene cuatro vecinos más cercanos $\mathcal{N}_{i,j} = \{(i-1,j), (i+1,j), (i,j-1), (i,j+1)\}$, un 'site' en las fronteras tiene tres vecinos y un 'site' en las esquinas tiene 2. El par (S, \mathcal{N}) constituyen

grafo donde S contiene los nodos y \mathcal{N} determina las ligas entre los nodos de acuerdo a la relación entre vecinos. Un clique c para (S, \mathcal{N}) se define como un subconjunto de ‘sites’ en S , y consiste tanto en un único ‘site’ $c = \{i\}$, como en un par de ‘sites’ vecinos $c = \{i, i'\}$, o una tripleta de ‘sites’ vecinos $c = \{i, i', i''\}$ y así sucesivamente. Las colecciones de cliques de un solo ‘site’, de un par de ‘sites’, de una tripleta de ‘sites’, se denotarán como c_1 , c_2 y c_3 , respectivamente, donde

$$C_1 = \{i \mid i \in \mathcal{S}\} \quad (4.13)$$

$$C_2 = \{\{i, i'\} \mid i' \in \mathcal{N}_i, i \in \mathcal{S}\} \quad (4.14)$$

y

$$C_3 = \{\{i, i', i''\} \mid i, i', i'', i \in \mathcal{S} \text{ son vecinos unos de otros}\} \quad (4.15)$$

Note que los ‘sites’ en un clique están ordenados, esto es, $\{i, i'\}$ no es el mismo clique que $\{i', i\}$. La colección de todos los cliques para (S, \mathcal{N}) es

$$C = C_1 \cup C_2 \cup C_3 \dots \quad (4.16)$$

donde \dots denota conjuntos de cliques mayores. El tipo de clique para (S, \mathcal{N}) de una ‘lattice’ regular está determinado por su tamaño, forma y orientación. La figura 4.1 d)-h), muestra tipos de clique para sistemas de primer y segundo orden de una ‘lattice’. Los cliques de un solo ‘site’, el par horizontal y el par vertical en (d) y (e), corresponden al sistema de vecindad de primer orden (a). El tipo de cliques para en sistema de vecindad de segundo orden (b), incluye no sólo a los que se muestran en (d) y (e) sino también a los cliques con pares de ‘sites’ en diagonal (f), los cliques con tripletas de ‘sites’ (g) y aquellos con cuadruplas de ‘sites’ (h). De esta manera, mientras el orden de los sistemas de vecindad se incrementa, el número de cliques crece rápidamente lo que involucra mayor gasto computacional.

4.5. MRF

Sea $F = \{F_1, \dots, F_m\}$ una familia de variables aleatorias definida en el conjunto S , en la que cada variable aleatoria F_i toma un valor f_i en \mathcal{L} . La familia F se conoce como campo aleatorio. Se utiliza la notación $F_i = f_i$ para denotar el evento en que F_i toma el valor f_i y la notación $(F_1 = f_1, \dots, F_m = f_m)$ para denotar el evento conjunto. Por simplicidad un evento conjunto se abrevia como $F = f$ donde $f = \{f_1, \dots, f_m\}$ es una configuración de F , lo cual corresponde a la realización del campo. Para un conjunto de etiquetas discreto \mathcal{L} , la probabilidad de que la variable aleatoria F_i tome el valor f_i , se

denota $P(F_i = f_i)$ o de manera abreviada $P(f_i)$, amenos que haya necesidad de utilizar notación más elaborada, y probabilidad conjunta denotada por $P(F = f) = P(F_1 = f_1, \dots, F_m = f_m)$ se expresa de manera abreviada como $P(f)$. Se dice que F es un Campo Aleatorio de Markov en S con respecto a un sistema de vecindad \mathcal{N} sí y sólo sí satisface las siguientes dos condiciones:

$$P(f) > 0, \quad \forall f \in \mathbb{F} \quad (\text{positividad}) \quad (4.17)$$

$$P(f | f_{S-\{i\}}) = P(f_i | f_{\mathcal{N}_i}) \quad (\text{Markovianidad}) \quad (4.18)$$

donde $S - \{i\}$ es el conjunto diferencia, $f_{S-\{i\}}$ denota el conjunto de etiquetas en los ‘sites’ de $S - \{i\}$ y

$$f_{\mathcal{N}_i} = \{f_{i'} | i' \in \mathcal{N}_i\} \quad (4.19)$$

es válida para el conjunto de etiquetas en los ‘sites’ vecinos a i . La positividad se asume por razones técnicas y puede satisfacerse en la práctica. Por ejemplo, cuando la condición de positividad se satisface, la probabilidad conjunta $P(f)$ de cualquier campo aleatorio está determinada solamente por las probabilidades condicionales locales. La markovianidad representa las características locales de F . La etiqueta en el ‘site’ depende solo de aquellas en los ‘sites’ vecinos. En otras palabras, solo las etiquetas vecinas tienen interacciones directas entre ellas. Existen dos aproximaciones para especificar un MRF, aquél en términos de probabilidades condicionales $P(f_i | f_{\mathcal{N}_i})$, y el que se define en términos de la probabilidad conjunta $P(f)$. Besag habla a favor de la probabilidad conjunta en viat de las desventajas de la aproximación vía la probabilidad condicional: Primero, no existen métodos disponibles para deducir la probabilidad conjunta a partir de las probabilidades condicionales. Segundo, las probabilidades condicionales por sí solas están sujetas a condiciones de consistencia altamente restrictivas. Tercero, la especificación natural del equilibrio de un proceso estadístico es en términos de la probabilidad conjunta y no de la distribución condicional de variables. Afortunadamente, un resultado teórico acerca de la equivalencia entre los MRF y la distribución de Gibbs provee los medios matemáticos para especificar la probabilidad conjunta de un MRF.

4.6. Campos Aleatorios de Gibbs (GRF, Gibbs Random Fields)

Se dice que un conjunto de variables aleatorias F es un GRF en S con respecto a \mathcal{N} , sí y sólo sí, sus configuraciones obedecen a distribuciones de

Gibbs. Una distribución de Gibbs toma la siguiente forma

$$P(f) = Z^{-1} \times e^{-\frac{1}{T}U(f)} \quad (4.20)$$

donde

$$Z = \sum_{f \in \mathbb{F}} e^{-\frac{1}{T}U(f)} \quad (4.21)$$

es una constante de normalización llamada función partición, T es una constante de temperatura generalmente con valor igual a 1 y $U(f)$ es la función de energía. La energía

$$U(f) = \sum_{c \in \mathcal{C}} V_c(f) \quad (4.22)$$

es la suma de potenciales de cliques $V_c(f)$ sobre todos los cliques posibles \mathcal{C} . El valor de $V_c(f)$ depende de la configuración local en el clique c . Obviamente, la distribución Gaussiana es un miembro especial de esta familia de distribuciones de Gibbs. Se dice que un GRF es homogéneo si $V_c(f)$ es independiente de la posición relativa del clique c en S . Se dice que es isotrópico si V_c es independiente de la orientación de c . Para calcular una distribución de Gibbs, es necesario evaluar la función de partición Z , que es la suma sobre todas las configuraciones posibles en \mathbb{F} . Como el número de elementos en \mathbb{F} para un \mathcal{L} discreto es combinatorio, la evaluación de Z es imposible aún para problemas de tamaños moderados. Para resolver este problema existen diversos métodos de aproximación que más adelante serán abordados. $P(f)$ mide la probabilidad de ocurrencia de una configuración o patrón particular f . Para configuraciones más probables se tienen menores energías. La temperatura T controla la forma de la distribución. Cuando la temperatura es alta, todas las configuraciones tienden a estar igualmente distribuidas. Cerca de la temperatura cero, la distribución se concentra alrededor de la energía mínima global. Dadas T y $U(f)$, es posible generar una clase de patrones, muestreando el espacio de configuración \mathbb{F} de acuerdo con $P(f)$. Para problemas de etiquetado discreto, el potencial del clique $V_c(f)$, puede especificarse por un número de parámetros. Por ejemplo, sea $f_c = (f_i, f_{i'}, f_{i''})$ la configuración local de un clique-triple $c = \{i, i', i''\}$, f_c toma un número finito de estados y por lo tanto $V_c(f)$ toma un número finito de valores. En ocasiones resulta conveniente expresar la energía de una distribución de Gibbs como la suma de varios términos, cada uno atribuido a los cliques de cierto tamaño, esto es,

$$U(f) = \sum_{\{i\} \in \mathcal{C}_1} V_1(f_i) + \sum_{\{i, i'\} \in \mathcal{C}_2} V_2(f_i, f_{i'}) + \sum_{\{i, i', i''\} \in \mathcal{C}_3} V_3(f_i, f_{i'}, f_{i''}) + \dots \quad (4.23)$$

Lo anterior implica una distribución de Gibbs homogénea por que V_1, V_2 y V_3 son independientes de la localización de i, i' e i'' . Un caso especial e importante es cuando solamente los cliques de tamaño menor o igual a dos se consideran. En este caso, la energía puede escribirse también como

$$U(f) = \sum_{i \in S} V_1(f_i) + \sum_{i \in S} \sum_{i' \in \mathcal{N}_i} V_2(f_i, f_{i'}) \quad (4.24)$$

Note que en el segundo término de la anterior igualdad $\{i, i'\}$ y $\{i', i\}$ son cliques distintos en C_2 por que los 'sites' de un clique están ordenados. La probabilidad condicional se puede escribir entonces como

$$P(f_i | f_{\mathcal{N}_i}) = \frac{e^{-[V_1(f_i) + \sum_{i' \in \mathcal{N}_i} V_2(f_i, f_{i'})]}}{\sum_{f_i \in \mathcal{L}} e^{-[V_1(f_i) + \sum_{i' \in \mathcal{N}_i} V_2(f_i, f_{i'})]}} \quad (4.25)$$

4.7. Equivalencia Markov-Gibbs

Un MRF se caracteriza por su propiedad local (Markovianidad), mientras que un GRF se caracteriza por su propiedad global (distribución de Gibbs). El teorema de Hammersley-Clifford establece la equivalencia de estas dos propiedades. El teorema dice que F es un MRF en S con respecto a \mathcal{N} si y sólo si, F es un GRF en S con respecto a \mathcal{N} . El valor práctico de este teorema es que provee una manera sencilla de especificar la probabilidad conjunta. Uno puede especificar la probabilidad conjunta $P(F = f)$ especificando las funciones de potencial de cliques $V_c f$ y escogiendo dichas funciones dependiendo del comportamiento deseado del sistema. De esta manera el teorema codifica el conocimiento a priori, o la preferencia existente, en las interacciones entre etiquetas. Como escoger la forma y los parámetros de las funciones potencial para codificar adecuadamente las restricciones es un área muy importante en el modelado de MRF. La forma de las funciones potencial determina la forma de la distribución de Gibbs. Cuando la totalidad de los parámetros involucrados en las funciones potencial están especificados, la distribución de Gibbs está definida completamente. Para calcular la probabilidad conjunta de un MRF, el cual es una distribución de Gibbs, es necesario evaluar la función partición. Como se trata de la suma de un número combinatorio de configuraciones en \mathbb{F} , el cálculo es generalmente imposible. La evaluación explícita puede evitarse en modelos de MRF basados en la probabilidad máxima cuando los parámetros que contiene $U(f)$ se conocen en su totalidad.

4.8. Modelos de Markov útiles

Las restricciones contextuales entre dos etiquetas son las restricciones de menor orden para expresar información contextual. Sin embargo, son muy utilizadas debido a su forma simple y su bajo costo computacional. Están incluidas en la energía de Gibbs como potenciales de cliques entre pares de 'sites'. Con potenciales de cliques de hasta dos 'sites' la energía se representa como

$$U(f) = \sum_{i \in S} V_1(f_i) + \sum_{i \in S} \sum_{i' \in \mathcal{N}_i} V_2(f_i, f_{i'}) \quad (4.26)$$

La expresión anterior es un caso especial de la ecuación 4.23, que se conoce como energía de segundo orden. Un GRF o MRF puede ser especificado con una selección apropiada de V_1 's y V_2 's. En el modelo logístico multinivel (multilevel logistic, MLL) o modelo MLL, existen $M (> 2)$ etiquetas discretas en el conjunto de etiquetas $\mathcal{L} = \{1, \dots, M\}$. El potencial de clique depende del tipo c (tamaño, forma y orientación posible) del clique y de la configuración local $f_c \triangleq \{f_i | i \in c\}$. Para cliques que contienen más de un 'site' ($\#c > 1$), los potenciales de cliques MLL están definidos por

$$V_c(f) = \begin{cases} \zeta_c & \text{si todos los 'sites' en } c \text{ tienen la misma etiqueta} \\ -\zeta_c & \text{de otro modo} \end{cases} \quad (4.27)$$

donde ζ_c es el potencial para cliques de tipo c ; para cliques de un solo 'site', el potencial depende de la etiqueta asignada al 'site'

$$V_c(f) = V_c(f_i) = \alpha_I \text{ si } f_i = I \in \mathcal{L}_d \quad (4.28)$$

donde α_I es el potencial para la etiqueta con valor I . Si asumimos un modelo MLL de segundo orden como el de la ecuación 4.26, donde sólo α (para cliques de un solo 'site') y β (para cliques de pares de 'sites') son diferentes de cero, la función potencial para cliques de pares de 'sites' se escribe como

$$V_2(f_i, f_{i'}) = \begin{cases} \beta_c & \text{si los 'sites' en el clique } \{i, i'\} = c \in C_2 \text{ tienen la misma etiqueta} \\ -\beta_c & \text{de otro modo} \end{cases} \quad (4.29)$$

donde β_c es el parámetro β para los cliques de tipo c y C_2 es el conjunto de cliques de pares de 'sites'. Una restricción contextual genérica de este mundo es la de la suavidad o lisura. A partir de la cual se asume que las propiedades físicas es un vecindario o espacio o en un intervalo de tiempo presentan alguna coherencia y generalmente no cambian abruptamente. Por ejemplo, la superficie de una mesa es plana y un evento temporal no cambia

abruptamente en un periodo pequeño. De hecho, siempre es posible encontrar regularidades de un fenómeno físico con respecto a ciertas propiedades. Es por ello que la suavidad es una de las hipótesis a priori más populares en visión de bajo nivel que ha sido desarrollada dentro de un marco general, llamado regularización, para una variedad de problemas de visión de bajo nivel. Las ecuaciones 4.27 y 4.29 del modelo MLL con ζ negativa y coeficientes β proporciona un método para construir términos de suavidad para etiquetas discretas sin orden. Cuando todas las etiquetas f_c de un clique c toman el mismo valor, lo cual significa que la solución f es localmente suave en c , se incurre en un potencial de clique o costo negativo; por otro lado, si las etiquetas no son todas las mismas, se incurre en un potencial positivo. Este tipo de modelo MLL, tiende entonces a generar una solución suave que prefiere etiquetas uniformes.

4.9. Vision basada en optimización

La optimización ha jugado un rol importante y esencial en visión por computadora. La mayoría de los problemas de visión se formulan como criterios de optimización, explícita o implícitamente. El uso extensivo de los principios de optimización se debe a la presencia de incertidumbre en los procesos de visión, como el ruido, la oclusión de la imagen sensada y ambigüedades en la interpretación visual. Las soluciones exactas o perfectas existen muy difícilmente. Soluciones inexactas pero óptimas (en algún sentido) se proponen en su lugar. Aunque existen controversias filosóficas y científicas acerca de la factibilidad de utilizar el criterio de Bayes para inferir o tomar decisiones, este criterio se encuentra entre los más populares en visión computacional y de hecho, MAP (maximum a posterior) es el criterio más popular de optimización de modelos basados en MRF's. El teorema de Hammersley-Clifford proporciona una manera conveniente para especificar la probabilidad conjunta a priori, resolviendo una de las dificultades del etiquetado MAP-MRF.

4.10. Etiquetado de MRF's por medio de Bayes

La estadística de Bayes es una teoría de importancia fundamental en estimación y toma de decisiones. De acuerdo con esta teoría cuando ambas, la distribución a priori y la función de probabilidad de un patrón, se conocen, el mejor estimador es el que etiqueta por medio de Bayes. La solución MAP es un caso especial del marco de Bayes utilizada en una gran cantidad de trabajos de visión. La teoría MAP-MRF se debe a Geman y Geman

principalmente, ya que desde que ellos publicaron su artículo se formularon numeroso problemas del área utilizando esta teoría para su solución.

4.11. Estimación de Bayes

En la estimación de Bayes el riesgo se minimiza para obtener la estimación óptima. El riesgo de un estimado, f^* , en la teoría de Bayes se define como

$$R(f^*) = \int_{f \in \mathbb{F}} C(f^*, f) P(f | d) df \quad (4.30)$$

donde d es la observación, $C(f^*, f)$ es una función de costo y $P(f | d)$ es la distribución a posteriori. Primero que nada, es necesario calcular la distribución a posteriori a partir de la distribución a priori y la probabilidad. De acuerdo con la regla de Bayes, la probabilidad a posteriori, puede calcularse utilizando la siguiente fórmula

$$P(f | d) = \frac{p(d | f)P(f)}{p(d)} \quad (4.31)$$

donde $P(f)$ es la probabilidad a priori de las etiquetas f , $p(d | f)$ es la función de densidad de probabilidad de las observaciones d , también conocida como función de probabilidad de f para d fija, y $p(d)$ es la densidad de d que es una constante cuando d se conoce. La función de costo $C(f^*, f)$ determina el costo del estimado f cuando el valor real es f^* . Se define de acuerdo a la preferencia del usuario. Dos opciones populares son la función de costo cuadrática

$$C(f^*, f) = \| f^* - f \|^2 \quad (4.32)$$

donde $\| a - b \|$ es la distancia entre a y b , y la función de costo $\delta(0-1)$,

$$C(f^*, f) = \begin{cases} 0 & \text{si } \| f^* - f \| \leq \delta \\ 1 & \text{cualquier otro caso} \end{cases} \quad (4.33)$$

donde $\delta > 0$ es cualquier constante pequeña. Para la función de costo δ el riesgo Bayes es

$$R(f^*) = \int_{f: \| f^* - f \| > \delta} P(f | d) df = 1 - \int_{f: \| f^* - f \| \leq \delta} P(f | d) df \quad (4.34)$$

Cuando $\delta \rightarrow 0$, lo anterior se aproxima como

$$R(f^*) = 1 - \kappa P(f|d) \quad (4.35)$$

donde κ es el volumen del espacio que contiene todos los puntos f para los cuales $\|f^* - f\| \leq \delta$. Minimizar la ecuación anterior es equivalente a maximizar la probabilidad a posteriori. Por lo tanto, el estimado con mínimo riesgo es

$$f^* = \arg \max_{f \in \mathbb{F}} P(f | d) \quad (4.36)$$

conocido como el estimado MAP. Como $p(d)$ en la ecuación 4.31 es una constante para una d fija, $P(f|d)$ es proporcional a la distribución conjunta

$$P(f | d) \propto P(f, d) = p(d | f)P(f) \quad (4.37)$$

De esta manera el estimado MAP se puede encontrar de manera equivalente con

$$f^* = \arg \max_{f \in \mathbb{F}} \{p(d | f)P(f)\} \quad (4.38)$$

Obviamente cuando la distribución a priori $P(f)$, es plana, el MAP es equivalente a la probabilidad máxima.

4.12. Etiquetado MAP-MRF

En etiquetado MAP-MRF, $P(f|d)$ es la distribución a posteriori de un MRF un paso importante en el etiquetado de MRFs por medio de Bayes es derivar esta distribución. A continuación se utiliza la formulación simple de una restauración de un MRF como ejemplo para ilustrar el etiquetado MAP-MRF. Si se asume que la superficie es plana entonces la distribución conjunta a priori es

$$P(f) = \frac{1}{Z} e^{-U(f)} \quad (4.39)$$

donde $U(f)$ es la energía a priori dada por

$$U(f) = \sum_i [f_i - f_{i-1}]^2 \quad (4.40)$$

Si se asume que la observación es el valor verdadero más ruido independiente Gaussiano, $d_i = f_i + e_i$, donde $e_i \sim N(\mu, \sigma^2)$, entonces la densidad de probabilidad es

$$p(d | f) = \frac{1}{\prod_{i=1}^m \sqrt{2\pi\sigma_i^2}} e^{-U(d|f)} \quad (4.41)$$

donde

$$U(d | f) = \sum_{i=1}^m \frac{(f_i - d_i)^2}{2\sigma_i^2} \quad (4.42)$$

es la energía de probabilidad. La probabilidad a posteriori es

$$P(f | d) \propto e^{-U(f|d)} \quad (4.43)$$

donde

$$\begin{aligned} U(f | d) &= U(d | f) + U(f) \\ &= \sum_{i=1}^m \frac{(f_i - d_i)^2}{2\sigma_i^2} + \sum_{i=1}^m (f_i - f_{i-1})^2 \end{aligned} \quad (4.44)$$

es la energía a posteriori. El estimado MAP puede encontrarse equivalentemente minimizando esta energía a posteriori

$$f^* = \underset{f}{\operatorname{arg\,mín}} U(f | d) \quad (4.45)$$

Donde los parámetros σ_i se conocen y por lo tanto $U(f|d)$ está completamente especificada, lo cual define completamente la solución del etiquetado MAP-MRF.

4.13. Resumen de la aproximación MAP-MRF

1. Defina un problema de visión como uno de los problemas de etiquetado presentados en las categorías LP1-LP4 y escoja una representación apropiada f del MRF.

2. Obtenga la energía a posteriori para definir la solución MAP al problema.

3. Encuentre la solución MAP. El proceso de calcular la energía a posteriori se resume a continuación:

1. Defina un sistema de vecindad \mathcal{N} en S y el conjunto C de cliques para \mathcal{N} .

2. Determine los potenciales a priori de los cliques $V_c(f)$ para obtener $U(f)$.

3. Calcule la energía de probabilidad $U(d|f)$.

4. Sume $U(f)$ y $U(d|f)$ para obtener la energía a posteriori $U(f|d)$.

El modelo a priori depende del tipo de la escena o superficie que esperamos.

En visión se utiliza generalmente modelo de Gibbs anteriormente propuesto.

El modelo de probabilidad depende de consideraciones físicas como el proceso del sensor (transformaciones, ruido, etc.). Para que este modelo este completo la totalidad de sus parámetros deben especificarse.

4.14. Modelos de observación

En visión de bajo nivel, una observación $d = d_1, \dots, d_m$ es un arreglo rectangular de valores de pixeles. En algunos casos la observación puede ser

escasa . Cada pixel toma un valor d_i de un conjunto D . Una observación d puede considerarse como una versión transformada y degradada de la realización de un MRF, f . La transformación puede incluir transformaciones geométricas pérdida de detalles (borrosidad) y la degradación puede deberse a factores aleatorios como el ruido. Esto determina la distribución condicional $P(d|f)$ o la probabilidad de f . Una observación general del modelo puede expresarse como

$$d = \varphi(B(f)) \odot e \quad (4.46)$$

donde B es un efecto que ocasiona borrosidad, φ es una transformación que puede ser o no lineal, determinística o probabilística, e es el ruido del sensor y \odot es un operador de adición o multiplicación. En la práctica, se asume un modelo de observación simple sin borrosidad con una transformación lineal y ruido aditivo Gaussiano independiente. Cada uno de los valores de los pixeles observados se asumen como la suma del valor verdadero de gris y ruido independiente Gaussiano

$$d_i = \varphi(f_i) + e_i \quad (4.47)$$

donde $\varphi(\cdot)$ es una función lineal y $e_i \sim N(0, \sigma_i^2)$. La distribución de probabilidad condicional de d con respecto a f , o la probabilidad de f es

$$P(d|f) = \frac{1}{\prod_i^m \sqrt{2\pi\sigma_i^2}} e^{-U(d|f)} \quad (4.48)$$

donde

$$U(d|f) = \sum_{i \in S} (\varphi(f_i) - d_i)^2 / [2\sigma_i^2] \quad (4.49)$$

es la energía de probabilidad. Obviamente la ecuación anterior es una forma especial de distribución de Gibbs cuya energía se debe a los cliques de un solo 'site' en el sistema de vecindad de orden cero, con potenciales de clique $[\varphi(f_i) - d_i]^2 / [2\sigma_i^2]$. Si el ruido también es homogéneo entonces las desviaciones $\sigma_i = \sigma$ son las mismas para toda $i \in S$. La función $\varphi(\dots)$ mapea una etiqueta f_i a un valor gris real donde f_i puede ser numérico o simbólico, continuo o discreto. Sin pérdida de generalidad podemos considerar que existe un único valor numérico para la etiqueta f_i y denotar $d_i = \varphi(f_i) + e_i$ simplemente como $d_i = f_i + e_i$. Entonces la energía de probabilidad se convierte en

$$U(d|f) = \sum_{i \in S} (f_i - d_i)^2 / [2\sigma^2] \quad (4.50)$$

para un ruido independiente de tipo Gaussiano.

4.15. Restauración y reconstrucción de imágenes

Las superficies constantes por tramos, o regiones homogéneas, pueden caracterizarse apropiadamente por medio de un modelo MLL homogéneo e isotópico. Para los cliques que contienen más de un ‘site’, los potenciales de los cliques se definen como

$$V_c(f) = \begin{cases} 0 & \text{si todos los sitios en } c \text{ tiene la misma etiqueta} \\ -\zeta_c & \text{cualquier otro caso} \end{cases} \quad (4.51)$$

donde $\zeta_c < 0$ es una constante dependiente de c . Que todos los ‘sites’ en c tengan la misma etiqueta, esto es, todas las $\{f_i | i \in c\}$ sean las mismas, significa la homogeneidad o suavidad de todas las etiquetas f en el clique c . Cualquier violación a la homogeneidad o suavidad se penaliza de manera que $-\zeta_c > 0$. Como las configuraciones más probables son aquellas con la $P(f)$ mayor, o menor $U(f)$, los valores del modelo MLL favorecen la homogeneidad de las f . Para cliques de un solo ‘site’, los potenciales de clique dependen de la etiqueta asignada al ‘site’

$$V_c(f) = V_1(f_i) = \alpha_l \text{ si } f_i = l \in \mathcal{L}_d \quad (4.52)$$

donde α_l es la penalización en contra de que f_i sea etiquetada como l . Mientras mayor sea α_l menos pixeles serán asignados con el valor l , ello controla el porcentaje de ‘sites’ etiquetados con l . Un caso especial ocurre cuando V_c es diferente de cero sólo para cliques de pares de ‘sites’ y cero para todos los demás. En este caso $V_c(f) = 0$ para $\#c > 2$ y para $\#c = 2$ es

$$V_c(f) = V_2(f_i, f_{i'}) = v_{20}[1 - \delta(f_i - f_{i'})] \quad (4.53)$$

donde $\delta(\cdot)$ es la función delta de Kronecker y v_{20} es la penalización en contra de etiquetas desiguales en cliques con dos ‘sites’. La energía a priori es la suma de todos los potenciales de cliques, esto es,

$$U(f) = \sum_{i \in \mathcal{S}} \sum_{i' \in \mathcal{N}_i} v_{20}[1 - \delta(f_i - f_{i'})] \quad (4.54)$$

donde $\sum_{i \in \mathcal{S}} \sum_{i' \in \mathcal{N}_i}$ es equivalente a $\sum_{\{i, i' \in C_2\}}$. En la restauración de superficies constantes a tramos, \mathcal{L} consiste de valores discretos. La idea es recuperar la configuración verdadera f a partir de la imagen observada d . Esto tiene que ver con segmentación de regiones ya que f particiona el conjunto de ‘sites’ de la ‘lattice’ en regiones mutuamente excluyentes. En el marco de MAP-MR, la f óptima es aquella que minimiza la energía posterior.

4.16. Obtención de la energía posterior

A partir de la serie de cuatro pasos propuesta anteriormente:

1. Definir un sistema de vecindad y el conjunto de cliques para él.
2. Determinar las funciones de potencial de cliques a priori en la distribución a priori de Gibbs, para el modelo a priori MLL es la ecuación 4.53
3. Determinar la energía de probabilidad a partir del modelo de observación. Asumir el modelo para ruido aditivo independiente Gaussiano. La función de probabilidad $p(f|d)$ toma la forma de la ecuación 4.50 , con f'_i s que toman valores discretos.
4. Sumar la energía a priori $U(f)$ y la energía de probabilidad $U(d|f)$ para obtener la energía a posteriori

$$E(f) = U(f|d) = \sum_{i \in S} (f_i - d_i)^2 / [2\sigma^2] + \sum_{i \in S} \sum_{i' \in \mathcal{N}_i} v_{20} [1 - \delta(f_i - f_{i'})] \quad (4.55)$$

Note que

$$\sum_{i' \in \mathcal{N}_i} [1 - \delta(f_i - f_{i'})] = \#\{f_{i'} \neq f_i | i' \in \mathcal{N}_i\} \quad (4.56)$$

es simplemente el número de ‘sites’ vecinos cuya etiqueta $f_{i'}$ es diferente a f_i . Se asume que los parámetros σ^2 y v_{20} se conocen.

4.17. Minimización de la energía

Como \mathcal{L}_d es discreto, minimizar la ecuación 4.55, es un problema combinatorio. La solución mínima f^* , es la imagen óptimamente restaurada en el espacio de configuración $\mathbb{F} = \mathcal{L}_d^m$. El algoritmo más simple consiste en descender el valor de la energía paso a paso y funciona como sigue. Se comienza con una configuración inicial $f^{(0)} \in \mathbb{F}$; para el ‘site’ i , se escoge la etiqueta nueva $f_i^{(t+1)}$ de entre todas las $f_i^{(t+1)} \in \mathcal{L}_d$, $f_i^{(t+1)} \neq f_i^{(t)}$, que minimice $E(f_i^{(t+1)}) - E(f_i^{(t)})$ localmente; la iteración continúa hasta que no es posible descender más la energía. Un ejemplo de este tipo de algoritmos es el algoritmo ICM (Iterated Conditional Modes). Este algoritmo encuentra el mínimo de energía localmente por lo que su desempeño depende en gran medida de la calidad del estimado inicial $f^{(0)}$. Si ese estimado inicial no es muy bueno, se requiere de una solución global y es necesario utilizar algoritmos de minimización global como el de Recocido Simulado (Simulated Annealing) por ejemplo.

4.18. Algoritmo ICM

Cuando el conjunto de etiquetas es discreto, minimizar la energía a posterior se convierte en un problema combinatorio. Como resulta difícil maximizar la probabilidad conjunta de un MRF, Besag, propuso un algoritmo determinístico llamado ICM que maximiza las probabilidades condicionales locales secuencialmente de la siguiente manera. Dados los datos d , y las etiquetas $f_{S-\{i\}}^{(k)}$, el algoritmo actualiza secuencialmente cada $f_i^{(k)}$ en $f_i^{(k+1)}$ maximizando $P(f_i|d, f_{S-\{i\}})$ la probabilidad condicional a posteriori, con respecto a f_i . Para calcular $P(f_i|d, f_{S-\{i\}})$ se asume que los componentes de la observación d_1, \dots, d_m , son condicionalmente independientes dado f y que cada d_i tiene la misma función de densidad condicional $p(d_i|f_i)$ que depende sólo de f_i . Esto es

$$p(d|f) = \prod_i p(d_i|f_i) \quad (4.57)$$

También se asume que f depende de las etiquetas en el vecindario local, lo que se conoce como Markovianidad. A partir de lo anterior y por medio del teorema de Bayes se tiene que

$$P(f_i|d, f_{S-\{i\}}) \propto p(d_i|f_i)P(f_i|f_{\mathcal{N}_i}) \quad (4.58)$$

Obviamente, $P(f_i|d_i, f_{\mathcal{N}_i}^{(k)})$ es más fácil de maximizar que $P(f|d)$, lo cual representa la ventaja de ICM. Ahora bien maximizar la ecuación 4.58 equivale a minimizar el potencial a posteriori correspondiente utilizando la siguiente regla

$$f_i^{(k+1)} \leftarrow \arg \min_{f_i} V(f_i|d_i, f_{\mathcal{N}_i}^{(k)}) \quad (4.59)$$

donde

$$V(f_i|d_i, f_{\mathcal{N}_i}^{(k)}) = \sum_{i' \in \mathcal{N}_i} V(f_i|f_{i'}^{(k)}) + V(d_i|f_i) \quad (4.60)$$

Por ejemplo, para la restauración discreta formulada anteriormente el potencial a posteriori es

$$V(f_i|d_i, f_{\mathcal{N}_i}) = (f_i^2 - d_i)/\sigma + v_{20} \sum_{i' \in \mathcal{N}_i} [1 - \delta(f_i - f_{i'})] \quad (4.61)$$

donde $\sum_{i' \in \mathcal{N}_i} [1 - \delta(f_i - f_{i'})] = \#\{f_{i'} \neq f_i | i' \in \mathcal{N}_i\}$ es el número de ‘sites’ vecinos cuyas etiquetas $f_{i'}$ difieren de f_i . Para \mathcal{L} discreto, $V(f_i|d_i, f_{\mathcal{N}_i})$ se evalúa para cada $f_i \in \mathcal{L}$ y la etiqueta que arroja el valor menor de $V(f_i|d_i, f_{\mathcal{N}_i})$ se

escoge como el valor para $f_i^{(k+1)}$. Cuando se aplica a cada i en turno, el proceso anterior define un ciclo de actualización de ICM. La iteración continúa hasta que se logra la convergencia, la cual se garantiza para actualizaciones en serie y es muy rápida. El resultado que se obtiene con ICM depende mucho del estimador inicial $f^{(0)}$. En un algoritmo ICM genuino, no es deseable que dos 'sites' vecinos se actualicen simultáneamente, Besag propuso un método de codificación que paraleliza las iteraciones y mediante el cual S se particiona en varios conjuntos de manera que ningún par de 'sites' analizados sean vecinos en un conjunto. De esta manera todas las f_i pueden actualizarse paralelamente mediante una sola codificación.

Capítulo 5

Regularización del campo de vectores de energía mediante Campos Aleatorios de Markov

5.1. Descripción

En este capítulo se presenta un método, basado en la utilización de Campos Aleatorios de Markov, para regularizar los campos de vectores de energía presentados en el capítulo 3, que resultan de la representación de imágenes mediante coeficientes polinomiales unidimensionales descrita en el capítulo 2.

5.2. Introducción

La regularización del campo de vectores de energía se hace necesaria debido a que en ocasiones no todos los vectores que pertenecen y representan a una misma estructura tienen una orientación adecuada. Se estima que esto puede ocurrir a causa del remanente de ruido, el cual puede continuar presente en la imagen aún después de la aplicación del proceso de restauración. Este ruido puede entonces, intervenir en el cálculo de los vectores de energía asociados a una imagen. Debido a que, en la mayoría de los casos, aproximadamente más del 50 % de los vectores, contenidos dentro de una vecindad o ventana de análisis, cuentan con el mismo ángulo de orientación de energía, es posible utilizar la teoría de Campos Aleatorios de Markov para efectuar cambios en la orientación y magnitud de los vectores (regularización) a partir

de la identificación del evento con mayor probabilidad de ocurrencia dentro de la vecindad o ventana de análisis.

5.3. Regularización

Nuestro problema, en términos de la teoría de Campos Aleatorios de Markov presentada en el capítulo 4, consiste en recuperar o estimar los vectores verdaderos $f = \{f_1, \dots, f_m\}$, a partir de los vectores observados $d = d_1, \dots, d_m$, es decir, los vectores calculados utilizando la representación de imágenes mediante coeficientes polinomiales unidimensionales. Cabe mencionar que en este trabajo se buscó regularizar únicamente la dirección de los vectores por lo que para efectos prácticos los conjuntos d y f están constituidos por los ángulos observados y ‘verdaderos’ o estimados respectivamente. La formulación anterior es posible debido a la estrecha dependencia entre los vectores que forman parte de una vecindad y que pueden pertenecer a una misma estructura, ya que un vector observado, perteneciente a una vecindad donde una mayoría de vectores presenta una misma orientación, tiene una probabilidad muy grande de contar con las mismas características que cuentan sus vectores vecinos, en este caso, una misma orientación. Para este caso el conjunto discreto de m sites $S = \{1, \dots, m\}$ donde $1, \dots, m$ son índices, es el conjunto formado por las localidades del campo de vectores de energía, cuyas dimensiones coinciden con las dimensiones de la imagen original. Ahora bien, como la etiqueta es un evento que puede ocurrir al site, para este caso el conjunto de etiquetas $\mathcal{L}_d = \{l_1, \dots, l_M\}$, es el conjunto de las orientaciones de los vectores que pueden existir en el campo de vectores de energía original. Esta formulación es similar a la que se propone para los problemas de restauración de imágenes corrompidas por ruido presentada en el capítulo 4, donde se supone que la imagen original es homogénea (al menos por tramos) y con base en ello se propone un conjunto de etiquetas posibles igual al conjunto de todas las intensidades posibles en la imagen; para una imagen donde cada pixel puede ser representado por 8 bits, esto implica un conjunto de etiquetas: $\mathcal{L}_d = \{0, \dots, 255\}$. En concreto, para nuestro problema el conjunto de ángulos de vectores observados y el de los ángulos de vectores ‘verdaderos’ o estimados es el mismo y comprende los diferentes ángulos que pueden presentarse en este tipo de análisis, tomando en cuenta lo señalado en el capítulo 2, por lo que el conjunto de etiquetas comprende los valores de dirección que a continuación se presentan: Rango de valores de dirección: $[90, \dots, 270]$ con incrementos de 22.5 unidades. Por lo anterior nuestro problema puede definirse como un problema de etiquetado donde es

necesario asignar una etiqueta, del conjunto de etiquetas \mathcal{L} , a cada uno de los sites en \mathcal{S} . Para poder hacer uso de la información contextual en nuestro análisis, es necesario definir un sistema de vecindad en \mathcal{S} , $\mathcal{N} = \{\mathcal{N}_i \mid \forall i \in \mathcal{S}\}$ donde \mathcal{N}_i es el conjunto de sites vecinos de i . Para nuestro problema este conjunto estará definido por el sistema de 8 conexidad, de segundo orden o de 8 sites vecinos. Ahora bien las colecciones de cliques C_2 que se analizarán consistirán únicamente de los pares de sites vecinos $c = \{i, i'\}$, de tal manera que: $C = C_2 = \{\{i, i'\} \mid i' \in \mathcal{N}_i, i \in \mathcal{S}\}$. Nuestro MRF, estará entonces definido por la familia de variables aleatorias $F = \{F_1, \dots, F_m\}$ definida en el conjunto \mathcal{S} , en la que cada variable aleatoria F_i toma un valor f_i en \mathcal{L} . De acuerdo con el teorema de Hammersley-Clifford que establece la equivalencia entre Campos Aleatorios de Markov y Gibbs, las configuraciones del conjunto de variables aleatorias F obedecen a distribuciones de Gibbs, las cuales tienen la siguiente forma:

$$P(f) = Z^{-1} \times e^{-\frac{1}{T}U(f)} \quad (5.1)$$

donde

$$Z = \sum_{f \in \mathbb{F}} e^{-\frac{1}{T}U(f)} \quad (5.2)$$

es una constante de normalización llamada función partición, T es una constante de temperatura generalmente con valor igual a 1 y $U(f)$ es la función de energía. La energía

$$U(f) = \sum_{c \in C} V_c(f) \quad (5.3)$$

es la suma de potenciales de cliques $V_c(f)$ sobre todos los cliques posibles C . Como se vió en el capítulo 4, el etiquetado MAP-MRF, basado en la teoría de estimación de Bayes, proporciona las herramientas necesarias para conocer el valor de $P(f)$ a partir de $U(f)$ en

$$P(f) = \frac{1}{Z} e^{-U(f)} \quad (5.4)$$

De acuerdo con la teoría de Bayes, es posible obtener una estimación óptima f^* , minimizando el riesgo, de tal manera que

$$f^* = \arg \max_{f \in \mathbb{F}} P(f \mid d). \quad (5.5)$$

Como la regla de Bayes establece que

$$P(f \mid d) = \frac{p(d \mid f)P(f)}{p(d)} \quad (5.6)$$

lo anterior equivale a

$$f^* = \arg \max_{f \in \mathbb{F}} \{p(d | f)P(f)\} \quad (5.7)$$

Ahora bien la probabilidad a posteriori $P(f | d) \propto e^{-U(f|d)}$ y $U(f | d) = U(d | f) + U(f)$, por lo que f^* también puede ser expresado como

$$f^* = \arg \min_f U(f | d) \quad (5.8)$$

lo que implica que debemos encontrar el valor de $U(f | d)$ para resolver nuestro problema. Comenzaremos por proponer las expresiones para calcular a $U(d | f)$ y $U(f)$. La primera puede calcularse utilizando un modelo de observación como el que se presentó en el capítulo 4, en el cual los valores observados d_i podían ser obtenidos en función de las f_i a través de $d_i = f_i + e_i$, donde e_i es el ruido del sensor y en este caso se consideraba independiente Gaussiano y aditivo. Como en nuestro campo de vectores de energía no está influyendo ningún ruido directamente del sensor la relación

$$U(d|f) = \sum_{i \in S} (f_i - d_i)^2 / [2\sigma^2] \quad (5.9)$$

es igual a $U(d|f) = 0$. En efecto el modelado de la perturbación que nuestro campo de vectores sufre depende de la remanente de ruido presente en la imagen original, por lo que sería conveniente utilizar un modelo de observación alternativo que nos permitiera conocer la manera en que esta remanente afecta el campo de vectores de energía para formular adecuadamente la dependencia entre las d_i y las f_i de nuestro problema. Por lo pronto consideraremos la suposición anterior válida para continuar con la solución a nuestro problema, lo cual equivale a suponer que no existe ruido afectando al campo de vectores calculado. Para calcular el valor de $U(f)$, nos apoyaremos de nueva cuenta en el modelo para la restauración y reconstrucción de imágenes constantes por tramos presentado en el capítulo 4, donde

$$V_c(f) = V_2(f_i, f_{i'}) = v_{20}[1 - \delta(f_i - f_{i'})] \quad (5.10)$$

y

$$U(f) = \sum_{i \in S} \sum_{i' \in \mathcal{N}_i} v_{20}[1 - \delta(f_i - f_{i'})] \quad (5.11)$$

Por lo que nuestro modelo para la regularización del campo de vectores de energía, de acuerdo con el etiquetado MAP-MRF, consistiría en encontrar la f óptima minimiza la energía posterior

$$f^* = \arg \min_f U(f | d) \quad (5.12)$$

donde

$$U(f | d) = U(f) = \sum_{i \in S} \sum_{i' \in \mathcal{N}_i} v_{20} [1 - \delta(f_i - f_{i'})] \quad (5.13)$$

Para poder minimizar esta última expresión, se utilizó el criterio de optimización ICM, también presentado en el capítulo 4, que consiste en descender el valor de la energía paso a paso apartir de una configuración inicial. En nuestro caso la configuración inicial consiste en la observación o campo de vectores de energía calculado en el capítulo 3, a través de la representación de imágenes mediante coeficientes polinomiales unidimensionales. En particular, en la configuración de las orientaciones o ángulos. A partir de esta configuración se efectúa un análisis de cada uno de los sites que en este caso alojan vectores, tomando en cuenta los valores de los ángulos de sus 8 vecinos más próximos. En este proceso de análisis, la selección del vector en turno se realiza mediante la generación de un número aleatorio que determina el site que deberá ser analizado. El análisis toma en cuenta los valores de los ángulos y las magnitudes presentes en los sites vecinos para identificar si, por lo menos 50 % de los ángulos vecinos son iguales, y si el vector con mayor magnitud del conjunto de vecinos se encuentra dentro de este 50 %. Si ambas condiciones se cumplen el valor del ángulo del vector analizado toma el nuevo valor del ángulo que más se repite en la vecindad. La justificación de los argumentos anteriores se basa en el hecho de que los vectores que tienen una magnitud mayor en el campo de vectores, son vectores con mayor energía 1D lo que denota que tienen más probabilidad de formar parte de un patrón 1D. Por esta razón se analiza el valor del ángulo con mayor frecuencia de aparición en la vecindad y además se comprueba que este ángulo esté asociado con el vector de magnitud mayor que pertenezca a este grupo de por lo menos 4 vectores. De esta manera, mediante esta aproximación se privilegia la máxima probabilidad de ocurrencia de un evento que en este caso es un valor de ángulo. Esto es,

$$f_i^{(k+1)} \leftarrow \arg \max_{f_i} P(f_i | d_i, f_{\mathcal{N}_i}^{(k)}) \quad (5.14)$$

ó equivalentemente,

$$f_i^{(k+1)} \leftarrow \arg \min_{f_i} V(f_i | d_i, f_{\mathcal{N}_i}^{(k)}) \quad (5.15)$$

donde

$$V(f_i | d_i, f_{\mathcal{N}_i}) = v_{20} \sum_{i' \in \mathcal{N}_i} [1 - \delta(f_i - f_{i'})] \quad (5.16)$$

y $\sum_{i' \in \mathcal{N}_i} [1 - \delta(f_i - f_{i'})] = \#\{f_{i'} \neq f_i \mid i' \in \mathcal{N}_i\}$ es el número de sites vecinos cuyas etiquetas $f_{i'}$ difieren de f_i , lo cual para nuestro caso se traduce en el número de sites vecinos cuyos ángulos difieren. Para completar un barrido y pasar a la siguiente iteración $k + 1$, es necesario analizar la totalidad de los sites. Los resultados obtenidos se presentaron en [9] y [10]. A continuación se presentan los resultados que se obtuvieron al aplicar la regularización a los campos de vectores de energía obtenidos en el capítulo 3



Figura 5.1: Campo de vectores de energía original de la sección 1

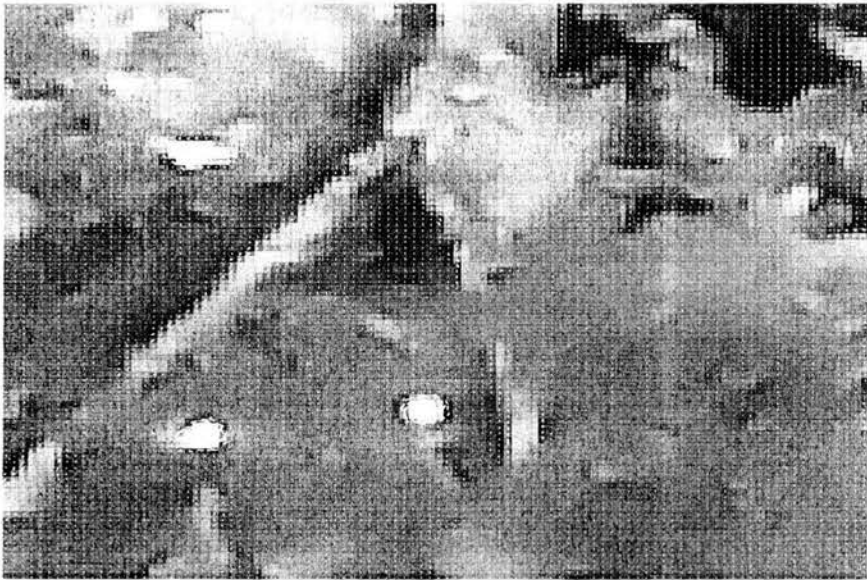


Figura 5.2: Campo de vectores de energía regularizado de la sección 1

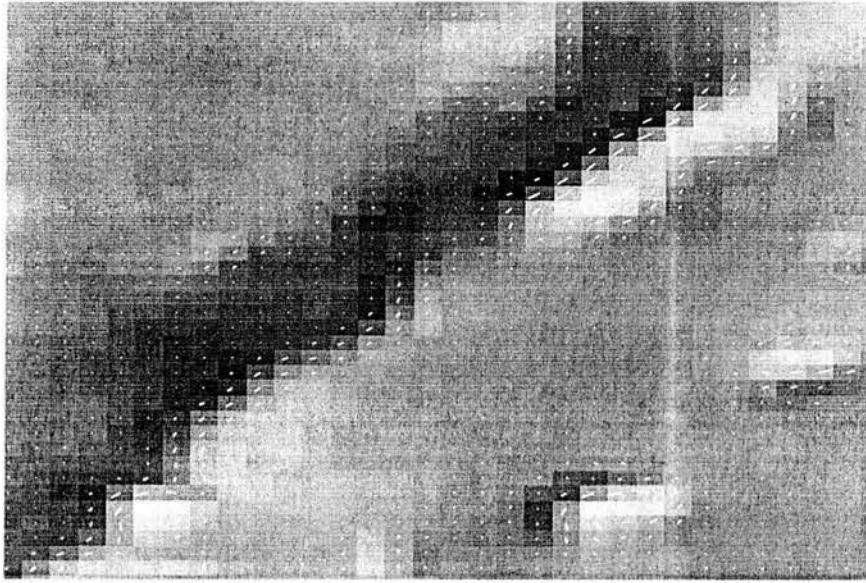


Figura 5.3: Recorte del Campo de vectores de energía original de la sección 1

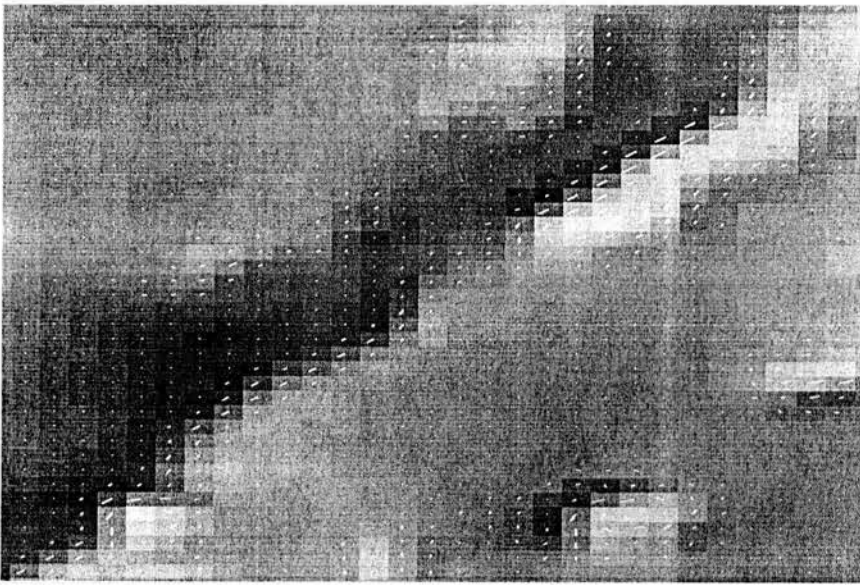


Figura 5.4: Recorte del Campo de vectores de energía regularizado de la sección 1

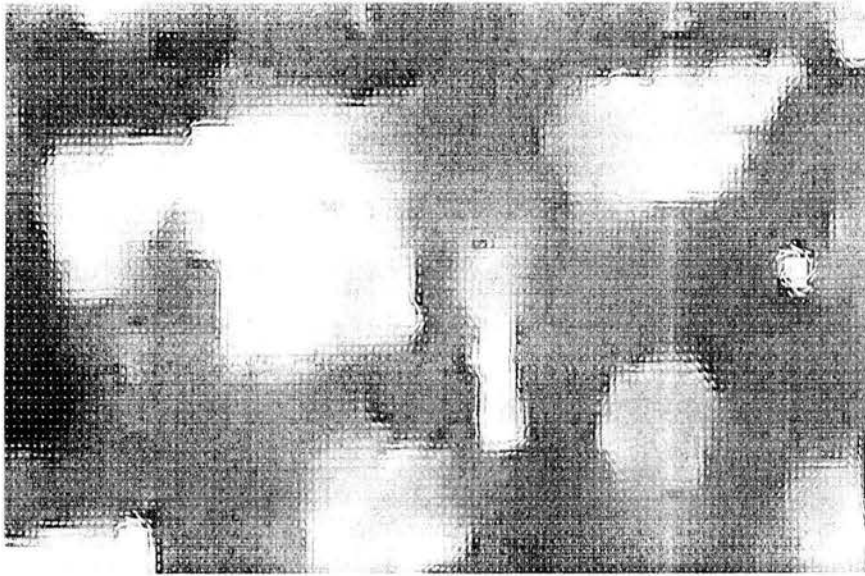


Figura 5.5: Campo de vectores de energía original de la sección 2



Figura 5.6: Campo de vectores de energía regularizado de la sección 2

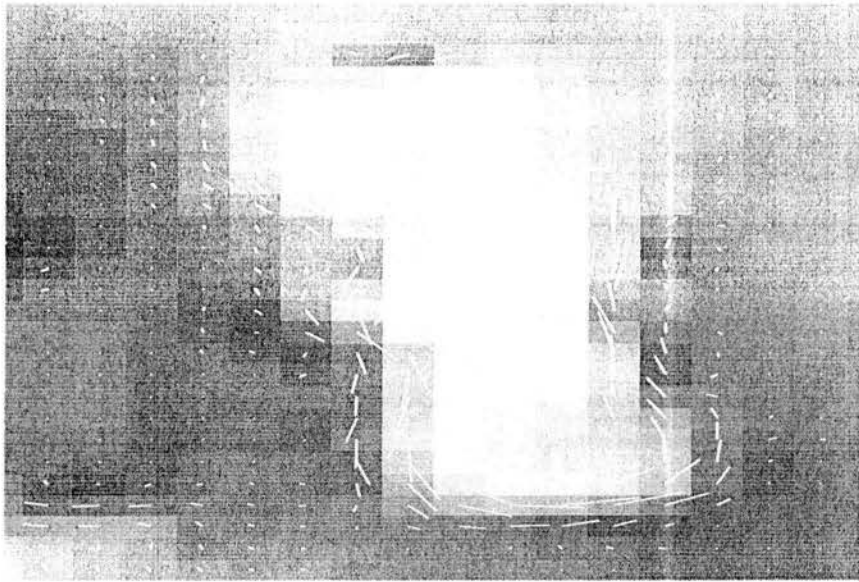


Figura 5.7: Recorte del Campo de vectores de energía original de la sección 2

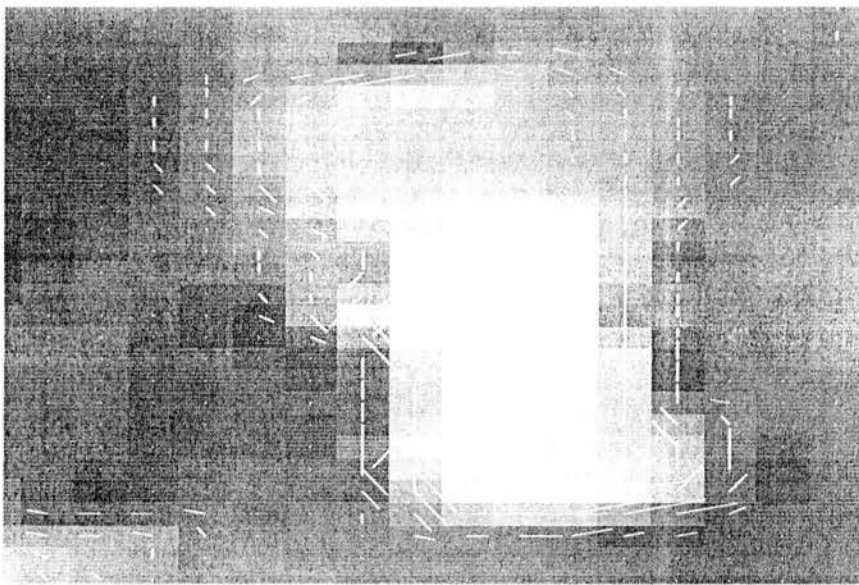


Figura 5.8: Recorte del Campo de vectores de energía regularizado de la sección 2

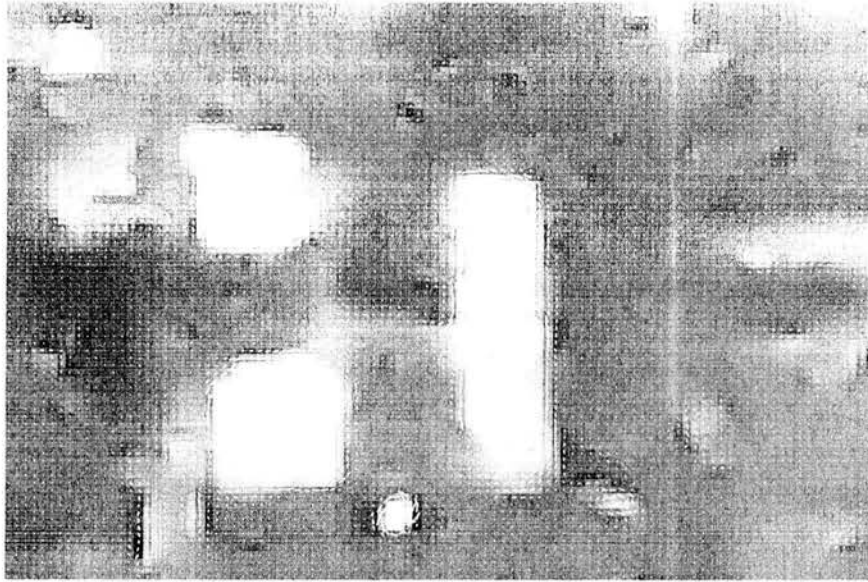


Figura 5.9: Campo de vectores de energía original de la sección 3

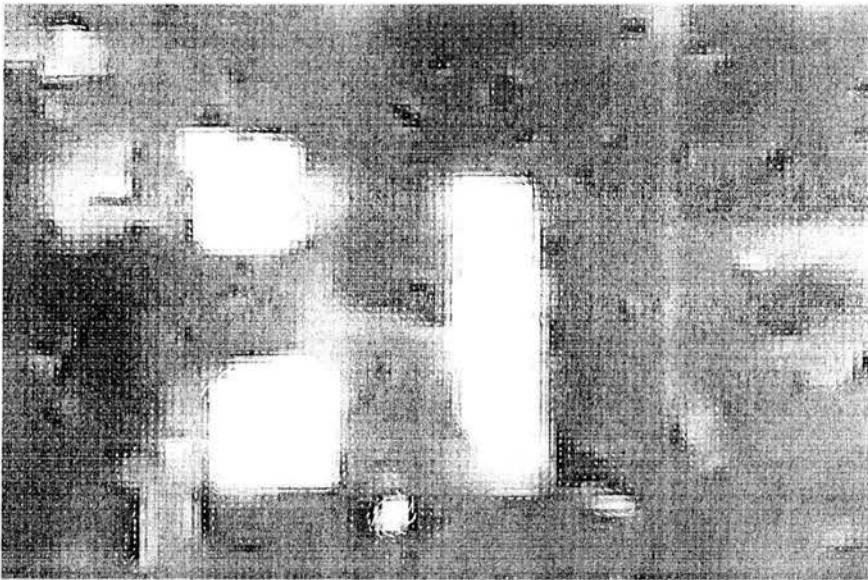


Figura 5.10: Campo de vectores de energía regularizado de la sección 3

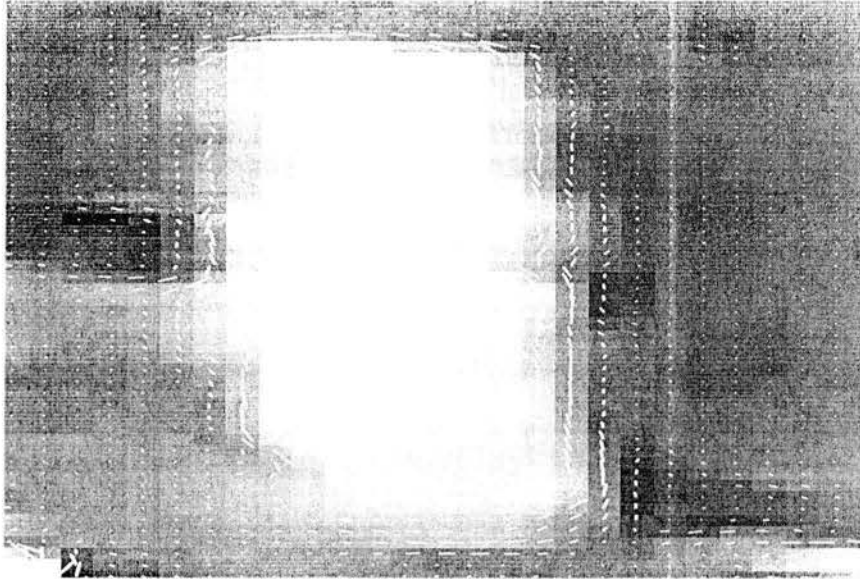


Figura 5.11: Recorte a) del Campo de vectores de energía original de la sección 3

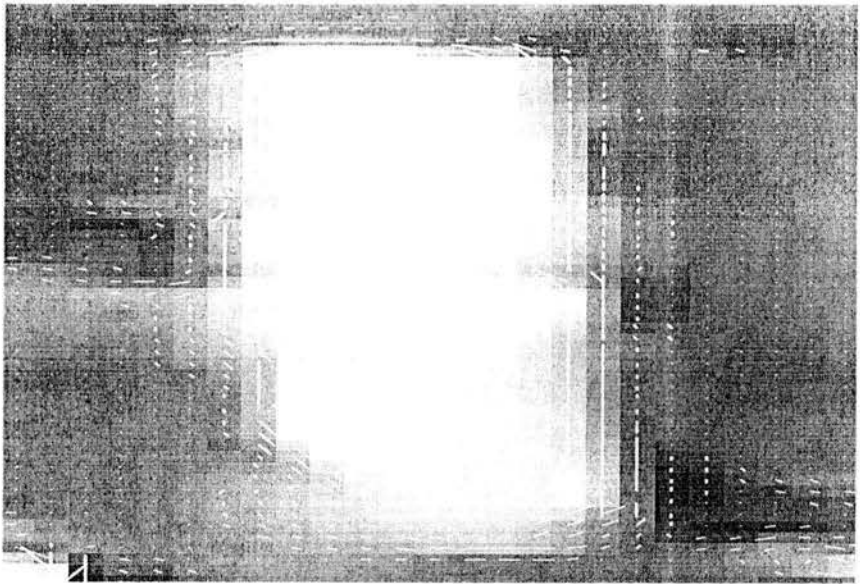


Figura 5.12: Recorte a) del Campo de vectores de energía regularizado de la sección 3

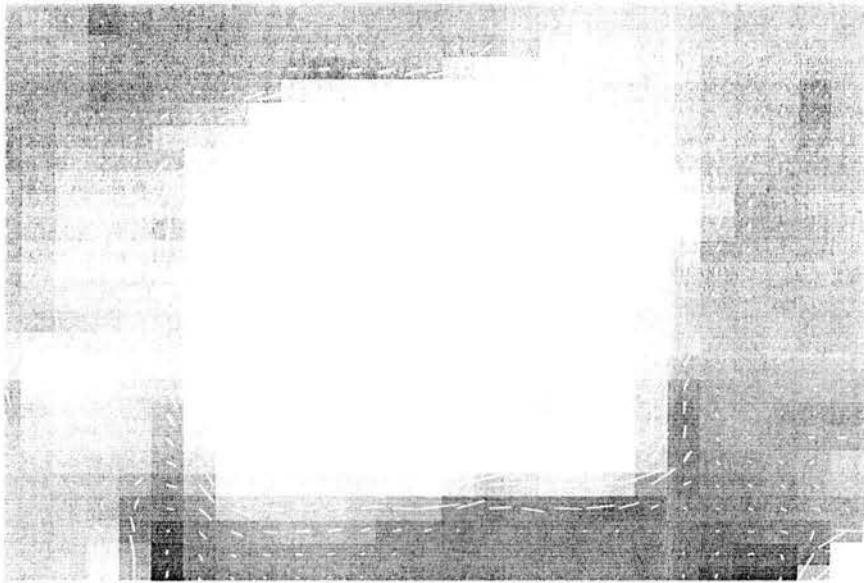


Figura 5.13: Recorte b) del Campo de vectores de energía original de la sección 3

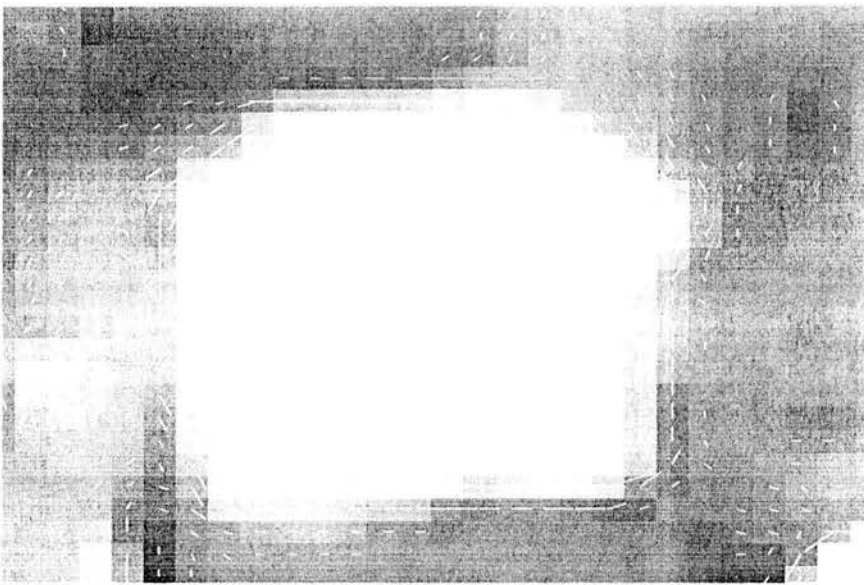


Figura 5.14: Recorte b) del Campo de vectores de energía regularizado de la sección 3

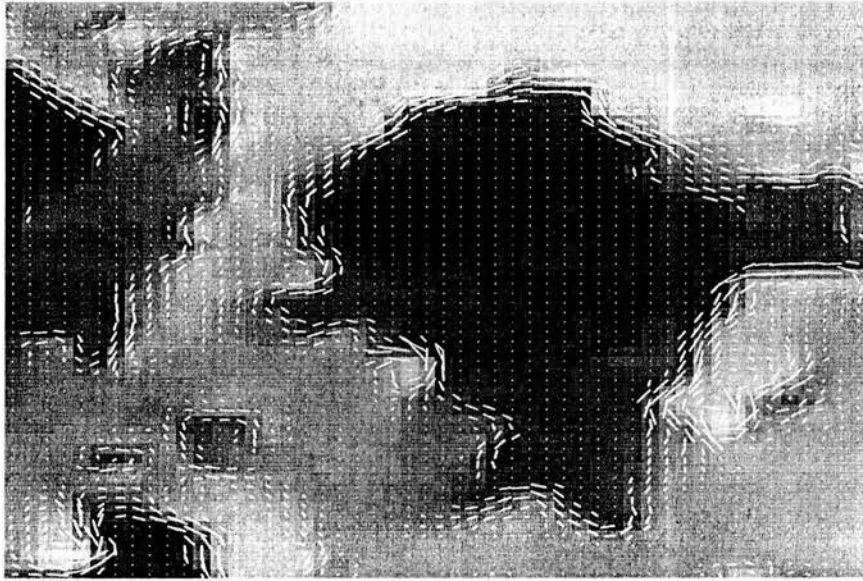


Figura 5.15: Campo de vectores de energía original de la sección 4

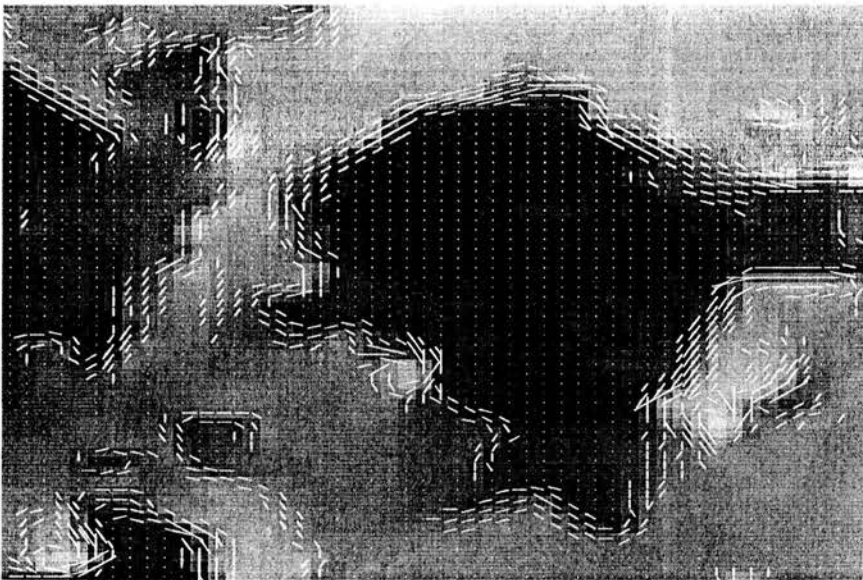


Figura 5.16: Campo de vectores de energía regularizado de la sección 4

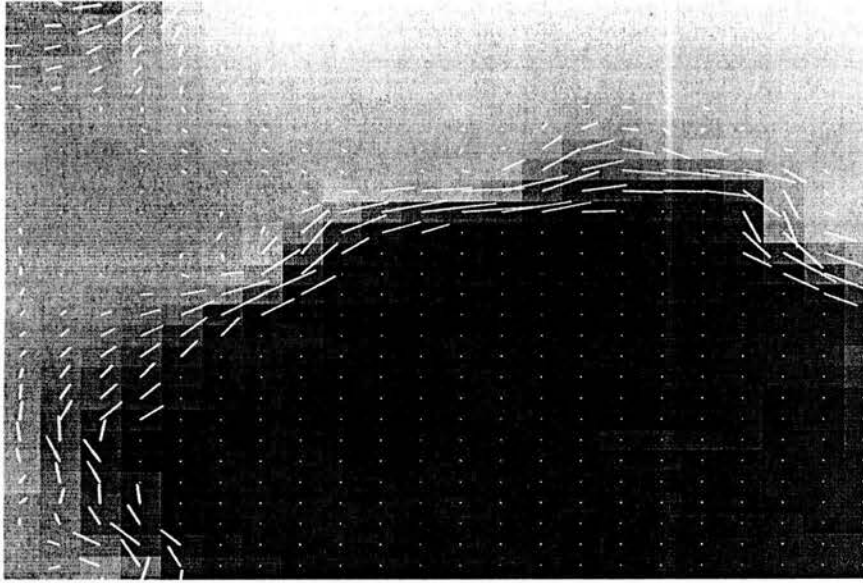


Figura 5.17: Recorte del Campo de vectores de energía original de la sección 4

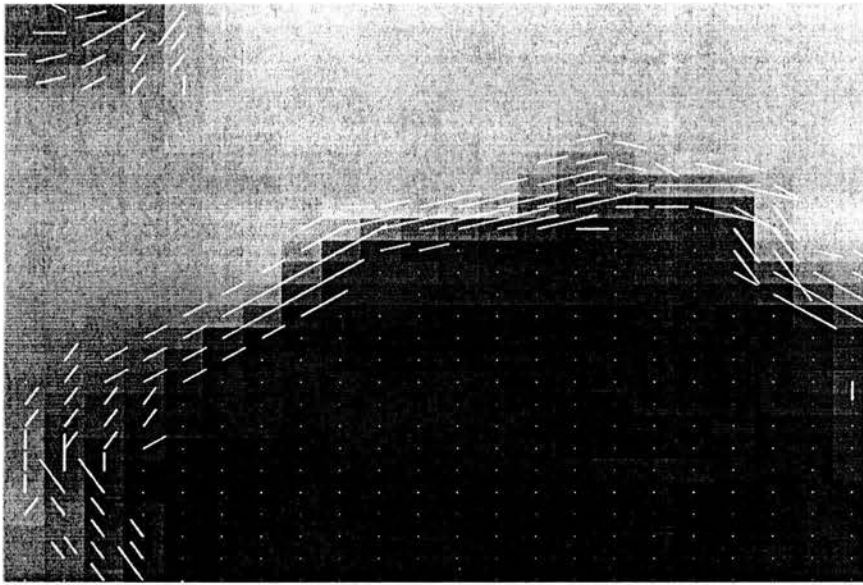


Figura 5.18: Recorte del Campo de vectores de energía regularizado de la sección 4

Capítulo 6

Conclusiones

Mediante este trabajo de investigación se comprobó por un lado, la viabilidad de aplicar la Transformada Orientada de Hermite a imágenes SAR, con el propósito de obtener el campo de vectores de energía asociado y la clasificación de patrones 0D, 1D y 2D; por otro lado, se mostró la conveniencia de regularizar el campo de vectores de energía utilizando la teoría de Campos Aleatorios de Markov. Ambas herramientas son importantes en el ámbito del Procesamiento Digital de Imágenes pues trabajan tomando en cuenta características fundamentales en las imágenes, como la relación con el sistema visual humano (filtros de Gaussianas) y la dependencia contextual. Lo anterior permite por un lado detectar con facilidad y certeza los patrones 1D, importantes estructuras en el análisis de imágenes, y por otro, realizar análisis con la finalidad de restaurar imágenes haciendo uso y priorizando la información de su entorno. La relevancia de este tipo de análisis radica en la amplia variedad de aplicaciones que pueden beneficiarse de éste, ya que los patrones 1D como líneas o bordes juegan un rol importante en visión temprana. De acuerdo con Martens [5], las estructuras locales 1D como los bordes representan una mayoría de las regiones no-constantes en imágenes y su importancia ha sido reconocida ampliamente en visión computacional, donde muchos algoritmos han sido diseñados específicamente para detectar y localizar estas estructuras. En el caso de imágenes de percepción remota, este análisis resulta particularmente útil en los estudios que utilizan técnicas como la Detección de Bordes, Segmentación y Clasificación. Un ejemplo es la planeación de infraestructura urbana, donde resulta necesario detectar la localización y delimitación de construcciones y caminos apoyando los proyectos de construcción de viviendas y vías de circulación eficaces. También en el monitoreo de recursos naturales donde es importante conocer la distribu-

ción de ríos y el tamaño de su cauce, lo cual en un momento dado puede ser útil en el ámbito de la agricultura, en la planeación de un nuevo terreno agrícola, ya que es posible preveer una adecuada irrigación acuífera. Como propuesta para estudios posteriores, sería importante estudiar la eficiencia del algoritmo de regularización de los vectores de energía propuesto pues los resultados presentados en este trabajo son únicamente cualitativos. En este sentido resultaría importante comprobar su eficiencia en imágenes sintéticas y a su vez comparar los resultados obtenidos con los que pueden obtenerse utilizando otros métodos de regularización de vectores. De esta manera obtendríamos resultados cuantitativos que nos permitirían respaldar y asegurar la eficiencia de nuestro método en aplicaciones diversas.

Bibliografía

- [1] López Quiroz Penélope, 'Procesamiento de imágenes SAR para la estimación del crecimiento de la zona urbana en el área metropolitana de la Ciudad de México, en el período 1995-1997', Tesis de Ingeniería en Computación, Universidad Nacional Autónoma de México, Facultad de Ingeniería, 2002.
- [2] Camarillo Sandoval M. Patricia, Varela Lopez Alma A., 'Restauración y segmentación de imágenes de hielo obtenidas mediante Radar de Apertura Sintética', Tesis de Ingeniería en Computación, Universidad Nacional Autónoma de México, Facultad de Ingeniería, 1998.
- [3] Yuen Zhou Héctor, 'Implementación distribuida de la estimación de flujo óptico mediante la transformada polinomial', Tesis de Ingeniería en Telecomunicaciones, Universidad Nacional Autónoma de México, Facultad de Ingeniería, 2003.
- [4] Martens, J.B., The Hermite Transform-Theory, IEEE transactions on Acoustics, Speech and Signal Processing; vol. 38 no.9, 1990; pp. 1595-1606.
- [5] Martens, J.B., The Hermite Transform-Applications, IEEE transactions on Acoustics, Speech and Signal Processing; vol. 38 no.9, 1990; pp. 1607-1618.
- [6] Silván Cárdenas José Luis, 'Compresión de imágenes basada en modelos gaussianos de Percepción Visual', Tesis de Maestría en Ingeniería Eléctrica, Universidad Nacional Autónoma de México, División de Estudios de Posgrado de la Facultad de Ingeniería, 2002.
- [7] Escalante-Ramírez Boris, Silván-Cárdenas José Luis., López-Quiroz Penélope, 'SAR-image classification with a directional-oriented discrete Hermite transform', Proceedings of SPIE, Image and Signal Processing for Remote Sensing VIII, vol.4885, 2003; pp. 234-245.

- [8] Li, S. Z., 'Markov Random Field Modeling in Computer Vision', Springer, Hong Kong, 1995.
- [9] Escalante-Ramírez Boris, Silván-Cárdenas José Luis., López-Quiroz Penélope, 'SAR-image classification with a directional-oriented discrete Hermite transform and Markov Random Fields', Proceedings of IEEE, International Geoscience and Remote Sensing Symposium, CD-ROM, 2003.
- [10] López-Quiroz Penélope, Escalante-Ramírez Boris, Silván-Cárdenas José Luis, 'SAR-image classification with a directional-oriented discrete Hermite transform and Markov Random Fields', Proceedings of SPIE, Image and Signal Processing for Remote Sensing IX, en imprenta.