



**UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO**

FACULTAD DE INGENIERÍA

**FUSIÓN DE DATOS Y SEGMENTACIÓN DE
IMÁGENES DE PERCEPCIÓN REMOTA**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

**INGENIERA EN TELECOMUNICACIONES
PRESENTA:
MARÍA ENRIQUETA BARILLA PÉREZ**

Y QUE PARA OBTENER EL TÍTULO DE:

**INGENIERO EN COMPUTACIÓN
PRESENTA:
JOSÉ ANTONIO FRANCO BELLO**

298575



DIRECTOR DE TESIS: DR. MIGUEL MOCTEZUMA FLORES

México, D.F.

2001



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 está 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.

DEDICATORIAS

José Antonio

En especial a mis padres,
Everth y José Antonio,
por todo su apoyo.

A mis hermanos Maribel, Gonzalo,
Jesús y Emmanuel.

A Ileana.

María Enriqueta

Con profundo amor a:

Mi buen Dios
Ma. Micaela y Fidel, mis padres
Concepción, Alex y Ana, mis hermanos
Mi Tía Meche, mi segunda madre
Miguel y Ezequiel, mis ángeles
Mis grandes amigos
México, mi patria

AGRADECIMIENTOS

Expresamos nuestro agradecimiento a:

Nuestro director de tesis, el Dr. Miguel Moctezuma Flores, por habernos encausado a adquirir los elementos académicos para desarrollar esta tesis.

El Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico ofrecido a través del Proyecto "Fusión de datos y clasificación contextual para análisis de escenas de percepción remota".

La Agencia Espacial Europea (ESA) por haber facilitado las imágenes de Radar de Apertura Sintética (SAR) de los satélites ERS a través del Proyecto AO-318.

PUBLICACIONES

El trabajo desarrollado en esta tesis ha generado parte de la información utilizada para la redacción de los siguientes artículos¹, presentados en Simposia Internacionales:

J. A. Franco, M. Moctezuma y F. Parmiggiani. "*Aproximación Contextual a la Segmentación de Fotografía Aérea*". Memorias del SIARP '99. La Habana, Cuba. Marzo 1999.

J. A. Franco, M. Moctezuma and F. Parmiggiani. "*Remote Sensing Images Segmentation: A mixed approach adapted for fine structure detection*". Memorias del SIARP 2000. Lisboa, Portugal. Septiembre 2000.

J. A. Franco, M. Moctezuma, M. E. Barilla, B. Escalante and F. Parmiggiani. "*SAR Images Filtering and Segmentation: A Multiresolution and Contextual Approach*". Memorias del IEEE IGARSS 2001. Sydney, Australia. Julio 2001.

¹ Estos artículos se anexan en el Apéndice A de esta tesis.

ÍNDICE

Dedicatorias	i
Agradecimientos	ii
Publicaciones	iii
Índice	iv
1 Introducción	1
1.1 Resumen.....	1
1.2 Contenido de la tesis	3
2 Principios de Percepción Remota.....	5
2.1 Introducción	5
2.2 Conceptos básicos	8
2.2.1 Radiación Electromagnética	8
2.2.2 Espectro Electromagnético	9
2.2.3 Sensores Remotos.....	12
2.2.3.1 Tipos de sensores	13
2.3 Sistemas de Percepción Remota	14
2.3.1 Percepción Remota Aérea	14
2.3.2. Percepción Remota Espacial	17

2.3.2.1 Satélites ERS 1 y 2	18
Radar de apertura sintética (SAR)	19
Imágenes SAR.....	21
2.3.2.2 Satélites SPOT 1, 2 y 4.....	23
2.3.2.3 Satélite IKONOS	25
3 Fundamentos.....	26
3.1 Procesamiento de Imágenes	26
3.2 Teoría de los Campos Aleatorios de Markov	30
3.2.1 Vecindades y cliques.....	30
3.2.1.1 Sistemas de vecindad.....	30
3.2.1.2 Cliques	31
3.2.2 Campos Aleatorios de Markov y Distribución de Gibbs.....	32
3.2.3 El modelo.....	34
3.2.4 Modelo de Ising	35
3.2.5 Modelo de Potts	35
3.2.6 Algoritmo de simulación y optimización.....	36
3.2.6.1 Algoritmo de simulación	36
Muestreador de Gibbs.....	37
Algoritmo de metrópolis a temperatura constante	37
3.2.6.2 Algoritmo de optimización.....	38
Algoritmo de Máximo A Posteriori.....	39
El recocido simulado (<i>Simulated Annealing</i>)	40
3.3 Transformada Polinomial	42
3.3.1 Introducción	42
3.3.2 Transformada Polinomial Unidimensional.....	43
3.3.3 Transformada Hermitiana Unidimensional	46
3.3.4 Transformada Polinomial en dos dimensiones	47
3.3.5 Transformada Hermitiana en dos dimensiones	50
4 Segmentación adaptiva de imágenes de percepción remota	51
4.1 Introducción	51
4.2 Segmentación de Imágenes de Percepción Remota.....	52
4.2.1 Segmentación supervisada y no supervisada	54
4.2.2 Segmentación contextual y no contextual	54
4.2.2.1 Segmentación no contextual	55
Algoritmos de mínima distancia	55
Algoritmos de máxima probabilidad.....	55
4.2.2.2 Segmentación contextual.....	56
Modos Condicionales Iterativos (ICM)	56
Modelo Markoviano	56
4.3 Aproximación Adaptiva a la Segmentación de Imágenes de Percepción Remota	57
4.4 Resultados	61
4.4.1 Segmentación de fotografías aéreas de alta resolución	62
4.4.2 Segmentación de imágenes SPOT	74
4.4.3 Segmentación de imágenes SAR	80

4.5 Conclusiones	88
5 Clasificación de zonas urbanas por Modelado Markoviano	89
5.1 Introducción	89
5.2 Parámetro de textura.....	90
5.2.1. Modelo 2D Isotrópico.....	91
5.2.1.1 El modelo	91
5.2.1.2 El método de estimación: Colas de cometa	92
5.2.2. Serie de modelos 1D	93
5.3 Binarización por Correlación	94
5.3.1 Introducción	94
5.3.2 El algoritmo.....	95
5.4 Resultados	96
5.4.1 Fotografías aéreas de alta resolución	97
5.4.1 Imagen SPOT pancromáticas	105
5.4.1 Imagen SAR	109
5.5 Conclusiones	113
6 Extracción de zonas urbanas por Fusión de Datos	114
6.1 Introducción	114
6.2 Máscara Urbana	116
6.2.1 Obtención de la imagen binarizada	116
6.2.2 Obtención de la imagen de textura binarizada.....	116
6.2.3 Proceso de fusión	117
6.3 Extracción de zonas urbanas	119
6.4 Resultados	119
6.4.1 Fotografías aéreas de alta resolución	120
6.4.2 Imagen SPOT pancromática	134
6.4.3 Imagen SAR	141
6.5 Conclusiones	148
7. Conclusiones Generales	149
Glosario.....	152
Referencias	154
Apéndice A	A-1
Aproximación Contextual a la Segmentación de Fotografía Aérea.....	A-2
Resumen.....	A-2
Introducción	A-2

Aproximación Estadística	A-3
Relajación Determinística	A-6
Conclusiones.....	A-8
Referencias.....	A-8
Remote Sensing Images Segmentation:	
A Mixed Approach Adapted for Fine Structure Detection.....	A-9
Abstract	A-9
Introduction.....	A-9
Non-contextual segmentation	A-10
Contextual segmentation.....	A-11
A mixed approach to image segmentation	A-12
Results.....	A-14
Comments and concluding remarks	A-18
References	A-18
SAR Images Filtering and Segmentation:	
A Multiresolution and Contextual Approach	A-19
Abstract	A-19
Introduction.....	A-19
Adaptive Noise-Reduction by Polynomial Transforms.....	A-20
Contextual Segmentation	A-21
Results.....	A-22
Conclusions	A-23
References	A-25
Apéndice B	B-1
Características de Cámaras Fotográficas	B-1
Características de Sistemas de Percepción Remota Espacial	B-2

1 INTRODUCCIÓN

1.1 Resumen

El conocimiento del medio urbano es para muchos una apreciación de la ciudad como una entidad física. Desde esta perspectiva, las técnicas de tratamiento de imágenes pueden aportar una ayuda precisa. Este conocimiento de la ciudad puede ser abordado de dos maneras distintas. El primero consiste en considerar a la ciudad como un solo y único objeto en oposición a los medios no urbanos (campo, bosques, mar, etc.). El segundo enfoque apunta a obtener un conocimiento más preciso del medio urbano en sí, distinguiendo diferentes tipos de estructuras urbanas en el interior de la ciudad.

Las dos aproximaciones recién presentadas son el resultado de una realidad urbana dual. Por una parte existe cierta heterogeneidad global, debido a la variedad de los materiales de construcción, que permite distinguir lo urbano de lo no urbano que, por lo general, es más homogéneo. Por otra parte, la existencia de cierta homogeneidad local nos va a permitir distinguir diferentes tipos de estructuras urbanas. Esta dualidad nos conduce a pensar que las técnicas utilizadas para distinguir las zonas urbanas de las no urbanas serán diferentes de aquellas utilizadas para caracterizar la ciudad intrínsecamente.

Esta tesis está dirigida a la extracción de regiones urbanas como un único objeto en contraposición a las regiones no urbanas. Para extraer dichas regiones obtenemos una máscara urbana (segmentación binaria), que sea lo más precisa posible.

El esquema de clasificación presentado en esta tesis fue aplicado a imágenes provenientes de distintos captores:

- Fotografía aérea de alta resolución
- Imagen SPOT pancromática
- Imagen ERS1-SAR

En los resultados finales se observa una clasificación de escenas urbanas, que permite una adecuada discriminación entre regiones urbanas y no urbanas. Esta discriminación se lleva a cabo mediante la superposición del mapa urbano sobre la imagen original, es decir, para las regiones no urbanas se asigna un valor único de nivel de gris (color negro), mientras que para las regiones urbanas se considera el nivel de gris de la imagen original.

El proceso utilizado para la extracción de zonas urbanas se detalla a continuación:

Como primer paso para la extracción de zonas urbanas, obtenemos una *máscara urbana* a partir del algoritmo de fusión de datos propuesto por Shan Yu [32].

La fusión se lleva a cabo con dos imágenes de entrada:

1. Una imagen binarizada que nos proporciona información muy aproximada acerca de los contornos de los elementos presentes en la escena en cuestión, así como información sobre las zonas urbanas, basada exclusivamente en los niveles de gris de la imagen original.

Con respecto al proceso de segmentación, en esta tesis proponemos un esquema adaptivo de segmentación, enfocado a la preservación de las estructuras finas de la escena y a una mejor definición de contornos, pero tratando de evitar los efectos de sobresegmentación. El algoritmo propuesto se basa en la adaptación de varios métodos clásicos de segmentación, que abarcan los dos principales enfoques de ésta: el no contextual (mínima distancia y máxima probabilidad) y el contextual (Campos Aleatorios de Markov, modelo de Potts y Recocido Simulado). Los resultados ofrecidos por este esquema de segmentación evidencian su buen desempeño, generando imágenes clasificadas mostrando regiones homogéneas, contornos bien definidos y respeto de las estructuras finas, así como una mínima presencia de píxeles aislados.

Debido a que existen diversos métodos de segmentación y binarización, es necesario hacer una evaluación de los resultados generados por algunos de los métodos de binarización existentes, con el fin de determinar la imagen que mejor se adecua para ser considerada en el proceso de fusión. Para evaluar dichos resultados se estima su confiabilidad a través del cálculo de un error global: a menor error global, mayor confiabilidad del resultado. El conocimiento ofrecido por una máscara o mapa urbano rudimentario es utilizado en este paso.

Aquí comprobamos la efectividad del método de segmentación propuesto, ya que la imagen binarizada por este método es la que tiene menor error global y, por tanto, es más confiable para ser considerada en el proceso de fusión.

2. Una imagen de temperatura binarizada, que nos dará información acerca de las texturas de los elementos presentes en la escena de interés, para disminuir el error que la segmentación genera al confundir zonas no urbanas con zonas urbanas, al considerar únicamente sus características espectrales.

La teoría de los Campos Aleatorios de Markov se utiliza para modelar la textura de la imagen mediante un campo Markoviano en 4 conexidad, lo que nos permite definir un parámetro característico de la textura urbana, el cual se denomina 'Parámetro de Temperatura', por lo que a la imagen que se obtiene se le conoce como 'Imagen de temperatura'. Esta imagen se caracteriza por tener valores altos en regiones heterogéneas (mayor temperatura), que son características de las zonas urbanas; mientras que en las regiones homogéneas, que por lo general caracterizan a las zonas no urbanas, presenta valores bajos (menor temperatura).

Una vez que se obtiene dicha imagen, procedemos a binarizarla. Esto con el objetivo de que, a partir de un umbral, los valores con mayor temperatura se asignen a una categoría urbana, mientras que los valores con menor temperatura son asignados a una categoría no urbana. El método de binarización utilizado es el de Correlación, ya que éste evalúa la correlación entre la imagen original y la imagen binarizada -para un rango de umbrales de binarización- y elige aquel valor que maximiza dicha correlación (umbral óptimo).

Una vez obtenidas la imagen segmentada y la imagen de texturas binarizada, se aplica el esquema de fusión de datos para incrementar la confiabilidad y precisión en la generación de la máscara urbana. Además de la información proporcionada por las dos imágenes recién mencionadas, el proceso de fusión considera también a la imagen original y al mapa urbano aproximado. El esquema de fusión calcula para cada *site* su coeficiente de confianza, tanto en la imagen segmentada como en la imagen de texturas binarizada, y considera para la imagen fusionada aquel valor que maximiza dicho coeficiente.

Por último, con base a la máscara generada, procedemos a extraer las regiones urbanas presentes en las escenas analizadas.

1.2 Contenido de la tesis

El trabajo desarrollado en esta tesis se encuentra dividido en los siguientes capítulos:

- Capítulo 1.** Presenta un panorama general del problema que se aborda, la motivación para resolverlo y las técnicas que se utilizaron como propuesta de solución.
- Capítulo 2.** Este capítulo muestra la teoría básica de la Percepción Remota y ofrece información relevante sobre los dispositivos utilizados en esta disciplina.
- Capítulo 3.** Aquí se presentan conceptos básicos de la teoría de los Campos Aleatorios de Markov y la Transformada Polinomial, los cuales son utilizados en capítulos posteriores.

- Capítulo 4.** En este apartado se propone un algoritmo adaptivo de segmentación de imágenes de percepción remota, basado en el uso combinado de varios conceptos utilizados por numerosos métodos clásicos de segmentación, los cuales abarcan los dos principales enfoques de esta área: el no contextual (mínima distancia y máxima probabilidad) y el contextual (Campos Aleatorios de Markov, modelo de Potts y Recocido Simulado).
- Capítulo 5.** Se utiliza el modelado Markoviano para efectuar una clasificación de imágenes urbanas a través del parámetro de textura denominado 'parámetro de temperatura'. La imagen obtenida por este parámetro de texturas es binarizada por correlación, para que a partir de la imagen obtenida se lleve a cabo la clasificación de escenas urbanas.
- Capítulo 6.** En este capítulo se obtiene una máscara urbana por medio de un esquema de fusión de datos y, a partir de ella, se realiza la extracción de zonas urbanas de las imágenes consideradas. El propósito de la fusión de datos es incorporar resultados parciales y complementarios. La fusión se basa en dos imágenes significativas: la primera es una imagen segmentada a través del algoritmo adaptivo propuesto en el capítulo 4; la segunda imagen es aquella que se obtuvo a través del análisis de textura en el capítulo 5. Con la fusión se disminuye el error generado por la mala interpretación de zonas no urbanas definidas a partir del resultado de la segmentación y, por otra parte, se preservan de mejor forma las estructuras finas y se obtiene una mejor estimación de contornos, aspectos que pueden verse afectados como resultado de un análisis basado exclusivamente en las características de la textura urbana.
- Capítulo 7.** Se presentan las conclusiones finales de este trabajo y se sugieren algunas alternativas futuras de investigación.

2 PRINCIPIOS DE PERCEPCIÓN REMOTA

2.1 Introducción

Volar ha sido desde épocas muy remotas uno de los sueños más intensamente anhelados por la Humanidad. Pese a la persistencia del intento, sólo en fechas recientes se ha dispuesto de los medios técnicos necesarios para hacer realidad este deseo. Desde ese momento, el ritmo de la innovación tecnológica ha sido vertiginoso, lo que nos ha permitido enriquecer notablemente nuestro conocimiento sobre el Planeta y sus habitantes.

Uno de los principales estímulos de esta aventura aérea ha sido la búsqueda de una nueva visión de los paisajes terrestres. El afán de remontar la limitada perspectiva de la visión humana es evidente ya desde los inicios de la aeronáutica, llegando a ser hoy en día una pieza clave de la investigación espacial.

La observación remota de la superficie terrestre constituye el marco de estudio de la Percepción Remota o Telemetría. Estos vocablos son una traducción latina del término inglés *Remote Sensing*, ideado a principios de los 60 para designar cualquier medio de observación remota, si bien se aplicó fundamentalmente a la fotografía aérea, principal sensor de aquel momento. En sentido amplio, la Percepción Remota no engloba sólo los procesos que permiten obtener una imagen, sino también su posterior tratamiento en el contexto de una determinada aplicación.

Si hemos definido la Percepción Remota espacial como aquella técnica que permite adquirir imágenes de la superficie terrestre desde sensores instalados en plataformas

distantes de ésta, estamos asumiendo que entre la Tierra y el sensor existe una interacción energética, ya sea por reflexión de la energía solar o de un haz energético artificial, ya por emisión propia. A su vez, es preciso que ese haz energético recibido por el sensor sea almacenado convenientemente, bien a bordo de satélites, bien en las estaciones receptoras, de tal manera que pueda interpretarse para una determinada aplicación. El análisis de la información recibida se incluye también –en sentido amplio– dentro del estudio de la Percepción Remota, aunque esté más allá de los procesos de adquisición propiamente dichos.

Un sistema de Percepción Remota incluye los siguientes elementos (figura 2.1):

- i. Fuente de energía que supone el origen de la radiación electromagnética que detecta el sensor. Puede tratarse de un foco externo a éste, en cuyo caso se habla de *Percepción Remota pasiva*, o de un haz energético emitido por el sensor, lo que se denomina *Percepción Remota activa*. La fuente de energía más importante es, obviamente, el sol.
- ii. Cubierta terrestre, formada por distintas masas de vegetación, suelos, agua o construcciones humanas, que reciben la señal energética procedente de (i) y la reflejan o emiten de acuerdo a sus características físicas.
- iii. Sistema sensor, compuesto por el sensor propiamente dicho, y la plataforma que lo alberga. Tiene como misión captar la energía procedente de las cubiertas terrestres, codificarla y grabarla o enviarla directamente al sistema de recepción.
- iv. Sistema de recepción - comercialización, en donde se recibe la información transmitida por la plataforma, se graba en un formato apropiado, y, tras las oportunas correcciones, se distribuye a los interesados.
- v. Intérprete, que convierte esos datos en información temática de interés, ya sea visual o digitalmente, de cara a facilitar la evaluación del problema en estudio.
- vi. Usuario final, encargado de analizar el documento fruto de la interpretación, así como de dictaminar sobre las consecuencias que de él se deriven.

El símil de la visión humana puede ayudar a entender el significado de estos componentes. El ojo humano (iii) recibe un haz energético procedente de los objetos exteriores (ii) por reflejo de la luz solar (i). Esta señal se transmite al cerebro (iv) que es capaz de formar una serie de imágenes sobre el mundo real que le circunda. El individuo que observa es a la vez intérprete (v) y usuario final (vi) de la imagen detectada, lo que le permite tomar decisiones sobre su propio comportamiento. En otras palabras, la visión humana forma un sistema de Percepción Remota –muy sofisticado por cierto–, puesto que nos permite caracterizar con mucho detalle los objetos que observamos.

Sin embargo, nuestra percepción visual presenta también algunas limitaciones. Por lo tanto, se restringe por la sensibilidad espectral de nuestras células sensoras, que sólo nos permiten apreciar un determinado tipo de energía, denominado, por esta razón, espectro visible. Otras formas energéticas, como el calor, no son observables visualmente, por lo que parece conveniente contar con unos 'ojos artificiales' que amplíen nuestras propias posibilidades. Por otro lado, el campo de visión humano está limitado por nuestras propias posibilidades.

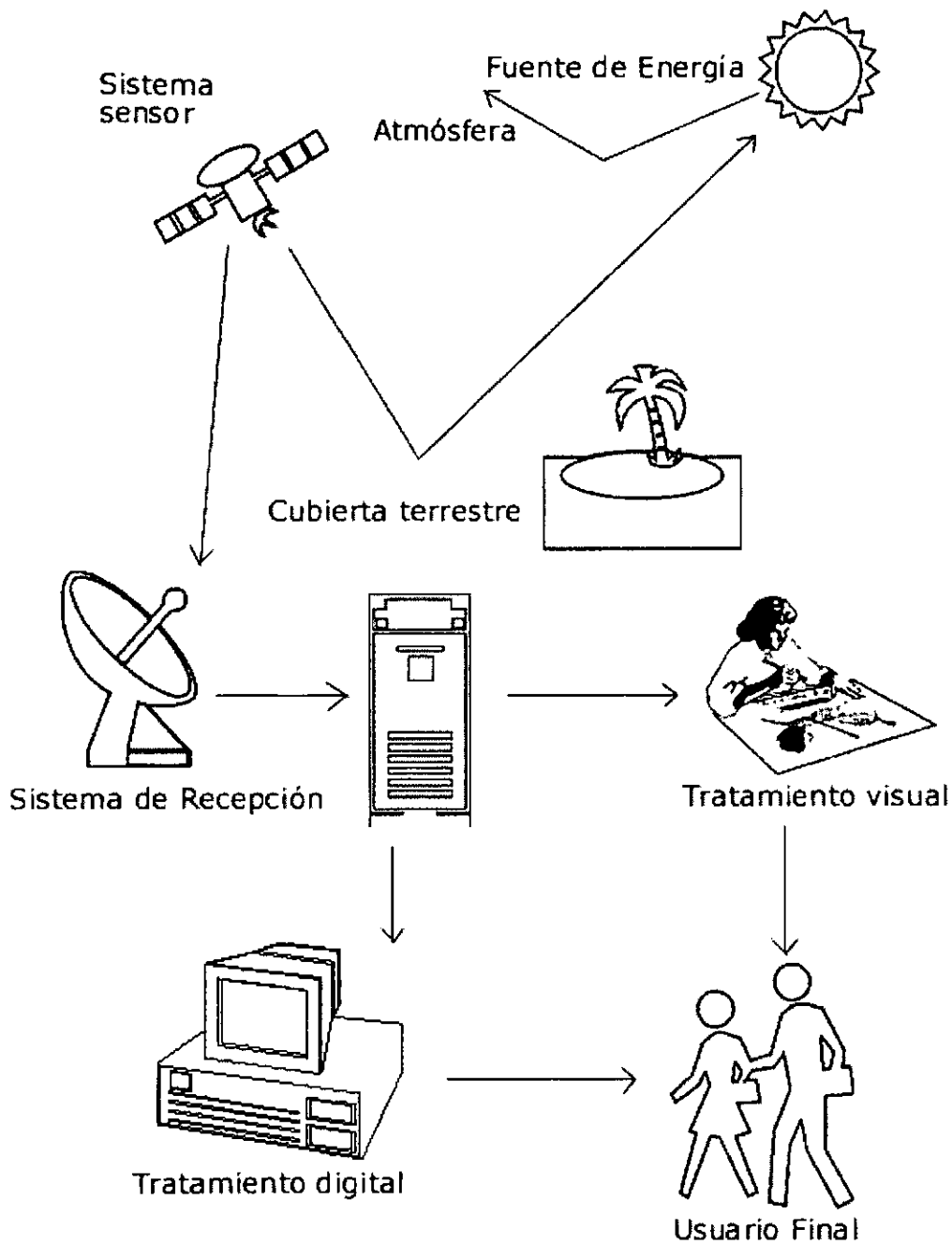


Figura 2.1

Componentes de un sistema de percepción remota.

Por lo tanto, el campo de visión humano está limitado por nuestra propia estatura, o a la que podemos adquirir desde observaciones naturales, por lo que cuenta con un pequeño radio de acción. Además, se trata de una percepción oblicua. Ambos factores dificultan observar fenómenos muy extendidos, como inundaciones o incendios donde las estimaciones visuales pueden ser imprecisas. Para disminuir estas limitaciones, se ha acudido a sensores artificiales, montados sobre plataformas situadas a una cierta

altitud. Con ellos se tiene acceso a tipos de energía no visibles (ultravioleta, infrarrojo, microondas) y, además, desde una nueva perspectiva, vertical y panorámica. Esta nueva información expande notablemente nuestro conocimiento del medio que nos rodea, facilitando la interpretación de múltiples procesos que afectan al Planeta.

El creciente empleo de distintos sensores de observación terrestre está originando no sólo una enorme cantidad de información, sino también una nueva forma de estudiar la superficie terrestre. La Percepción Remota, junto a las técnicas añejas de almacenamiento y proceso de la información geográfica, permiten disponer de una masa ingente de datos sobre el territorio, algunos de ellos accesibles en tiempo real. Para aprovechar mejor estas técnicas, el recurrir a las computadoras resulta casi inevitable. Una vez salvado el necesario entrenamiento, estos equipos amplían notablemente nuestra capacidad de análisis, ayudan a elaborar una interpretación rápida y económica del problema bajo estudio, a la vez que permiten abordar análisis integrados, difícilmente viables con otros medios tradicionales. En definitiva, y gracias a nuestros recursos, podemos cada día poner más énfasis en el planteamiento de modelos y el análisis de resultados –tareas éstas donde la inteligencia humana resulta insustituible– librándola de otras labores mecánicas (acopio y ordenamiento de datos) que han consumido tradicionalmente las mejores fuerzas del investigador.

2.2 Conceptos básicos

2.2.1 Radiación Electromagnética

Históricamente, las propiedades de la radiación electromagnética se han explicado por dos teorías aparentemente contrapuestas: aquella que la concibe como un haz ondulatorio (Huygens, Maxwell), y aquella otra que la considera como una sucesión de unidades discretas de energía, fotones o cuantos, con masa igual a cero (Plank, Einstein). Actualmente, parece que las dos teorías se pueden compaginar, pues se ha demostrado que la luz puede comportarse de acuerdo a ambos planteamientos.

Según la teoría ondulatoria, la energía electromagnética se transmite de un lugar a otro siguiendo un modelo armónico y continuo, a la velocidad de la luz y conteniendo dos campos de fuerzas ortogonales entre sí: eléctrico y magnético (figura 2.2).

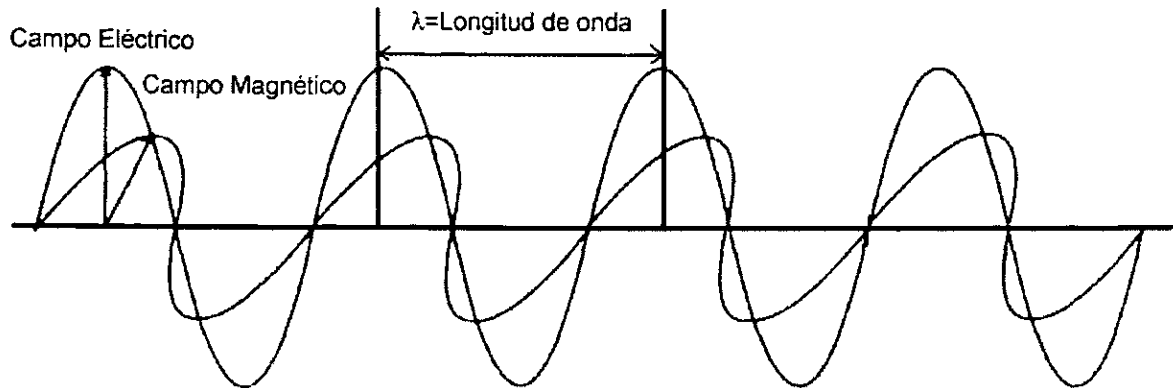


Figura 2.2
Esquema de una onda electromagnética.

Las características de este flujo energético pueden describirse por dos elementos, longitud de onda (λ) y frecuencia (F). La primera hace referencia a la distancia entre dos picos sucesivos de una onda, mientras que la frecuencia designa el número de ciclos pasando por un punto fijo en una unidad de tiempo. Ambos elementos están inversamente relacionados:

$$c = \lambda F$$

donde c indica la velocidad de la luz (3×10^8 m/s), λ expresa la longitud de onda (habitualmente en *micrómetros*, $1\mu\text{m} = 10^{-6}$ m) y F la frecuencia (*Hertz*, ciclos por segundo). En definitiva, a mayor longitud de onda, menor frecuencia y viceversa, por lo que basta indicar un sólo termino para identificar propiamente el tipo de energía relacionado.

Gracias a la teoría cuántica, podemos calcular la cantidad de energía transportada por un fotón, siempre que se conozca su frecuencia:

$$Q = h F$$

donde Q es la energía radiante de un fotón (en Joules), F la frecuencia y h la constante de Plank (6.63×10^{-34} Js). Sustituyendo, podemos así mismo expresar:

$$Q = h (c/\lambda)$$

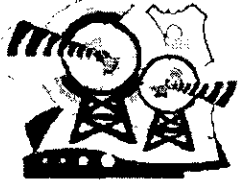
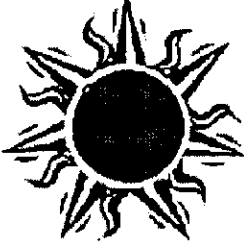
Lo que significa, en definitiva, que a mayor longitud de onda –o menor frecuencia- el contenido energético será menor y viceversa.

2.2.2 Espectro Electromagnético

De las fórmulas anteriores, se deduce que podemos definir cualquier tipo de energía radiante en función de su longitud de onda o frecuencia. Aunque la sucesión de valores de longitud de onda es continua, suelen establecerse una serie de bandas en donde la radiación electromagnética manifiesta un comportamiento similar. La organización de

estas bandas de longitudes de onda o frecuencia se denomina espectro electromagnético.

El espectro electromagnético comprende desde las longitudes de onda mas cortas (rayos gamma, rayos X), hasta las kilométricas (figura 2.3).

Región	Nombre	Rango Específico
<p style="text-align: center;">Ondas de Radio</p>  <p style="text-align: center;">$10^8 - 10^{-3} \text{ m} / 3 - 3 \times 10^{10} \text{ Hz}$</p>	Ultra Low Frequencies (ULF) Frecuencias Ultra Bajas	3 - 30 Hz
	Extremely Low Frequencies (ELF) Frecuencias Extremadamente Bajas	30 - 300 Hz
	Voice Frequencies (VF) Frecuencias de voz	300 Hz - 3 KHz
	Very Low Frequencies (VLF) Frecuencias Muy Bajas	3 - 30 KHz
	Low Frequencies (LF) Frecuencias Bajas	30 - 300 KHz
	Medium Frequencies (MF) Frecuencias Medias	300 KHz - 3 MHz
	High Frequencies (HF) Frecuencias Altas	3 - 30 MHz
	Very High Frequencies (VHF) Frecuencias Muy Altas	30 - 300 MHz
	Ultra High Frequencies (UHF) Frecuencias Ultra Altas	300 MHz - 3 GHz
	Super High Frequencies (SHF) Frecuencias Super Altas	3 - 30 GHz
	Extremely High Frequencies (EHF) Frecuencias Extremadamente Altas	30 - 300 GHz
	<p style="text-align: center;">Infrarrojo</p>  <p style="text-align: center;">$10^{-3} - 10^{-6} \text{ m} / 10^{11} - 10^{14} \text{ Hz}$</p>	Lejano
Medio		30 - 3 μm
Cercano		3 - 0.75 μm




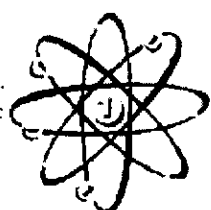
Región	Nombre	Rango Específico
Visible 	Rojo	770 - 622 nm
	Naranja	622 - 597 nm
	Amarillo	597 - 577 nm
	Verde	577 - 492 nm
	Azul	492 - 455 nm
	Violeta	455 - 390 nm
Ultravioleta  ZONO	UV - A	400 - 320 nm
	UV - B	320 - 280 nm
	UV Vacío	Menos de 200 nm
Rayos X 	Rayos X	$10^{-9} - 10^{-11}$ m ($10^{17} - 10^{19}$ Hz)
Rayos Gama 	Rayos Gama	$10^{-11} - 10^{-13}$ m ($10^{19} - 10^{21}$ Hz)

Figura 2.3

Espectro Electromagnético.

Desde el punto de vista de la Percepción Remota, conviene destacar una serie de bandas espectrales, que son las más frecuentemente empleadas en la tecnología actual. Su denominación y amplitud varían, según distintos autores, si bien la terminología más común es la siguiente:

- **Espectro visible (0.4 a 0.7 μ m):** Se denomina así por tratarse de la única radiación electromagnética que pueden percibir nuestros ojos, coincidiendo con las longitudes de onda en donde es máxima la radiación solar. Suelen distinguirse tres bandas elementales, que se denominan azul (492 - 455 nm), verde (577 - 492 nm) y rojo (770 - 622 nm), en razón de los colores primarios que nuestros ojos perciben a esas longitudes de onda.

- **Infrarrojo próximo (0.7 - 1.3 μm):** A veces se denomina también infrarrojo reflejado o fotográfico, puesto que parte de él puede detectarse a través de filmes dotados de emulsiones especiales. Resulta de especial importancia por su capacidad para discriminar masas vegetales y concentraciones de humedad.
- **Infrarrojo medio (1.3 - 8 μm):** En donde se entremezclan los procesos de reflexión de la luz solar y de emisión de la superficie terrestre. Resulta idóneo para estimar el contenido de humedad en la vegetación y para la detección de focos de alta temperatura.
- **Infrarrojo lejano o térmico (8 - 14 μm):** Incluye la porción emisiva del espectro terrestre, en donde se detecta el calor proveniente de la mayor parte de las cubiertas terrestres.
- **Microondas (a partir de 1 μm):** Son de gran interés por ser un tipo de energía bastante transparente a las nubes.

2.2.3 Sensores Remotos

Los sistemas remotos son sistemas fotográficos u óptico-electrónicos capaces de detectar y registrar, en forma de imágenes, el flujo de energía radiante reflejado o emitido por objetos distantes.

Un flujo de radiación electromagnética, al propagarse por el espacio, puede interactuar con superficies u objetos, siendo reflejado, absorbido o emitido por dichas superficies u objetos. Las variaciones que estas interacciones producen en el flujo considerado dependen fuertemente de las propiedades físico-químicas de los elementos en la superficie.

Una cámara fotográfica con flash puede servir como ejemplo de un sistema sensor: cuando el sistema de la cámara es activado, el flash es accionado y emite radiación. La radiación fluye hacia el objeto y es reflejada por éste, hacia el sistema óptico de la cámara. Entonces la radiación reflejada es registrada sobre el plano de la película, que constituye un detector fotoquímico de radiación. Una imagen del patrón de radiación es formada en la película y después es desarrollada químicamente.

Cualquier sistema sensor presenta los siguientes componentes necesarios para captar la radiación electromagnética (figura 2.4):

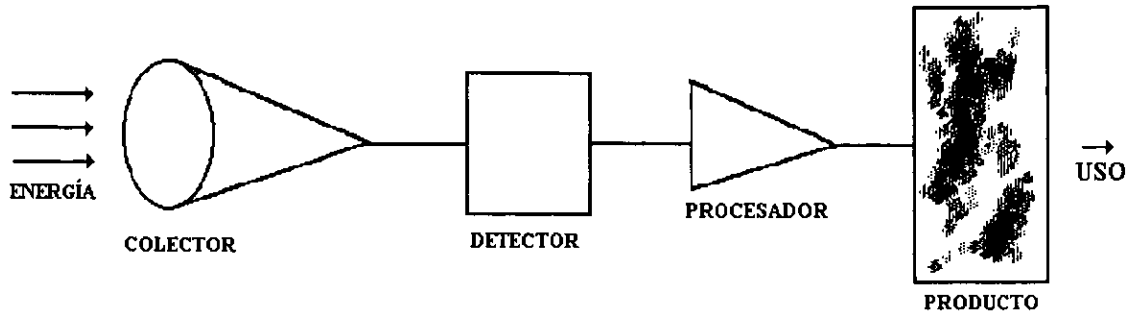


Figura 2.4

Componentes del sistema sensor.

donde:

colector – recibe la energía a través de una lente, espejo, antenas, etc.

detector – capta la energía colectada de una determinada franja del espectro.

procesador – la señal registrada es sometida a un procesamiento (revelado, ampliación, etc.) a través del cual se obtiene el producto.

producto – contiene la información necesaria para el usuario.

2.2.3.1 Tipos de sensores

Los sensores pueden ser clasificados en función de la fuente de energía o en función del tipo de producto que se obtiene de ellos.

En función de la fuente de energía:

- a) **Pasivos:** no poseen una fuente propia de radiación. Miden la radiación solar reflejada. Por ejemplo, los sistemas fotográficos.
- b) **Activos:** poseen su propia fuente de radiación electromagnética, trabajando en franjas restringidas del espectro. Por ejemplo, los radares.

En función del tipo de producto:

- a) **No imageadores:** no generan una imagen de la superficie observada. Por ejemplo, los radiómetros (salida en dígitos o gráficos) y los espectrorradiómetros (firma espectral). Son esenciales para la adquisición de información minuciosa sobre el comportamiento espectral de los objetos de la superficie terrestre.
- b) **Imageadores:** se obtiene como resultado una imagen de la superficie observada. Suministran información sobre la variación espacial de la respuesta espectral de la superficie observada.
 - b.1. Sistema de cuadro ('framing systems'): adquieren la imagen de la totalidad de la escena en un mismo instante.
 - b.2. Sistema de barrido ('scanning systems'): la imagen se obtiene de forma secuencial, mediante un barrido de la escena de interés.
 - b.3. Sistema fotográfico.

Los imageadores no fotográficos (sistema de imageamiento por barrido) se originaron para cubrir la laguna dejada por los problemas inherentes al uso del más difundido sensor óptico, la cámara fotográfica. Ésta, a pesar de presentar condiciones más fáciles de operación y costos, tiene una limitación en captar la respuesta espectral, debido a que las películas cubren solamente el espectro del ultravioleta próximo al infrarrojo lejano. También este tipo de sensor limita las horas de sobrevuelo, ya que debido a fenómenos atmosféricos no es posible observar frecuentemente el suelo a grandes altitudes.

Como los datos de estos sensores no fotográficos son colectados en forma de señales eléctricas, estos datos podrán ser fácilmente transmitidos para estaciones distantes, donde un procesamiento electrónico hará el análisis discriminatorio.

2.3 Sistemas de Percepción Remota

De acuerdo a la plataforma donde se ubique el sensor, se distinguen dos grandes tipos de sistemas de Percepción Remota:

- Aérea
- Espacial

2.3.1 Percepción Remota Aérea

Se considera Percepción Remota Aérea aquella que se obtiene a partir de las plataformas:

- Globo
- Avión
- Helicóptero

Los sensores más usados son:

- Cámara Fotográfica
- Cámara Multibanda
- Cámara Digital
- Cámara de Video
- Barredor Multiespectral
- Barredor Térmico

Las fotografías aéreas son las imágenes de percepción remota más comúnmente utilizadas. Son usadas para el análisis de todo tipo de recursos naturales, cartografía y estudios de aproximación en los sectores público y privados.

Las ventajas básicas que ofrece esta técnica son:

- Permite tener una vista de los detalles de la superficie en estudio en su contexto espacial a menor escala que la imagen satelital. Esto la transforma en una herramienta muy útil para trabajar en áreas pequeñas.
- Flexibilidad. Por el hecho de usar plataformas aéreas para embarcar los sensores de imagen, es muy fácil rediseñar el plan de trabajo según las necesidades que vayan surgiendo.
- La fotografía nos da una visión instantánea de procesos bajo condiciones dinámicas (inundaciones, derrames de petróleo, tráfico, incendios). Si bien este tipo de trabajo se realiza muy bien con imágenes satelitales, la fotografía aérea permite desplazarse al lugar de interés sin demoras. De esta forma se dispone de un medio complementario al uso de plataformas satelitales especialmente apto para cubrir aquellos procesos dinámicos con períodos de cambio inferiores al de los satélites (catástrofes).

El sensor usado en esta técnica es la película fotográfica. Existen diferentes tipos de película pensados para diferentes tipos de problemas.

Las películas a las que estamos más acostumbrados son aquellas que son capaces de 'ver' la imagen en forma similar a como lo hace el ojo humano. Estas películas, con algunas características adicionales, son usadas en fotografía aérea y permiten analizar y resolver muchos problemas. En el siguiente ejemplo podemos ver el tipo de detalles que se ven al volar sobre la superficie terrestre a 2500m de altura (figura 2.5).

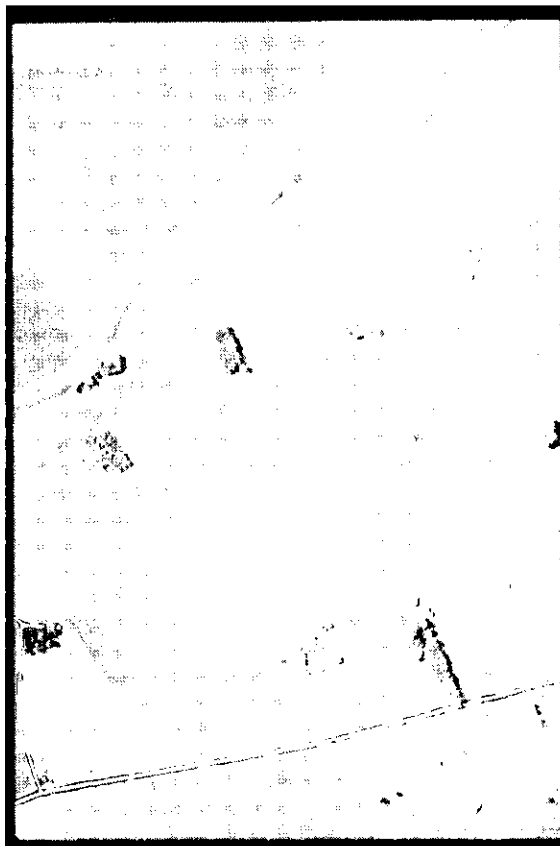


Figura 2.5

*Campos cercanos al Río Negro (Uruguay)
Color Normal.*

Para otros tipos de problema es necesario recurrir a películas que 'ven colores' en donde el ojo humano no ve nada. Estamos hablando en este caso de la película sensible a la radiación infrarroja. El ejemplo siguiente corresponde a la misma imagen, pero usando una película sensible al Infrarrojo Cercano. Esta foto es en falso color, ya que no podemos hablar de color en el infrarrojo. Lo que se hizo fue asignar en la película el color rojo a la radiación infrarroja cercana reflejada por la superficie en estudio (figura 2.6).



Figura 2.6

*Campos cercanos al Río Negro (Uruguay)
Falso Color Infrarrojo.*

Otra característica importante es que esta película filtra completamente el color azul, eliminando el efecto de dispersión de luz atmosférica, conocido como *haze* [43]. Este efecto disminuye el contraste de la imagen y le da una tonalidad azul grisáceo. Si comparamos ambas fotos vemos la eliminación de este efecto en la fotografía infrarroja (figura 2.7).

En esta figura es posible ver y analizar los detalles (por ejemplo, heterogeneidades en las áreas cultivadas) que se presentan en la fotografía infrarroja, en comparación con su correspondiente versión en color real.



Figura 2.7

Eliminación del efecto haze en la fotografía infrarroja [43].

Cuando hablamos de fotografía infrarroja debemos de tener cuidado de no confundirla con la fotografía térmica. En esta técnica también estamos trabajando con radiación infrarroja, pero de diferente tipo a la que es sensible la película infrarroja que se usó en este ejemplo. En esta película queda registrado un rango de radiación infrarroja que da información acerca de la naturaleza química de los objetos fotografiados. Algunos de los trabajos posibles a realizar con esta técnica son:

- Agricultura de precisión (*Precision Farming*).
- Clasificación de cultivos.
- Detección y seguimiento de enfermedades.
- 'Estrés' fisiológico de los cultivos.
- Cálculo de áreas forestales y análisis de estado sanitario.
- Índice de defoliación de cultivos por plagas.

2.3.2. Percepción Remota Espacial

Se considera Percepción Remota Espacial aquella que se obtiene a partir de las plataformas que pueden ser naves espaciales tripuladas, estaciones orbitales o satélites autónomos, que giran alrededor de la Tierra en órbitas polares, oblicuas o ecuatoriales.

Sus ventajas son: la visión sinóptica, la periodicidad en la toma de datos, el sincronismo con el Sol y el bajo costo por hectárea de los datos generados.

Algunos de los sistemas satelitales más comúnmente usados son:

- ERS-1 y 2 (ESA)
- SPOT 1, 2, 3 y 4 (Francia)
- IKONOS
- LANDSAT 5 y 7 (USA)
- RADARSAT (Canadá)
- JERS-1 (Japón)

2.3.2.1 Satélites ERS 1 y 2

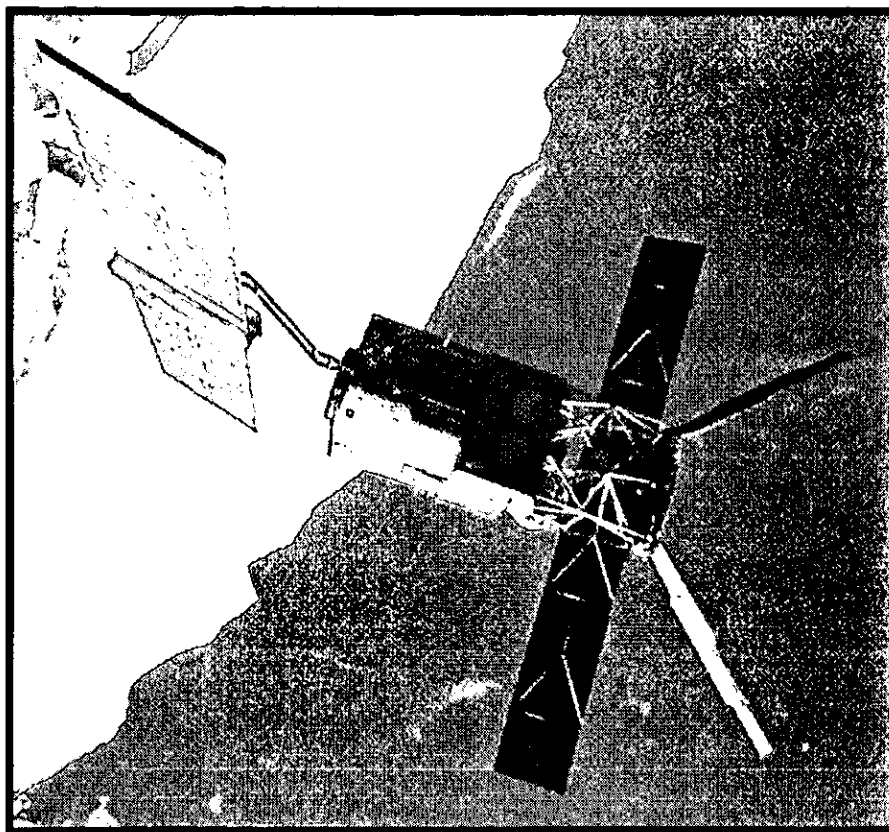


Figura 2.8

Satélite ERS-2.

En julio de 1991 la Agencia Espacial Europea lanzó el ERS²-1, el más grande y más sofisticado satélite de vuelo libre construido a lo largo y ancho de Europa, llevando con él un centro de sensores activos de microondas.

² ERS - *ESA Remote Sensing Satellite*

El programa ERS-1 fue diseñado para servir a una gran variedad de usuarios con un amplio rango de productos y servicios. En respuesta a esto y otros objetivos de misiones muy cambiantes, las instalaciones en tierra del ERS-1 combinan facilidades distribuidas y centralizadas. La pieza clave entre el usuario y la parte de cobro por los servicios se encuentra en facultad de la ESA³-ESRIN⁴ en Italia, mientras que el monitoreo y control del satélite son responsabilidad del ESOC⁵ en Alemania. Tomando ventaja de los centros especiales existentes, las *ESA Processing and Archiving Facilities* han sido establecidas en Francia, Alemania, Italia y el Reino Unido, los cuales manejan los datos recibidos por las estaciones terrenas de la ESA y la red de otras estaciones distribuidas a lo largo del mundo.

Los principales usos del ERS-1 son el estudio de los océanos, de los glaciares y en meteorología. Actualmente se están recolectando datos de las regiones polares del planeta y de los océanos del sur. Además, la capacidad para producir imágenes de alta resolución en todos los climas recolectará información valiosa de áreas continentales y zonas costeras. El ERS-1 es tanto un sistema experimental como un sistema operacional, puesto que ha demostrado los nuevos conceptos y tecnologías de los equipos e instalaciones de espacio y tierra. Esto también ilustrará algunos requerimientos operacionales para la rápida entrega de datos que proveen de contribuciones meteorológicas significativas, pronóstico del estado del mar y monitoreo de la distribución de los hielos oceánicos.

El sistema ERS-1 utiliza tecnología de microondas avanzada para obtener medidas e imágenes independientemente de las nubes existentes en la atmósfera terrestre y de las condiciones de la luz solar. En comparación con los sistemas satelitales contemporáneos, el ERS-1 es único en la medición de ciertos parámetros, incluyendo aquellos del estado del mar, ondas en la superficie del mar, tierra y hielo bajo cualquier clima. También mide la temperatura de la superficie del mar con mayor exactitud que cualquiera de los sistemas espaciales actuales.

La misión del satélite ERS-2 comprende segmentos de espacio y tierra similares a los del ERS-1 con mejoras/adhesiones a fin de proveer imágenes ópticas, medidas de la concentración del ozono en la estratósfera y tropósfera, así como la determinación exacta de la posición orbital del satélite y la posición de los puntos de referencia en tierra (fijamente geodésico).

El sensor que utilizan ambos satélites es el SAR⁶.

Radar de apertura sintética (SAR)

El Radar de Apertura Sintética difiere del Radar de Apertura Real principalmente en el hecho que se usa una antena de longitud efectiva, sólo que se sintetiza al usar procesamiento de señales en vez de usar una antena físicamente grande. El propósito es mejorar la resolución en la dirección de vuelo.

³ ESA - *European Space Agency*

⁴ ESRIN - *European Space Research Institute*

⁵ ESOC - *European Space Operation Center*

⁶ SAR - *Synthetic Aperture Radar*

Para lograr esto, es necesario producir la imagen de radar en dos pasos por separado. A continuación se hace una explicación geométrica del proceso, asumiendo procesamiento y grabación análogos. Primero se genera una grabación fotográfica durante la adquisición de datos. La grabación es parecida a una fotografía convencional. Un segundo paso, un proceso de transformación, sustituye la grabación preliminar (la cual se conoce como filme de datos, filme de señal o filme de la historia de fase) en un producto fotográfico final llamado filme de imagen o filme de mapa.

En el primer paso (adquisición de datos) la antena transmite y recibe pulsos que son necesariamente coherentes (esto es, las crestas o valles de una onda son predecibles) en contraste con el sistema SLR⁷. Cada pulso produce una línea de imagen a lo largo de la dirección de movimiento del filme, de la misma forma que con el sistema de apertura real.

En el sistema SAR, siempre está presente una onda coherente de referencia. Cada señal transmitida o recibida se compara con esta señal de referencia sumándose a ella. A lo anterior se le llama patrón de interferencia. El grabado producido en la película durante la adquisición de datos, la película de señal o historia de datos, difiere del sistema de imágenes de radar de 'fuerza bruta' en dos aspectos:

1. Los pulsos de radar transmitidos no se enfocan tan cercanamente sino fielmente extendidos.
2. El grabado incluye no sólo a los ecos de radar en sí, sino la suma de los ecos y de una onda de referencia coherente. Ésta es la razón por la que un radar es llamado coherente.

Para el intérprete de fotografía convencional, los resultados de la película de señal representan una masa de grises sin significado. Es útil recordar que el ancho de la señal en dirección en rango es simplemente el resultado de la longitud del pulso. El rango no es afectado por este proceso.

Además, la antena es objeto de movimiento, claramente no es en cada posición, pero existe el movimiento. Como resultado de ello, los ecos recibidos tienen una frecuencia diferente a la onda de referencia.

El segundo paso es la transformación de la señal en una imagen de mapa-correlación de la imagen SAR. La luz normal emitida por un bulbo eléctrico está compuesta de muchas ondas con longitud de onda de alrededor de 30 cm. La relación existente entre las diferentes fases de cada componente no puede predecirse. En luz coherente, sin embargo, la relación entre las fases de dos ondas es constante. La fase concierne al estado de la onda, esto es, una cresta o un valle ocurren en la onda en un determinado momento y lugar. Si se hace incidir luz coherente a través de la película de datos, la difracción de la luz en las señales grabadas enfocarían la luz en una imagen única apilada detrás de la película de datos.

Existe un número variado de puntos de vista, pero esencialmente equivalentes, que explican el procesamiento de la fase de historia para el sistema SAR. Tres de ellos son:

- Adición de vectores

⁷ SLR - *Side Looking Radar*

La adición de vectores caracteriza la vista geométrica en la aproximación de la antena sintética. Matemáticamente, esta aproximación se describe como la aplicación de las funciones de peso de retardo de fase en la señal recibida en cada 'nodo' de la antena sintética, que coherentemente lista todos los regresos para cada elemento con resolución particular en cada uno de los nodos.

- Correlación

Puede usarse en los datos recibidos usando una función de referencia de historia de fase para cada elemento de resolución.

- Filtrado de frecuencia

Se refiere a procesar el contenido de frecuencias de los datos recibidos. Esto se logra comparando una función de historia de fase con las señales grabadas, lo anterior en el dominio de la frecuencia –en vez del dominio del tiempo– que se obtiene al aplicar la transformada de Fourier en el dominio del tiempo.

Imágenes SAR

Las imágenes SAR son una subdivisión de las imágenes *Side-Looking Radar*. Existe una pequeña similitud entre el principio de la fotografía convencional y el de éste último tipo de imágenes. Las fotografías son tomadas con cámaras con sistemas de lentes que utilizan la luz visible y radiación cercana a la infrarroja. Además con el resultado de una exposición instantánea. Las imágenes de radar se forman con microondas en un modo cinemático. El rango que se utiliza para crear imágenes de radar se extiende de longitudes de onda de menos de 1 cm a longitudes de muchos metros. Se considerarán las longitudes de onda mayores para usos donde la capacidad de penetración de la radiación electromagnética es de valor particular y la resolución geométrica, además de la imagen de ruido, son de importancia menor.

El elemento 'óptico' de un sistema de radar es una antena. Una lente no puede enfocar radiaciones con longitudes de onda de microondas, pero una antena sí. De manera similar a los lentes, se trabajará con la energía transmitida o recibida con la propiedad de la forma de la antena.

Existen dos tipos principales de imágenes de radar en este caso: las de escaneo circular PPI⁸ y las imágenes *Side-Looking*. Las imágenes PPI limitan sus aplicaciones al monitoreo de aeropuertos y tráfico marino o aéreo.

Entre las imágenes *Side-Looking* existe una división principal, en base a dos diferentes principios de operación. De gran interés es el radar de apertura sintética SAR debido a su habilidad de crear imágenes de mayor resolución que los menos complejos radares de apertura real o de 'fuerza bruta'. Este simple sistema se denota usualmente como SLAR⁹ o SLR¹⁰.

⁸ PPI es el acrónimo de *Plan Position Indicator*

⁹ SLAR es el acrónimo de *Side-Looking Airbone Radar*

El término RADAR¹¹ ha sido utilizado de forma genérica para clasificar los sistemas que operan en la región de frecuencias del microondas.

La supresión de ruido es una tarea de preprocesamiento de imágenes SAR muy importante. En principio, la caracterización de imágenes SAR involucra modelar los efectos de muchos parámetros de los sistemas SAR. Entre ellos se encuentran los niveles de potencia de señal y ruido, longitud de onda, geometría del terreno, geometría de los objetos hechos por el hombre, estadísticas de ruido, efecto de propagación atmosférica, y polarización transmitida y recibida. También se incluyen las operaciones debidas al procesamiento y compensadores por error de fase en el procesador de historia de fase (PHP) o correlador de imágenes SAR, tal como los errores por movimientos en la plataforma satelital.

El terreno varía en el macrorrelieve, rugosidad o microrelieve y en las constantes dieléctricas correspondientes. Los objetos hechos por el hombre pueden tener formas complejas que además son construidas con una gran variedad de materiales de diferentes constantes dieléctricas. En ambos casos, las propiedades geométricas y eléctricas afectan la reflectividad y por lo tanto la imagen captada por el radar del objeto.

Las estadísticas de señal a ruido son importantes en el contraste de la imagen. Por otro lado, la longitud de onda y la polarización afectan la reflectividad y transmitividad del objeto. El *speckle* es un efecto coherente que es consecuencia de la presencia de dispersiones severas múltiples dentro de la misma zona tomada, resultando así un efecto acumulativo. Generalmente este efecto existe en radares de apertura real y sintética. En sistemas SAR el uso de sistemas de luz incoherente reduce el efecto del *speckle*, pero a costa de reducir la resolución. Este mismo proceso reduce otros efectos de ruido diferentes al *speckle*. Los efectos atmosféricos pueden reducir la coherencia de las señales en un sistema SAR, desenfocando y algunas veces creando falsos detalles.

¹⁰ SLR es el acrónimo de *Side-Looking Radar*

¹¹ RADAR es la abreviatura de *Radio Detection And Ranging*

2.3.2.2 Satélites SPOT 1, 2 y 4

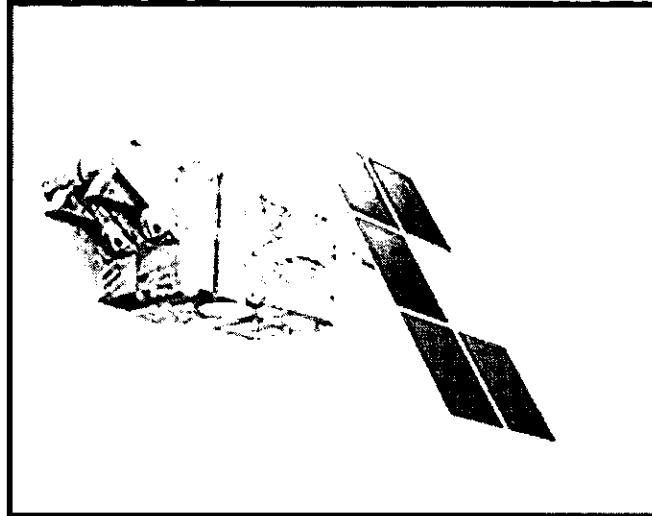


Figura 2.9
Satélite SPOT 4.

El programa SPOT (*Système Provoaire d'Observation de la Terre*), fue desarrollado por el CNES (*Centre National d'Études Spatiales*), en Francia, con la colaboración de Bélgica y Suecia.

Gracias a los satélites SPOT 1, SPOT 2, SPOT 3 y SPOT 4 el sistema ha estado operando por cerca de 15 años.

- SPOT 1 - fue lanzado el 22 de Febrero de 1986 y todavía captura datos.
- SPOT 2 - fue lanzado el 22 de Enero de 1990 y todavía es operacional.
- SPOT 3 - fue lanzado el 26 de Septiembre de 1993. Un accidente ocurrió en el SPOT 3, el 14 de Noviembre de 1997, y éste dejó de funcionar.
- SPOT 4 - lanzado el 24 de Marzo de 1998.
- El grupo de trabajo de ingeniería para desarrollar el SPOT 5 ha iniciado de tal manera que el satélite pueda ser lanzado a principios del año 2002 para asegurar la continuidad del servicio.

Dado el incremento de la demanda de las imágenes SPOT, sobre todo en el hemisferio norte en las épocas de siembra, el SPOT 2 y el SPOT 1 todavía operarán en conjunto con el SPOT 4.

Las características únicas de los SPOT –alta resolución, imágenes estéreo y capacidad de volver a visualizar– permite coleccionar datos en áreas de especial interés para varias aplicaciones como la cartografía, la agricultura, el ambiente, etc.

Los satélites SPOT han permitido capturar, desde 1986, más de 7.5 millones de imágenes, las cuales han sido archivadas, lo cual constituye un récord no igualado en nuestro planeta.

La percepción remota basada en satélites no es sólo un segmento clave para la observación de la tierra, sino también una fuente principal de información geográfica. Los datos SPOT han llegado a ser esenciales para un gran número de usuarios quienes miran la continuación de la familia SPOT para probar qué tan fiables y operacionales llegan a ser.

La órbita SPOT es polar, circular y sincronizada con el sol y en fase. La órbita es polar porque, en conjunto con la rotación de la tierra alrededor del eje polar, la inclinación del plano orbital (98 grados) permite al satélite volar sobre cada punto de la tierra durante un ciclo de 26 días. La órbita es circular porque si las imágenes de diferentes ubicaciones son satisfactorias para la comparación, éstas deben adquirirse a la misma altitud. Entonces, la órbita debe ser circular, o tener una altitud constante relativa a la superficie de la tierra. La órbita está en sincronía con el sol porque, debido a la comparación válida de las imágenes de una zona adquirida en diferentes días, depende de la similitud de las condiciones de iluminación, el plano orbital debe formar un ángulo constante relativo a la dirección del sol. Esto último permite asegurar que el satélite vuela sobre un punto dato en la misma hora local. Por último, la órbita está en fase porque, cada 26 días, cada satélite SPOT vuela sobre los mismo puntos de la tierra. Durante este período de tiempo, éstos hacen un número entero de revoluciones siguiendo un ciclo completo. El mismo patrón es entonces repetido una y otra vez.

Los satélites SPOT utilizan un sensor denominado HRV¹². Cada satélite contiene dos sensores de este tipo. La posición de cada entrada de los espejos del HRV puede ser comandado desde la base de control para observar la región de interés no necesariamente vertical bajo el satélite. Luego entonces, el HRV ofrece una capacidad de vistas oblicuas. El ángulo de vista puede ser ajustado entre los ± 27 grados, relativos a la vertical. Se emplean dos modos de imageamiento, la pancromática (P) y multispectral (XS). Ambos HRVs pueden operar en cualquiera de los dos modos, simultánea o individualmente.

¹² HRV - *High Resolution Visible*

2.3.2.3 Satélite IKONOS

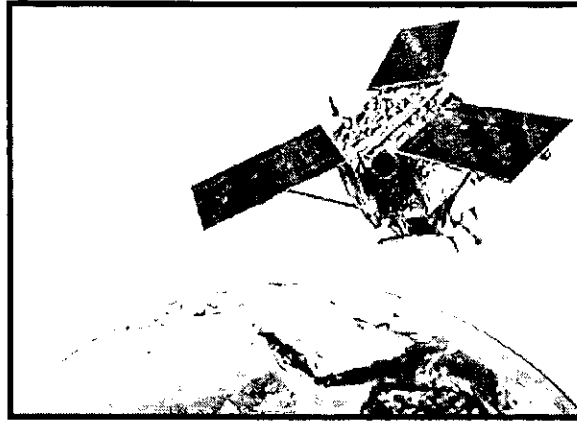


Figura 2.10

Satélite IKONOS.

IKONOS es un satélite que fue lanzado el 24 de septiembre de 1999. Este satélite es capaz de producir imágenes que pueden ser usadas en varias aplicaciones del mercado incluyendo planeación urbana, monitoreo ambiental, mapeo, ocurrencia de desastres naturales, planeación de redes de telecomunicaciones, exploración de gas y petróleo y monitoreo de la agricultura.

IKONOS es un satélite con la posibilidad de obtener imágenes de 1 metro de resolución pancromática (0.45 - 0.90 micrómetros), 4 metros en imágenes multiespectrales (R, G, B) y un metro en color (Banda 1 [Azul]: 0.45-0.53 μm , Banda 2 [Verde]: 0.52-0.61 μm , Banda 3 [Rojo]: 0.64-0.72 μm y Banda 4 [Infrarrojo cercano]: 0.77-0.88 μm). El tamaño de las escenas es especificado por los usuarios.

Los datos son colectados en 11 bits por píxel, pero las imágenes finales pueden ser entregadas en 11 bits por píxel (en 2 bytes) o en 8 bits (un byte) por píxel.

Este satélite envía las imágenes capturadas hacia las estaciones base en la tierra a una tasa comparable a ver 50 canales de televisión simultáneamente.

La cámara que utiliza el satélite tiene la potencia de un telescopio de 30 pies (9.14 metros) de largo y contiene un conjunto de espejos alineados. Estos espejos fueron diseñados para ser lo más ligeros posible. Además de que la cámara está a una temperatura constante de 68° F.

3 FUNDAMENTOS

3.1 Procesamiento de Imágenes

Se puede decir que los primeros intentos de manipulación, almacenamiento y transmisión de imágenes surgieron en 1920, cuando se desarrolló un sistema de transmisión de fotografías a través de cable trasatlántico usando códigos telegráficos. Este sistema permitía una codificación de la imagen en cinco niveles de gris, y luego fue mejorado a 15 niveles de gris en 1929. Dicho sistema redujo la transmisión de imágenes, con calidad para publicación en periódicos, de dos semanas (tiempo requerido para el envío por barco) a alrededor de quince minutos.

Un segundo momento importante se da al final de los años cincuenta y consiste en la necesidad de analizar imágenes derivadas de las actividades ligadas a la guerra fría: la fotografía aérea y espacial. Las primeras imágenes cercanas de un cuerpo celeste diferente de la tierra fueron tomadas por la Unión Soviética a principios de 1959, cuando la sonda espacial Luna 3 fotografió el lado oscuro de la luna. Las fotografías fueron reveladas dentro de la sonda espacial y posteriormente se hizo una transmisión tipo facsimilar. La calidad de las imágenes era bastante mala, pero después de procesarlas digitalmente se pudo realizar un aceptable atlas del lado oscuro de la luna.

Estas técnicas tuvieron un renacimiento en 1964 con el procesamiento de las imágenes recibidas de la sonda estadounidense Ranger 7, que transmitía cerca de la luna. Este procesamiento era básicamente la corrección de distintos tipos de distorsión producida por las cámaras de TV. Algunas de las distorsiones que se eliminaron fueron: la distorsión geométrica producida por la diferencia en velocidades de barrido entre el vidicon en la sonda y el tubo reproductor en la Tierra; la no linealidad fotométrica debida a una respuesta irregular en el tubo; y un ruido oscilatorio debido a la

contaminación de la señal de televisión por los sistemas electrónicos de la sonda. Afortunadamente, estos tres problemas se pudieron solucionar y a partir de esta época y con el impulso fundamental del *Jet Propulsion Laboratory*, el procesamiento digital de imágenes ha cobrado gran importancia en diversos ámbitos de la vida actual.

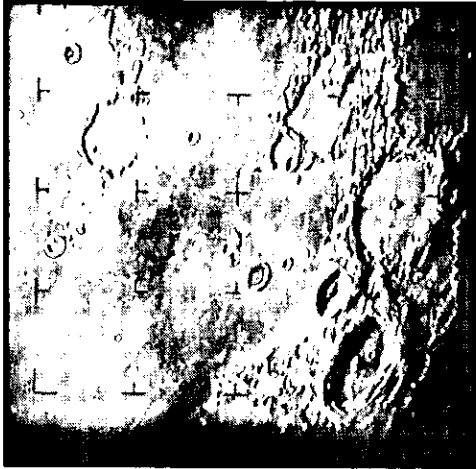


Figura 3.1

Imagen tomada por el Ranger 7 el 31 de julio de 1964, 17 minutos antes de estrellarse sobre la superficie lunar (fuente: Nasa).

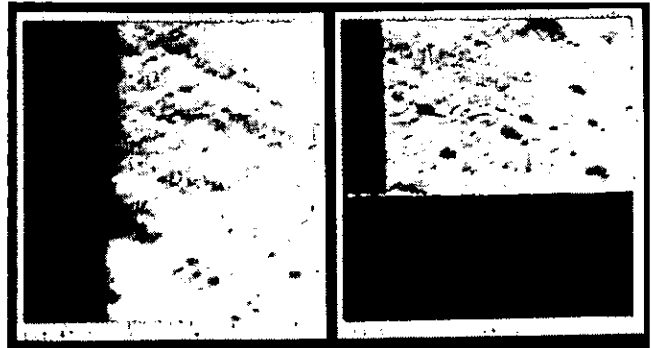


Figura 3.2

Ultimas dos imágenes tomadas a altitudes de 1000 y 500 m sobre la superficie lunar. Las imágenes están cortadas porque la sonda se estrelló antes de terminar la transmisión. Sobre estas imágenes el ruido oscilatorio es muy visible (fuente: Nasa).

La foto siguiente muestra una imagen de Plutón, procesada después de haber sido adquirida por el telescopio espacial Hubble después de su reparación. Este ejemplo muestra las capacidades de la combinación del procesamiento óptico y digital para ver un objeto 500 veces más pequeño que la tierra, a 6000 millones de kilómetros de distancia.



Figura 3.3

Plutón y su satélite, Charon tomados desde el telescopio espacial Hubble (fuente: NASA).

La imagen anterior es la de mayor resolución de este cuerpo celeste, ya que no se han enviado sondas para el estudio de este planeta. Se puede observar claramente a Charon, el satélite de Plutón, que tiene la característica de tener un tamaño muy grande en relación con el planeta al que acompaña.

Dentro del procesamiento de imágenes se pueden distinguir tres tipos de procesamiento para mejorar las características o para evaluar algunos aspectos estadísticos de la escena que se esté analizando. A continuación se describen estos métodos.

En el **procesamiento óptico** se emplean arreglos de lentes para mejorar las características de la imagen. Algunos ejemplos típicos son los anteojos que se emplean para dotar de visión normal a las personas con deficiencias en este rubro y los procesos empleados en fotografía, incluyendo las lentes de las cámaras y las de las ampliadoras fotográficas. En este último ejemplo se debe mencionar que el trabajo en el cuarto oscuro es un ejemplo muy claro del empleo de varias técnicas para mejorar las características de las imágenes.

Un ejemplo del procesamiento óptico es el mejoramiento de la calidad de las imágenes telescópicas. En la imagen siguiente, en el recuadro inferior derecho, se muestra la imagen de una región de la galaxia M100 tomada con el telescopio Hale, de 200 pulgadas, en el monte Palomar en California.

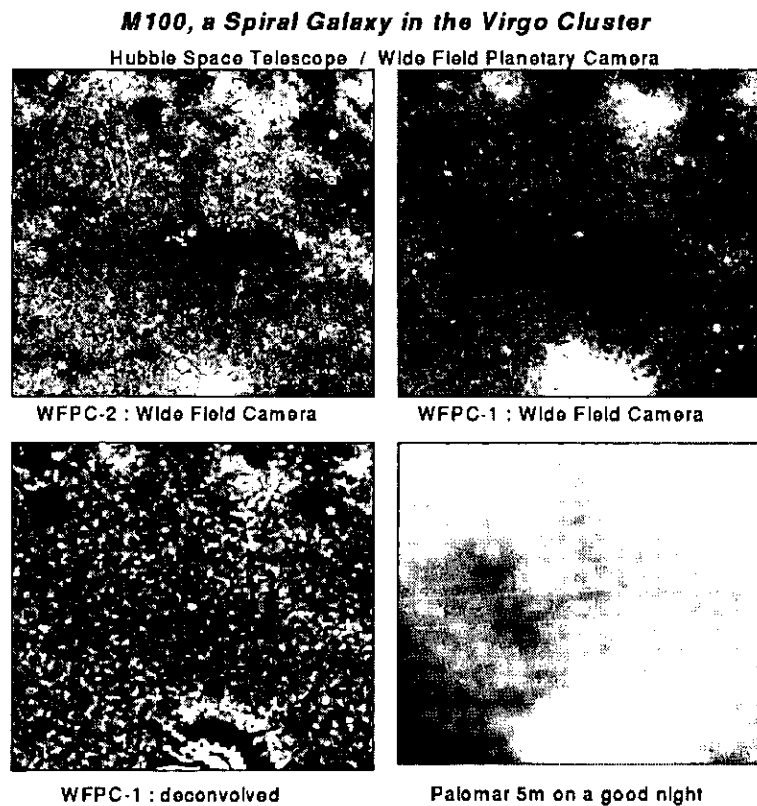


Figura 3.4

Comparación entre distintas modalidades de visualización del mismo segmento estelar.

Este telescopio permite obtener la máxima calidad posible en observaciones terrestres. Se tiene una resolución de un poco menos de un segundo de arco, en una noche con condiciones mejores que el promedio para observaciones terrestres. En el cuadrante superior derecho se muestra la misma región de la galaxia M100, tomada por el telescopio espacial Hubble en noviembre de 1993, unos días antes de su reparación. Este telescopio, que orbita la tierra a más de 400 km. de altitud, evita que las variaciones atmosféricas interfieran con la adquisición de imágenes de objetos lejanos. Sin embargo, al ponerse en marcha en 1990 se observó que el juego de espejos en su interior tenía una importante aberración esférica que no se había detectado anteriormente. A pesar de esto, la resolución de la imagen es mucho mejor que aquellas tomadas desde la tierra, lo que permitió el uso de este telescopio a pesar de los errores de aberración ya mencionados. Sin embargo, sólo se podían ver objetos relativamente brillantes. El cuadrante inferior izquierdo corresponde a la misma imagen, después de una reconstrucción asistida por computadora. Aunque los algoritmos empleados pueden realzar secciones borrosas, no se pueden observar las regiones con poca intensidad luminosa. Adicionalmente, se introducen errores en el brillo en diversas regiones de la imagen, como se puede apreciar cuando esta imagen se compara con la imagen tomada con el sistema óptico corregido del telescopio espacial. Esto impide obtener información cuantitativa. En el cuadrante superior izquierdo se muestra la misma región de la galaxia M100, tomada después de la corrección óptica de las aberraciones esféricas del telescopio espacial en 1993.

El procesamiento analógico consiste en el uso de sistemas eléctricos para alterar las imágenes. Ejemplos de esto son las imágenes por televisión, donde existen los controles de contraste e intensidad de éstas.

El procesamiento digital consiste en hacer el mapeo de una imagen a puntos definidos discretamente, a los cuales se les asigna un par de coordenadas y un valor de intensidad. La alteración de los valores de intensidad por medio de una computadora permite efectuar con gran facilidad operaciones de realce y de análisis de la imagen. La imagen siguiente muestra los resultados de efectuar un proceso de realce y restauración en una fotografía deteriorada. A la imagen original se le hizo un filtrado pasa-altas, de manera que el efecto de 'borrosidad' de la imagen se minimiza. Por otro lado, se ajustó el nivel de intensidad luminosa a partir de la especificación del histograma, a fin de mejorar el contraste de la imagen. Por último, las secciones deterioradas se restauraron por medio de un proceso manual que incluyó el reemplazo de secciones dañadas por secciones cercanas.



Figura 3.5

Imagen original e imagen después de la restauración.

El Procesamiento Digital de Imágenes se basa en diversas teorías, algunas de las cuales pueden ser: los Campos Aleatorios de Markov, la Transformada Polinomial, los Algoritmos Genéticos, las Redes Neuronales, las Ondículas (Wavelets), la Lógica Difusa, etc.

En esta trabajo se hace uso de los Campos Aleatorios de Markov y la Transformada Polinomial, principalmente. Dichas teorías se enuncian a continuación.

3.2 Teoría de los Campos Aleatorios de Markov

3.2.1 Vecindades y cliques

Muchos problemas de visión pueden ser catalogados como problemas de etiquetado, en los cuales la solución a un problema es un conjunto de etiquetas asignadas de acuerdo a las características de los *pixeles* de una imagen. El etiquetado es también la representación natural para el estudio de los Campos de Markov.

El problema de etiquetado se especifica en términos de un conjunto de *sites* y un conjunto de etiquetas. Un *site* se define como un punto o una región en el espacio euclidiano, por ejemplo, un punto o una región de la imagen tal como una esquina, un segmento de línea, etc.

3.2.1.1 Sistemas de vecindad

Para un *site* en particular (i,j) , v_{ij} es identificada como la vecindad del *site*.

v_{ij} está definida como vecindad si:

1. $(i,j) \notin v_{ij}$

Los *pixeles* que pertenecen a la vecindad de (i,j) no incluyen al *pixel* analizado.

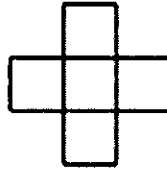
2. Si $(m,n) \in v_{ij}$ entonces $(i,j) \in v_{mn}, \forall (m,n) \in L$

Reciprocidad, es decir, si un *pixel* (i,j) tiene una vecindad v_{ij} , para todos los *pixeles* de esa vecindad se tiene que (i,j) forma parte de sus respectivas vecindades.

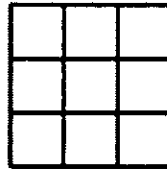
A la colección de todos los subconjuntos de L correspondientes a una definición específica de vecindad se le conoce como sistema de vecindad.

$$v = \{v_{ij} : (i, j) \in L, v_{ij} \subset L\}$$

En el sistema de vecindades de primer orden, también llamado sistema de 4 vecinos, cada *site* (central) tiene cuatro vecinos como se muestra en la figura siguiente.



En el sistema de vecindades de segundo orden, también llamado sistema de 8 vecinos cada *site* (central) tiene 8 vecinos.



Nótese que los *sites* sobre o cerca de la frontera tienen menos vecinos.

Un sistema homogéneo se define como:

$$v = \{(m, n) \in L : 0 < (m - i)^2 + (n - j)^2 \leq d^2\}$$

donde i, j, m, n y d son enteros siendo d el orden de la vecindad.

3.2.1.2 Cliques

Un clique C es el *site* (*pixel*) mismo o bien un conjunto de *sites*, todos vecinos los unos con los otros.

Un *clique* par (L, v) es un subconjunto de L tal que:

1. C consiste en un solo *pixel*, o
2. Para $(i, j) \neq (k, l)$, $(i, j) \in C$ y $(k, l) \in C$, implica que $(i, j) \in v_{k,l}$

La figura siguiente (figura 3.6) muestra los *cliques* asociados a cada sistema de vecindad. La figura (a) muestra el sistema de vecindad de primer orden, mientras que los incisos (b) y (c) muestran los sistemas de segundo y tercer orden, respectivamente. Las figuras (d) y (e) muestran los *cliques* asociados al sistema de primer orden. Las figuras (d) a la (h) muestran los *cliques* asociados al sistema de segundo orden.

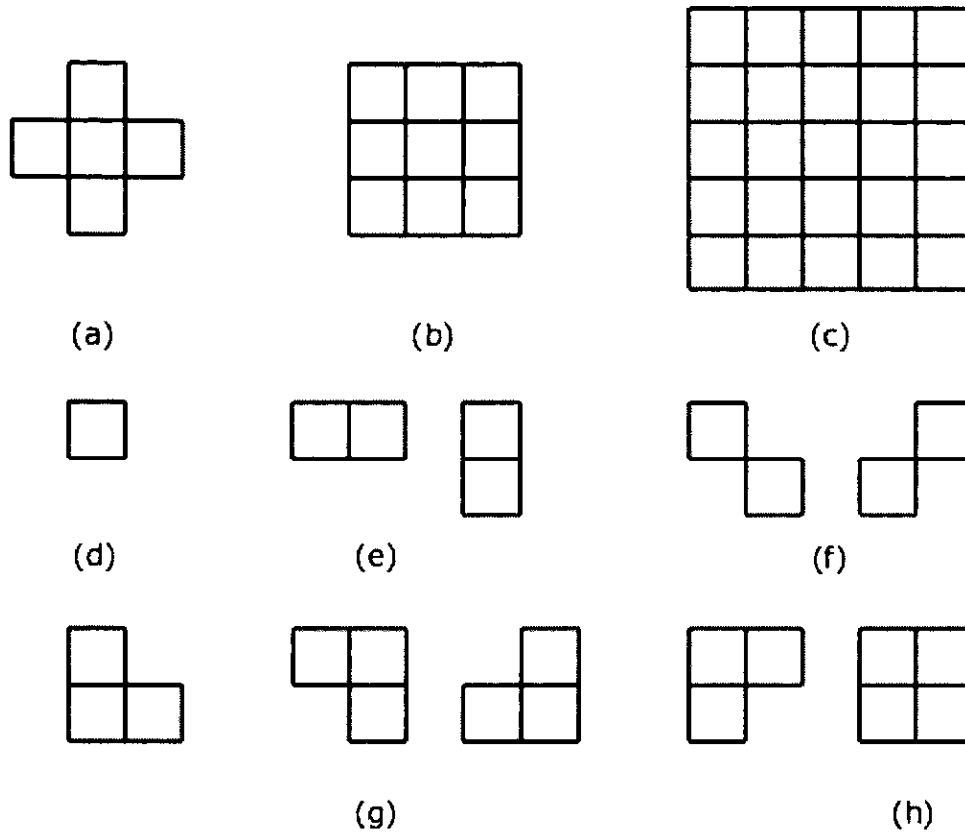


Figura 3.6
Cliques y vecindad en una rejilla regular de sites.

3.2.2 Campos Aleatorios de Markov y Distribución de Gibbs

Sea $F = \{F_1, \dots, F_m\}$ una familia de variables aleatorias definidas sobre el sistema de vecinos S donde cada F_i toma el valor f en el conjunto discreto de etiquetas L . A la familia F se le conoce como campo aleatorio.

Se dice que F es un campo aleatorio de Markov en S con respecto a un sistema de vecindades N si y sólo si se satisfacen las siguientes dos condiciones:

$$P(f) > 0, \forall f \in F \tag{3.1}$$

Positividad

$$P(f_i | f_{S-i}) = P(f_i | f_{N_i}) \tag{3.2}$$

$$P(X_{ij} = x_{ij} | X_{mn} = x_{mn}, (m,n) \neq (i,j)) = P(X_{ij} = x_{ij} | X_{mn} = x_{mn}, (m,n) \in \eta_{ij})$$

Propiedad de Markov

Donde $S - \{i\}$ es el conjunto de diferencia, $f_{S-\{i\}}$ denota el conjunto de etiquetas en los sites en $S - \{i\}$, $f_{N_i} = \{f_i | i' \in N\}$, η_{ij} es la vecindad del *pixel* (i,j) , Q es el conjunto de sites y x_{ij} está contenido en Q .

Esto es, un campo de Markov X tiene la propiedad de que la distribución condicional de X en un *pixel* particular, dados los valores de todos los demás *pixeles* en la imagen completa, es igual a la distribución condicional dados sólo los valores de sus vecinos.

La propiedad de positividad también es muy importante. Bajo esta propiedad el campo de Markov X puede tener una distribución de Gibbs. La distribución de Gibbs es una medida de probabilidad de una propiedad de un *pixel* en particular, derivada de las propiedades de sus vecinos y tiene la siguiente forma:

$$P(x_{ij} = x) = \frac{1}{Z} \exp\left\{-\frac{1}{T} \sum_{C \in c} V_c(x)\right\} \quad (3.3)$$

donde $x \in N$, C es el conjunto de *cliques* relacionado con el *pixel* (i,j) , $V_c(x)$ es la función de potencial asociada con el *clique* C , T es la constante de control tomada de temperatura que puede considerarse igual a 1 si no se requiere de ninguna consideración especial y Z es la constante de normalización definida por:

$$Z = \sum_x \exp\left\{-\frac{1}{T} \sum_{C \in c} V_c(x)\right\} \quad (3.4)$$

En la práctica, un campo aleatorio puede asumirse generalmente como homogéneo. En otras palabras, $V_c(x)$ es irrelevante respecto a la posición del *clique* en la red L y depende sólo del tipo y de los valores de los *pixeles* en el *clique* analizado.

Los campos de Markov que se utilizan en el procesamiento de imágenes adquieren una forma discreta, esto es, el campo aleatorio X consiste sólo de variables aleatorias discretas con K valores tomados de Q . Este campo de Markov aleatorio discreto se conoce algunas veces como un campo logístico multinivel. La pieza clave de éste es que se le asocia un parámetro con cada tipo de *clique*.

Siendo uno de sus componentes el *pixel* bajo consideración; el potencial para un *clique* de dos *pixeles* se define como

$$\begin{aligned} V_c(x) &= -\beta && \text{si } x_{mn} = x \text{ y } (m,n) \in C \\ V_c(x) &= \beta && \text{en cualquier otro caso} \end{aligned}$$

donde el parámetro β es una constante positiva. De acuerdo con la definición de la distribución de *Gibbs*, x es el índice del tipo de región a determinar para el *pixel* considerado y puede tomar cualquier valor en Q . A fin de lograr una decisión, es necesario evaluar todas las probabilidades que un *pixel* adquiere en un tipo de región y después tomar la máxima de ellas. Para esto, es suficiente analizar los atributos del *pixel* central y la de sus *pixeles* vecinos. Por esta razón, la siguiente fórmula, con $k=1, 2, \dots, k$, es preferible y más clara.

$$\begin{aligned} V_c(x) |_{x=q_k} &= -\beta && \text{si } x_{mn} = q_k \text{ y } (m,n) \in C \\ V_c(x) |_{x=q_k} &= \beta && \text{en cualquier otro caso} \end{aligned}$$

$$P(x_{ij} = x) |_{x=q_k} = \frac{1}{Z} \exp \left\{ -\frac{1}{T} \sum_{C \in c} V_c(x) |_{x=q_k} \right\} \quad (3.5)$$

con

$$Z = \sum_{k=1}^K \exp \left\{ -\frac{1}{T} \sum_{C \in c} V_c(x) |_{x=q_k} \right\} \quad (3.6)$$

3.2.3 El modelo

Como se mencionó anteriormente, para los propósitos de segmentación de imágenes, las características de la imagen deben modelarse de tal manera que ésta pueda dividirse en K regiones diferentes, cada una siguiendo una distribución normal caracterizada por una media μ_k y una varianza σ_k^2 . Teniendo en cuenta la distribución de Gibbs del proceso región y haciendo $T=1$, se tiene una nueva medida de probabilidad de que el *pixel* (i,j) pertenezca a la región q_k y es como sigue:

$$P(x_{ij} = q_k | Y_{ij}) \propto \exp \left\{ -\frac{1}{\sigma_k^2} (Y_{ij} - \mu_k)^2 - \sum_{C \in c} V_c(x) |_{x=q_k} \right\} \quad (3.7)$$

Este modelo se compone de dos partes. La primera depende de las estadísticas de los datos y de lo cerca que la intensidad de la región se encuentre de los datos. La segunda depende de las propiedades de los *pixeles* vecinos y el central. De acuerdo con la definición potencial, es claro que si el número de *pixeles* pertenecientes a los K tipos de regiones en la vecindad es exactamente el mismo, el segundo término en la ecuación anterior tomará el mismo valor para todos los tipos de regiones. Esto no cambiará el orden relativo de los valores de probabilidad y los atributos del *pixel* bajo consideración serán determinados solamente por los datos. En contraste, si éstos números son diferentes entre ellos, el valor de probabilidad correspondiente al máximo ganará el mayor incremento y entonces el *pixel* central tendrá más probabilidades de ser asignado al tipo de región relacionado. Si el valor del término dependiente de los datos llegara a ser despreciable, por ejemplo, si β tomara valores muy grandes, entonces, el atributo del *pixel* será determinado principalmente por el atributo de sus vecinos. En otras palabras, el primer término tomará efecto sólo cuando el número de *pixeles* pertenecientes a los K tipos de regiones en la vecindad del *pixel* central difieran uno de otro, y harán el atributo de cada *pixel* en la segmentación de la imagen más consistente con la mayoría de sus vecinos. Finalmente, tomando la probabilidad logarítmica de la ecuación anterior puede obtenerse fácilmente una clasificación de tipo de máxima probabilidad.

3.2.4 Modelo de Ising

El modelo de más bajo orden utilizado para transmitir información es el que utiliza dos etiquetas solamente. Es muy empleado debido a su simplicidad y bajo costo computacional. Se codifican en la energía de Gibbs como potenciales de cliques de parejas de sites. Con los potenciales de cliques de hasta dos sites, la función de energía toma la siguiente forma:

$$U(f) = \sum_{i \in S} V_1(f_i) + \sum_{i \in S} \sum_{i' \in N_i} V_2(f_i, f_{i'}) \quad (3.8)$$

donde $\sum_{i \in S}$ es equivalente a $\sum_{\{i\} \in C_1}$ y $\sum_{i \in S} \sum_{i' \in N_i}$ a $\sum_{\{i, i'\} \in C_2}$.

Cuando $V_1(f_i) = f_i G_i(f_i)$ y $V_2(f_i, f_{i'}) = \beta_{i, i'} f_i f_{i'}$ donde $G_i(\cdot)$ son funciones arbitrarias y $\beta_{i, i'}$ son constantes que reflejan la interacción entre el par de sites i e i' , la energía es:

$$U(f) = \sum_{\{i\} \in C_1} f_i G_i(f_i) + \sum_{\{i, i'\} \in C_2} \beta_{i, i'} f_i f_{i'}$$

La ecuación anterior es conocida como *auto modelo*. Las funciones de auto modelo pueden clasificarse de acuerdo a las suposiciones hechas en las funciones individuales f_i 's.

Se dice que un auto modelo es un modelo autológico, si las f_i 's toman valores en el conjunto de etiquetas discreto $L = \{0, 1\}$. La energía correspondiente toma entonces la siguiente forma:

$$U(f) = \sum_{\{i\} \in C_1} \alpha_i(f_i) + \sum_{\{i, i'\} \in C_2} \beta_{i, i'} f_i f_{i'} \quad (3.9)$$

Cuando N es el sistema de vecindad más cercano en una red (los cuatro vecinos más cercanos en dos dimensiones o los dos vecinos más cercanos en una dimensión), el modelo autológico se reduce al modelo de Ising.

3.2.5 Modelo de Potts

En el caso de imágenes de varios niveles de gris, es posible modelar las interacciones entre los *pixeles* a efecto de segmentar en varias clases:

Siendo a y b sites de una vecindad en 4 conexidad:

$$V(a, b) = -\beta \quad \text{si} \quad f_a = f_b$$

$$V(a,b) = \beta \quad \text{si} \quad f_a \neq f_b$$

o lo que es lo mismo:

$$V(a,b) = -K\delta(f_a, f_b)$$

donde δ representa el símbolo de kronecker.

Así, el modelo de Potts puede expresarse como:

$$V(a,b) = -K\delta(f_a, f_b) \quad \text{donde } k = 2\beta$$

3.2.6 Algoritmo de simulación y optimización

3.2.6.1 Algoritmo de simulación

- Síntesis de un campo de Gibbs.
- Muestreador de Gibbs.

Se trata de un algoritmo iterativo basado en la construcción a partir de un estado inicial cualquiera, por ejemplo una cadena de Markov $\Pi(0)$ de configuraciones sucesivas en las cuáles se espera encontrar 'buenas' propiedades de convergencia. Para los casos de los algoritmos de simulación y de optimización se considera una serie de sites $S^{(n)}$, donde:

$S^{(n)}$ $n \in \mathbb{N}$ tal que $\forall s \in S$, s es 'visitado' frecuentemente en un número infinito de veces.

El procedimiento es el siguiente:

1. Barrido de la imagen.
2. En cada uno de los *pixeles/sites* se calculan las probabilidades locales en función de la vecindad y de la estructura Markoviana del campo.
3. En función de las posibilidades condicionales locales, efectuar un sorteo aleatorio de estado, por ejemplo del nivel de gris de la etiqueta, etc.
4. El procedimiento es repetido en todos los *pixeles* de todas las líneas de la imagen, esto es, para todos los sites del campo de Markov.

El barrido de toda la imagen (de la totalidad del campo de Markov) debe ser repetido un gran número de veces para obtener buenos resultados.

Muestreador de Gibbs

Algoritmo:

- Inicialización de la simulación: $\mathcal{X}^{(0)}$.
- En cada etapa (n):
 - Escoger el site $r = \mathcal{S}^{(n)}$. Su valor es $\mathcal{X}_r^{(n)}$ y su configuración de vecindad es $V_r^{(n)}$.
 - Sortear aleatoriamente su nuevo valor $\mathcal{X}_r^{(n+1)}$, según una probabilidad de transición 'Gibbsiana':

$$\Pi(\mathcal{X}_r^{(n)} \rightarrow \mathcal{X}_r^{(n+1)}) = P(\mathcal{X}_r = \mathcal{X}_r^{(n+1)} | V_r^{(n)})$$

Para toda la configuración inicial $\eta \in \Omega$ y toda configuración $\omega \in \Omega$ se tiene:

$$\lim_{n \rightarrow \infty} P((\mathcal{X}^{(n)} = \omega | \mathcal{X}^{(0)} = \eta)) = \prod(\omega) \quad (3.10)$$

La expresión anterior se identifica como el teorema de relajación o muestreador de Gibbs, donde,

η =configuración inicial

ω =configuración n

Ω =conjunto de configuraciones

esto es, cuando n tiende al infinito la configuración ω sigue siendo una cadena de Markov.

$\prod(0)$ corresponde a la medida de probabilidad definida por una cadena de Markov. Con ello, existe la convergencia hacia una realización dada, con una probabilidad que sigue la distribución de Gibbs global.

En teoría, para asegurar un correcto muestreo de la imagen, siguiendo una ley de probabilidad, es necesario efectuar un número infinito de barridos. En la práctica, pueden emplearse dos criterios para finalizar el algoritmo:

- a) Por el aspecto visual del resultado.
- b) Por la variación de la tasa de cambio entre la imagen n y la n-1.

Algoritmo de metrópolis a temperatura constante

Se trata también de un método de relajación en el que se efectúan barridos sobre la imagen, y en cada *pixel* se tiene la posibilidad de efectuar cambios en el estado del *pixel*. Estos cambios involucran una variación de la energía del campo $U(x)$.

Algoritmo:

- Inicialización de la solución $x^{(0)}$.
- En cada etapa (n):
- Escoger al site $r=s^{(n)}$. Su valor es $x_r^{(n)}$ y su configuración de vecindad es $V_r^{(n)}$.
- Efectuar el sorteo de una variable de distribución uniforme $\xi \in \Omega$.
- Calcular la variación de la energía local asociada.

$$\Delta U = U(\xi | V_r^{(n)}, Y) - U(x_r^{(n)} | V_r^{(n)}, Y)$$

en este último paso existen dos eventualidades:

- a) $\Delta U \leq 0$: puesto que la energía disminuye, el cambio es aceptado incondicionalmente, entonces $x_r^{(n+1)} = \xi$.
- b) $\Delta U > 0$: el cambio es aceptado o rechazado efectuando un sorteo aleatorio con las probabilidades respectivas,

$$p = \prod(x_r^{(n+1)} = \xi) = \exp(-\Delta U)$$

donde X es la imagen original y Y la imagen de clases.

Sea ξ_2 la variable obtenida en el sorteo aleatorio, $x_r^{(n+1)} = \xi$ si $\xi_2 < p$.

Comentarios:

Se observa que para toda $\eta \in \Omega$ y toda $\omega \in \Omega$,

$$\lim_{n \rightarrow \infty} P((x^{(n)} = \omega | x^{(0)} = \eta) = \prod(\omega) \tag{3.11}$$

ΔU corresponde a la variación tanto de la energía local como de la energía global asociada a la transición $x_r^{(n+1)} = \xi$. Tal caso resulta de la estructura Markoviana de la distribución a posteriori $P(Y = y | X = x)$ y de la estructura de $V(y | x)$ bajo la forma de la suma de los potenciales y cliques.

3.2.6.2 Algoritmo de optimización

Habiendo definido ya una distribución de probabilidad a posteriori $U(y | x)$ existe un algoritmo de optimización, es decir para la estimación de y.

Algoritmo de Máximo A Posteriori

Uno de los grandes intereses de los métodos Markovianos reside en la posibilidad que ellos ofrecen de poder efectuar estimaciones en el sentido MAP¹³. Aquí interesa maximizar la ley de probabilidad de la función buscada (segmentación, restauración, etc.), condicionada a las medidas hechas (en la imagen).

Por Bayes:

$$P(X | Y) = \frac{P(Y | X)P(X)}{P(Y)} \quad (3.12)$$

$P(X | Y)$ describe enteramente al proceso de adquisición de la imagen.
 $P(X)$ es el modelo Markoviano, y puede ser, en este caso, de Potts o de Ising.

El método de MAP consiste efectivamente en encontrar la configuración x que maximice la distribución *a posteriori*:

$$x = \arg \max P(X = x | Y = x) = \arg \min U(x | y) \quad (3.13)$$

En la práctica, se utiliza el algoritmo de Metrópolis a temperatura variable a fin de minimizar la energía $U(x|y)$. El algoritmo general es el siguiente:

- Inicialización de la solución: $x^{(0)}$.
- Dar un conjunto de temperaturas tendientes a cero:

$$\lim_{n \rightarrow \infty} T_n = 0$$

- Para cada iteración (n):

Escoger el site: $r = s^{(n)}$. Su valor es $x_r^{(n)}$ y su configuración de vecindad es $V_r^{(n)}$.

Crear una variable aleatoria uniforme $\xi \in \Omega$.
 Calcular la variación de energía local asociada:

$$\Delta U = U(\xi | V_r^{(n)}, Y) - U(x_r^{(n)} | V_r^{(n)}, Y)$$

$$x_r^{(n+1)} = x_r^{(n)} \quad \text{si } \Delta U \leq 0$$

$$x_r^{(n+1)} = \xi \quad \text{con la probabilidad:}$$

$$\pi(x_r^{(n)} \rightarrow \xi) = \exp\left[-\frac{\Delta U}{T_n} - \text{si } \Delta U > 0\right]$$

¹³ MAP es el acrónimo de *Maximum A Posteriori*

En la práctica se simula la probabilidad $\pi = \exp\left(-\frac{\Delta U(n)}{T}\right)$ efectuando el sorteo aleatorio de una variable $\eta \in [0,1]$. Si $0 \leq \eta \leq p$, se efectúa la transición $X_r^{(n+1)} = \xi$. En caso contrario $X_r^{(n+1)} = X_r^{(n)}$.

El recocido simulado (*Simulated Annealing*)

Los campos de Markov, además de ser un conjunto de métodos para el análisis y la representación de imágenes, están ligados a un método específico de optimización: el recocido simulado. Este algoritmo permite resolver el problema de estimación en el sentido del MAP. Además, el recocido simulado asegura la convergencia hacia un óptimo global del problema.

En la práctica se utiliza un algoritmo en Metrópolis de temperatura variable con el fin de minimizar la energía $U(x|y)$. El objetivo del problema de optimización consiste en encontrar un elemento de Ω_0 donde:

$$\Omega_0 = \{\omega \in \Omega \mid \omega \text{ es de energía mínima.}\}$$

- a) Si U^* es la energía máxima de todas las configuraciones y U_* es la energía mínima, la desviación entre las dos energías es $\Delta U = U^* - U_*$.
- b) Al inicializar un muestreador de Gibbs con una temperatura T_n , que varía después de cada barrido, mediante el siguiente esquema:

- $T_n \rightarrow 0$ cuando $n \rightarrow \infty$.
- $T_n \geq \frac{N\Delta U}{\log(n)}$ para $n \geq 2$ donde N es el número de sites en la imagen.

- c) Para toda configuración inicial η (de la imagen) y para toda configuración ω , se tiene,

$$\lim_{n \rightarrow \infty} P(x^{(n)} = \omega \mid x^{(0)} = \eta) = \prod \omega$$

Algoritmo:

1. Inicialización aleatoria de la solución: $x^{(0)}$.
Escoger una temperatura inicial $T^{(0)}$ suficientemente elevada.
2. En cada etapa (n):
 - Escoger el site $r = s^{(n)}$, su valor es $x_r^{(n)}$ y su configuración de vecindad es $V_r^{(n)}$.

- Efectuar el sorteo aleatorio uniforme de una variable $\xi \in \Omega$.
3. Criterio Metrópolis: calcular la variación de energía local asociada:

$$\Delta U = U(\xi | V_r^{(n)}, Y) - U(x_r^{(n)} | V_r^{(n)}, Y)$$

dos eventualidades pueden surgir:

- Si $\Delta U \leq 0$, $x_r^{(n+1)} = \xi$
- Si $\Delta U > 0$, $x_r^{(n+1)} = \xi$ con probabilidad

$$P(x_r^{(n+1)} \rightarrow \xi) = \exp\left(-\frac{\Delta U}{T_n}\right)$$

4. Serie de temperaturas decrecientes hacia 0: $\lim_{n \rightarrow \infty} T_n = 0$.

Comentarios:

- $\Delta U = U(\xi | V_r^{(n)}, Y) - U(x_r^{(n)} | V_r^{(n)}, Y)$ representa tanto la variación local de energía, como la variación global de la energía asociada a la transición $x_r^{(n+1)} = \xi$.
- La convergencia del algoritmo hacia el mínimo global de la función de energía, independientemente de la configuración inicial se da cuando:

$$T^n > \frac{c}{\log(n+1)}$$

para n grande, donde c es una constante que depende de la expresión de la función de la energía a minimizar $U(x|y)$.

- En la práctica, se simula la probabilidad $p = \exp\left(-\frac{\Delta U^{(n)}}{T}\right)$ efectuando el sorteo aleatorio uniforme de una variable $\eta \in [0,1]$.

Entonces, si $0 \leq \eta \leq p$, $x_r^{(n+1)} = \xi$, en caso contrario, $x_r^{(n+1)} = x_r^{(n)}$.

- En la práctica, los esquemas de descenso de temperatura son del tipo $T^{(n)} = \frac{c}{\log(n+1)}$ e incluso $T^{(n+1)} = \alpha \cdot T^{(n)}$.

Cuando T es mayor a un valor determinado entonces del paso 4 volver al paso 2.

La temperatura inicial $T^{(0)}$ debe ser tan alta que virtualmente todas las transiciones propuestas son aceptadas. Tal valor no es fácil de determinar ya que está relacionado con los valores máximos y mínimos de la función de energía a

minimizar. Teóricamente $T^{(n)} \rightarrow 0$ no más rápido que $\frac{c}{\text{Ln}(k)}$ para una constante c independiente de k . Sin embargo, en este esquema el descenso resulta demasiado lento.

3.3 Transformada Polinomial

3.3.1 Introducción

Para muchas aplicaciones en codificación de imágenes de visión por computadora, así como en el caso de la percepción visual, se requiere que los datos que componen la imagen, que son una matriz de valores de intensidades, sean interpretados dentro de un marco visual con significado. Generalmente se está de acuerdo en que se requiere alguna forma de procesamiento espacio-temporal local de los datos originales para tal propósito. Éste tipo de procesamiento involucra dos decisiones importantes. En primer lugar, para llevar a cabo el procesamiento local, usualmente la imagen se multiplica por una ventana. El tamaño de la ventana establece el tamaño del conjunto de puntos de la imagen que contribuyen en este primer paso de procesamiento. La forma de la ventana determina el peso relativo de cada punto contribuyente de la imagen. A fin de describir completamente la imagen, este procesamiento local debe ser representado por un número suficiente de posiciones de la ventana. La forma, tamaño y espaciamiento de la ventana deben seleccionarse. En segundo lugar, para cada posición de la ventana debe llevarse a cabo el procesamiento. Como cada elección específica de procesamiento implica la búsqueda de patrones específicos, seleccionar este proceso es equivalente a fijar los patrones virtuales que sean considerados más relevantes *a priori*.

Es muy difícil hacer elecciones óptimas para la ventana y procesamiento basados únicamente en argumentos teóricos. En este contexto, el sistema visual humano se utiliza frecuentemente como referencia. Así pues, es interesante revisar rápidamente técnicas de procesamiento de imágenes disponibles con respecto a la elección que ha sido hecha por la ventana y el procesamiento.

Una clase de ventana muy frecuentemente usada es el cuadrado, que es un tipo de ventana que no se traslapa. Son usadas, por ejemplo, en TC¹⁴, y cuantización de vector de imagen (VQ) y representa la forma más simple de subdividir una imagen. Una desventaja importante es que, a bajas tasas, aparece el efecto de bloques (el cual resulta molesto a un observador humano). La DPCM¹⁵, por otro lado, utiliza alguna forma de patrón de escaneo junto con una ventana de predicción que sólo incluye *pixeles* precedentes. De aquí, que sólo parte de la información vecina se incluye en el

¹⁴ TC es el acrónimo de *Transform Coding*.

¹⁵ DPCM es el acrónimo de *Diference pulse code modulation*.

análisis de la imagen. Las ventanas anteriores representan fielmente las subdivisiones artificiales del dominio del píxel y tiene relevancia perceptual pequeña. Muchas de las técnicas de codificación de imágenes de la segunda generación, tales como los codificadores de pirámide (PC), codificadores de sub-banda (SBC), la transformada cortes y las expansiones de Gabor, usan ventanas que se superponen.

Un parámetro importante de la ventana es su tamaño (o escala). La selección del tamaño apropiado de una ventana es en sí un problema fundamental. Por otro lado, a fin de facilitar altas reducciones de datos, el tamaño de la ventana debe ser suficientemente grande. Pero, a su vez, la complejidad del análisis dentro de cada ventana se incrementa con el tamaño de ésta. Existen dos aproximaciones posibles al problema. Primero, puede seleccionarse una ventana de tamaño fijo y hacer un análisis dentro de cada ventana que incluya todos los patrones visuales de interés. Segundo, se puede limitar la complejidad del análisis que se desarrollará en cada ventana y subsecuentemente determinar el tamaño necesario de la ventana para describir localmente a la imagen con suficiente exactitud. Aquí, en lugar de restringir el procesamiento a una escala, se repite el mismo procesamiento a escalas múltiples y subsecuentemente se utilizan las salidas de esta etapa para seleccionar la escala óptima en cada posición. Existe evidencia de que el sistema visual humano utiliza este principio 'escala-espacio'.

Si se usan ventanas de diferentes tamaños, entonces el espaciamiento de éstas se toma usualmente proporcional a su tamaño. En muchas de las estructuras de pirámide el espaciamiento y tamaño de la ventana se incrementa en pasos de dos. Algunos artículos muestran cómo este incremento exponencial en el tamaño y espaciamiento es no sólo preferible desde el punto de vista de la teoría de la información [1,2], sino también resulta ser computacionalmente eficiente.

3.3.2 Transformada Polinomial Unidimensional

La transformada polinomial es una técnica de descomposición de señales, en la cual éstas son aproximadas localmente con polinomios. En este apartado se introducen las ideas básicas en una dimensión, con señales analógicas.

El análisis para la transformada polinomial involucra dos pasos. En el primero la señal original $L(x)$ se estudia de manera local multiplicándola por una ventana $V(x)$. Una descripción completa de la señal requiere que el proceso anterior se repita en un número suficiente de posiciones de la ventana. Se considerará el caso de ventanas equidistantes.

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

$$W(x) = \sum_k V(x - kT) \quad (3.14)$$

con repetición periódica. La función de peso es periódica con periodo T . Considerando que $W(x)$ es diferente de cero para todo x , se obtiene:

$$L(x) = \frac{1}{W(x)} \sum_k L(x) \cdot V(x - kT) \tag{3.15}$$

tal que se pueda garantizar que las señales locales $L(x) \cdot V(x - kT)$ para todas las posiciones distintas de kT contengan suficiente información de la señal original.

El segundo paso consiste en aproximar la señal localizada dentro de la ventana $V(x - kT)$ mediante una expansión polinomial. Como funciones base para la expansión polinomial se toman los polinomios $G_n(x)$ de grado n , que son ortonormales con respecto a $V^2(x)$, es decir:

$$\int_{-\infty}^{\infty} V^2(x) G_m(x) G_n(x) dx = \delta_{mn} \tag{3.16}$$

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

$$G_n(x) = \frac{1}{\sqrt{M_{n-1} M_n}} \begin{vmatrix} c_0 & c_1 & \dots & c_n \\ c_1 & c_2 & \dots & \\ \dots & \dots & \dots & \\ c_{n-1} & c_n & \dots & c_{2n-1} \\ 1 & x & \dots & x^n \end{vmatrix} \tag{3.17}$$

donde el determinante $M_n(x)$ se define como

$$M_n = |c_{i+j}|_{i,j=0,\dots,n}, M_{-1} = 1 \tag{3.18}$$

y

$$c_n = \int_{-\infty}^{\infty} x^n V^2(x) dx \tag{3.19}$$

es el momento de orden n .

Bajo condiciones muy generales, para la señal $L(x)$ se obtiene:

$$V(x - kT) \left| L(x) - \sum_{n=0}^{\infty} L_n(kT) \cdot G_n(x - kT) \right| = 0 \tag{3.20}$$

con

$$L_n(kT) = \int_{-\infty}^{\infty} L(x) \cdot G_n(x - kT) V^2(x - kT) dx \quad (3.21)$$

Por ejemplo, si $L(x)$ es analítica y finita para todo x , queda garantizada la convergencia de la expansión en serie en (3.20) para la mayoría de las ventanas. De aquí que el error de aproximación entre una señal y un polinomio puede hacerse suficientemente pequeño, tomando un grado de la expansión polinomial suficientemente grande. Esto implica que la descripción de la señal local $L(x) \cdot V(x - kT)$ puede reducirse a un número finito de coeficientes polinomiales $L_n(kT)$. La señal de energía dentro de la ventana puede expresarse en términos de los coeficientes de expansión, como se muestra,

$$\int_{-\infty}^{\infty} L^2(x) \cdot V^2(x - kT) dx = \sum_{-\infty}^{\infty} L_n^2(kT) \quad (3.22)$$

Esta es la generalización del Teorema de Parseval con polinomios ortonormales. Combinando (3.15) y (3.20) se obtiene la siguiente expansión para la señal completa.

$$L(x) = \sum_{n=0}^{\infty} \sum_k L_n(kT) \cdot P_n(x - kT) \quad (3.23)$$

donde

$$P_n(x) = \frac{G_n(x) V(x)}{W(x)} \quad (3.24)$$

La ecuación (3.27) implica que los coeficientes $L_n(kT)$ pueden derivarse de la convolución de la señal $L(x)$ con las funciones de filtrado,

$$D_n(x) = G_n(-x) V^2(-x)$$

seguido por muestras en múltiplos de T .

A este mapeo de la señal original $L(x)$ en coeficientes polinomiales $L_n(kT)$ se le llama transformada polinomial 'directa'. La reconstrucción de la señal a partir de esos coeficientes está dada por (ec. 3.23) y se le llama transformada polinomial inversa, la cual consiste en interpolar los coeficientes $\{L_n(kT)\}$, k entero con la función patrón $P_n(x)$ y sumar sobre todos los órdenes de n .

3.3.3 Transformada Hermitiana Unidimensional

A fin de ilustrar las propiedades de la transformada polinomial, se hará el análisis sobre un caso especial muy importante de ventana que es la Gaussiana, dada por:

$$V(x) = \frac{1}{\sqrt{\pi\sigma}} \exp(-x^2/2\sigma^2) \quad (3.25)$$

donde el factor de normalización es tal que $V^2(x)$ tiene energía unitaria. Los polinomios ortogonales asociados con $V^2(x)$ son conocidos como polinomios hermitianos y de aquí el nombre de transformada hermitiana.

Las razones para utilizar este tipo de ventana son muchas. Primero, la teoría matemática es conocida y por lo tanto las propiedades de la transformada pueden derivarse fácilmente. Segundo, las ventanas Gaussianas separadas por dos veces la desviación estándar σ son un buen modelo para los campos receptivos encontrados en experimentos psicológicos. Tercero, la transformada hermitiana involucra filtros derivados de Gaussianas. Han sido ya muy utilizadas en el modelado psicológico del sistema visual humano. Además, las ventanas Gaussianas minimizan el producto de incertidumbres en los dominios espacial y de la frecuencia, lo que es una propiedad interesante en el análisis de imágenes.

La función de peso es periódica, a continuación se muestra el desarrollo y el resultado.

$$w(x) = \frac{T}{\sqrt{2\pi\sigma}} \sum_k \int_{-\infty}^{\infty} \exp\left[-\frac{(x-kT)^2}{2\sigma^2} - j\omega x\right] dx = T \sum_k e^{-j\omega kT} e^{-(\omega\sigma)^2/2} \quad (3.26)$$

$$W(x) = \frac{\sqrt{2\sqrt{\pi\sigma}}}{T} w(x) \quad (3.27)$$

con,

$$w(x) = 1 + 2 \sum_{k=1}^{\infty} \exp\left[-\frac{1}{2}\left(k\frac{2\pi\sigma}{T}\right)^2\right] \cdot \cos\left(k\frac{2\pi\sigma}{T}\right) \quad (3.28)$$

El contraste de esta función viene determinado por el parámetro de muestreo $\tau = T/\sigma$. Generalmente se desea limitar el número de descomposiciones locales, especialmente en codificación. Así pues, es conveniente hacer el parámetro de muestreo tan grande como sea posible.

Los filtros determinan que la información será expresada en los coeficientes de la transformada hermitiana. Por lo tanto, las propiedades principales de la transformada hermitiana están determinadas por estas funciones. De la expresión general se obtiene:

$$D_n(x) = \frac{(-1)^n}{\sqrt{2^n n!}} \cdot \frac{1}{\sigma\sqrt{2\pi}} H_n\left(\frac{x}{a}\right) e^{-x^2/2\sigma^2} \quad (3.29)$$

Las funciones patrón $P_n(x)$ son necesarias para resintetizar la señal original a partir de los coeficientes hermitianos. Están dados por la siguiente expresión:

$$P_n(x) = \frac{T}{\sqrt{2^n n!}} \cdot \frac{1}{\sigma\sqrt{2\pi}} H_n\left(\frac{x}{a}\right) e^{-x^2/2\sigma^2} / w(x) \quad (3.30)$$

3.3.4 Transformada Polinomial en dos dimensiones

La transformada polinomial puede generalizarse fácilmente en dos dimensiones. Dada una 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 dados por,

$$\int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} V^2 G_{m,n-m}(x, y) G_{j,i-j}(x, y) dx dy = \delta_{ni} \delta_{mj} \quad (3.31)$$

Donde $n=0, 1, \dots, \infty$; $m=0, 1, \dots, n$; $j=0, 1, \dots, y$.

La descomposición de señales 2D en polinomios locales es:

$$L(x, y) = \sum_{n=0}^{\infty} \sum_{m=0}^n \sum_{(p,q) \in S} L_{m,n-m}(p, q) \cdot P_{m,n-m}(x-p, y-q) \quad (3.32)$$

donde (p, q) varía sobre todo el conjunto de coordenadas de la rejilla S . La única condición para la red de muestreo es que la función de peso

$$W(x, y) = \sum_{(p,q) \in S} V(x-p, y-q) \quad (3.33)$$

sea diferente de cero para todas las coordenadas (x, y) .

Los coeficientes polinomiales $L_{m,n-m}(p, q)$ se obtienen de la convolución de la imagen con las funciones de filtrado

$$D_{m,n-m}(x, y) = G_{m,n-m}(-x, -y) V^2(-x, -y) \quad (3.34)$$

seguidas del muestreo de la salida en (p, q) sobre S .

Las funciones patrón usadas para interpolar los coeficientes polinomiales se define como

$$P_{m,n-m}(x,y) = G_{m,n-m}(x,y)V(x,y)/W(x,y) \quad (3.35)$$

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

El análisis de una imagen tendría como objetivo descomponer la señal en patrones que son perceptualmente importantes. Es bien conocido, especialmente en percepción visual y en computación visual, que los patrones locales de una dimensión tales como los lados y líneas juegan un papel central en la visión. A continuación se mencionará cómo se pueden encontrar los mejores patrones locales óptimos con ayuda de la transformada polinomial. Usando el criterio del error cuadrático se desea minimizar:

$$E^2 = \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} [k(x \cdot \cos \theta + y \cdot \text{sen} \theta) - L(x,y)]^2 V^2(x,y) dx dy \quad (3.36)$$

sobre todos los patrones 1D k y ángulos θ .

Defínase la ventana 1D

$$V_{\theta}^2(u) = \int_{-\infty}^{\infty} V^2(u \cos \theta - v \text{sen} \theta, u \text{sen} \theta + v \cos \theta) dv \quad (3.37)$$

como la proyección de la función $V^2(x,y)$ sobre el eje que forma un ángulo θ con el eje x . La ventana es independiente de la orientación si $V^2(x,y)$ es rotacionalmente simétrica. Se puede expandir el patrón $K(u)$ sobre la base $\{F_{n,\theta}(u), n=0,1,\dots\}$ de polinomios ortonormales sobre $V_{\theta}^2(u)$

$$V_{\theta}(u) \left[K(u) - \sum_{n=0}^{\infty} K_{n,\theta} \cdot F_{n,\theta}(u) \right] = 0 \quad (3.38)$$

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

$$K_{n,\theta} = \sum_{k=0}^n \sum_{l=0}^k L_{l,k-l} \cdot h_{n,\theta}(l,k-l) \quad (3.39)$$

para los coeficientes patrón 1D, donde

$$h_{n,\theta}(l,k-l) = \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} F_{n,\theta}(x \cos \theta + y \text{sen} \theta) \cdot G_{l,k-l}(x,y) \cdot V^2(x,y) dx dy \quad (3.40)$$

es una función ángulo completamente determinada por $V^2(x, y)$.

Los polinomios ortonormales $F_{n,\theta}(u)$ y la función ángulo pueden determinarse sin conocimiento explícito de $V_\theta(u)$. En realidad, la ecuación 3.17 implica que sólo los momentos

$$c_{n,\theta} = \int_{-\infty}^{\infty} u^n V_\theta^2(u) du \quad (3.41)$$

son necesarios para describir completamente los polinomios ortonormales.

El error de aproximación 1D

$$E^2 = \sum_{k=0}^{\infty} \sum_{l=0}^k L_{l,k-l}^2 - \sum_{n=0}^{\infty} k_{n,\theta}^2 \quad (3.42)$$

puede minimizarse sobre el ángulo θ minimizando la energía direccional

$$\sum_{n=0}^{\infty} K_{n,\theta}^2 \quad (3.43)$$

donde $K_{n,\theta}$ se determina por los coeficientes 2D. En la práctica, los primeros términos en esta medida de energía direccional son suficientes para hacer una buena estimación de la dirección óptima. Hay que hacer notar que la energía direccional se encuentra mediante una simple combinación de los coeficientes polinomiales 2D. Computacionalmente esto es más eficiente que usar filtros para calcular la energía en cada dirección, especialmente si se hace con un gran número de direcciones.

Posteriormente, se localiza la orientación óptima de los contornos maximizando el contraste direccional de la imagen, C_θ definida como,

$$C_\theta = \log \left[1 + \sum_{n=1}^N K_{n,\theta}^2 \right] \quad (3.44)$$

En cada posición donde es encontrado un contorno, sólo los coeficientes 1D con mayor contraste se almacenan en el vector $K_{N,\theta_{opt}}$.

El proceso de discriminación por orientación es, por sí mismo, un proceso de restauración de contornos, puesto que sólo los patrones unidimensionales que conforman un contorno se preservan. Todos los demás patrones orientados en diferentes direcciones, tales como el ruido, son descartados. La restauración se lleva a cabo mediante la transformada polinomial inversa.

3.3.5 Transformada Hermitiana en dos dimensiones

Un caso especialmente interesante de la transformada polinomial de dos dimensiones aparece cuando la función ventana es separable, esto es, $V(x,y)=V(x)V(y)$ y la rejilla de muestreo es cuadrada. Las funciones de filtrado y patrón son también separables y, por lo tanto, su implantación es muy eficiente. Por ejemplo, los coeficientes polinomiales se encuentran al realizar la convolución de la imagen con las funciones de filtrado $D_m(x)D_{n-m}(y)$ donde $D_m(x)$ es la función de filtrado 1D para la ventana $V(x)$ seguida por un muestreo de la salida en direcciones horizontal y vertical en múltiplos del espaciamiento de las muestras T .

Las ventanas Gaussianas de dos dimensiones tienen la propiedad única de que ambas variables pueden separarse y además existe simetría rotacional. Las funciones de filtrado son separables tanto de manera polar como espacial. La transformada de Fourier de $D_m(x)D_{n-m}(y)$ expresada en coordenadas polares $\omega_x = \omega \cos \theta$ y $\omega_y = \omega \sin \theta$

$$d_m(\omega_x)d_{n-m}(\omega_y) = g_{m,n-m}(\theta) \cdot d_n(\omega) \quad (3.45)$$

donde $d_n(\omega)$ es la transformada de Fourier de la función de filtrado hermitiana 1D $D_n(\omega)$ siendo r la coordenada radial y

$$g_{m,n-m}(\theta) = \sqrt{\frac{n!}{m!(n-m)!}} \cos^m(\theta) \cdot \sin^{n-m}(\theta) \quad (3.46)$$

expresa la sensibilidad direccional del filtro. De aquí, que al incrementar n , el análisis de las frecuencias radiales superiores es mejor, esto es, una mayor resolución espacial, de manera similar al caso 1D.

4 SEGMENTACIÓN ADAPTIVA DE IMÁGENES DE PERCEPCIÓN REMOTA

4.1 Introducción

La investigación científica en la observación de la tierra a partir de imágenes de percepción remota permite un efectivo monitoreo y manejo de sus recursos. En este dominio, la segmentación de imágenes es una herramienta útil para la obtención de información temática, lo que permite actualizaciones sencillas en cartografía y distribución urbana.

La segmentación de imágenes de percepción remota puede aproximarse en diversas formas: desde los clasificadores basados en el concepto simple e intuitivo de la mínima distancia, hasta algoritmos basados en teorías ampliamente estudiadas, como lo son los clasificadores basados en los Campos Aleatorios de Markov y en la distribución de Gibbs. El primer enfoque y –hablando de forma general– los clasificadores no contextuales tienen una desventaja inherente: al considerar únicamente las características espectrales de la imagen, cada *pixel* es procesado de forma totalmente independiente a su vecindad, lo cual puede llevar a la obtención de imágenes segmentadas con demasiados *pixeles* aislados. Como el segundo enfoque sí toma en cuenta la vecindad de cada *pixel*, genera imágenes con regiones homogéneas, sin embargo, una desventaja de estos algoritmos es la eliminación eventual de estructuras finas bien definidas de la escena, al fusionarlas con regiones adyacentes.

Los algoritmos de segmentación pueden ser utilizados de forma combinada para obtener mejores resultados. Con este objetivo en mente, aquí proponemos un esquema adaptivo (no contextual – contextual) enfocado a la preservación de las estructuras finas y a una mejor detección de contornos, pero tratando de evitar los efectos de sobresegmentación. El algoritmo propuesto está basado en la adaptación de varios métodos clásicos que abarcan los dos principales enfoques de la segmentación: el no contextual (mínima distancia y máxima probabilidad) y el contextual (MRF, modelo de Potts y Recocido Simulado). Los pasos seguidos en este planteamiento se describen brevemente a continuación:

- ♦ Obtención de los centros de clase a través de un algoritmo de mínima distancia.
- ♦ Aproximación de las funciones de probabilidad, basados en la desviación estándar y la media de cada clase.
- ♦ Segmentación de la imagen por medio de un algoritmo de máxima probabilidad, pero definiendo un criterio de rechazo. Los *pixeles* rechazados se asignan a una nueva clase, denominada 'clase de rechazo'.
- ♦ Finalmente, los *pixeles* rechazados en el paso anterior son reclasificados mediante un clasificador Markoviano, utilizando un esquema de Recocido Simulado basado en el modelo de Potts.

Este planteamiento ofrece muy buenos resultados al trabajar con fotografías aéreas o imágenes SPOT pancromáticas. Sin embargo, al aplicarlo a imágenes SAR, la presencia de *speckle*, y en general la pobre calidad de la imagen, representó una seria desventaja. En este caso fue necesario incluir una etapa de filtrado para poder utilizar las ventajas que ofrecen las imágenes SAR en conjunción con el algoritmo de segmentación propuesto. Para esta etapa se utilizó un algoritmo adaptivo de filtrado basado en la transformada polinomial, el cual provee imágenes SAR con menor incidencia de ruido, al tiempo que preserva los detalles finos de la imagen y provoca una degradación mínima de los contornos.

Los resultados ofrecidos por el esquema de segmentación propuesto evidencian su buen desempeño, generando imágenes clasificadas mostrando regiones homogéneas, contornos bien definidos y respeto de las estructuras finas, así como una mínima presencia de *pixeles* aislados. Además, es un algoritmo fácil de implementar, ofreciendo ejecuciones rápidas y parámetros variables.

4.2 Segmentación de Imágenes de Percepción Remota

Para gran parte de los usuarios de imágenes de percepción remota, la clasificación supone la fase culminante del procesamiento digital de la imagen. Como resultado de la clasificación se obtiene otra imagen, de iguales dimensiones a la original, pero con la importante diferencia de que el nivel de gris que define a cada *pixel* representa una etiqueta que identifica la clase a la cual fue asignado dicho *pixel*. A partir de una imagen clasificada se puede generar fácilmente cartografía temática.

Es importante distinguir entre los conceptos de segmentación y clasificación de imágenes. Aunque ambos tienen como finalidad el dividir una imagen en regiones características, la clasificación es un término más concreto. La clasificación supone un extractor de características, cuyo propósito es reducir los datos mediante la determinación de propiedades que distinguen elementos de la imagen (por ejemplo, regiones urbanas, acuíferas, agrícolas, etc.). Así, el problema de clasificación es básicamente particionar el espacio de características en regiones –una región para cada categoría– y relacionar cada una de esas categorías con elementos o parámetros físicos. El espacio de características contiene básicamente texturas y valores de nivel de gris. Un espacio de características más complicado podría obtenerse al considerar propiedades geométricas, o bien, al considerar los niveles de gris de más de un canal.

En nuestra terminología, la segmentación puede ser vista como un caso especial de la clasificación en el cual las clases suelen definirse de modo relativo a las características específicas de la escena a clasificar, es decir, la segmentación no busca una definición absoluta de cada clase, que pudiera aplicarse a cualquier imagen, sino más bien una caracterización relativa, válida para una determinada imagen y un territorio concreto. Con este planteamiento no es preciso conocer con detalle las condiciones de adquisición de la imagen: basta con identificar en la imagen misma las clases a discriminar, sin pretender que esa identificación sea extrapolable a otras situaciones. Así, no se trata de establecer un modelo entre medidas digitales y parámetros físicos, sino entre medidas digitales, para una imagen determinada. En este sentido, la segmentación supone un enfoque más cualitativo, si bien, por tratarse de un proceso numérico, incorpora un aparato estadístico ciertamente importante.

La segmentación implica categorizar una imagen. En términos estadísticos esto supone reducir la escala de medida de una variable continua (los niveles de gris detectados originalmente por el sensor) a una escala nominal o categórica (clase). La apariencia visual puede resultar muy similar entre ambas imágenes, ya que tienen el mismo tamaño y pueden visualizarse en niveles de gris; sin embargo, es importante considerar que no pueden aplicarse ciertas operaciones estadísticas a una imagen segmentada, ya que, pese a ser digital, no es una variable cuantitativa, sino cualitativa.

El proceso de segmentación tiene como objetivo particionar la imagen en regiones tales que:

- ♦ La segmentación debe ser completa (es decir, cada *pixel* debe pertenecer a una región).
- ♦ Los *pixeles* de una región deben de estar conectados.
- ♦ Las regiones deben de ser disjuntas.

Los primeros ensayos de segmentación digital de imágenes de satélite se basaron exclusivamente en los niveles de gris de los *pixeles* que forman la imagen (aproximación no contextual). Estos niveles de gris son una traducción digital de un flujo energético recibido por el sensor para una determinada banda del espectro. Consecuentemente, una segmentación basada exclusivamente en esta característica puede denominarse propiamente segmentación espectral. Aún cuando es muy rica esta información, no siempre es capaz de discriminar entre clases, ya que en una escena puede haber distintas clases con un comportamiento espectral similar. En este caso una opción ampliamente utilizada es la de incorporar a la segmentación el contexto espacial en el que aparece cada *pixel* (aproximación contextual).

La segmentación se inicia definiendo cada una de las categorías que pretenden discriminarse. En el caso de una segmentación basada en los niveles de gris de los *pixeles* (valores numéricos), esta definición también tiene que ser numérica; esto es, se trata de obtener el nivel de gris o, mejor aún, el rango de niveles de gris que identifica a cada clase que interviene en la segmentación.

Diversos factores introducen una cierta dispersión en torno al comportamiento espectral medio de cada escena. Esto implica que las distintas categorías no se definen por un solo nivel de gris, sino por un conjunto de niveles de gris más o menos próximos entre sí. En términos de su segmentación, esto supone que existe una cierta dispersión en torno al nivel de gris medio de cada clase, más o menos significativa, según las clases que se consideren y la propia complejidad de la escena original. Por esto, la fase de entrenamiento resulta compleja: se trata de definir con rigor cada una de las clases que se pretende discriminar, teniendo en cuenta su propia variabilidad en la zona de estudio.

Este objetivo se logra seleccionando una muestra de *pixeles* de la imagen, que representen adecuadamente a las categorías de interés. A partir de estos *pixeles* puede calcularse el nivel de gris medio y la variabilidad para cada una de las clases consideradas.

4.2.1 Segmentación supervisada y no supervisada

Tradicionalmente, se han dividido los métodos de segmentación en dos grupos: supervisado y no supervisado, de acuerdo a la forma en que son obtenidas las estadísticas de entrenamiento. El método supervisado parte de un conocimiento previo del terreno, a partir del cual se seleccionan las muestras para cada una de las categorías. Por otra parte, el método no supervisado procede a una búsqueda automática de grupos de valores homogéneos dentro de una imagen; en este caso corresponde al usuario la labor de encontrar las correspondencias entre estos grupos y sus categorías de interés.

Así, el método supervisado pretende definir clases informacionales, mientras que el no supervisado tiende a identificar las clases espectrales presentes en la imagen. Ninguno de los dos métodos proporciona una solución inmediata a todos los problemas que implica una segmentación. Por un lado, el método supervisado podría forzar al algoritmo a discriminar categorías que no tengan un claro significado espectral, mientras que el método no supervisado puede proporcionar resultados de difícil interpretación, poco relacionados con las necesidades del usuario final.

4.2.2 Segmentación contextual y no contextual

Una vez definidas las clases que pretenden considerarse en el proceso de segmentación, el siguiente paso es llevar a cabo la segmentación en sí. Desde este

punto de vista, los métodos de segmentación pueden dividirse en dos grandes categorías: contextuales y no contextuales. Aquí mencionaremos brevemente las principales características de cada uno de estos métodos, y describiremos algunos de los enfoques 'clásicos' pertenecientes a cada categoría. Esta descripción no pretende ser exhaustiva, sino centrarse en aquellos conceptos que son de relevancia en la propuesta del algoritmo adaptivo de segmentación que aquí presentamos.

4.2.2.1 Segmentación no contextual

Este enfoque a la segmentación de imágenes no toma en cuenta la información de la vecindad del *pixel* que está siendo clasificado, lo que puede llevar a problemas de sobresegmentación, resultando en imágenes segmentadas con una alta incidencia de *pixeles* aislados. Dos esquemas clásicos utilizados para llevar a cabo una segmentación no contextual son aquellos basados en los conceptos básicos de mínima distancia y máxima probabilidad.

Algoritmos de mínima distancia

El concepto de mínima distancia se obtiene intuitivamente del hecho de que la similitud entre elementos en un espacio dado puede ser expresado en términos de su proximidad a cada uno de los centros de clase. Estos clasificadores son especialmente efectivos cuando las clases muestran un grado de variabilidad 'razonablemente pequeño', es decir, cuando los elementos de cada clase se agrupan estrechamente en torno a su media.

Los algoritmos de mínima distancia son simples y ofrecen ejecuciones rápidas, pero como sólo consideran las propiedades espectrales del *pixel* que está siendo analizado, las imágenes clasificadas de esta forma pueden presentar muchos *pixeles* aislados, resultando en una segmentación con efectos de 'nieve'.

Algoritmos de máxima probabilidad

Consideran que las propiedades espectrales de los *pixeles* que pertenecen a cierta clase siguen una ley de probabilidad; consecuentemente, cada clase puede ser descrita mediante una función de probabilidad. En breve, esto significa que estas funciones reflejan la distribución real de cada clase, lo que nos permite determinar la probabilidad de pertenencia a una clase dada para cada uno de los *pixeles* que componen la escena. El cálculo se realiza para todas las categorías consideradas en el proceso de segmentación y los *pixeles* son asignados a aquella clase que maximiza la función de probabilidad.

Los algoritmos de máxima probabilidad son más complejos y requieren mayor cantidad de cálculos que los clasificadores de mínima distancia y, aunque ofrecen mejores resultados, los efectos de sobresetimación no se evitan del todo.

4.2.2.2 Segmentación contextual

Los clasificadores contextuales comúnmente requieren un mayor tiempo de procesamiento que los métodos no contextuales, y son aplicados de forma iterativa para minimizar una función no convexa de energía. Dos clasificadores contextuales muy conocidos son el ICM¹⁶ (un algoritmo de relajación determinista) y el modelo Markoviano.

Modos Condicionales Iterativos (ICM)

Propuesto por Besag [4], es el algoritmo determinista más popular y, dado que no incluye un parámetro de temperatura, no es un recocido. El algoritmo ICM requiere de una configuración inicial 'buena' (la calidad del resultado final depende en gran forma de esto) y garantiza una rápida convergencia –en aproximadamente 10 iteraciones– hacia el mínimo local.

No se sabe como asegurar una buena configuración inicial, pero, para la segmentación de imágenes, el clasificador de máxima probabilidad (MLC¹⁷) se adopta usualmente como el método de inicialización para el algoritmo ICM.

Modelo Markoviano

Utilizando MRF y estimación Máxima a Posteriori (MAP), la minimización de las funciones de energía puede ser alcanzada por recocido simulado (SA¹⁸). Los algoritmos de SA alcanzan (teóricamente) un mínimo global, pero requieren una gran cantidad de cálculos y, por lo tanto, mucho tiempo de ejecución, lo cual es una gran desventaja.

El recocido simulado es controlado por un parámetro llamado 'temperatura' (T), la cual decrece secuencialmente. Para un valor alto de T , pueden permitirse grandes incrementos de energía; conforme T decrece, sólo se aceptan incrementos pequeños de energía; y alrededor de la T de congelamiento no se permite ningún incremento de energía. Este criterio permite al algoritmo evitar mínimos locales y alcanzar el mínimo global [23].

¹⁶ ICM – *Iterated Conditional Modes*

¹⁷ MLC – *Maximum Likelihood Classifier*

¹⁸ SA – *Simulated Annealing*

Como el proceso para alcanzar el óptimo global es muy lento –la temperatura decrece de forma mínima después de cada iteración– generalmente se implementan aproximaciones del algoritmo de recocido simulado y, como consecuencia, la convergencia hacia el óptimo global no se puede garantizar. El proceso de regularización generalmente adopta el siguiente esquema para decrementar T : $T^{(t)} = kT^{(t-1)}$, donde k toma valores entre 0.8 y 0.99, permitiendo una convergencia más rápida.

Las imágenes clasificadas mediante segmentación contextual muestran regiones homogéneas y una baja incidencia de *pixeles* aislados pero, bajo este planteamiento, estructuras finas bien definidas de la imagen pueden ser eliminadas eventualmente, al ser fusionadas a regiones contiguas.

4.3 Aproximación Adaptiva a la Segmentación de Imágenes de Percepción Remota

Hemos descrito brevemente las principales ventajas y desventajas de algunos de los principales clasificadores contextuales y no contextuales. Ahora nuestro objetivo es definir un algoritmo adaptivo enfocado a la obtención de imágenes clasificadas con regiones homogéneas, pero conservando las estructuras finas que estén bien definidas.

Cuando utilizamos el término ‘estructuras finas bien definidas’ nos referimos a aquellas estructuras pequeñas que claramente se distinguen de las regiones que las circundan, es decir, sus características espectrales se diferencian claramente y, por lo general, se encuentran muy cercanas a alguno de los centros de clase, lo cual significa que son representativas de la clase a la cual pertenecen y, por tanto, deben ser preservadas.

Estas estructuras por lo general son segmentos lineales que al momento de ser adquirida la imagen quedan representadas por líneas de no más de tres *pixeles* de grosor, las cuales en una aproximación puramente contextual pueden ser fusionados a las regiones contiguas. Las figuras 4.1 y 4.2 ejemplifican claramente este concepto.

El algoritmo que proponemos está basado en una adaptación de varios métodos clásicos de segmentación, y busca tomar las principales ventajas de cada uno de ellos. Los conceptos y teorías en los que se basa nuestra propuesta son los siguientes:

- ♦ Mínima distancia
- ♦ Máxima probabilidad *a posteriori* (MAP)
- ♦ Campos aleatorios de Markov
- ♦ Modelo de Potts
- ♦ Recocido Simulado



Figura 4.1

Fotografía aérea de una región urbana de la Ciudad de México.

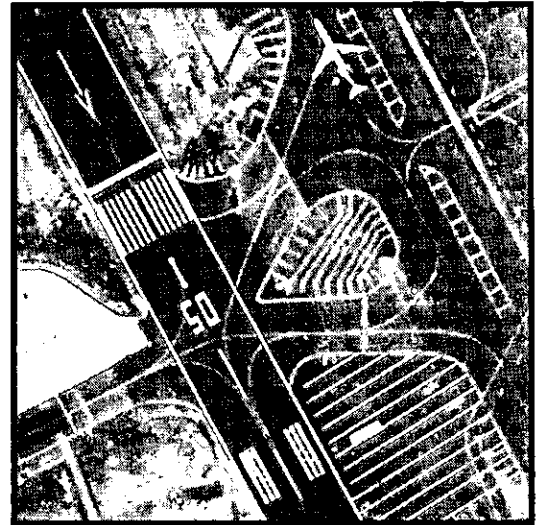


Figura 4.2

Zona de pistas del aeropuerto internacional de la Ciudad de México.

Nuestra propuesta está compuesta de cuatro pasos principales, los cuales se describen detalladamente a continuación.

I. Obtención de los centros de clase mediante un algoritmo no contextual

En este caso decidimos emplear el ampliamente conocido algoritmo de k -medias, el cual ofrece ejecuciones rápidas y no requiere de una etapa de entrenamiento, para obtener una imagen en k clases. El algoritmo de k -medias se basa en el concepto de mínima distancia y consiste en los siguientes pasos:

1. Inicialización: Determinación arbitraria de los k centros de clase iniciales.
2. Para cada iteración:
 - 2.1. Barrer secuencialmente la imagen. Para cada *pixel* analizado:
 - 2.1.1. Calcular la 'distancia' entre su respectivo nivel de gris y cada uno de los centros de clase.
 - 2.1.2. Asignar el *pixel* a la clase que minimiza dicha distancia.
 - 2.2. Calcular los nuevos centros de clase (promedio del nivel de gris de todos los *pixeles* asignados a cada una de las clases).
 - 2.3. Si todos los centros de clase recién calculados son iguales a los anteriores, el algoritmo ha alcanzado la convergencia y se finaliza; en caso contrario, volver a iterar utilizando los nuevos centros de clase.

II. Aproximación de las funciones de probabilidad condicional, basados en la desviación estándar y en la media de cada clase

En esta etapa utilizamos la desviación estándar (una medida de la dispersión o variación en una distribución) y la media de cada una de las clases para crear las funciones de probabilidad conjunta que describen la distribución normal de cada clase. Estas funciones se generaron utilizando la función lineal mostrada en la figura 4.3.

Se decidió utilizar la desviación estándar como parámetro para la creación de estas funciones, ya que al estar directamente relacionada con el comportamiento específico

de cada una de las clases ayuda a obtener una mejor aproximación de la función de probabilidad conjunta para cada una de ellas.

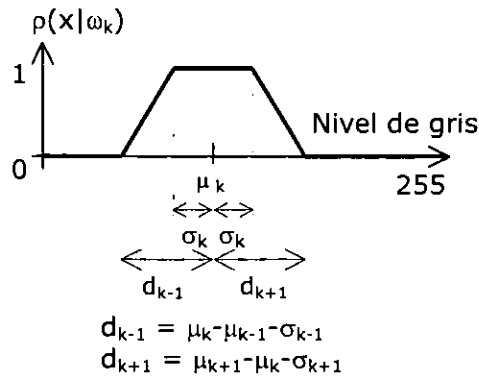


Figura 4.3

Modelo utilizado para aproximar la función de probabilidad condicional de cada clase.

III. Segmentación de la imagen mediante un algoritmo de máxima probabilidad a posteriori (MAP), pero definiendo un criterio de rechazo

Una vez que se definen las funciones de probabilidad para cada clase, el siguiente paso es llevar a cabo una nueva segmentación basada en el concepto de máxima probabilidad. Aquí definimos un parámetro llamado el 'umbral de rechazo', de tal manera que los *pixeles* cuya máxima probabilidad de pertenencia a una clase dada es menor que este umbral, son asignadas a una nueva clase $K+1$ (clase no definida o clase de rechazo), en caso contrario, son asignados de forma definitiva a la clase que maximiza su función de probabilidad.

Esta etapa consiste en los siguientes pasos:

1. Inicialización: Aproximación de las funciones de probabilidad condicional de cada clase y cálculo de las probabilidades *a priori*.
2. Barrido secuencial de la imagen. Para cada *pixel* analizado:
 - 2.1. Calcular su probabilidad *a posteriori* de pertenencia a cada una de las clases consideradas.
 - 2.2. Obtener el valor que maximiza dicha probabilidad de pertenencia.
 - 2.3. Si la probabilidad máxima de pertenencia a alguna de las clases es mayor que el umbral definido, el *pixel* es asignado de forma definitiva a la clase correspondiente; en caso contrario, es asignado a la clase $k+1$ o clase de rechazo.

IV. Reasignación de los pixeles incluidos en la clase de rechazo mediante un clasificador por MRF

El último paso fue llevar a cabo una reclasificación, pero considerando sólo los *pixeles* asignados a la clase de rechazo. Para este paso empleamos el modelo Markoviano de Potts, considerando una vecindad de segundo orden (η_2), la cual sólo considera los ocho vecinos más cercanos al *pixel* central. Las funciones lineales que describen las estadísticas de los datos se definieron siguiendo el esquema mostrado en la figura 4.4. Se aplicó un proceso de recocido simulado para optimar la función de energía.

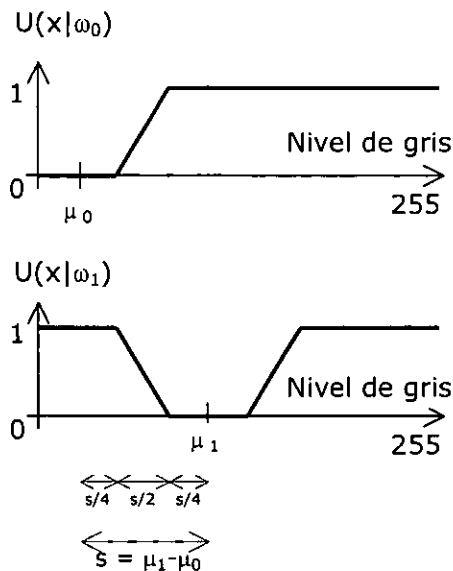


Figura 4.4

Funciones de energía de la imagen (estadísticas de los datos) para el clasificador Markoviano.

El algoritmo se describe a continuación:

1. Inicialización:

- 1.1.** Inicialización aleatoria de la solución, utilizando un generador de números aleatorios uniformemente distribuidos, pertenecientes al grupo finito de clases $\Omega = \{0, 1, \dots, K-1\}$.
- 1.2.** Selección de una temperatura inicial $T^{(0)}$ suficientemente alta. ($T=2$ para este modelo)
- 1.3.** $\beta[\text{actual}] = 0.3$
- 1.4.** $\beta[\text{elección}] = 0.3$

2. Para cada iteración:

2.1. Barrer secuencialmente la imagen. Para cada pixel analizado (r):

- 2.1.1.** Su valor actual es *clase_actual* y su configuración de vecindad es η_2 .
- 2.1.2.** Generar una variable aleatoria uniformemente distribuida (*clase_elección*) perteneciente a Ω .
- 2.1.3. Análisis en 8-conexidad**
 - 2.1.3.1.** $\text{probabilidad}[\text{actual}] = 0$
 - 2.1.3.2.** $\text{probabilidad}[\text{elección}] = 0$
 - 2.1.3.3.** Para cada vecino (etiquetado como clase x) de r :
 - Si $\text{clase_actual} = x$:
 $\text{probabilidad}[\text{actual}] = \text{probabilidad}[\text{actual}] - \beta[\text{actual}]$
 - Si $\text{clase_elección} = x$:
 $\text{probabilidad}[\text{elección}] = \text{probabilidad}[\text{elección}] - \beta[\text{elección}]$
- 2.1.4. Variación de la energía local asociada al site r :**
 - 2.1.4.1.** $\text{probabilidad}[\text{actual}] = \text{probabilidad}[\text{actual}] + U[\text{actual}]$
 - 2.1.4.2.** $\text{probabilidad}[\text{elección}] = \text{probabilidad}[\text{elección}] + U[\text{elección}]$
- 2.1.5. Asignación de clase para el site r**
 - Si $\text{probabilidad}[\text{elección}] \leq \text{probabilidad}[\text{actual}]$:
 $r = \text{clase_elección}$

en caso contrario:

$$p = \exp -[(\text{probabilidad}[\text{elección}] - \text{probabilidad}[\text{actual}]) / T^{(n)}]$$

ξ_2 = variable aleatoria de distribución uniforme en el intervalo [0,1].

Si $\xi_2 < p$:

$$r = \text{clase_elección}$$

2.2. Decrementar la temperatura: $T^{(n+1)} = 0.95 * T^{(n)}$

2.3. Regresar al paso 2 hasta completar el número especificado de iteraciones.

4.4 Resultados

Para evaluar el funcionamiento de nuestro algoritmo utilizamos 3 distintos tipos de imágenes, que presentan características muy particulares. Todas las imágenes empleadas se encuentran en 256 niveles de gris y tienen un tamaño de 512x512 *pixeles*. Los tipos de imágenes utilizados son:

- ♦ Fotografía aérea de alta resolución (50 cm/*pixel*)
- ♦ Imagen SPOT pancromática (10 m/*pixel*)
- ♦ Imagen SAR (30 m/*pixel*)

En cada una de las imágenes seleccionadas se buscó la presencia de estructuras finas contenidas en regiones muy homogéneas, para evidenciar el buen funcionamiento del algoritmo propuesto en la preservación de dichas estructuras. De igual forma, el contar con elementos de estructura regular dentro de las imágenes seleccionadas fue de utilidad para mostrar que la parte contextual de nuestra propuesta no provoca una degradación notable de los contornos.

Para tener una mejor referencia sobre la calidad de los resultados obtenidos, las mismas imágenes fueron segmentadas utilizando de forma directa diversos métodos ampliamente utilizados:

- ♦ K-medias
- ♦ Esquema Bayesiano clásico
- ♦ Modos Condicionales Iterativos (ICM)
- ♦ Esquema Markoviano clásico

Los resultados obtenidos con cada tipo de imagen empleado y las comparaciones respectivas con los métodos arriba mencionados se presentan a continuación.

4.4.1 Segmentación de fotografías aéreas de alta resolución

La primera imagen que se consideró para la aplicación del algoritmo de segmentación propuesto es una fotografía aérea pancromática de alta resolución que muestra una escena urbana de la Ciudad de México (figura 4.5). En esta imagen, las regiones de transición –que deben ser rechazadas en el paso 3 del algoritmo propuesto– son identificadas fácilmente, así como diferentes estructuras finas, las cuales deben preservarse.



Figura 4.5

Fotografía aérea de alta resolución que muestra una región urbana de la Ciudad de México (imagen original en 256 niveles de gris, 512x512 píxeles).

Para el proceso de segmentación definimos 4 clases típicas, basados en las características espectrales de las estructuras representativas. Las principales estadísticas de cada una de las clases consideradas se muestran en la tabla 4.1.

	ω_0	ω_1	ω_2	ω_3
Media de la clase	66	107	164	210
Desviación estándar	11	14	15	14
Probabilidad <i>a priori</i>	0.3282	0.2380	0.2067	0.2271

Tabla 4.1

Estadísticas para la imagen segmentada en 4 clases.

En base a estos datos, aproximamos la función de probabilidad conjunta para cada clase y llevamos a cabo la reclasificación de la imagen mediante un clasificador de máximo *a posteriori* (MAP), los píxeles cuya máxima probabilidad fue menor que 0.8 (el *umbral de rechazo*) fueron etiquetados como clase 5.



Figura 4.6

Imagen que muestra la clase de rechazo (en blanco) después del paso III del algoritmo propuesto (sección 4.3).

Como se puede observar, muchas de las zonas de transición fueron asignadas a la clase de rechazo, y las estructuras finas claramente diferenciadas no fueron alteradas en gran escala. También es importante hacer notar que la clase de rechazo no sólo aloja a píxeles aislados, sino también a grupos homogéneos de algunos cientos de píxeles, por lo cual incluso regiones relativamente extensas pueden ser reasignadas, dependiendo de su entorno. Los píxeles rechazados representan el 20.15% de la imagen.

La imagen definitiva en 4 clases se muestra en la figura 4.7. Resulta evidente que las zonas de transición se volvieron más homogéneas mientras que las estructuras finas fueron preservadas casi en su totalidad. Una observación especial en esta etapa es el bajo requerimiento de tiempo para el procesamiento de la imagen, el cual fue de aproximadamente el 25% del tiempo requerido para un esquema Markoviano clásico con minimización por recocido simulado, ya que sólo los píxeles de la clase de rechazo fueron considerados en esta etapa.



Figura 4.7

*Resultado del algoritmo de segmentación propuesto,
la imagen se segmentó en 4 clases.*

Para hacer un análisis más detallado de los resultados obtenidos, a continuación se presenta una serie de imágenes que corresponden a una ventana de 256x256 píxeles extraída de la imagen original, la cual fue segmentada en 4 clases mediante cuatro algoritmos distintos. Los resultados se presentan en falso color para hacer más evidentes las diferencias entre cada uno de los resultados obtenidos.

En la ventana original (figura 4.8) tenemos dos regiones de especial interés para nuestro análisis. Una de ellas es la zona de transición en la esquina inferior derecha; la otra es la región localizada al centro, donde el concepto de estructura fina bien definida es claramente ejemplificado.

La figura 4.9 muestra en color blanco la clase de rechazo después de aplicar el paso 3 del algoritmo de segmentación propuesto. En esta imagen se observa que un alto porcentaje de los píxeles pertenecientes a la zona de transición antes mencionada fueron asignados a la clase de rechazo. Por otra parte, el conjunto de estructuras lineales en el centro de la imagen logra mantenerse bien definido, lo que permite que al llegar a la parte contextual de nuestro algoritmo no sean degradadas en gran escala.

En la imagen definitiva, segmentada en 4 clases con el algoritmo propuesto –mostrada en las figuras 4.7 y 4.10– se observa como las regiones de transición, que fueron asignadas a la clase de rechazo, se vuelven homogéneas al considerar un enfoque contextual para su reclasificación. En lo referente a las estructuras finas del centro de la imagen, también terminaron por definirse mejor al finalizar el proceso adaptivo de segmentación. En general, podemos afirmar que para la imagen seleccionada el resultado de la segmentación exhibe regiones homogéneas, contornos bien definidos, y respeto de las estructuras finas presentes en la imagen. Aunque esta propuesta también puede permitir la presencia de píxeles aislados –pero en una tasa mucho menor que los clasificadores no contextuales– muestra un muy buen desempeño, ya que cumple con los objetivos que se consideraron al desarrollarlo.

En la figura 4.11 se muestra el resultado obtenido al utilizar un planteamiento Markoviano clásico. Este esquema genera imágenes segmentadas de buena calidad, con regiones homogéneas y una buena definición de contornos cuando se trata de estructuras relativamente grandes. Sin embargo, las estructuras finas de la imagen pueden verse afectadas al ser fusionadas eventualmente a regiones adyacentes, principalmente cuando dichas estructuras se encuentran contenidas en una región claramente homogénea, como en la imagen utilizada en este caso.

El clasificador Bayesiano clásico (figura 4.12) ofrece ejecuciones rápidas y preserva prácticamente en su totalidad las estructuras finas de la imagen a clasificar, sin embargo, permite la existencia de muchos píxeles aislados, especialmente en las zonas de transición, por lo cual los contornos no se definen claramente.

El último resultado de segmentación que se presenta en la figura 4.13 es el obtenido por el algoritmo de Modos Condicionales Iterativos (ICM), el cual es de especial utilidad cuando se busca tener regiones lo más homogéneas posibles, lo cual en este ejemplo resulta contraproducente, ya que las estructuras que se pretende preservar son eliminadas casi en su totalidad.

Otra forma de analizar las diferencias entre los algoritmos de segmentación estudiados es mediante una serie de estadísticas básicas. De esta forma, en la tabla 4.2 se puede observar cómo el algoritmo ICM provoca grandes cambios en la asignación de clases: el número de píxeles asignados a la clase 2 –a la cual pertenecen principalmente las

estructuras finas del centro de la imagen- decrece 4% (alrededor de 10,500 píxeles) con referencia a nuestra propuesta, como consecuencia de la casi absoluta eliminación de estas estructuras.

La tabla 4.3 muestra a su vez el porcentaje en que difieren las imágenes obtenidas con los diversos métodos de segmentación. Aquí resulta interesante observar que la menor diferencia del resultado obtenido con nuestra propuesta se da en relación con el resultado del esquema Markoviano clásico, lo cual confirma lo que se infiere de un análisis visual de los resultados para esta imagen: la diferencia principal entre el resultado obtenido con nuestro esquema y el del modelo Markoviano clásico radica en la preservación de las estructuras finas de la escena segmentada.



Figura 4.8

Subimagen original [256x256x256].



Figura 4.9

Imagen mostrando la clase de rechazo (en blanco).

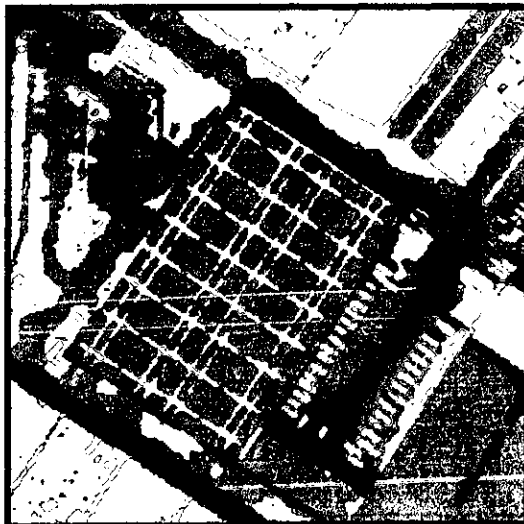


Figura 4.10

Segmentación en 4 clases utilizando el algoritmo propuesto.



Figura 4.11

Resultado de un clasificador Markoviano clásico.



Figura 4.12

Resultado de un clasificador Bayesiano clásico.

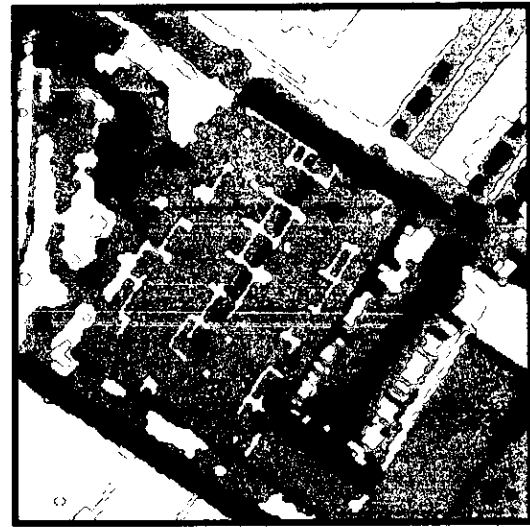


Figura 4.13

Resultado del algoritmo ICM (Modos Condicionales Iterativos).

	ω_0 [%] negro	ω_1 [%] azul	ω_2 [%] rojo	ω_3 [%] amarillo
K-medias	32.62	23.83	20.33	23.22
Bayes clásico	34.05	22.72	19.46	23.77
ICM	37.94	21.76	16.74	23.56
MRF clásico	33.60	23.19	20.01	23.20
Esquema propuesto	31.83	24.84	20.75	22.98

Tabla 4.2

Porcentaje de píxeles asignados a cada clase, de acuerdo al algoritmo de segmentación empleado.

	K-medias	Bayes Clásico	ICM	MRF Clásico	Esquema Propuesto
K-medias	0	1.53	14.33	10.50	8.11
Bayes clásico	1.53	0	14.02	10.86	8.44
ICM	14.33	14.02	0	8.62	12.64
MRF clásico	10.50	10.86	8.62	0	7.31
Esquema propuesto	8.11	8.44	12.64	7.31	0

Tabla 4.3

Porcentaje de píxeles asignados a distintas clases al comparar los resultados obtenidos con los diferentes métodos de segmentación analizados.

La segunda fotografía aérea que se consideró para la aplicación del algoritmo propuesto es la mostrada en la figura 4.14. Esta imagen es de especial interés, ya que incluye una cantidad considerable de segmentos lineales cuyo grosor fluctúa entre 1 y 3 píxeles. La mayoría de estos segmentos se diferencia claramente de su entorno, por lo cual resulta de gran utilidad para la evaluación de nuestro algoritmo.

Una región a la cual debemos prestar especial atención es la que se encuentra en el extremo derecho de la ventana (en color claro), ya que en ella se aprecian tres aspectos que son de relevancia en nuestro planteamiento: una parte muy homogénea (parte superior de la estructura), una estructura fina lineal (formado por 4 pequeños segmentos en color oscuro, por lo que se diferencian claramente y deben ser preservados por el algoritmo), y una zona de transición en la parte inferior (la cual debe ser fusionada con el resto de la región).



Figura 4.14

Fotografía aérea de alta resolución (50 cm/píxel) que muestra una parte de la zona de pistas del aeropuerto internacional de la Ciudad de México [256 niveles de gris, 512x512 píxeles].

Dadas las características espectrales de la imagen utilizada, se decidió segmentar en 3 clases. La clase 0 representa las regiones más oscuras de la imagen (lo que es la pista en si), las estructuras finas se encuentran distribuidas entre las clases 1 y 2. Las estadísticas de las clases consideradas se presentan en la tabla 4.4.

	ω_0	ω_1	ω_2
Media de la clase	31	103	223
Desviación estándar	15	27	27
Probabilidad a priori	0.7748	0.1238	0.1014

Tabla 4.4

Estadísticas para la imagen segmentada en 3 clases.

La figura 4.15 muestra la clase de rechazo después del paso III del algoritmo adaptivo propuesto. En este caso el umbral considerado fue de 0.8 (con relación a la probabilidad condicional máxima de cada pixel).

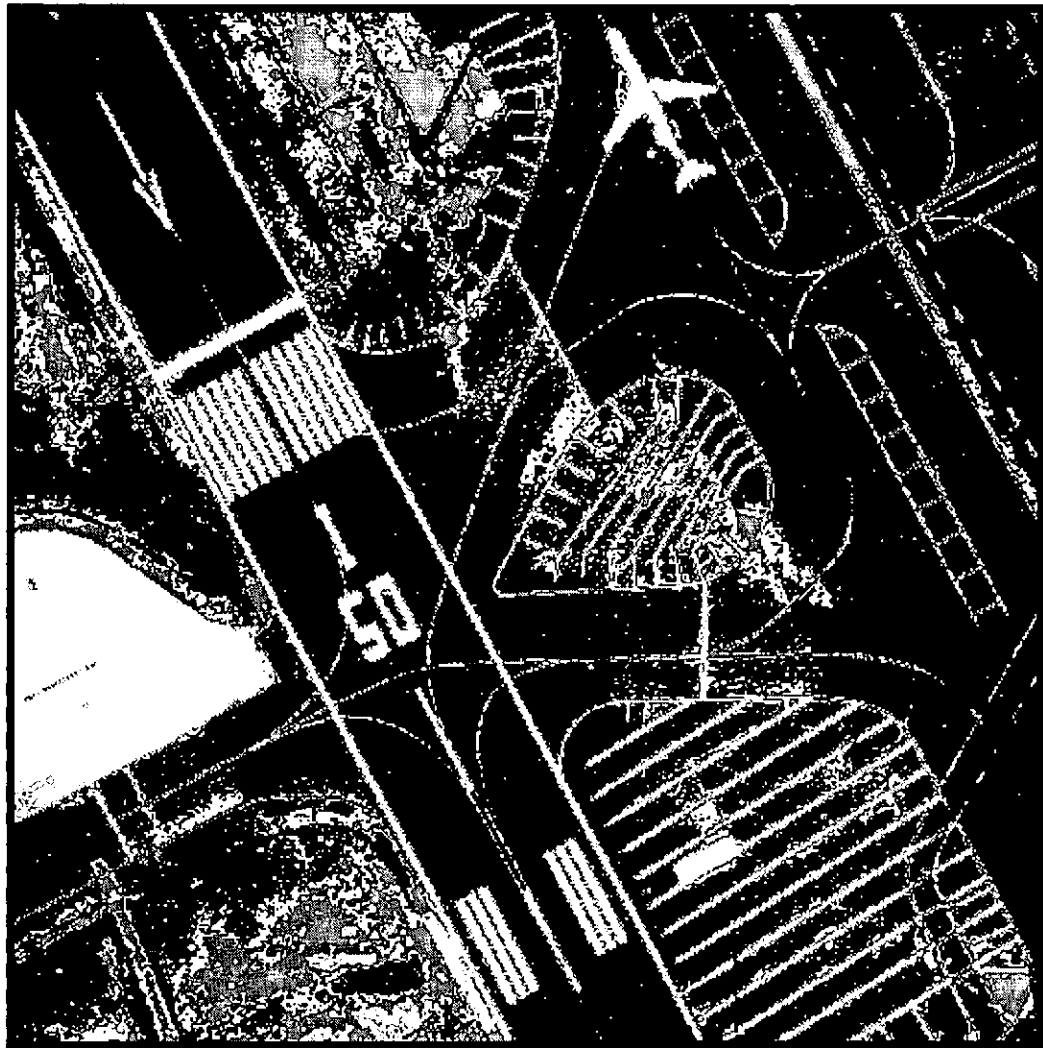


Figura 4.15

Imagen que muestra la clase de rechazo en color blanco, el número de pixeles rechazados representa el 13.52% de la imagen.

La imagen definitiva, segmentada en 3 clases, se muestra en la figura 4.16. También en este caso la mayoría de los segmentos lineales se preservaron de forma casi completa. Aquí podemos evaluar el desempeño del algoritmo en base a la región que se mencionó anteriormente (la zona clara a la derecha de la imagen), donde se observa que –aunque resultaron un poco degradados– los cuatro segmentos lineales que se habían mencionado anteriormente aún se distinguen. Estos segmentos fueron preservados también por las aproximaciones no contextuales (K-medias, MAP), mientras que en las contextuales (esquema Markoviano clásico, ICM) fueron fusionadas completamente a la región adyacente.

Por otra parte, en lo que se refiere a las zonas de transición presentes en la imagen (definidas claramente por la clase de rechazo mostrada en la imagen anterior), al ser reasignadas por un esquema contextual dieron como resultado regiones mucho más homogéneas. De hecho, la presencia de píxeles aislados en la imagen segmentada es mínima y se observa una buena definición de contornos, como es el caso de la silueta del avión.



Figura 4.16

*Imagen definitiva de la segmentación en 3 clases
utilizando el algoritmo propuesto.*

A continuación se presenta una serie de ventanas en falso color, a partir de las cuales se hará un análisis comparativo entre los resultados obtenidos mediante diversos algoritmos de segmentación. La ventana de estudio (figura 4.17) tiene una dimensión de 256x256 píxeles, y presenta una variedad de elementos de interés, como lo son la región de la esquina inferior izquierda, que es muy heterogénea; las estructuras lineales en la parte derecha de la imagen, y la gran región homogénea que representa la pista principal.

La clase de rechazo está formada principalmente por elementos de las regiones poco homogéneas y por regiones de transición en los contornos de ciertas estructuras. Como puede observarse en la figura 4.18, las estructuras lineales quedan en su mayor parte asignadas a una clase definitiva desde esta etapa del proceso de segmentación. En las principales zonas homogéneas (que corresponden a la clase 0, en color negro) se llegan a apreciar algunos píxeles que cayeron en la clase de rechazo, pero al estar aislados es un hecho que en la reasignación serán fusionados a su entorno.

El resultado final –figura 4.19– confirma los comentarios anteriores, al mostrar regiones homogéneas en la esquina inferior izquierda de la imagen; así como preservación de las principales estructuras lineales; sólo aquellas que no estaban muy bien definidas (las que atraviesan la pista principal) aparecen de forma discontinua, lo que tiene como resultado en ocasiones la presencia de algunos píxeles aislados. Por otra parte, todas las regiones pertenecientes a la clase 0 son muy homogéneas, y las líneas que las delimitan están muy bien definidas.

Al igual que en el caso de la imagen anterior, el esquema Markoviano clásico (figura 4.20) genera una degradación considerable de las estructuras finas de la imagen, aunque no en una forma tan marcada como lo hace el algoritmo ICM (figura 4.22), el cual prácticamente elimina todas las estructuras finas pertenecientes a la clase 1 (en color azul) que tienen un grosor promedio de dos píxeles. Como se muestra en la tabla 4.5, la diferencia en el porcentaje de píxeles asignados a la clase 1 entre nuestro algoritmo y el ICM es de aproximadamente 10% (alrededor de 26,000 píxeles), lo cual marca una gran diferencia en cuanto a la calidad del resultado.

Un caso especial en esta serie de resultados es el ofrecido por el esquema Bayesiano (figura 4.21), que para este ejemplo resulta particularmente bueno. Esto se debe a que la clase que predomina en la imagen es la 0 (en color negro), lo que hace que la probabilidad *a priori* para esta clase sea mucho mayor que la de las otras dos (ver tabla 4.4), lo que provoca que las estructuras finas que no tienen un nivel de gris muy cercano a alguno de los centros de clase no sean muy bien preservadas (lo que en este caso va en detrimento de las estructuras finas pertenecientes a la clase 1, ya que no se diferencian tan claramente del resto de la escena) pero, por otra parte, las regiones pertenecientes a la clase 0 resultan ser muy homogéneas.

Comparando una vez más el resultado ofrecido por nuestro esquema con el que provee el esquema Markoviano clásico, encontramos que en esta ocasión la diferencia es de 6.05%, la cual resulta menor que la obtenida para la primer imagen procesada. Aunque –debido a una mayor presencia de estructuras finas en este segundo ejemplo– la diferencia en el análisis visual podría parecer mayor para esta segunda imagen, no debemos perder de vista que tanto la clase 0 como la clase 2 incluyen a las regiones mejor definidas de la imagen (lo que las hace muy homogéneas), y el asignar los píxeles pertenecientes a dichas clases no representa mayor complejidad, por lo que ambos algoritmos las asignan de forma muy semejante. La diferencia está dada principalmente por los píxeles pertenecientes a la clase 2, a la cual corresponden las

estructuras más finas de la escena segmentada. Estas estructuras están rodeadas principalmente por píxeles pertenecientes a la clase 0, lo que provoca que en los métodos contextuales sean fusionadas a dicha clase.

La tabla 4.5 clarifica el hecho recién mencionado: para todas las clases el porcentaje de píxeles asignados a la clase 2 (en color amarillo) es de alrededor del 10%, mientras que las que muestran un alto grado de variación son la 0 y la 1, que van de 75% a 85% y de 4% a 14%, respectivamente.



Figure 4.17

Subimagen original [256x256x256].

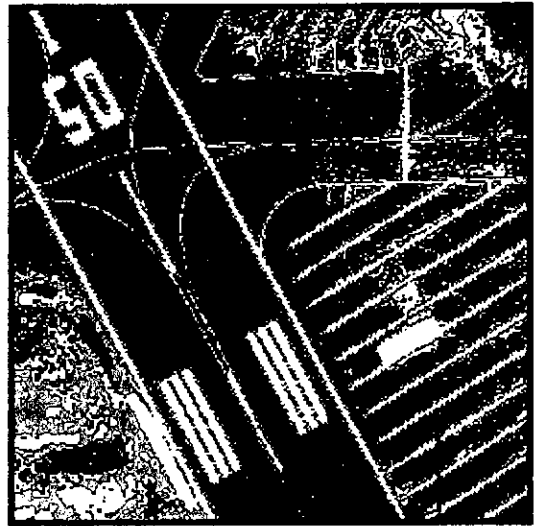


Figure 4.18

Imagen que muestra la clase de rechazo (en blanco).

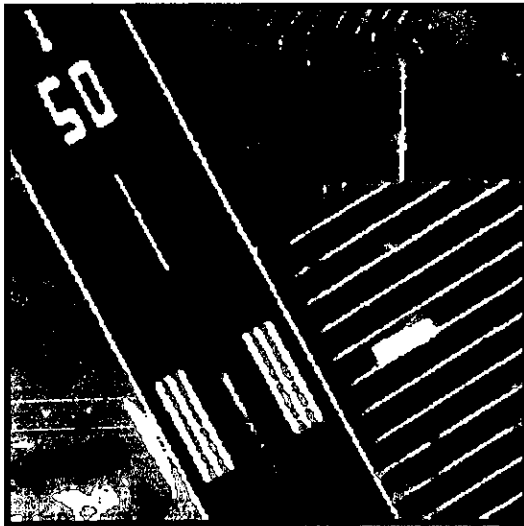


Figura 4.19

Segmentación definitiva en 3 clases aplicando el esquema propuesto.

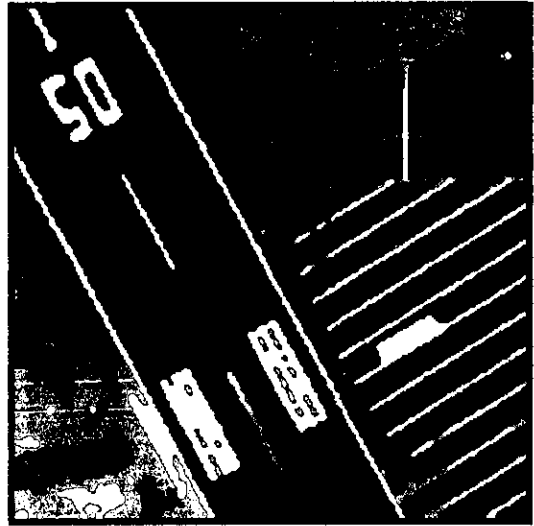


Figure 4.20

Resultado obtenido mediante un clasificador Markoviano clásico.

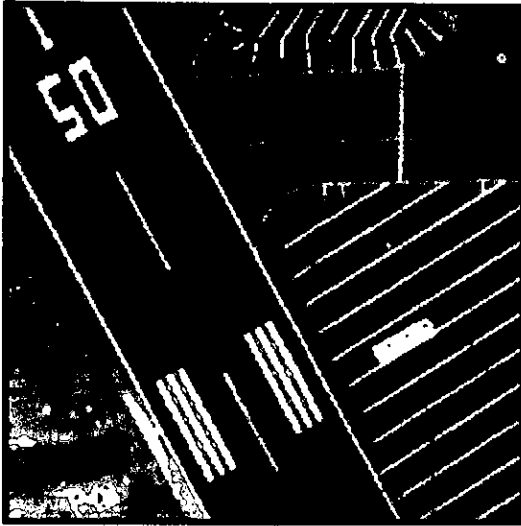


Figure 4.21

Resultado del esquema Bayesiano clásico.

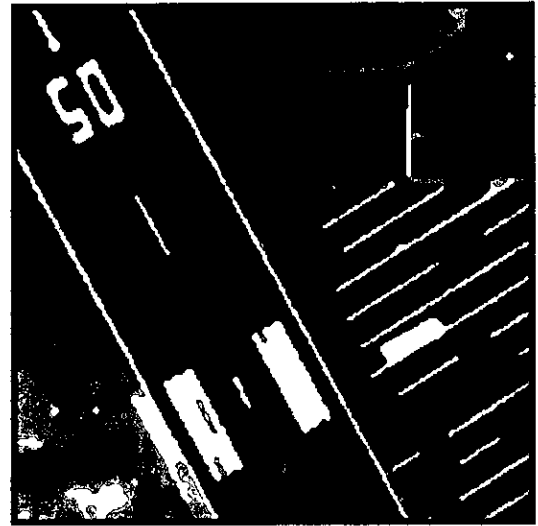


Figure 4.22

Resultado del algoritmo de Modos Condicionales Iterativos.

	ω_0 [%] negro	ω_1 [%] azul	ω_2 [%] amarillo
K-medias	77.48	12.38	10.14
Bayes clásico	80.81	9.15	10.04
ICM	85.77	4.30	9.93
MRF clásico	79.81	9.31	10.88
Esquema propuesto	75.14	14.68	10.18

Tabla 4.5

Porcentaje de pixeles asignados a cada clase, de acuerdo al algoritmo de segmentación empleado.

	K-medias	Bayes Clásico	ICM	MRF Clásico	Esquema Propuesto
K-medias	0	3.43	9.72	5.84	4.41
Bayes clásico	3.43	0	6.98	5.98	7.36
ICM	9.72	6.98	0	6.47	11.47
MRF clásico	5.84	5.98	6.47	0	6.05
Esquema propuesto	4.41	7.36	11.47	6.05	0

Tabla 4.6

Porcentaje de pixeles asignados a distintas clases al comparar los resultados obtenidos con los diferentes métodos de segmentación analizados.

4.4.2 Segmentación de imágenes SPOT

Una vez evaluado nuestro algoritmo utilizando fotografías aéreas, decidimos aplicarlo a imágenes de percepción remota que ofrecieran una menor resolución. Para esto elegimos una imagen SPOT pancromática, que ofrece una resolución de 10 metros por pixel. La ventana seleccionada muestra una escena urbana de la ciudad de México, la cual incluye la zona del aeropuerto, que en este caso representa la región más homogénea de la imagen. En esta imagen las estructuras finas están representadas principalmente por las calles y algunas otras estructuras regulares de menor tamaño (como las de color claro que se observan en la zona de pistas del aeropuerto).



Figura 4.23

Imagen SPOT pancromática con una resolución de 10 metros por pixel [512x512x256].

Dadas las características espectrales de la imagen, se decidió segmentarla en 3 clases, con la finalidad de que se hiciera una buena distinción de las calles (clase 0), mientras que la clase 2 representa principalmente las zonas habitacionales. La tabla 4.7 muestra las estadísticas básicas para cada una de las clases consideradas.

	ω_0	ω_1	ω_2
Media de la clase	128	163	202
Desviación estándar	15	10	18
Probabilidad <i>a priori</i>	0.3459	0.5235	0.1306

Tabla 4.7

Estadísticas para la imagen SPOT segmentada en 3 clases.

En este caso el porcentaje de píxeles rechazados fue de 12.07%. La clase de rechazo se encuentra muy distribuida en lo que son las zonas habitacionales, que es donde se presenta una constante variación en los niveles de gris.



Figura 4.24

Imagen obtenida después del paso III del esquema propuesto, la clase de rechazo se muestra en blanco. El umbral de rechazo considerado fue 0.8.

En la imagen definitiva en 3 clases se observa una buena definición de las calles principales, aunque las calles menores –que son difíciles de identificar incluso en la imagen original– se fusionaron a las zonas de residencias. Esto se explica fácilmente, ya que una calle 'secundaria' de dos carriles tiene un ancho menor a los 10 metros, por lo que en la imagen no pueden ser detectadas de manera correcta. Por otra parte, la zona del aeropuerto presenta zonas homogéneas y una buena definición de los contornos de las pistas, que es otro de los objetivos con que se desarrolló el algoritmo aquí presentado.

Si bien las regiones habitacionales muestran una gran variación en cuanto a sus características espectrales, en la imagen se observa que la incidencia de píxeles aislados es en realidad baja, esto a comparación con los resultados ofrecidos por algoritmos no contextuales. Esa es una de las ventajas del esquema propuesto: únicamente las regiones de transición muy heterogéneas son enviadas a la clase de rechazo para su posterior reasignación mediante un enfoque contextual, lo que las hace más homogéneas.

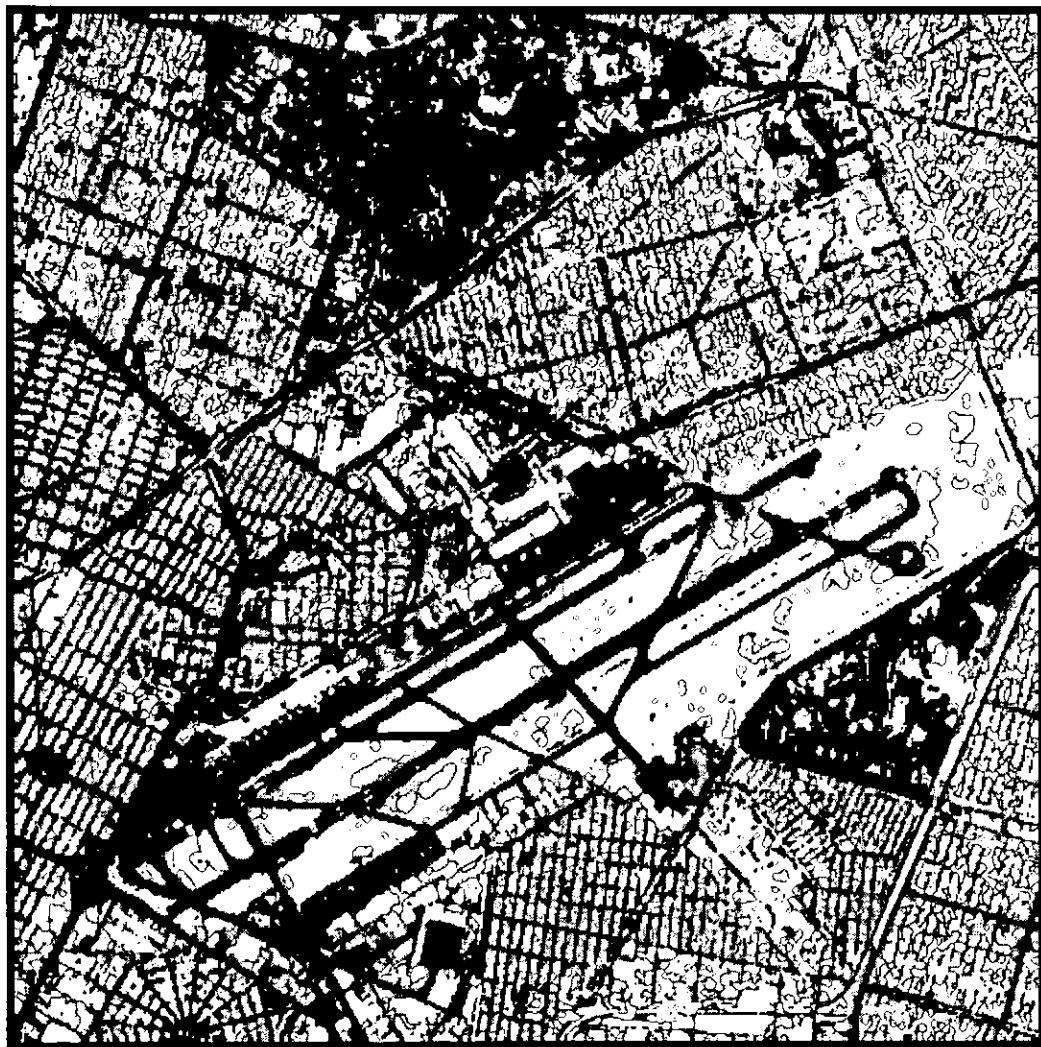


Figura 4.25

Imagen segmentada en 3 clases mediante el algoritmo propuesto.

Para el análisis más detallado de los resultados obtenidos se consideró la región mostrada en la figura 4.26 (256x256 píxeles), que incluye parte de la zona de pistas del aeropuerto y una porción de regiones habitacionales con tres calles paralelas que se distinguen fácilmente del resto de la escena.

Como ya se indicó, la clase de rechazo se encuentra distribuida de forma prácticamente homogénea sobre todas las zonas habitacionales (figura 4.27), y –para la ventana que estamos analizando– en los contornos de la zona de pistas del aeropuerto. Dada esta distribución de la clase de rechazo, no se espera que toda una zona resulte ser mucho más homogénea después de hacer la reclasificación mediante el esquema Markoviano, pero si se espera una importante disminución de los píxeles aislados (región habitacional) y una mejor definición de contornos (región del aeropuerto).

El resultado final –figura 4.28– muestra que las consideraciones anteriores eran correctas, ya que la incidencia de píxeles aislados puede considerarse baja, mientras que por otra parte las regiones claramente homogéneas de la imagen analizada quedaron muy bien definidas. De igual forma, las principales calles de la escena se pueden distinguir con facilidad.

La aplicación directa del esquema Markoviano clásico da como resultado una imagen con una mejor definición de contornos y regiones muy homogéneas, pero esto mismo resulta contraproducente en lo que se refiere a la mejor identificación de las calles, ya que algunas de ellas son fusionadas en parte a las regiones contiguas, lo que las hace aparecer de forma discontinua. El resultado obtenido mediante este esquema se muestra en la figura 4.29.

Dadas las características de la imagen, lo que lleva a un predominio en la escena de la clase 1 (en color azul, con probabilidad *a priori* del 52%), resulta que en la aproximación Bayesiana también se ve perjudicada la detección de calles ya que, para algunas porciones de ellas, al obtener su probabilidad *a posteriori* (el producto de su probabilidad condicional por la probabilidad *a priori* de la clase considerada) esta resulta ser mayor para la clase 1, por lo cual son asignadas a ella por el estimador MAP, cuyo resultado se muestra en la figura 4.30.

El resultado del algoritmo ICM, mostrado en la figura 4.31, deja en evidencia una vez más que este enfoque puramente contextual resulta de especial interés para segmentar imágenes mostrando grandes regiones con poca variación espectral, o bien, cuando no se pretende preservar demasiados detalles de la imagen original. Pero cuando por el contrario se busca preservar la mayor cantidad de detalles finos de la imagen (como es nuestro caso) el algoritmo ICM representa poca utilidad, ya que estos se pierden casi en su totalidad.

En imágenes de este tipo, donde abunda la presencia de pequeñas estructuras, el algoritmo propuesto ofrecerá resultados más cercanos a los ofrecidos por métodos no contextuales (K-medias, esquema Bayesiano) que a aquellos obtenidos con algoritmos contextuales (esquema Markoviano, ICM), como se observa en la tabla 4.9. Esta diferencia, en relación a los algoritmos no contextuales, está dada principalmente por la eliminación de píxeles aislados en las zonas de contornos; mientras que en relación a los algoritmos contextuales está dada por el respeto que hace nuestro algoritmo de las estructuras finas.

Una observación interesante para las imágenes en las que predomina claramente una clase (como en este caso la clase 1, y para la fotografía aérea anterior la clase 0) es que provoca un efecto similar en los algoritmos contextuales y en los probabilísticos: la clase con mayor presencia resulta 'beneficiada' por estos algoritmos, por lo cual al final de la segmentación el número de píxeles asignados a dicha clase aumenta, lo cual va en perjuicio de las estructuras finas de la imagen, que por lo general pertenecen a una clase no predominante en la imagen. Analizando las tablas respectivas (4.6 y 4.9) se observa cómo el algoritmo propuesto reduce el porcentaje de píxeles asignados a la clase predominante, mientras que aumenta el de aquellos asignados a la clase a la cual pertenecen las estructuras finas que se pretende preservar.

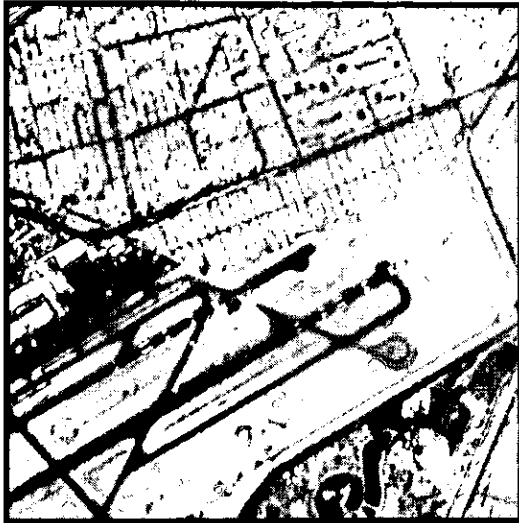


Figura 4.26

Subimagen original [256x256x256].



Figura 4.27

Imagen mostrando la clase de rechazo (en blanco).



Figura 4.28

Resultado de la segmentación en 3 clases con el esquema propuesto.

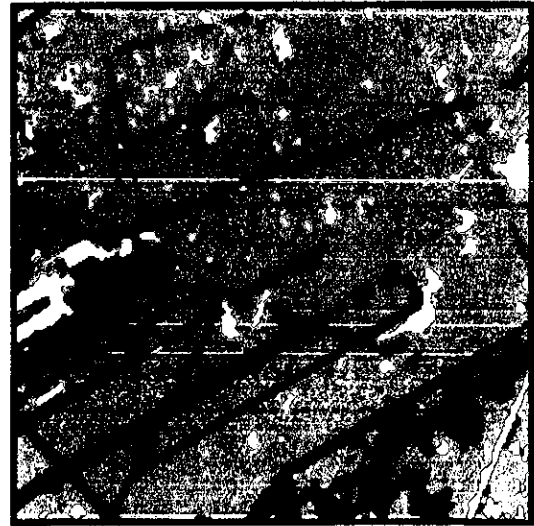


Figura 4.29

Resultado de un clasificador Markoviano Clásico.



Figura 4.30

Resultado de un clasificador Bayesiano clásico.



Figura 4.31

Resultado del esquema ICM.

	ω_0 [%] negro	ω_1 [%] azul	ω_2 [%] amarillo
K-medias	34.59	52.35	13.06
Bayes clásico	32.05	58.94	9.01
ICM	28.92	66.38	4.70
MRF clásico	31.32	60.46	8.22
Esquema propuesto	39.01	44.69	16.30

Tabla 4.8

Porcentaje de pixeles asignados a cada clase, de acuerdo al algoritmo de segmentación empleado.

	K-medias	Bayes Clásico	ICM	MRF Clásico	Esquema Propuesto
K-medias	0	6.60	20.01	14.29	7.79
Bayes clásico	6.60	0	14.90	11.93	14.34
ICM	20.01	14.90	0	8.74	25.25
MRF clásico	14.29	11.93	8.74	0	18.02
Esquema propuesto	7.79	14.34	25.25	18.02	0

Tabla 4.9

Porcentaje de pixeles asignados a distintas clases al comparar los resultados obtenidos con los diferentes métodos de segmentación analizados.

ESTA TESIS NO SALE
DE LA BIBLIOTECA

4.4.3 Segmentación de imágenes SAR

Para evaluar nuestro algoritmo con imágenes del tipo SAR se utilizó la imagen mostrada a continuación, la cual tiene una resolución de 30 metros por píxel. En esta imagen se consideraron 4 clases para la segmentación, en base a sus características espectrales. Dos de esas clases (0 y 1) corresponden a la zona de pistas del aeropuerto, y las dos restantes (clases 2 y 3) a las zonas habitacionales circundantes. El objetivo de esta segmentación es definir bien la zona de pistas, así como las avenidas principales de la zona mostrada.

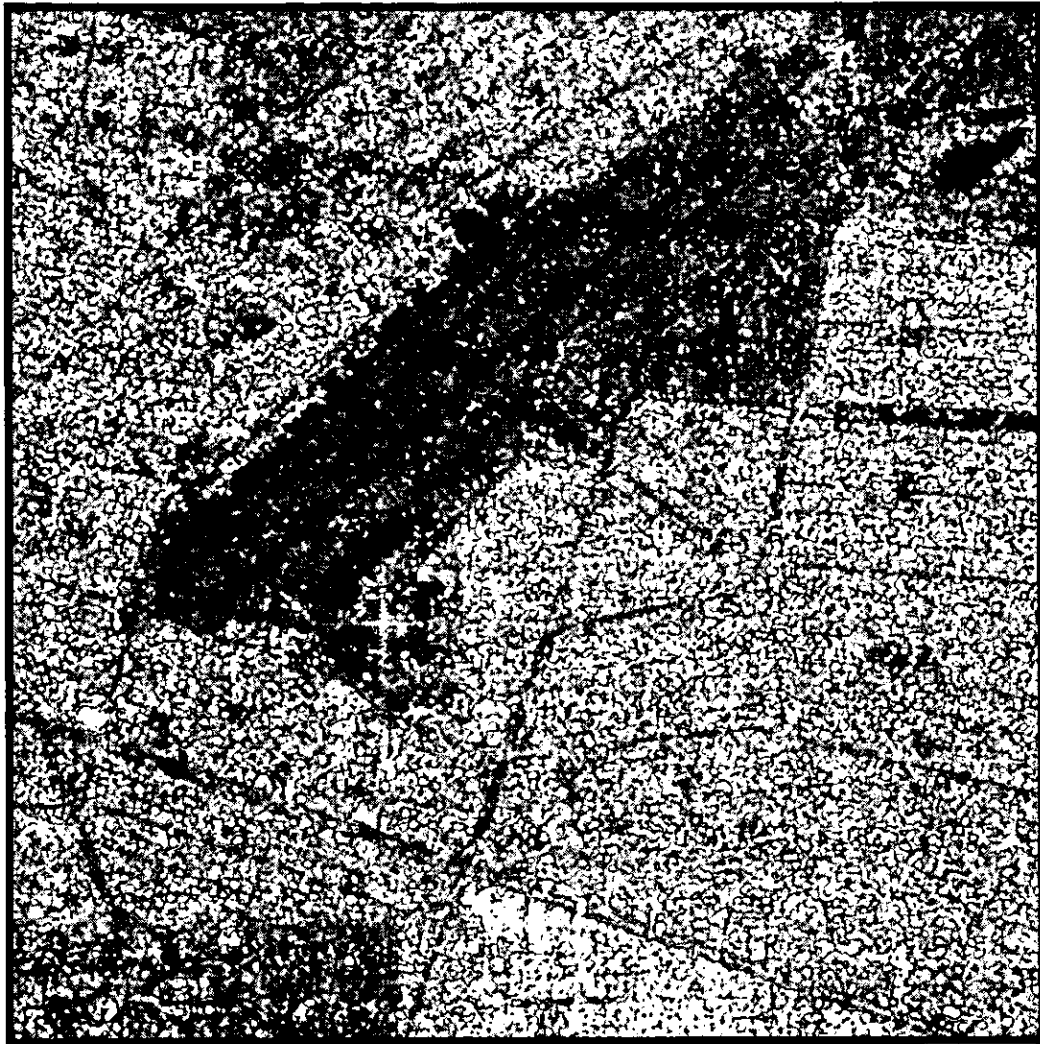


Figura 4.32

Imagen SAR adquirida por el satélite ERS-1 el 19 de septiembre de 1997, la escena corresponde a la zona urbana que rodea al aeropuerto internacional de la Ciudad de México [512x512 píxeles, 256 niveles de gris].

Las imágenes SAR son de especial interés para la obtención de información temática sobre regiones extensas. Este tipo de imágenes tienen varias ventajas bien conocidas, como lo es la posibilidad de obtener imágenes independientemente del estado del tiempo. Desafortunadamente, la presencia de *speckle* y, en general, la pobre calidad de las imágenes SAR representa una seria desventaja al aplicar de forma directa algún método de extracción de información. De hecho, muchas veces es necesario incorporar datos externos de referencia (obtenidos *in situ*) para garantizar la calidad de los resultados.

La imagen 4.33 muestra el resultado obtenido al aplicar nuestro algoritmo de forma directa sobre la imagen original, lo que resultó en una imagen segmentada con una muy alta incidencia de píxeles aislados. El mismo resultado se obtiene con los otros métodos no contextuales aplicados. En el caso de la imagen analizada, aún los algoritmos contextuales presentaron problemas para ofrecer regiones muy homogéneas. Debido a esto, el extraer información a partir de estos resultados representa una tarea complicada y podría generar información poco confiable.

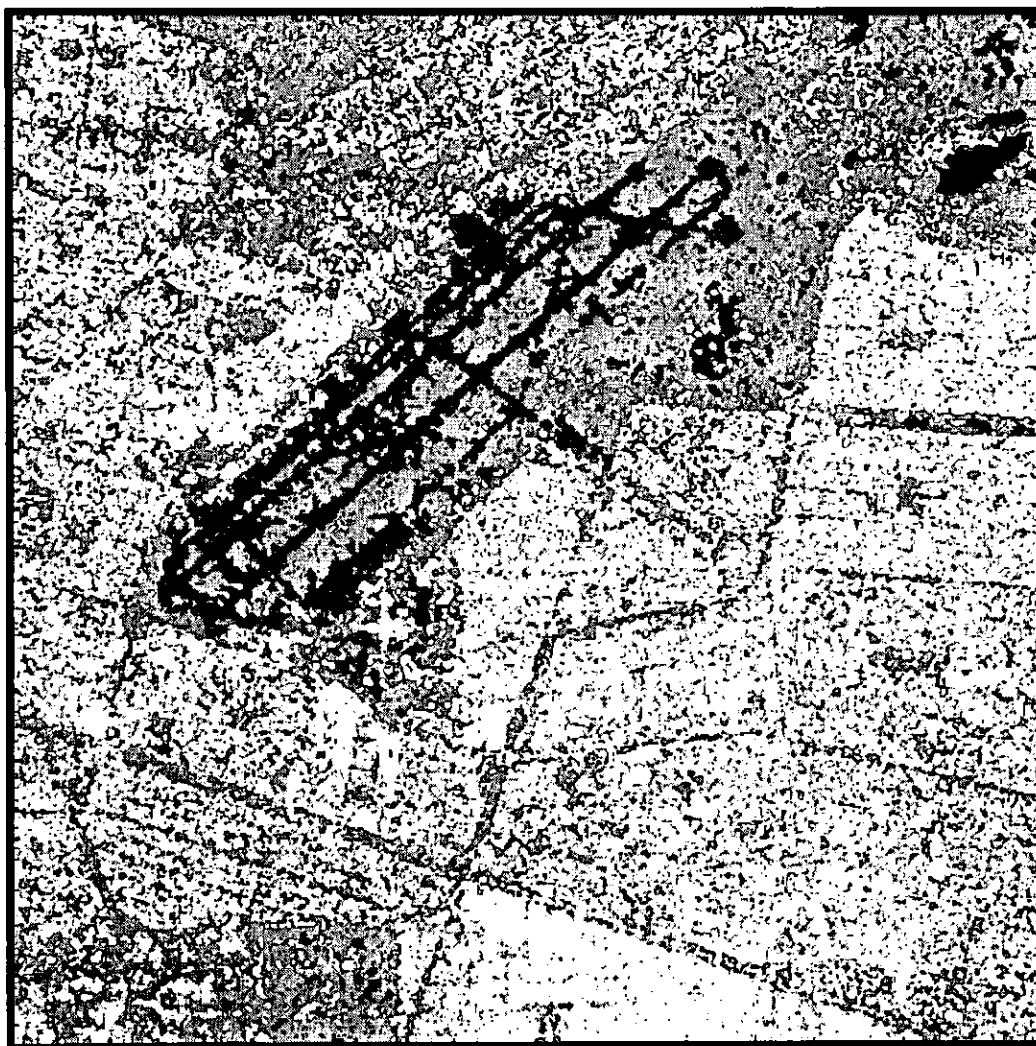


Figura 4.33

Resultado obtenido al aplicar el esquema de segmentación propuesto a la imagen SAR original. Se consideraron 4 clases de interés.

Resulta evidente que debe incluirse una etapa de filtrado para aprovechar las ventajas que ofrecen las imágenes SAR. En este contexto, se han desarrollado varios filtros para eliminar los efectos del *speckle*; sin embargo, en la mayoría de los casos, estos algoritmos también afectan información que puede ser de relevancia para el tipo de estudio que se pretende realizar a partir de estas imágenes.

En nuestro caso decidimos utilizar una propuesta de filtrado basada en la transformada polinomial [7]. Este algoritmo se basa en una técnica sensitiva direccional que adapta la tasa de reducción de ruido a la información de la imagen, es decir, a la presencia de características importantes de la imagen y a las estadísticas locales del ruido. La calidad del resultado se mejora al detectar la posición y orientación de los contornos de la imagen, y al considerar la naturaleza multiplicativa del *speckle* en este tipo de imágenes. La imagen filtrada mediante este algoritmo se presenta en la figura 4.34.



Figura 4.34

Imagen filtrada mediante el algoritmo adaptivo de reducción de ruido multiplicativo mediante transformadas polinomiales, se consideró una reducción de ruido del 40%.

Dadas las características de la imagen, en este caso no resultó conveniente el inicializar los centros de clase mediante el método de K-medias, por lo que se emplearon 4 ventanas de entrenamiento para definir cada una de las medias de clase. De estas cuatro ventanas se consideraron dos en la zona de pistas y dos en las zonas habitacionales. Las medias obtenidas y sus estadísticas básicas se muestran aquí.

	ω_0	ω_1	ω_2	ω_3
Media de la clase	30	52	98	143
Desviación estándar	6	9	12	26
Probabilidad a priori	0.0702	0.2756	0.3428	0.3114

Tabla 4.10

Estadísticas para la imagen SAR segmentada en 4 clases.

La figura 4.35 muestra la clase de rechazo. Se observa que desde esta etapa las zonas habitacionales se presentan como regiones homogéneas, y que la clase de rechazo está representada principalmente por agrupaciones de varios píxeles.



Figura 4.35

Clase de rechazo (en blanco) para la imagen SAR considerada.

La imagen 4.36 muestra la imagen final obtenida al aplicar el algoritmo de segmentación propuesto a la imagen filtrada mediante el esquema antes mencionado. En este caso desde el paso anterior (la definición de la clase de rechazo) se podía asegurar una mejor calidad del resultado obtenido al utilizar la imagen filtrada, ya que la mayor parte de las regiones urbanas fueron asignadas de forma definitiva desde la etapa no contextual del algoritmo propuesto, y ya se presentaban como regiones más homogéneas a las obtenidas a partir de la imagen sin filtrar.

Aquí es importante hacer notar que, sí por una parte el resultado muestra regiones mucho más homogéneas (especialmente en lo que se refiere a zonas habitacionales), por otra parte las estructuras que nos interesaba delimitar de forma clara (las pistas del aeropuerto y las principales avenidas) no fueron degradadas de forma muy marcada al someter la imagen original a los procesos adaptivos de filtrado y segmentación, ya que ambos esquemas tienen como objetivo el preservar, en el mayor grado posible, este tipo de estructuras, por lo que al ser utilizados en conjunción ofrecen resultados de buena calidad.

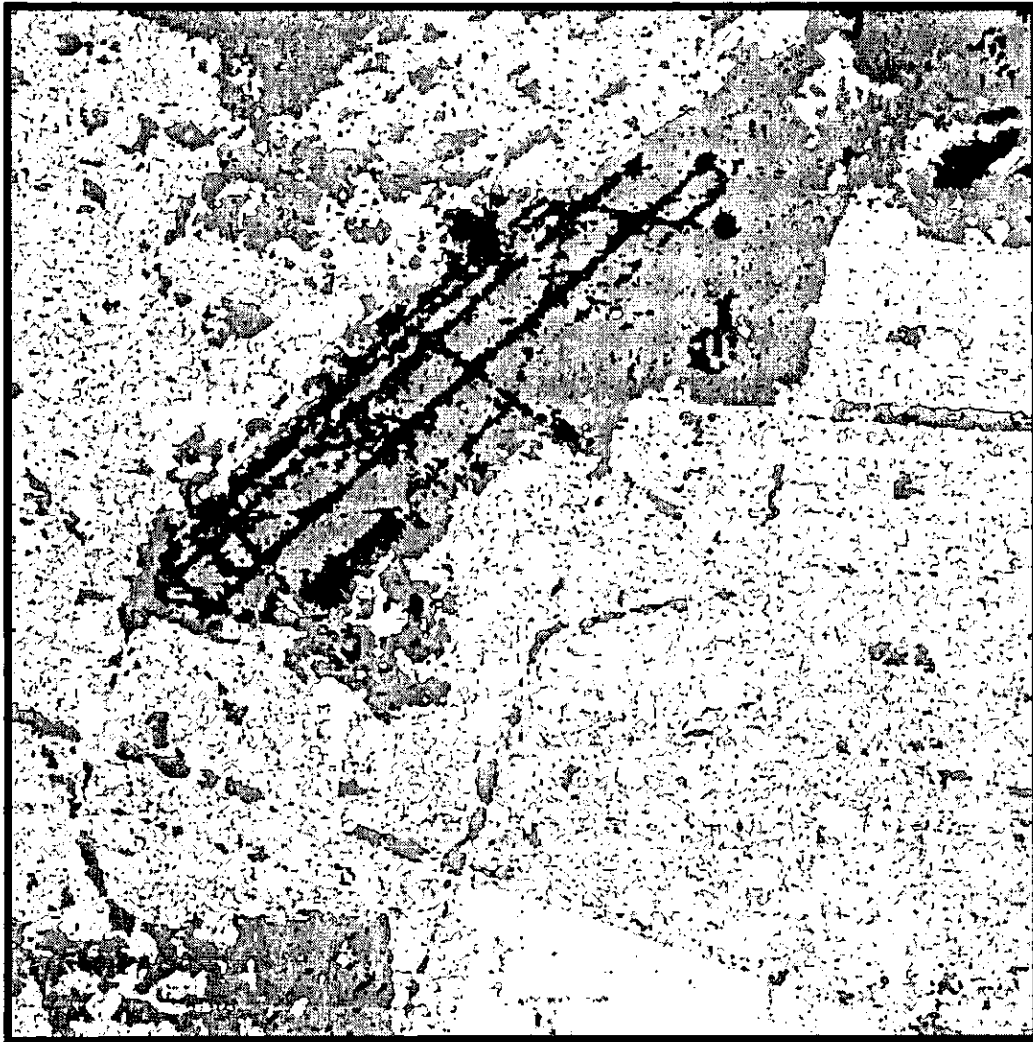


Figura 4.36

Resultado final de la segmentación en 4 clases mediante el algoritmo propuesto, utilizando la imagen filtrada.

La segmentación de imágenes SAR no es una tarea sencilla, ya que dadas sus muy peculiares características de resolución y presencia de *speckle* es muy difícil el obtener regiones homogéneas que resulten de utilidad para la extracción de información. Por otra parte, si se aplica un algoritmo de filtrado para reducir los efectos del ruido, se corre el riesgo de perder información relevante al obtener una imagen demasiado 'suavizada', lo cual nos llevaría a la generación de imágenes segmentadas con regiones muy homogéneas, pero con una mínima preservación de detalles.

Aquí hemos utilizado de forma combinada un algoritmo de filtrado y un esquema de segmentación que tienen como objetivo común el preservar la mayor cantidad de detalles de interés de la imagen procesada. Los resultados obtenidos se confrontan con aquellos ofrecidos por varios métodos clásicos de segmentación. Aquí debe hacerse notar que para la aplicación de estos últimos se utilizó también la imagen filtrada, ya que lo que se pretende evaluar es exclusivamente la funcionalidad del algoritmo adaptivo de segmentación propuesto.

La ventana que se consideró para este análisis comparativo se muestra en la figura 4.37, aquí se observa la zona de pistas del aeropuerto (que en la imagen segmentada está representada por las clases 0 y 1) y algunas zonas habitacionales, que se muestra como regiones muy heterogéneas. La imagen filtrada por el método adaptivo de reducción de ruido multiplicativo se muestra en la figura 4.38. Se puede notar que las zonas habitacionales se presentan ahora más homogéneas.

En la ventana que muestra la clase de rechazo (figura 4.39) se observa claramente que la mayor parte de los píxeles que forman parte de las pistas del aeropuerto son asignados de forma definitiva a la clase correspondiente (clase 0), con lo que se garantiza su preservación. La clase de rechazo representa el 19.34% de la imagen. Para este paso se consideró, de igual forma, un umbral de rechazo de 0.8.

La figura 4.40 muestra la imagen definitiva en 4 clases obtenida mediante nuestra propuesta. Si bien en las zonas habitacionales se observa aún cierta incidencia de píxeles aislados, es un hecho que la calidad de la segmentación mejoró mucho al incorporar la etapa de filtrado. La imagen obtenida por la aplicación directa de nuestro algoritmo sobre la imagen original (figura 4.41) evidencia la influencia negativa del *speckle* en el proceso de segmentación.

Respecto a los resultados obtenidos mediante esquemas clásicos no existe gran variación respecto a los casos antes considerados. El resultado del esquema Bayesiano clásico –figura 4.42– muestra una vez más una alta incidencia de píxeles aislados, aunque preserva los detalles de la imagen. La segmentación mediante la aproximación Markoviana genera una imagen de clases que presenta regiones homogéneas, pero no ofrece una muy buena definición de la zona de pistas. Este resultado se muestra en la figura 4.43. Finalmente, el esquema ICM (figura 4.44) es una vez más el que provoca una mayor degradación de los detalles finos de la imagen, al presentar regiones demasiado homogéneas.

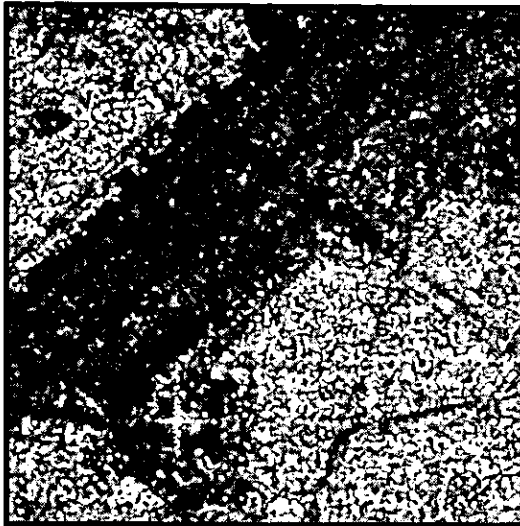


Figura 4.37
Subimagen original.



Figura 4.38
Subimagen filtrada.



Figura 4.39
Imagen que muestra la clase de rechazo (en blanco).

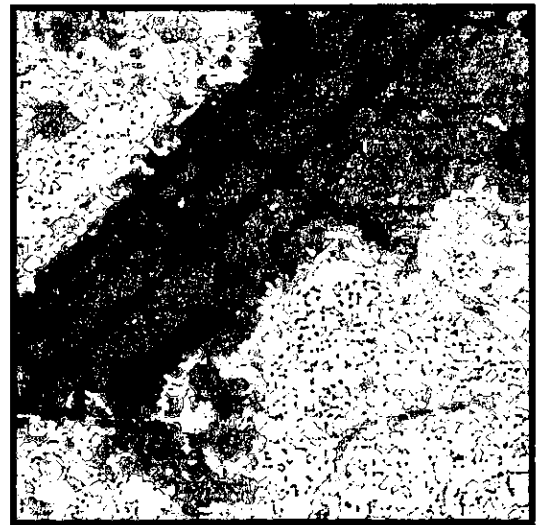


Figura 4.40
Imagen segmentada en 4 clases con el esquema propuesto.

	ω_0 [%] negro	ω_1 [%] azul	ω_2 [%] rojo	ω_3 [%] amarillo
K-medias	7.05	27.56	34.36	31.03
Bayes clásico	5.45	28.33	35.70	30.52
ICM	4.31	29.11	37.18	29.40
MRF clásico	5.57	27.22	35.32	31.89
Esquema propuesto	5.59	25.75	31.22	37.44

Tabla 4.11

Porcentaje de píxeles asignados a cada clase, de acuerdo al algoritmo de segmentación empleado.



Figura 4.41

Imagen (sin filtrar) segmentada con el esquema propuesto.

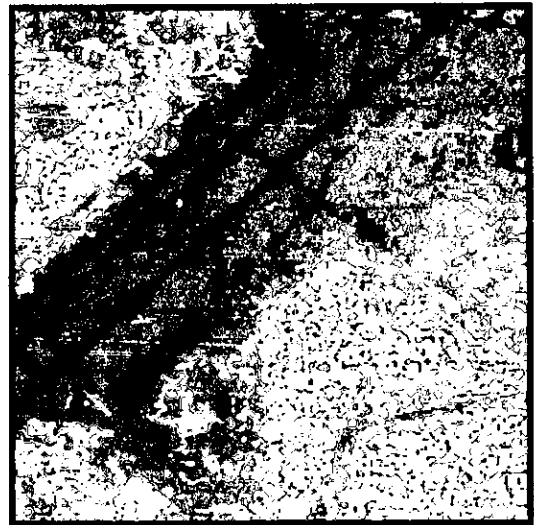


Figura 4.42

Resultado del esquema Bayesiano Clásico.

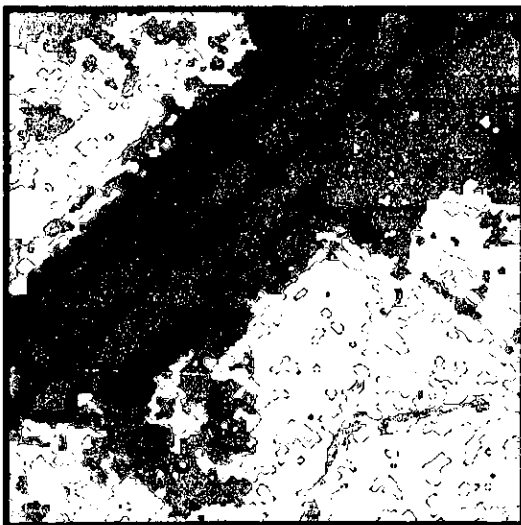


Figura 4.43

Resultado del modelo Markoviano clásico.

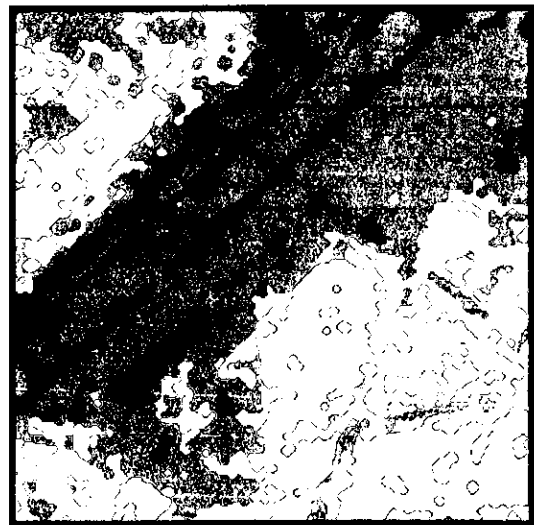


Figura 4.44

Resultado del algoritmo ICM.

	K-medias	Bayes Clásico	ICM	MRF Clásico	Esquema Propuesto
K-medias	0	3.00	22.25	16.51	13.74
Bayes clásico	3.00	0	20.77	16.22	13.48
ICM	22.25	20.77	0	10.25	20.81
MRF clásico	16.51	16.22	10.25	0	15.02
Esquema propuesto	13.74	13.48	20.81	15.02	0

Tabla 4.12

Porcentaje de diferencia entre los diferentes resultados obtenidos.

4.5 Conclusiones

En este capítulo hemos propuesto un algoritmo adaptivo de segmentación, el cual cuenta con dos etapas: una no contextual y una contextual, y tiene como objetivo el ofrecer imágenes segmentadas con regiones homogéneas, al tiempo que preserva las estructuras finas de la imagen. Para determinar la forma en que serán asignados los píxeles se hace una aproximación de las funciones de probabilidad conjunta a partir de los centros de clase y la desviación estándar de cada una de las clases consideradas, definiendo después un *umbral de rechazo*. Los píxeles rechazados son clasificados entonces mediante un esquema contextual.

Los resultados obtenidos al aplicar el algoritmo propuesto a diferentes tipos de imágenes –las cuales muestran características muy particulares– muestran su buen funcionamiento, ya que en las imágenes de prueba se obtuvieron resultados con las características esperadas, es decir, imágenes con regiones homogéneas, con una baja incidencia de píxeles aislados, pero preservando en el mayor grado posible las estructuras o detalles finos de la escena analizada.

De esta forma, una de las conclusiones a las que podemos llegar es que el utilizar la desviación estándar como parámetro para la definición de las funciones de probabilidad conjunta es de especial utilidad, ya que debido a esto el comportamiento de cada clase va a ser distinto, y eso se refleja en el proceso de segmentación.

Por otra parte, mediante los análisis comparativos de los resultados obtenidos con el método adaptivo propuesto contra aquellos obtenidos por métodos de segmentación clásicos, pudimos observar que el considerar la probabilidad *a priori* en un esquema de segmentación puede ir en perjuicio de las estructuras finas presentes en la escena analizada, ya que la mayoría de los casos estas pertenecen a clases que no predominan en la imagen.

Otro aspecto a considerar es que al procesar imágenes SAR con nuestro algoritmo es esencial el incluir una etapa de filtrado, ya que el efecto del *speckle* nos lleva a la obtención de imágenes con una muy alta incidencia de píxeles aislados. Aquí se observó que el uso combinado de nuestro algoritmo con un método adaptivo de reducción de ruido multiplicativo (el cual también tiene como objetivo el preservar los detalles finos de la imagen) conduce a la obtención de imágenes segmentadas con buena calidad, lo que facilita las tareas de extracción de información.

Finalmente, podemos afirmar que el esquema de segmentación propuesto cumplió con las expectativas que se tenían en mente al momento de desarrollarlo, aunque obviamente aún es perfectible.

5 CLASIFICACIÓN DE ZONAS URBANAS POR MODELADO MARKOVIANO

5.1 Introducción

Los trabajos efectuados a partir de imágenes de percepción remota se dirigen en su mayor parte al estudio de los bosques, de glaciares, y de la agricultura. La segmentación, seguida de la clasificación de imágenes de percepción remota que contienen zonas urbanas, representa una tarea particularmente difícil de realizar. La mayoría de las veces, la información aportada únicamente por el nivel de gris es por mucho insuficiente.

Entre las diversas aproximaciones utilizadas, la mayoría hace uso de los parámetros de textura. En la mayoría de los trabajos recientes sobre la extracción urbana mediante un análisis de la textura, los autores no proponen nuevos parámetros. La novedad proviene principalmente de los algoritmos de segmentación o de estimación de parámetros, sabiendo que estos últimos son comúnmente los mismos. Así, Gouinaud [18] propone una comparación de diversos parámetros de textura utilizados para la clasificación de imágenes satelitales ópticas. Gouinaud muestra claramente que no existe un parámetro 'ideal' sino más bien un conjunto de parámetros 'útiles'. Algunos de ellos son aptos para distinguir, de forma relativamente correcta, lo urbano de lo no urbano (índice de temperatura [10]), pero fallando en cuanto a la distinción de los diferentes tipos de zonas urbanas; otros teniendo capacidades inversas (índice del tamaño), y otros que fallan en todos los casos (índice de brillantez).

Entre los trabajos efectuados en este dominio están aquellos realizados por Houzelle [20] quien fusiona la información obtenida con la ayuda de dos captoreos (SPOT y SAR). Las aglomeraciones urbanas tienen la particularidad de aparecer como regiones brillantes en las imágenes de radar. Así, mediante una simple definición de umbrales puede determinar máscaras de aprendizaje para el cálculo de los parámetros de las zonas urbanas. Además, considerando la distancia respecto a los centros de clase de esas zonas, determina máscaras de aprendizaje para los parámetros de otras zonas. La clasificación se realiza con la ayuda de un clasificador Gaussiano clásico multivariable. Houzelle utiliza 4 atributos de textura (dos ligados al histograma local, una medida local del Laplaciano y una medida de la orientación del gradiente).

Haralick [19] propone parámetros calculados a partir de la matriz de co-ocurrencia (entropía, energía, correlación, etc.) los cuales son usados para segmentar imágenes aéreas, así como imágenes satelitales LANDSAT conteniendo una o más clases urbanas. Sin embargo, Gouinaud [18], así como Baraldi y Parmiggiani [3] señalan la inadecuación de esos parámetros para el análisis del medio urbano a partir de imágenes SPOT. Esto se debe al hecho de que las imágenes SPOT (particularmente las imágenes del futuro captor SPOT-5) tienen una mejor resolución que las imágenes LANDSAT y, por lo tanto, a esta resolución la textura urbana es de escala variable.

Otras aproximaciones se basan en el modelado de la imagen. Los parámetros del modelo, cuando son estimados sobre la zona urbana, constituyen entonces los parámetros de textura [10]. Entre los diferentes esquemas para realizar el modelado de la imagen citaremos la aproximación Markoviana.

Los campos de Markov son ampliamente utilizados en el análisis de imágenes. En este apartado se utilizan para modelar la textura de la imagen por medio de un campo Markoviano en 4 conexidad, lo que nos permite definir un parámetro característico de la textura urbana. Este parámetro se denomina 'Parámetro de Temperatura', por lo cual a la imagen que se obtiene se le denomina 'Imagen de temperatura'. Esta imagen se caracteriza por tener valores altos (mayor temperatura) en las zonas heterogéneas, que por lo general representan a las regiones urbanas; y valores bajos (menor temperatura) en las zonas homogéneas, que caracterizan a las regiones no urbanas.

Una vez que se obtiene dicha imagen, procedemos a binarizarla. Esto con el objetivo de que, a partir de un umbral, las zonas con mayor temperatura se identifiquen como regiones urbanas, mientras que las zonas con menor temperatura se consideran como regiones no urbanas. El método de binarización elegido es el de correlación, ya que éste evalúa el grado de semejanza (correlación) entre la imagen original y la binarizada, y determina el umbral de binarización que maximiza dicha correlación (umbral óptimo).

5.2 Parámetro de textura

En esta parte presentamos un parámetro característico de las texturas urbanas. Este parámetro se define a partir de los trabajos de X. Descombes [10].

5.2.1. Modelo 2D Isotrópico

5.2.1.1 El modelo

Descombes propone un nuevo parámetro de textura [10] utilizado a fin de extraer una máscara urbana. Este parámetro, llamado temperatura y denotado T, está fundado en el modelado de la imagen mediante un esquema Markoviano Gaussiano en 4 conexidad (ecuación 5.1). Las zonas urbanas, que son regiones con una fuerte varianza y poco correlacionadas, tienen una temperatura más elevada que las zonas de campos agrícolas. Lo interesante del parámetro propuesto es que está directamente relacionado a la definición de un modelo Markoviano Gaussiano en 4 conexidad. En lo que concierne al modelo isotrópico, Descombes muestra en una primer instancia que la ley condicional $P(X_s | X_r \in V_s)$ no depende más que de la media m_s de los cuatro vecinos lo que favorece la robustez de los estimadores (ecuación 5.2)

$$P(X_s | X_r \in V_s) = \frac{1}{Z_{V_s}} \exp\left(-\beta \left(\sum_{r \in V_s} (X_s - x_r)^2 + \lambda (X_s - \mu)^2 \right)\right) \quad (5.1)$$

$$P(X_s | X_r \in V_s) = P(X_s | m_s) \quad (5.2)$$

$\beta = \frac{1}{T}$, λ y μ son los parámetros del modelo

V_s es la vecindad del *site* s

Z_{V_s} es la constante de normalización

En una segunda instancia, Descombes muestra que la probabilidad condicional local es una ley normal, definida de la siguiente manera:

$$P(X_s | X_r \in V_s) \equiv N\left(\frac{4m_s + \mu\lambda}{4 + \lambda}, \frac{1}{2\beta(4 + \lambda)}\right) \quad (5.3)$$

La varianza condicional $\sigma_{P(X_s|m_s)}^2 = \frac{1}{2\beta(4 + \lambda)}$ es estimada a partir del método de 'colas de cometa'.

Sin embargo, utilizado únicamente para extraer una máscara urbana, este parámetro conduce a la confusión de viñedos, invernaderos y zonas urbanas, como se muestra en la figura 5.1. Esta imagen muestra claramente el interés que se tiene al considerar de forma separada las diferentes direcciones. Estas 8 direcciones se representan en la figura 5.3.



Figura 5.1

Imagen de Toreilles (Francia), zona de invernaderos y viñedos [10].

5.2.1.2 El método de estimación: Colas de cometa

Para estimar la varianza condicional empleamos el método de las colas de cometa.

La meta es estimar $\sigma_{P(X_s|m_s)}^2 = \frac{1}{2\beta(4 + \lambda)}$

Estimamos entonces la matriz de probabilidades condicionales. Para hacer esto, para cada *site* s , incrementamos el punto de la matriz en el cual las coordenadas corresponden respectivamente al valor X_s del *site* y al valor m_s de la media de sus vecinos (figura 5.2). Después, para cada una de las distribuciones $P(X_s|m_s)$ estimamos la varianza condicional $\sigma_{P(X_s|m_s)}^2$ mediante el estimador empírico. La estimación de la varianza se efectúa sobre una ventana centrada en el *pixel* X_s .

Se considera entonces el problema del tamaño de la ventana, que debe ser lo suficientemente grande para tener estadísticas fiables y suficientemente pequeña para evitar al máximo las mezclas de texturas. En este método de estimación, el problema es resuelto en parte ponderando el estimador de la varianza por el número de muestras que han servido para obtenerlo. Así, obtenemos bordes bien definidos.

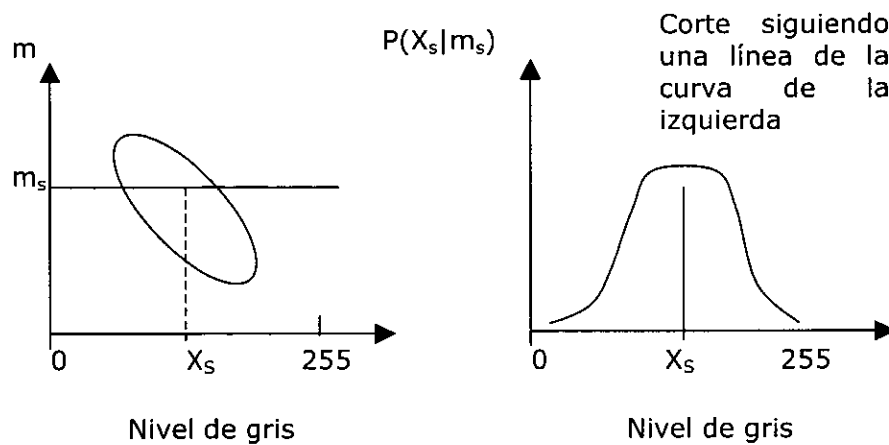


Figura 5.2

Colas de cometa.

5.2.2. Serie de modelos 1D

En lo que nos concierne, hemos tenido en cuenta la anisotropía considerando una serie de modelos 1D tomados en las 8 direcciones principales (figura 5.3; debe hacerse notar que las direcciones NOo y SEe son la misma). En este caso, la probabilidad condicional para el modelo definido en la dirección d es la siguiente:

$$P(X_s | X_r \in V_s^d) = P(X_s | m_s^d) \tag{5.4}$$

$$P(X_s | X_r \in V_s) = \frac{1}{Z_{V_s^d}} \exp\left(-\beta^{(d)} \left(\sum_{r \in V_s^d} (X_s - x_r)^2 + \lambda^{(d)} (X_s - \mu)^2\right)\right) \tag{5.5}$$

$$P(X_s | X_r \in V_s) \equiv N\left(\frac{2m_s^d + \mu\lambda^{(d)}}{2 + \lambda^{(d)}} + \frac{1}{2\beta^{(d)}(4 + \lambda^{(d)})}\right) \tag{5.6}$$

d representa la dirección considerada

m_s^d representa la media de los dos vecinos del site s en la dirección d

Así, en la dirección (NOo/SEe) la probabilidad condicional local es:

$$P(X_s | X_r \in V_s^{(d=NOo/SEe)}) = P(X_s | (X(NOo) + X(SEe)) / 2) \tag{5.7}$$

Nuestros parámetros de textura son las varianzas condicionales estimadas por el método de colas de cometa en las 8 direcciones:

$$\sigma_d^2 = \frac{1}{2\beta^{(d)}(2 + \lambda^{(d)})} \tag{5.8}$$

Como se muestra en la figura 5.3, si consideramos la dirección (E/O) los *pixeles* vecinos son situados a una distancia 1 del *pixel* s. Por el contrario, si consideramos la dirección (NOo/SEe) los *pixeles* vecinos se encuentran a una distancia $\sqrt{5}$ del *pixel* s. Con la finalidad de corregir el 'error' introducido por esas distancias y de poder obtener los parámetros calculados en las 8 direcciones, se normalizan los parámetros considerados.

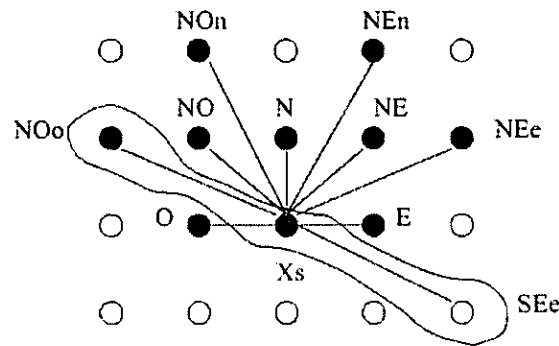


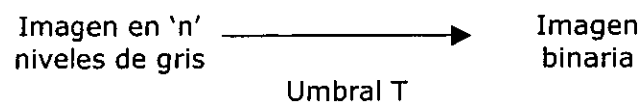
Figura 5.3

Representación de las 8 direcciones tomadas en cuenta.

5.3 Binarización por Correlación

5.3.1 Introducción

Los métodos de binarización (representación de una imagen en n niveles de gris mediante otra imagen en sólo dos niveles) se basan en la elección de un umbral, el cual es idealmente un valle en un histograma bimodal.



En la mayoría de los casos, este valle es difícil de detectar de forma precisa. Peor aún, existen imágenes con histogramas unimodales o multimodales (con varios valles) en donde la tarea de la elección de un umbral se vuelve más compleja.

Se han desarrollado varios métodos para facilitar la elección de un umbral de binarización, basándose en varios parámetros del histograma. El método de

binarización por correlación se basa en evaluar la correlación (grado de semejanza) entre la imagen original y la imagen binarizada para cierto umbral. El método determina entonces el umbral que maximiza dicha correlación.

5.3.2 El algoritmo

Para la obtención del umbral óptimo, consideremos X , que representa el conjunto posible de niveles de gris g de la imagen original, mientras que Y representa los nuevos niveles de gris de la imagen binaria. Estos niveles son determinados por el umbral alto y bajo ($\mu_0(T)$ y $\mu_1(T)$) de la imagen original, dados por:

$$\mu_0(T) = \frac{\sum_{g=0}^T g \cdot p_g}{\sum_{g=0}^T p_g} \quad (5.9)$$

$$\mu_1(T) = \frac{\sum_{g=T+1}^n g \cdot p_g}{\sum_{g=T+1}^n p_g} \quad (5.10)$$

Donde $g=0, 1, 2, \dots, n$ son los niveles de gris de la imagen original y p_g su probabilidad de aparición.

La correlación entre la imagen original y la imagen binaria se expresa en función de cada uno de los posibles valores del umbral T (con $0 \leq T \leq 255$):

$$\rho_{xy}(T) = \frac{E_{xy}(T) - E_x E_y(T)}{[V_x V_y(T)]^{1/2}}, 0 \leq T \leq n \quad (5.11)$$

Donde

$$E_y(T) = \sum_{g=0}^T \mu_0(T) \cdot p_g + \sum_{g=T+1}^n \mu_1(T) \cdot p_g \quad (5.12)$$

$$E_x = \sum_{g=0}^n g \cdot p_g \quad (5.13)$$

Para los momentos de segundo orden

$$E_{yy}(T) = \sum_{g=0}^T \mu_0^2(T) \cdot p_g + \sum_{g=T+1}^n \mu_1^2(T) \cdot p_g \quad (5.14)$$

$$E_{xx} = \sum_{g=0}^n g^2 \cdot p_g \quad (5.15)$$

$$E_{xy}(T) = \sum_{g=0}^T g \cdot \mu_0(T) \cdot p_g + \sum_{g=T+1}^n g \cdot \mu_1(T) \cdot p_g \quad (5.16)$$

Y las varianzas

$$V_x = E_{xx} - [E_x]^2 \quad (5.17)$$

$$V_y(T) = E_{yy}(T) - [E_y(T)]^2 \quad (5.18)$$

El umbral T , retenido sobre los 256 posibles valores, deberá ser el que maximiza la función de correlación, siendo éste el mejor umbral global para la binarización de la imagen original.

5.4 Resultados

Para evaluar los resultados generados por el parámetro de temperatura y la binarización por correlación (para la definición de máscaras urbanas), hemos utilizado 3 tipos de imágenes con características muy particulares. Todas las imágenes empleadas se encuentran en 256 niveles de gris y tienen un tamaño de 512x512 *pixels*. Los tipos de imágenes utilizados son:

- ♦ Fotografía aérea de alta resolución (50 cm/*pixel*)
- ♦ Imagen SPOT pancromática (10 m/*pixel*)
- ♦ Imagen SAR (30 m/*pixel*)

Estas imágenes fueron seleccionadas para evaluar los diferentes resultados arrojados a partir de imágenes con diferente resolución y características muy específicas. Los resultados obtenidos con cada tipo de imagen empleado se presentan a continuación.

5.4.1 Fotografías aéreas de alta resolución

Se consideraron dos ventanas de fotografía aérea de alta resolución, las cuales presentan diferentes características de interés que nos permitirán evaluar el parámetro de textura urbana.



Figura 5.4

Fotografía aérea de alta resolución que muestra una región urbana del oriente de la Ciudad de México (Ventana 1, imagen original en 256 niveles de gris, 512x512 píxeles).

La primer imagen es una fotografía aérea pancromática de alta resolución que muestra una escena urbana de la Ciudad de México (figura 5.4). En esta imagen podemos distinguir que a la izquierda se encuentran unos campos deportivos, en el centro se

presentan construcciones de vías del transporte colectivo Metro, y a la derecha tenemos bloques de construcciones de casas y edificios, con calles entre ellos.



Figura 5.5

Imagen de temperatura. Las zonas homogéneas aparecen en color oscuro (baja temperatura), mientras que las heterogéneas están representadas por las zonas claras de la imagen (alta temperatura).

Para cada *pixel* estimamos entonces la varianza condicional a partir del método de las colas de cometa, esto en las 8 direcciones, y obtenemos el promedio de dichas varianzas para cada *pixel*. La mayoría de los medios urbanos se caracterizan por una fuerte varianza en todas las direcciones. Así, las construcciones realizadas por el hombre, tales como las casas y edificios, tienen una fuerte varianza. Por otra parte, las zonas de campo o de bosque, e incluso carreteras y calles, presentan una pequeña varianza en todas las direcciones.

En la imagen obtenida de la evaluación del parámetro de temperatura (figura 5.5) podemos ver que los campos deportivos que aparecen a la izquierda en la imagen original (figura 5.4) se toman efectivamente como zonas no urbanas. Como veremos más adelante, cuando esta ventana de la Ciudad de México es analizada solamente con

un algoritmo de segmentación, los campos deportivos resultan clasificadas como zonas urbanas. Esto se debe a que los niveles de gris son muy altos, y se pueden confundir con los niveles de gris de los edificios. La imagen de temperatura nos ayuda a disminuir ese error.

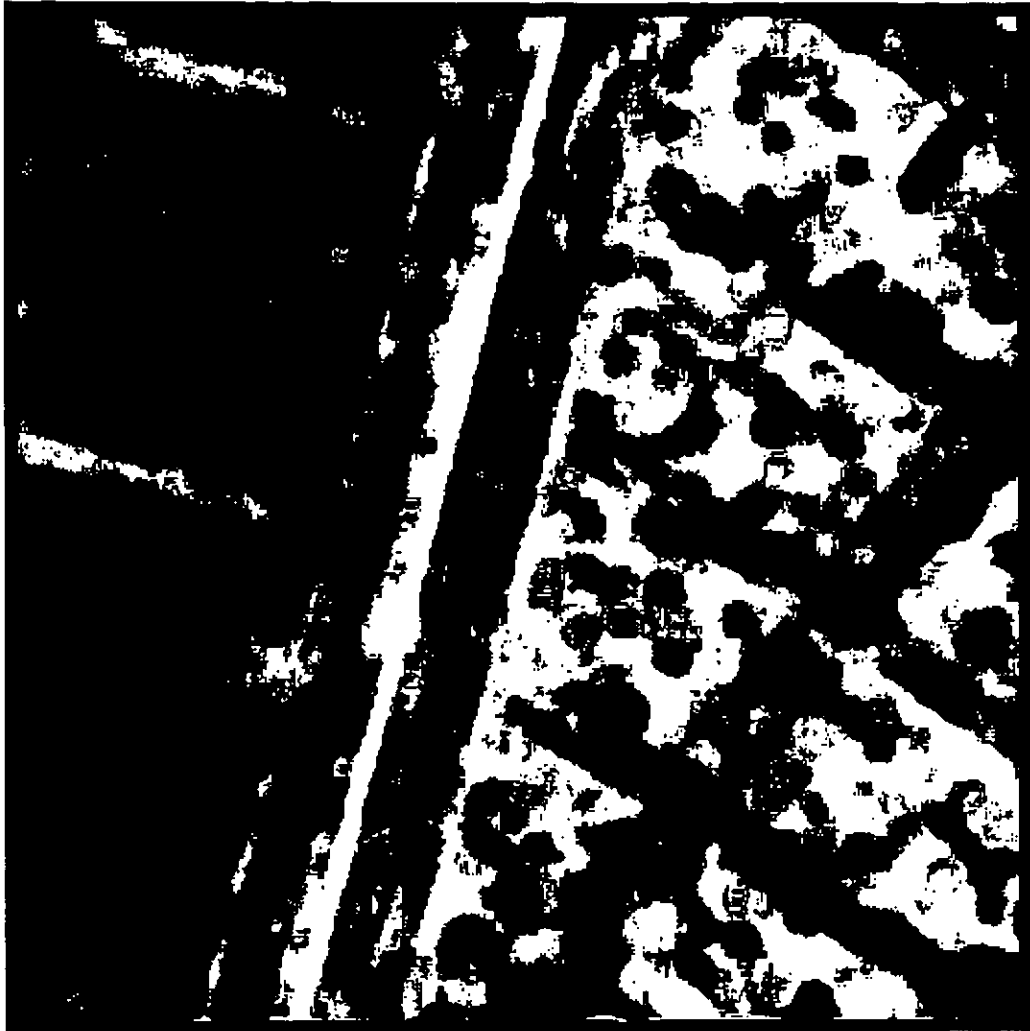


Figura 5.6

Resultado de la imagen de temperatura binarizada por correlación.

Para poder separar las zonas urbanas de las no urbanas a partir de la imagen de temperatura obtenida al aplicar el método de las colas de cometa, binarizamos la imagen por medio del método de binarización por correlación. El valor de nivel de gris que maximizó la función de correlación para esta imagen fue de 142.

Las regiones claras de la imagen binarizada aproximan las zonas urbanas, mientras que las regiones oscuras aproximan a las no urbanas.



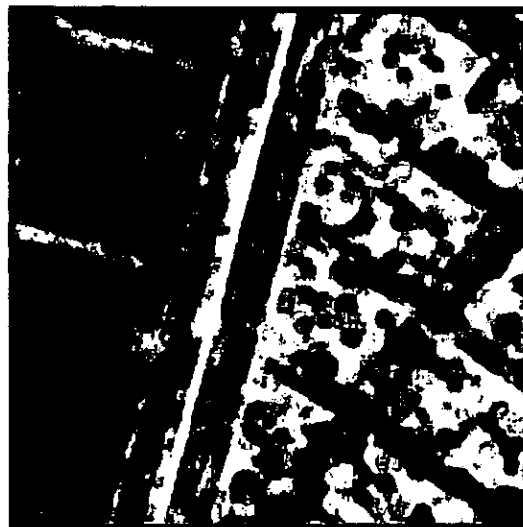
(a)

Imagen original de la Ciudad de México.



(b)

Imagen de temperatura.



(c)

Imagen de temperatura binarizada por correlación.

Figura 5.7

Obtención de la imagen de temperatura para la fotografía aérea de alta resolución.

Para poder tener una visión general de los resultados obtenidos, se presenta la imagen original, la imagen de temperatura y la imagen de temperatura binarizada por correlación (figura 5.7), con el fin de hacer más evidentes las diferencias entre las regiones clasificadas.

La segunda imagen es también una fotografía aérea pancromática de alta resolución, que muestra una escena urbana de la Ciudad de México (figura 5.8). En esta imagen podemos distinguir que en la parte superior izquierda se encuentra una zona boscosa muy extensa; mientras que en la parte inferior derecha aparece una región urbana compuesta por casas y edificios, en la cual se distinguen asimismo varios conjuntos de zonas arboladas.



Figura 5.8

Fotografía aérea de alta resolución que muestra una región urbana de la zona sur de la Ciudad de México (Ventana 2, imagen original en 256 niveles de gris, 512x512 píxeles).

Sobre esta imagen calculamos su parámetro de temperatura, es decir, para cada *pixel* estimamos la varianza condicional a partir del método de las colas de cometa, esto en las 8 direcciones, y obtenemos el promedio de dichas varianzas para cada pixel.



Figura 5.9

Imagen de temperatura para la segunda fotografía aérea considerada.

En la imagen obtenida de la evaluación del parámetro de temperatura (figura 5.9) podemos ver que la región boscosa de la mitad superior izquierda de la imagen original (figura 5.8) tiene valores bajos de temperatura, es decir, que en esa zona tenemos una varianza pequeña, al igual que en las regiones arboladas que se encuentran inmersas en la parte urbana de la escena considerada.

En la imagen de temperatura, la región urbana (parte inferior derecha de la imagen), compuesta por los edificios y casas, muestra una alta varianza, lo cual se refleja en una alta temperatura (regiones claras).

Para poder separar las zonas urbanas de las no urbanas a partir de la imagen de temperatura, obtenida mediante el método de las Colas de Cometa, binarizamos la imagen con el esquema de Binarización por Correlación. El valor de nivel de gris que maximizó la función de correlación para esta imagen fue de 90.

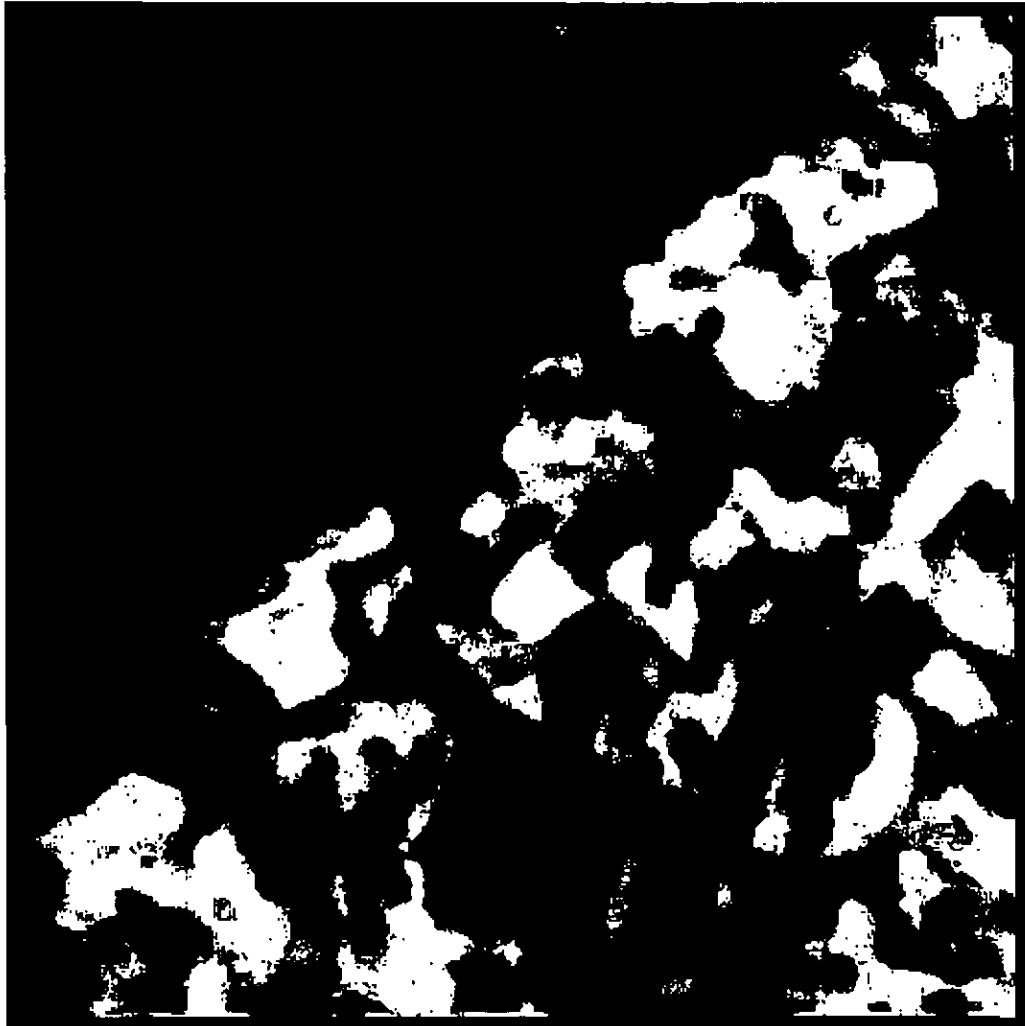


Figura 5.10

Resultado de la imagen de temperatura binarizada por correlación.

Se puede distinguir, en la imagen de temperatura binarizada, que la zona boscosa y las agrupaciones de árboles presentes en la parte urbana de la imagen analizada son clasificadas como regiones no urbanas. Por otra parte, las zonas de construcciones son clasificadas como regiones urbanas, de acuerdo a lo esperado (figura 5.10).



(a)

Imagen original de la ciudad de México.



(b)

Imagen de temperatura.



(c)

Imagen de temperatura binarizada por correlación

Figura 5.11

Obtención de la imagen de temperatura para la fotografía aérea de alta resolución.

La figura 5.11 permite tener una vista general de los resultados obtenidos. Se presenta la imagen original, la imagen de temperatura y la imagen de temperatura binarizada, con el fin de hacer más evidentes las diferencias entre las regiones clasificadas.

5.4.1 Imagen SPOT pancromáticas

Después de mostrar los resultados obtenidos al aplicar el algoritmo de texturas a fotografías aéreas de alta resolución, se decidió aplicar este mismo algoritmo a imágenes de percepción remota que ofrecen una menor resolución, como lo son las imágenes SPOT y las imágenes SAR.



Figura 5.12

Imagen SPOT pancromática, con una resolución de 10 metros por pixel [512x512x256].

Para la imagen SPOT (figura 5.12) la ventana seleccionada muestra una escena urbana de la Ciudad de México que incluye la zona del aeropuerto y el bosque de Aragón, que en este caso representan las regiones más homogéneas de la imagen (región no

urbana). La región urbana está representada por las zonas habitacionales que se encuentran en las inmediaciones del aeropuerto.

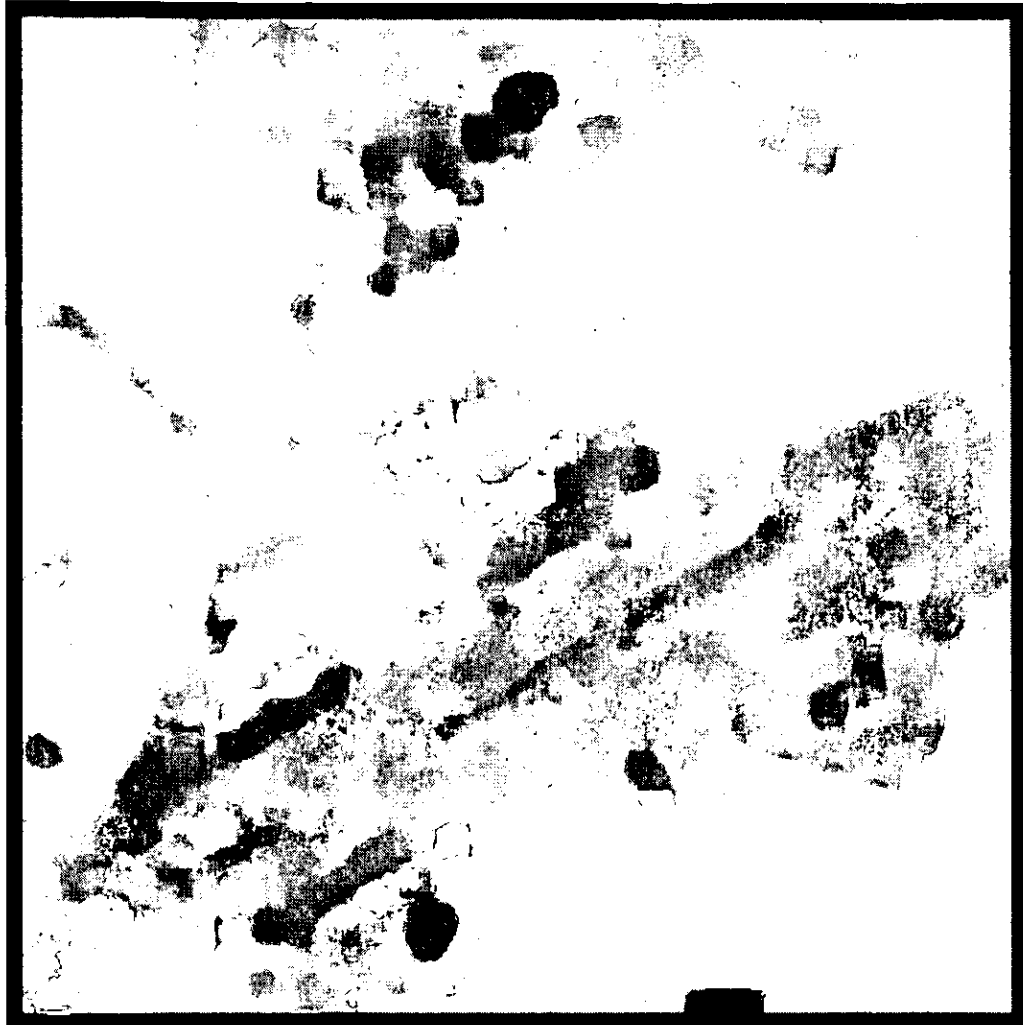


Figura 5.13

Imagen de temperatura para la imagen SPOT.

Para la obtención de la imagen de temperatura se calcula la varianza condicional para cada *pixel*, a partir del método de las colas de cometa y, además, se calcula el promedio de dichas varianzas para cada *pixel*.

En la imagen obtenida (figura 5.13) se puede apreciar que la zona de pistas del aeropuerto, que está al centro de la imagen original (figura 5.12), tiene valores bajos de temperatura (poca varianza) al igual que la zona del Bosque de Aragón.

La región urbana, compuesta por las zonas habitacionales que rodean al aeropuerto, aparece como una región con alta varianza, la cual se visualiza en colores claros en la imagen de temperatura.

Resulta importante resaltar que las zonas más claras que aparecen alrededor de las pistas del aeropuerto se refieren a las instalaciones de este mismo, como bodegas, oficinas, salas de espera, zona comercial, estacionamiento, etc.

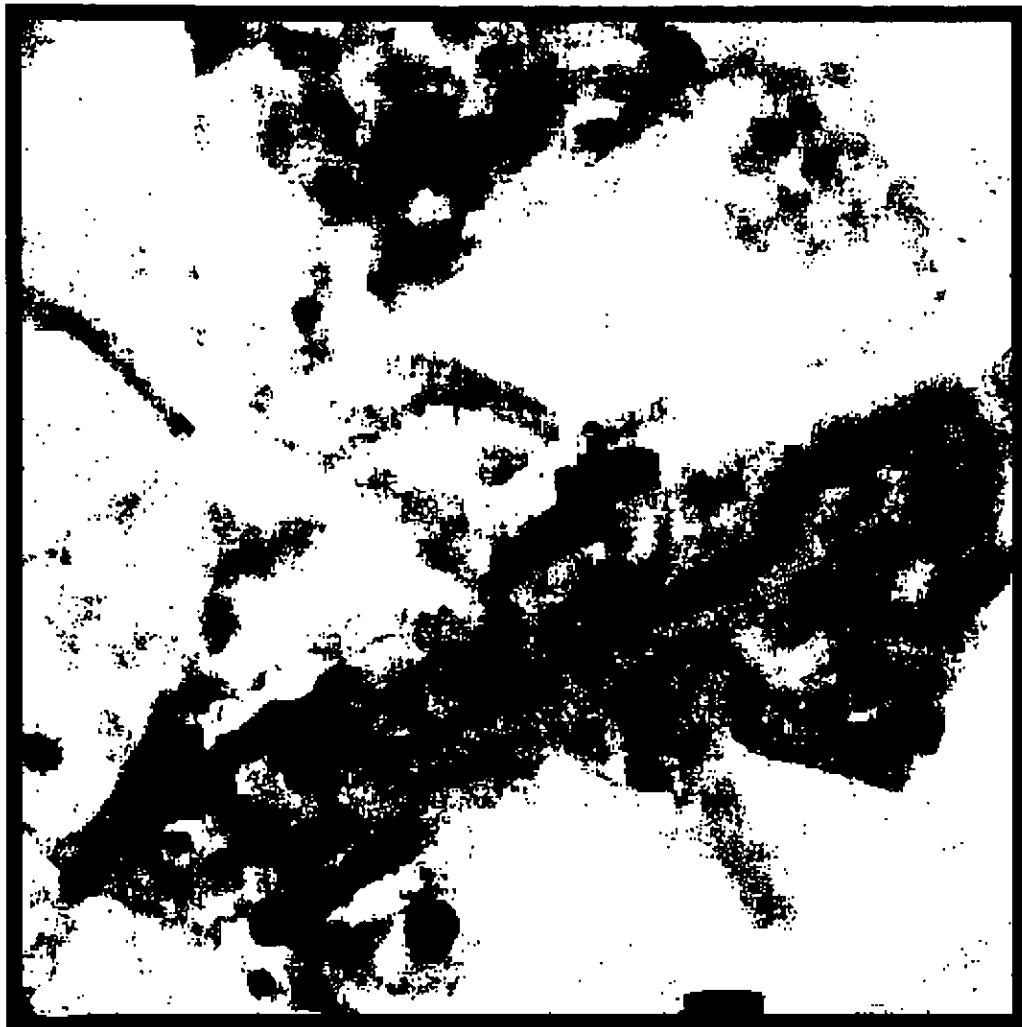


Figura 5.14

Resultado de la imagen de temperatura binarizada por correlación.

La imagen de temperatura fue binarizada por medio del método de binarización por correlación para discriminar las zonas urbanas de las no urbanas. El valor de nivel de gris que maximizó la función de correlación para esta imagen fue de 155.

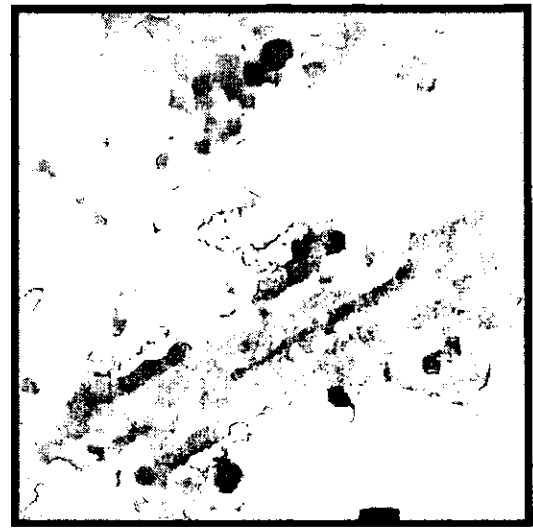
En la figura 5.14 se muestra la imagen SPOT clasificada por este método. Podemos distinguir cómo el Bosque de Aragón y las pistas del aeropuerto pertenecen a la zona no urbana (regiones oscuras) y las colonias circundantes se clasifican como zonas urbanas (regiones claras).

Es importante hacer notar que, debido a la resolución de las imágenes SPOT (10 m/pixel), las calles que están dentro de las colonias no son significativas, de tal manera que también se consideran como zonas urbanas.



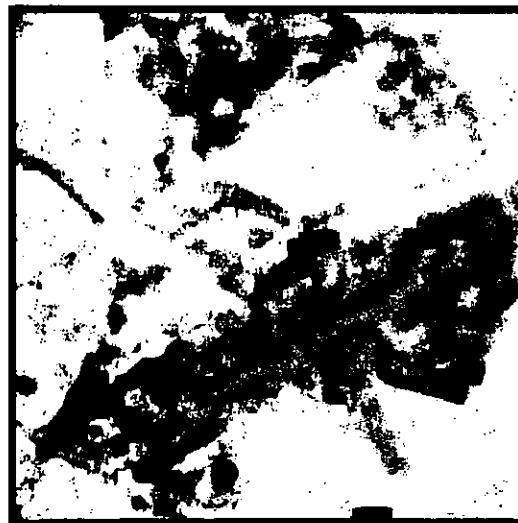
(a)

Imagen SPOT pancromática con una resolución de 10 metros por píxel.



(b)

Imagen de temperatura.



(c)

Imagen de temperatura binarizada por correlación.

Figura 5.15

Obtención de la imagen de temperatura para la imagen SPOT.

La figura 5.15 muestra la imagen original, la imagen de temperatura y la imagen de temperatura binarizada por correlación, con el fin de hacer más evidentes las diferencias entre las regiones clasificadas mediante el parámetro de temperatura.

5.4.1 Imagen SAR

Para calcular el parámetro de textura (temperatura) y la binarización (urbano – no urbano) en imágenes del tipo SAR, se utilizó la imagen mostrada en la figura 5.16, con una resolución de 30 metros por *pixel*. Esta imagen se sometió a una etapa previa de filtrado, utilizando un esquema adaptivo basado en la Transformada Polinomial.



Figura 5.16

Imagen SAR adquirida por el satélite ERS-1 el 19 de septiembre de 1997, la escena corresponde a la zona urbana que rodea el aeropuerto internacional de la Ciudad de México [512x512 pixeles, 256 niveles de gris].

El objetivo de la clasificación final es distinguir la zona de pistas del aeropuerto y las avenidas principales como regiones no urbanas (zonas homogéneas).

Se calculó el parámetro de temperatura para la imagen de la figura 5.16. Podemos distinguir en los resultados que las pistas del aeropuerto, el fragmento del Bosque de Aragón (parte superior de la imagen) y las avenidas principales presentan una baja varianza, lo cual asocia a estas regiones con las zonas no urbanas.

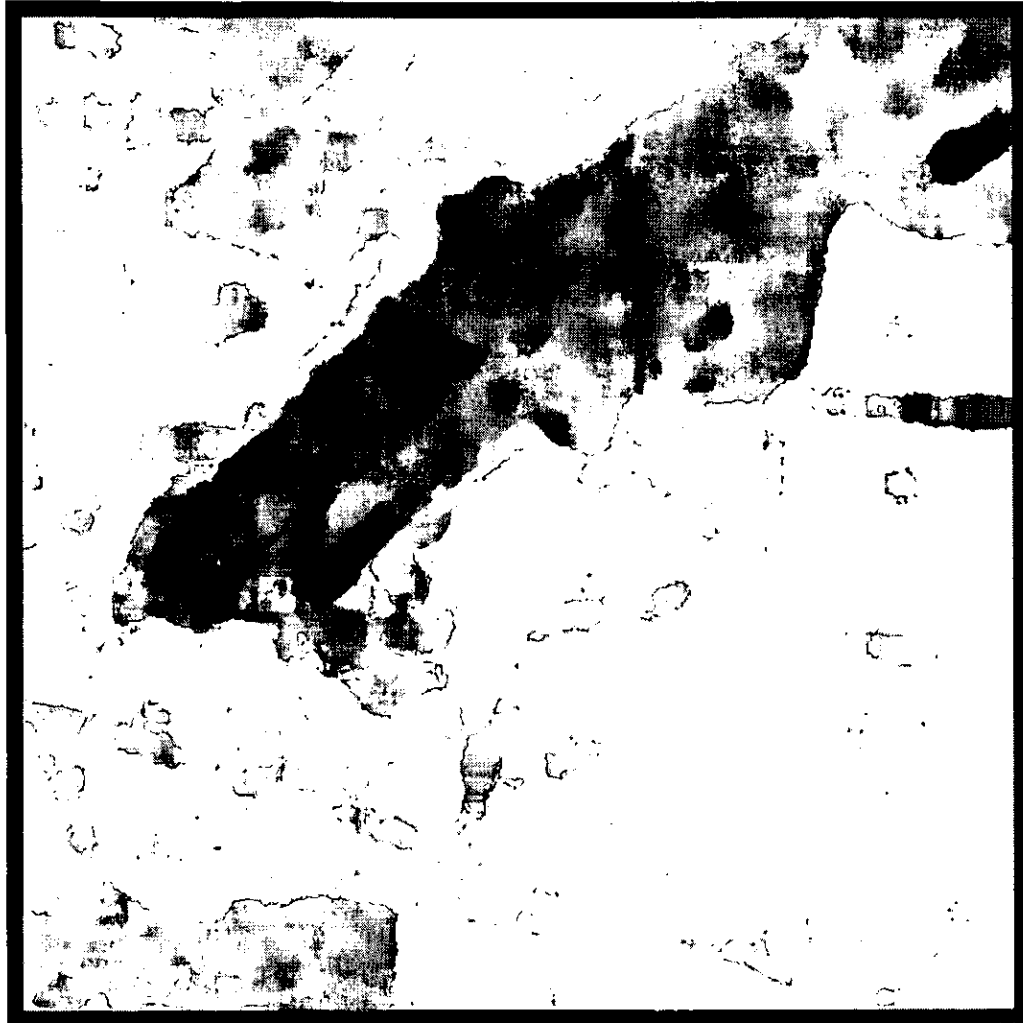


Figura 5.17

Imagen de temperatura para la imagen SAR considerada.

Por el contrario, las zonas habitacionales que rodean el conjunto del aeropuerto presentan una fuerte varianza (regiones claras de la imagen de temperatura), por lo cual se consideran como regiones urbanas.

Para poder clasificar la imagen SAR en zonas urbanas y no urbanas, tomando en cuenta el parámetro de textura, binarizamos por correlación la imagen de temperatura obtenida. El valor de nivel de gris que maximizó la función de correlación para esta imagen fue de 117.



Figura 5.18

Binarización por correlación de la imagen de temperatura.

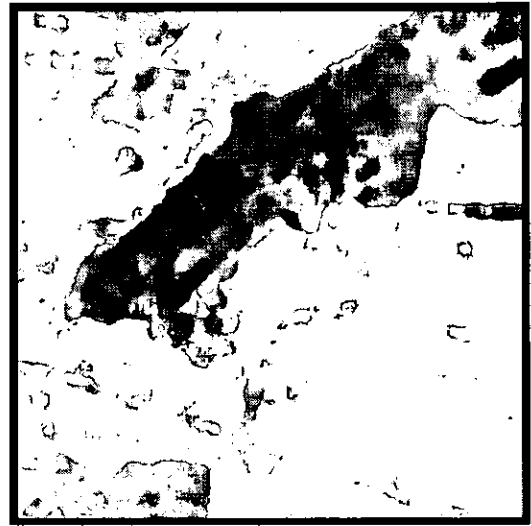
Las regiones blancas de la imagen binarizada aproximan las zonas urbanas, y las regiones oscuras a las zonas no urbanas. Nuevamente distinguimos que las pistas del aeropuerto y el fragmento del Bosque de Aragón son clasificadas como regiones no urbanas, mientras que las colonias son consideradas como urbanas.

Al igual que en las imágenes SPOT, en las imágenes SAR (que tienen una resolución aún más baja) las calles que están dentro de las zonas habitacionales no son significativas, de tal manera que también se consideran como parte de las zonas urbanas. Solamente se distinguen aquellas avenidas de dimensiones considerables.



(a)

Imagen SAR original, mostrando una escena de la Ciudad de México.



(b)

Imagen de temperatura.



(c)

Imagen de temperatura binarizada por correlación.

Figura 5.19

Obtención de la imagen de temperatura para la imagen SAR.

Una vez más se muestra una vista general de los resultados obtenidos. Se presentan (figura 5.19) la imagen original, la imagen de temperatura y la imagen de temperatura binarizada por correlación, con el fin de facilitar el análisis visual de los resultados obtenidos.

5.5 Conclusiones

En este capítulo hemos representado la información proporcionada por los niveles de gris de una imagen pancromática en una información característica de la textura urbana, esto a través del parámetro de temperatura Markoviana.

Esta aproximación nos parece interesante, en la medida que permite hacer la distinción entre la zona urbana, que corresponde a una fuerte varianza en todas las direcciones en lugares construidos, y la zona no urbana, que presenta poca varianza.

Debemos hacer notar que la calidad de los resultados obtenidos es proporcional a la calidad del tipo de imagen considerado. Como se puede observar, los resultados generados para las imágenes de fotografía aérea de alta resolución son, por mucho, más aproximados que aquellos obtenidos para las imágenes SAR, no obstante que para este último tipo de imágenes se consideró una etapa previa de filtrado.

Sin embargo, debemos reconocer que el parámetro de temperatura representa un gran aporte en los procesos de segmentación binaria cuya aplicación conduce hacia resultados de clasificación interpretables como urbano y no urbano. Este parámetro tiene la gran ventaja de que, al tratar con textura y no con niveles de gris, los errores que son clásicos en binarización (de confundir zonas no urbanas con zonas urbanas, debido a la eventual presencia de altos niveles de gris), se reducen en la imagen de temperatura, lo cual nos hace confiar en estos resultados para su uso posterior en la fusión de datos.

6 EXTRACCIÓN DE ZONAS URBANAS POR FUSIÓN DE DATOS

6.1 Introducción

Cuando se considera a la ciudad como una entidad física, las técnicas de tratamiento de imágenes pueden aportar una excelente ayuda. El conocimiento de la ciudad se puede abordar de dos maneras distintas. La primera consiste en considerar a la ciudad como un único objeto en oposición a los medios no urbanos. La meta es, entonces, extraer una máscara urbana. La segunda consiste en obtener un conocimiento más preciso del medio urbano en sí, diferenciando las estructuras urbanas en el interior de la ciudad.

Esto conlleva a una realidad urbana dual. Por un lado, dado que existe cierta heterogeneidad debida la variedad de los materiales de construcción, se puede distinguir lo urbano de lo no urbano (más homogéneo). Por otro lado, la existencia de cierta homogeneidad local nos va a permitir distinguir diferentes tipos de estructuras urbanas. Esta dualidad nos conduce a pensar que las técnicas utilizadas para diferenciar las zonas urbanas de las no urbanas serán diferentes de aquellas utilizadas para caracterizar la ciudad intrínsecamente. En lo que nos concierne, nos enfocaremos a la extracción de zonas urbanas tomando en cuenta su heterogeneidad .

Los métodos utilizados hasta el presente en el dominio de la extracción de una máscara urbana a partir de imágenes de percepción remota están, en su mayor parte, fundados en un análisis de la textura de la imagen. La mayoría de las veces la información aportada únicamente por el nivel de gris es por mucho insuficiente. Por

esta razón, es preciso recurrir a métodos que nos permitan corregir errores emitidos por la segmentación para dicha clasificación.

En este trabajo proponemos extraer zonas urbanas de los siguientes tipos de imágenes de percepción remota:

- Fotografía aérea de alta resolución
- Imagen SPOT
- Imagen SAR

Procedimiento:

Paso 1: Como primer paso para la extracción de zonas urbanas, se emplean los esquemas de segmentación y análisis de texturas descritos en los capítulos 4 y 5 de esta tesis.

La fusión se lleva a cabo con dos imágenes de entrada:

3. Una imagen binarizada que nos proporcione información muy aproximada acerca de los contornos de los elementos presentes en la escena en cuestión. Además, la imagen binarizada también proporciona información de zonas urbanas basada en los niveles de gris, esto bajo la consideración de que las construcciones hechas por el hombre presentan altos niveles de gris.

Para poder obtener la imagen binarizada, es necesario hacer una evaluación de los resultados arrojados por algunos de los diversos métodos de binarización con el fin de obtener la imagen que mejor se adecue para ser considerada en el proceso de fusión. Para evaluar los resultados de binarización, Shan Yu [32] propone estimar la confiabilidad de los resultados a través del cálculo de un error global; a menor error global, mayor confiabilidad del resultado. El conocimiento de una máscara o mapa rudimentario es usado en este primer paso.

4. Una imagen de temperatura binarizada que nos dará información acerca de las texturas de los elementos presentes en la escena de interés para disminuir el error que la segmentación genera al confundir zonas no urbanas con zonas urbanas por sus niveles de gris.

La información de textura la obtendremos a través del Modelado Markoviano presentado en el capítulo 5. La imagen resultante es binarizada por correlación para generar la imagen de temperatura binarizada que será considerada en el proceso de fusión.

Paso 2: Se considera el problema de la fusión de estos datos para realizar una segmentación binaria.

Paso 3: Por último, con base a la máscara urbana, procedemos a extraer las regiones urbanas presentes en las escenas analizadas.

6.2 Máscara Urbana

La máscara urbana que utilizaremos para la extracción de zonas urbanas se obtiene a partir de la fusión de dos imágenes que nos proporcionan información de interés. Por una parte, necesitamos una imagen binarizada para obtener de ella el análisis de zonas urbanas y no urbanas a partir de sus niveles de gris y además para obtener información de los contornos de los elementos presentes en la escena de interés. Por otra parte, necesitamos una imagen que nos proporcione información de texturas. En este caso, utilizaremos la imagen que se obtiene del parámetro de temperatura propuesto por X. Descombes [10] descrito en el capítulo 5.

A continuación explicamos los tres pasos a seguir para la obtención de la máscara urbana, es decir, la obtención de la imagen binarizada, la obtención de la imagen de textura y el proceso de fusión de estas imágenes.

6.2.1 Obtención de la imagen binarizada

Dado que existen variados métodos de segmentación y binarización, debemos elegir la imagen que más se adecue a nuestras necesidades. Para esto, Shan Yu [32] propone evaluar la confiabilidad de los algoritmos a través de un error global, cuya obtención se describe a continuación.

Para estimar la confiabilidad de los resultados del análisis de una imagen, se deben calcular dos medidas de error: una medida de error local (ecuación 6.3) para cada *pixel* y una medida de error global para los resultados totales (ecuación 6.4). Lo que sigue después es simplemente la acumulación de lo anterior sobre todos los *sites* de la imagen. La medida de error global determina si un resultado es suficientemente bueno para ser usado en un proceso de fusión. Para aquellas imágenes seleccionadas para considerarse en el proceso de fusión, la medida de error local determina cómo cada *pixel*, en resultados diferentes, debe ser tomado en cuenta en el proceso de fusión: el *pixel* con una medida de error local pequeña tiene una gran confiabilidad, por tanto, tiene más influencia en el proceso de fusión.

De esta manera, evaluamos el error global para los resultados arrojados por diferentes algoritmos de segmentación y binarización, y elegimos aquel que tenga el menor error global.

6.2.2 Obtención de la imagen de textura binarizada

La imagen de textura que utilizaremos será la descrita en el capítulo 5, la cual se basa en el cálculo del parámetro de temperatura Markoviana propuesta por X. Descombes [10].

6.2.3 Proceso de fusión

Consideremos el problema de clasificación de imagen. El conocimiento de una máscara o mapa rudimentario es usado en este paso.

Consideremos que:

s denota un *site* de la imagen.

S denota el conjunto de *sites* de toda la imagen.

$L = \{l_i, 1 \leq i \leq N\}$ denota el conjunto de clases (o etiquetas).

$M(s)$ denota la clase del *site* s de acuerdo al mapa.

$I(s)$ la clase del *site* s dado por el resultado del análisis de una imagen.

Se asume que $M(s) = l_i, I(s) = l_j$.

El cambio de un *site* de la clase l_i (en el mapa) a la clase l_j (en el resultado del análisis de la imagen) puede incurrir en un error que podemos llamar un 'riesgo': $x_{l_i, l_j} \geq 0$ ($x_{l_i, l_j} = 0$ si $i = j$). Estimamos el 'riesgo' usando información de ambos mapas y el resultado del análisis de la imagen, así como algunos conocimientos comunes acerca del modelo del mundo. Observe que un *site* de una imagen no puede ser analizado sin ser relacionado con sus vecinos.

Consideremos que:

s' denota un *site* de la vecindad V_s del *site* s , y

$d(s, s')$ la distancia entre s y s'

Podemos calcular las dos medidas siguientes:

$$a(s) = \sum_{s' \in V_s} \frac{1\{M(s') = l_j\}}{d(s, s')} \quad (6.1)$$

$$b(s) = \sum_{s' \in V_s} \frac{1\{I(s') = l_j\}}{d(s, s')} \quad (6.2)$$

donde $1_{\{s\}}$ es el indicador de la función. $a(s)$ es el número de *sites* en V_s los cuales son clasificados como l_j en el mapa, inversamente medidos por sus propias distancias del *site* s . Esto representa el 'soporte' que el *site* s obtiene de sus vecinos en el mapa para tomar l_j como su clase. $b(s)$ representa el 'soporte' que el *site* s obtiene del resultado del análisis de la imagen para esta decisión. El tamaño de la vecindad depende de la resolución de la imagen. Entre mayor sea la resolución, la vecindad puede ser de mayor tamaño.

La medida de error local $E(s)$ de un *site* s está entonces definida como:

$$E(s) = \frac{x_{i,j}}{ra + (1+b)}, r \geq 1 \quad (6.3)$$

Aún cuando los resultados del análisis de la imagen y el mapa puedan incluir incertidumbre e imprecisión, nosotros creemos más en el mapa que en el resultado del análisis de la imagen. Esto está reflejado por el coeficiente r .

Como se dijo anteriormente, $x_{i,j}$ representa el 'riesgo' que sucede cuando un *site* cambia su clase de l_i (en el mapa) a l_j (en el resultado del análisis de la imagen). El valor de $x_{i,j}$ de la clase específica l_i y l_j . Generalmente hablando, el 'riesgo' de cambiar de una clase de un objeto natural (por ejemplo el bosque) a una clase de un objeto 'hecho por el hombre' (como una carretera) es menor que en el caso inverso. Por ejemplo, el 'riesgo' de un *site* de clase 'campo' de acuerdo al mapa cambiara a la clase 'carretera' en el resultado del análisis de la imagen es menor que el *site* de clase 'ciudad' en el mapa llegara a ser de clase 'bosque' en el resultado del análisis de la imagen.

La medida de error global es la acumulación del error local a través de todos los sites de la imagen:

$$GE = \sum_{s \in S} E(s) \quad (6.4)$$

Una vez que tenemos la medida de error global del resultado del análisis de una imagen, podemos decidir si el resultado es suficientemente bueno para ser usado en el proceso de fusión. Para hacer esto, simplemente tomamos p como el mejor resultado entre todos los disponibles. Otra forma de hacer esto es usar un umbral t : sólo aquellos cuyo $GE < t$ son retenidos para el proceso de fusión. Pero no es trivial determinar el valor de t .

Aunque el resultado seleccionado para ser utilizado en el proceso de fusión pueda tener un alto valor de confiabilidad, no todos los *sites* en este resultado tienen la misma confiabilidad (esto también puede ser visto del proceso de estimación del error). Luego entonces, podemos calcular un coeficiente de confianza para cada *site* del resultado basado en la medida de error local para este *site*:

$$C(s) = 1 - \frac{E(s)}{\sum_{l_i, l_j \in L} x_{l_i, l_j}} \quad (6.5)$$

donde el denominador es para la normalización de $E(s)$ así que $0 \leq C(s) \leq 1$.

La decisión de la clase final de un *site* depende de la acumulación del coeficiente de confianza del mismo *site* a través de todos los resultado retenidos:

$$l(s) = \arg \max_{l_i \in L} \sum_{k=1}^p C_k(s) I_{\{l_k(s)=l_i\}} \quad (6.6)$$

donde $I_k(s)$ representa el resultado usado como la k-ésima entrada al proceso de fusión, y $C_k(s)$ el coeficiente de confiabilidad del *site* s en este resultado.

Nótese que en la decisión tomada sobre la clase final de un *site* se toman en cuenta no sólo las medidas de error local, sino también las medidas de error global de estos resultados – sólo 'buenos' resultados son considerados en el proceso de fusión.

6.3 Extracción de zonas urbanas

Una vez que hemos obtenido la máscara urbana, procedemos a extraer la información que nos interesa (aquella que describe exclusivamente a las zonas urbanas). Esto lo realizamos mediante un producto punto a punto entre la imagen original y la máscara urbana obtenida mediante la fusión de datos.

6.4 Resultados

Al igual que en los capítulos anteriores, hemos utilizado 3 tipos de imágenes. Las imágenes empleadas se encuentran en 256 niveles de gris y tienen un tamaño de 512x512 *pixeles*. Los tipos de imágenes utilizados son:

- ◆ Fotografía aérea de alta resolución (50 cm/*pixel*)
- ◆ Imagen SPOT pancromática (10 m/*pixel*)
- ◆ Imagen SAR (30 m/*pixel*)

Estas imágenes fueron seleccionadas para evaluar los diferentes resultados arrojados a partir de imágenes con diferente resolución y con características particulares.

Los resultados obtenidos se muestran en los pasos siguientes:

1. Obtención de la máscara urbana.
 - Obtención de la imagen binarizada
 - Obtención de la imagen de textura binarizada
 - Fusión de las dos imágenes anteriores
2. Extracción de zonas urbanas.

En el proceso de fusión se utilizaron los parámetros r , x_{i_1, j_2} y x_{i_2, j_1} con los siguientes valores: $r = 2$, $x_{i_1, j_2} = 20$, $x_{i_2, j_1} = 1$. Además, se utilizó un sistema de vecindad de segundo orden, es decir, de 8 *pixeles* en la ventana de 3x3 centrada en el *site* s .

Los resultados del análisis de cada tipo de imagen empleado y las comparaciones respectivas se presentan a continuación.

6.4.1 Fotografías aéreas de alta resolución

Para la extracción de zonas urbanas en fotografías aéreas de alta resolución, se utilizaron dos ventanas de la Ciudad de México.

La primer imagen es una fotografía aérea pancromática de alta resolución que muestra una escena urbana de la Ciudad de México (figura 6.1). En esta imagen, podemos distinguir que a la izquierda se encuentran unos campos deportivos, en el centro se presentan construcciones de vías del transporte colectivo Metro y, a la derecha tenemos construcciones de casas y edificios con calles entre ellas.



Figura 6.1

Fotografía aérea de alta resolución que muestra una región urbana del oriente de la Ciudad de México (Ventana 1, imagen original en 256 niveles de gris, 512x512 pixeles)

Obtención de la imagen binarizada: Calculamos el error global de las imágenes que surgen de algunos algoritmos de segmentación y binarización. Dichas imágenes son:

- Imagen segmentada en 2 clases con el método adaptivo de segmentación presentado en el capítulo 4 (figura 6.2).
- Imagen segmentada inicialmente en 4 clases con el método adaptivo de segmentación. Posteriormente el número de clases se reduce a 2: se fusionan las clases 1 y 2 en una clase, y las 3 y 4 en otra (figura 6.3).
- Imagen segmentada en 2 clases por el método de k-medias (figura 6.4).
- Imagen binarizada por el método de binarización por correlación (figura 6.5).

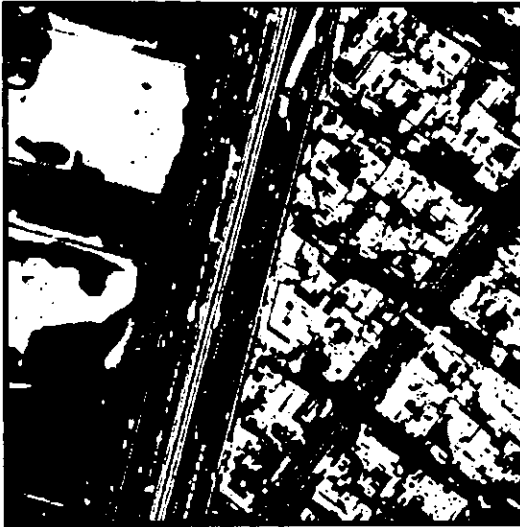


Figura 6.2

Fotografía aérea (ventana 1) segmentada en 2 clases mediante el método adaptivo de segmentación.

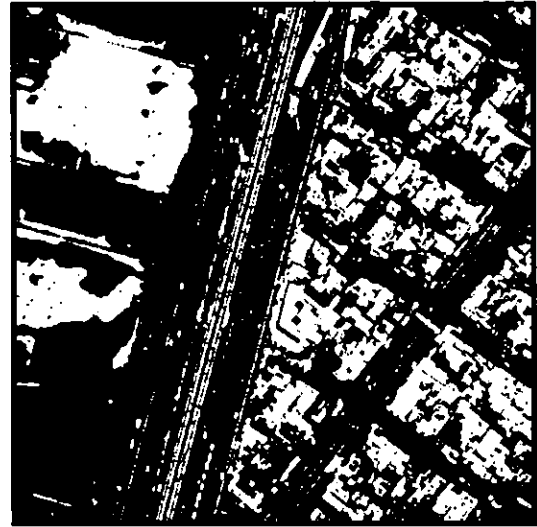


Figura 6.3

Fotografía aérea (ventana 1) segmentada de 4 a 2 clases por el método adaptivo de segmentación.

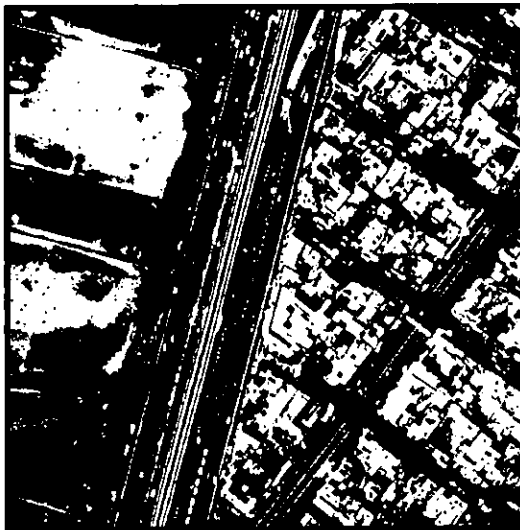


Figura 6.4

Fotografía aérea (ventana 1) segmentada en 2 clases por k-medias.

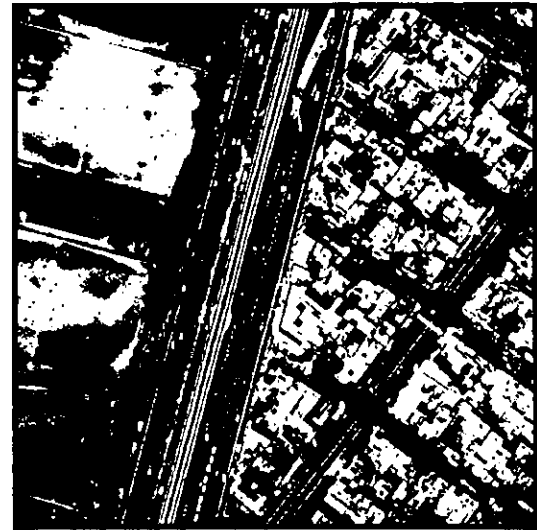


Figura 6.5

Fotografía aérea (ventana 1) binarizada por correlación.

Los valores de error global son normalizados con respecto al menor de ellos (GE_n). Los resultados obtenidos se presentan en la tabla 6.1.

Imagen	GE_n
Imagen segmentada en 2 clases con el método adaptivo de segmentación	1.0000
Imagen segmentada de 4 a 2 clases con el método adaptivo de segmentación	1.0483
Imagen segmentada en 2 clases por el método de k-medias	1.0471
Imagen binarizada por correlación	1.0419

Tabla 6.1

Valores de error global para cada imagen considerada.

De estos valores podemos observar que la imagen con menor error global es aquella binarizada mediante el método adaptivo de segmentación, de tal manera que esa es la imagen considerada para la fusión.

Obtención de la imagen de textura: Esta imagen se obtiene a partir del parámetro de temperatura propuesto por X. Descombes [10] y descrito en el capítulo 5. La imagen de textura se presenta en la figura 6.6. Esta imagen es binarizada por correlación para hacer la clasificación de las regiones urbanas de la imagen. La imagen binarizada se presenta en la figura 6.7.

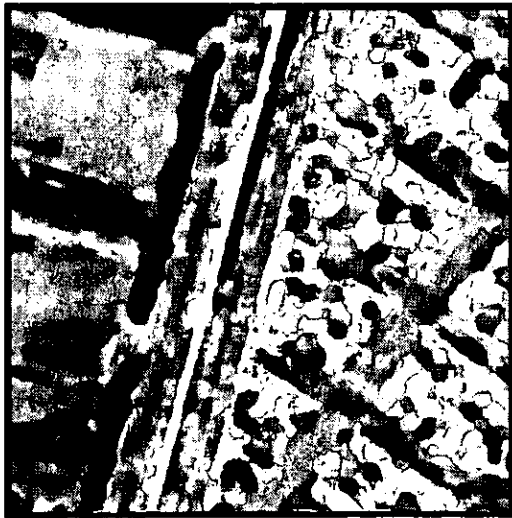


Figura 6.6

Imagen de temperatura de la fotografía aérea (ventana 1).

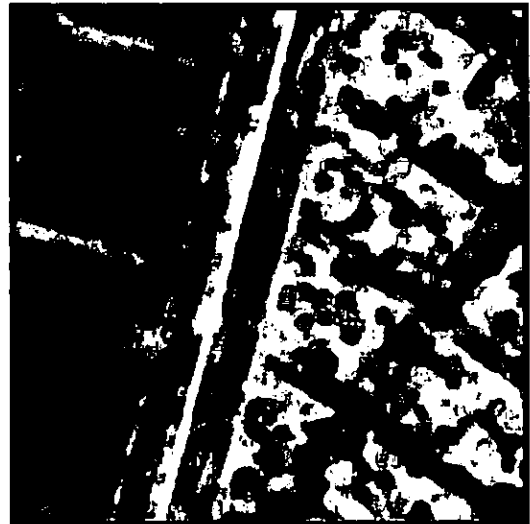


Figura 6.7

Imagen de temperatura binarizada por correlación de la fotografía aérea (ventana 1).

Proceso de fusión: El proceso de fusión lo llevamos a cabo con dos imágenes de entrada obtenidas anteriormente, es decir, la imagen original segmentada en dos clases por el método adaptivo de segmentación (figura 6.8) y la imagen de temperatura binarizada (imagen 6.9).

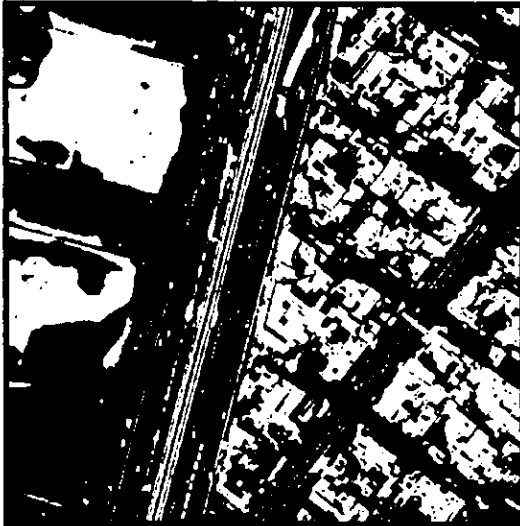


Figura 6.8

Fotografía aérea (ventana 1) segmentada en 2 clases mediante el método adaptivo de segmentación.

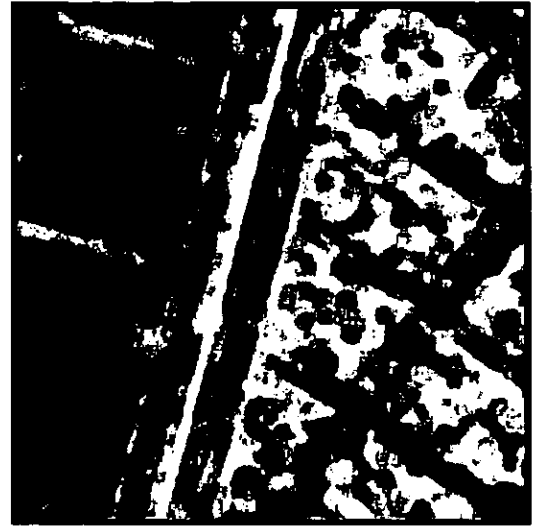


Figura 6.9

Imagen de temperatura binarizada por correlación de la fotografía aérea (ventana 1).

Utilizamos un mapa rudimentario de las regiones urbanas y no urbanas (figura 6.10) creado a partir de la imagen original.

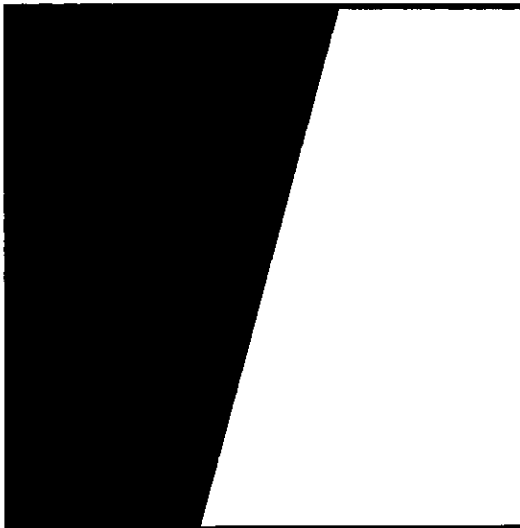


Figura 6.10

Mapa urbano de la fotografía aérea (ventana 1).

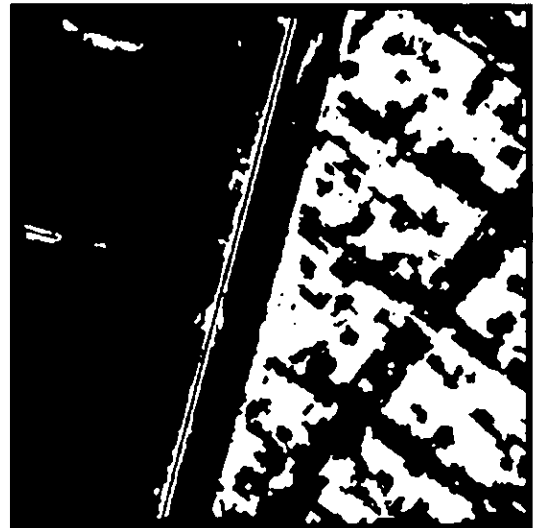


Figura 6.11

Máscara urbana de la fotografía aérea (ventana 1).

La máscara urbana obtenida después del proceso de fusión se presenta en la figura 6.11.



Figura 6.12

Fotografía aérea (ventana 1).

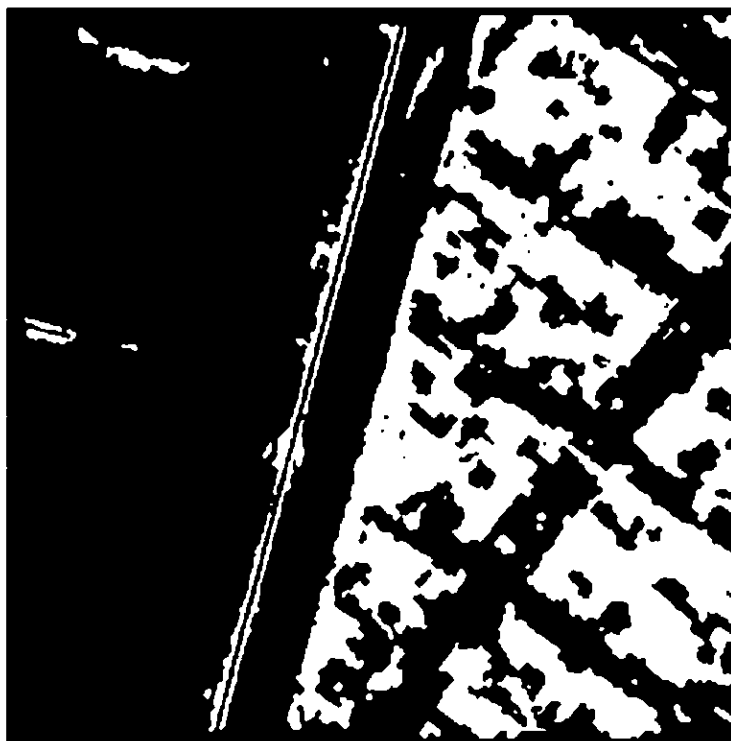


Figura 6.13

Mapa urbano de la fotografía aérea (ventana 1).

En la página anterior mostramos la ventana 1 de las fotografías aéreas de alta resolución (figura 6.12) con su máscara urbana (figura 6.13). El objetivo de mostrarlas en una misma página es para tener una vista general de la máscara urbana resultante.

Los campos deportivos que se encuentran en el lado izquierdo de la imagen son clasificados dentro de la región no urbana aun cuando el método adaptivo de segmentación había clasificado esta zona como urbana por sus niveles de gris. Sin embargo debemos notar que el error no se eliminó por completo, aún quedaron dos regiones pequeñas clasificadas como urbanas dentro de los campos deportivos cuando estas regiones no existen. Las calles que se encuentran dentro de la región urbana fueron clasificadas como regiones no urbanas.

Como podemos notar, la máscara se aproxima bien a la zona urbana de la imagen. Sin embargo, existen varias 'sombras' dentro de esta región que nuestro algoritmo está detectando como regiones no urbanas. Las vías del transporte colectivo Metro también fueron identificadas por nuestro algoritmo como regiones urbanas lo cual es correcto, porque estas instalaciones son objetos creados por el hombre.

En la página siguiente mostramos nuevamente la ventana 1 de las fotografías aéreas de alta resolución (figura 6.14) con la superposición de la máscara urbana (figura 6.15) obtenida por nuestro algoritmo. El objetivo de mostrarlas en una misma página es para tener una vista general de la extracción de zonas urbanas.

Si comparamos la imagen original (figura 6.14) con la que obtuvimos con la superposición de las zonas urbanas, vemos que las zonas no urbanas son clasificadas correctamente, salvo las sombras de la región urbana y las regiones sobre los campos deportivos que mencionamos anteriormente.



Figura 6.14

Fotografía aérea (ventana 1).

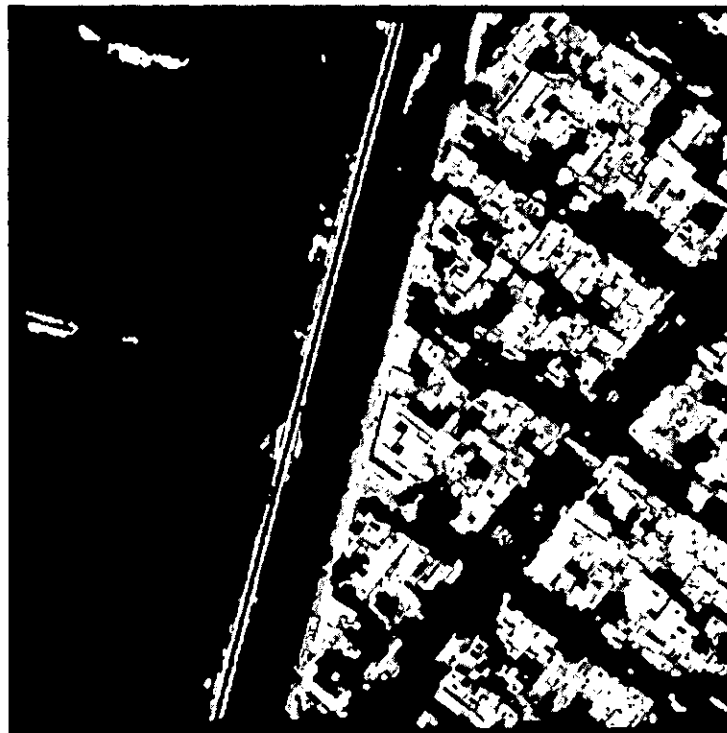


Figura 6.15

*Resultado final: Máscara urbana de la fotografía aérea (ventana 1)
con superposición de la zona urbana.*

La segunda imagen es también una fotografía aérea pancromática de alta resolución que muestra una escena urbana de la Ciudad de México (figura 6.16). En esta imagen podemos distinguir que en la esquina superior izquierda se encuentran una zona boscosa y, en la parte inferior derecha, aparece una región urbana compuesta por casas y edificios, pero además, se distinguen varios conjuntos de copas de árboles.



Figura 6.16

Fotografía aérea de alta resolución que muestra una región urbana del oriente de la Ciudad de México (Ventana 2, imagen original en 256 niveles de gris, 512x512 píxeles).

Obtención de la imagen binarizada: Calculamos el error global de las imágenes que surgen de algunos algoritmos de segmentación y binarización. Dichas imágenes son:

- Imagen segmentada en 2 clases con el método adaptivo de segmentación (figura 6.17).
- Imagen segmentada inicialmente en 4 clases con el método adaptivo de segmentación. Posteriormente el número de clases se reduce a 2: se fusionan las clases 1 y 2 en una clase, y las 3 y 4 en otra (figura 6.18).
- Imagen segmentada en 2 clases por el método de k-medias (figura 6.19).
- Imagen binarizada por correlación (figura 6.20).



Figura 6.17

Fotografía aérea (ventana 2) segmentada en 2 clases mediante el método adaptivo de segmentación.

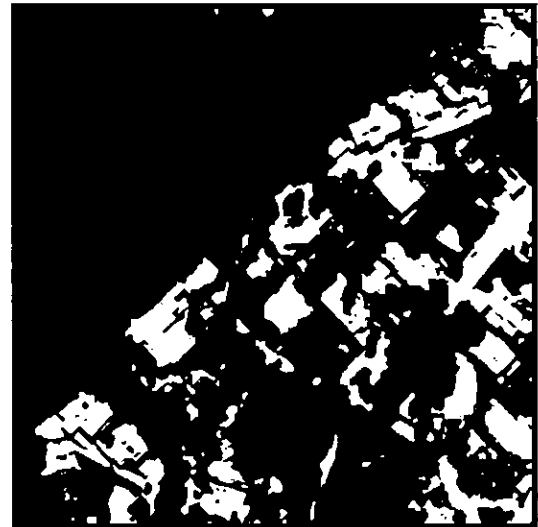


Figura 6.18

Fotografía aérea (ventana 2) segmentada de 4 a 2 clases por el método adaptivo de segmentación.



Figura 6.19

Fotografía aérea (ventana 2) segmentada en 2 clases por k-medias.

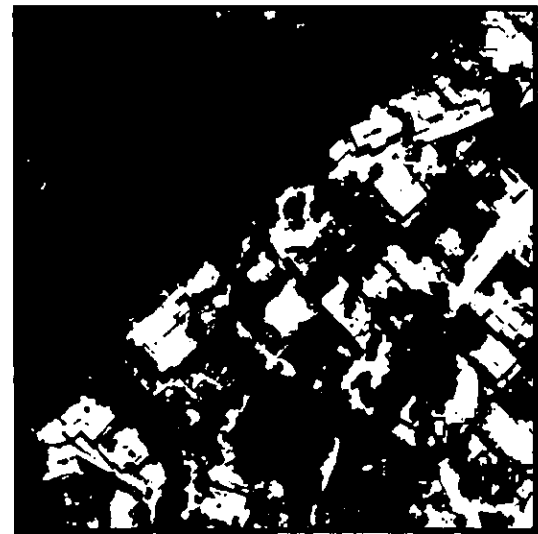


Figura 6.20

Fotografía aérea (ventana 2) binarizada por correlación.

Los valores de error global son normalizados con respecto al menor (GE_n). Los resultados obtenidos se presentan en la tabla 6.2.

Imagen	GE_n
Imagen segmentada en 2 clases con el método adaptivo de segmentación	1.0000
Imagen segmentada de 4 a 2 clases con el método adaptivo de segmentación	1.0775
Imagen segmentada en 2 clases por el método de k-medias	1.0731
Imagen binarizada por correlación	1.0685

Tabla 6.2

Valores de error global para cada imagen considerada.

De estos valores podemos observar que la imagen con menor error global es aquella binarizada mediante el método adaptivo de segmentación, de tal manera que esa es la imagen considerada para la fusión.

Obtención de la imagen de textura: Esta imagen se obtiene a partir del parámetro de temperatura propuesto por X. Descombes [10] y descrito en el capítulo 5. La imagen de textura se presenta en la figura 6.21. Esta imagen es binarizada por correlación para hacer la clasificación de las regiones urbanas de la imagen. La imagen binarizada se presenta en la figura 6.22.



Figura 6.21

Imagen de temperatura de la fotografía aérea (ventana 2).

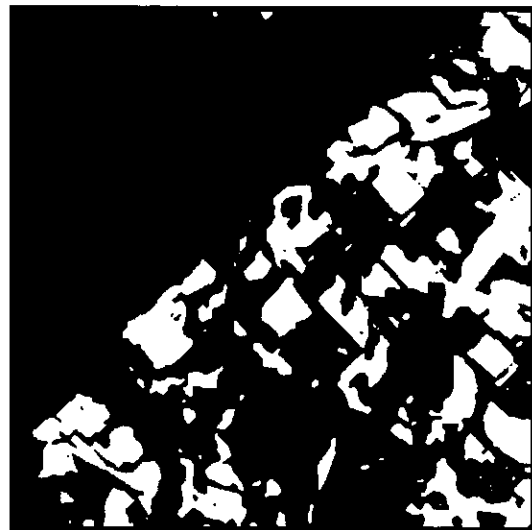


Figura 6.22

Imagen de temperatura binarizada por correlación de la fotografía aérea (ventana 2).

Proceso de fusión: El proceso de fusión lo llevamos a cabo con dos imágenes de entrada obtenidas anteriormente, es decir, la imagen original segmentada en dos clases por el método adaptivo de segmentación (figura 6.23) y la imagen de temperatura binarizada (imagen 6.24).

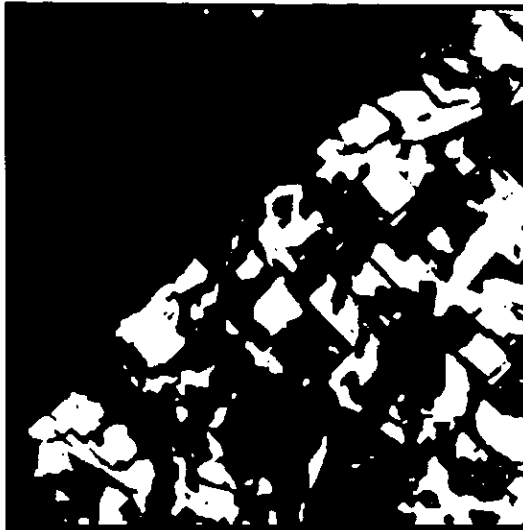


Figura 6.23

Imagen de temperatura binarizada por correlación de la fotografía aérea (ventana 2).

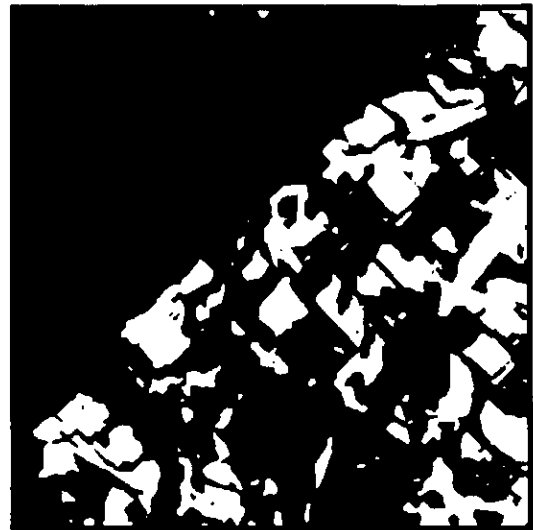


Figura 6.24

Fotografía aérea (ventana 2) segmentada en 2 clases mediante el método adaptivo de segmentación.

Utilizamos un mapa rudimentario de las regiones urbanas y no urbanas (figura 6.25) creado a partir de la imagen original.

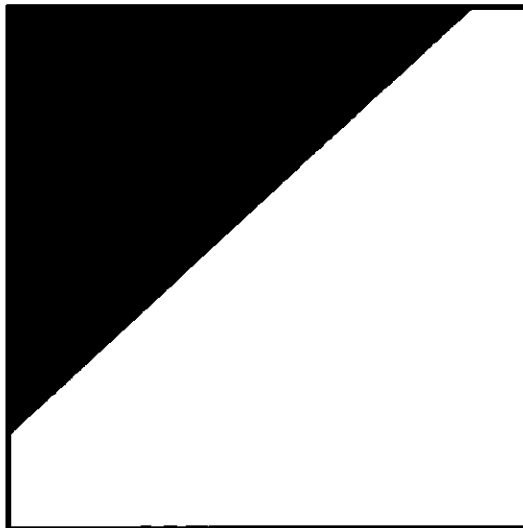


Figura 6.25

Mapa urbano de la fotografía aérea (ventana 2).

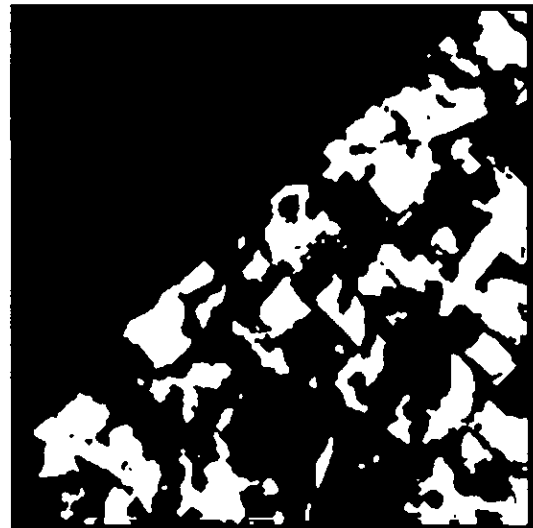


Figura 6.26

Máscara urbana de la fotografía aérea (ventana 2).

La máscara urbana obtenida después del proceso de fusión se presenta en la figura 6.26.

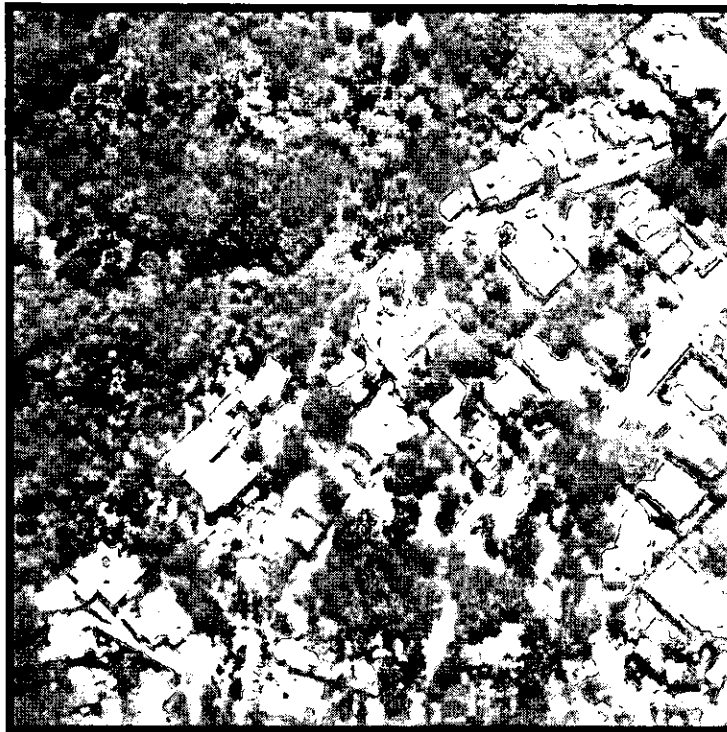


Figura 6.27

Fotografía aérea (ventana 2).

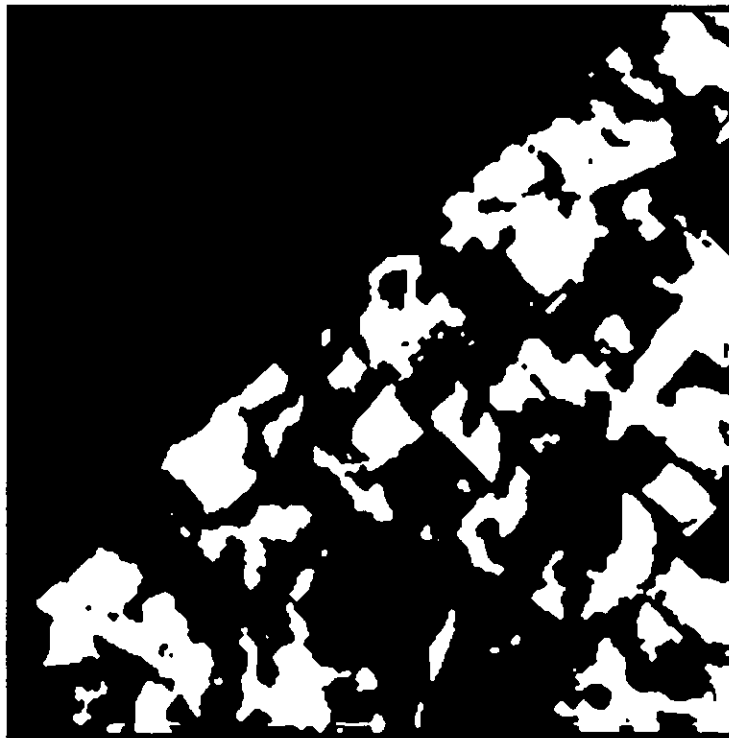


Figura 6.28

Máscara urbana de la fotografía aérea (ventana 2).

En la página anterior mostramos la ventana 2 de las fotografías aéreas de alta resolución (figura 6.27) con su máscara urbana (figura 6.28). El objetivo de mostrarlas en una misma página es para tener una vista general de la máscara urbana resultante.

Esta fotografía aérea nos muestra resultados muy satisfactorios en la clasificación de zonas urbanas y no urbanas. Como podemos notar, la región boscosa de la esquina superior izquierda es identificada como una zona urbana. Además, los árboles que se encuentran entre las construcciones de la zona urbana (esquina inferior derecha) también son clasificados como regiones no urbanas.

La máscara urbana se aproxima por mucho a la región urbana real de la imagen.

En la página siguiente mostramos nuevamente la ventana 2 de las fotografías aéreas de alta resolución (figura 6.29) con la superposición de la máscara urbana (figura 6.30) obtenida por nuestro algoritmo. El objetivo de mostrarlas en una misma página es para tener una vista general de la extracción de zonas urbanas.

Si comparamos la imagen original (figura 6.29) con la que obtuvimos con la superposición de las zonas urbanas, vemos éstas son clasificadas correctamente.

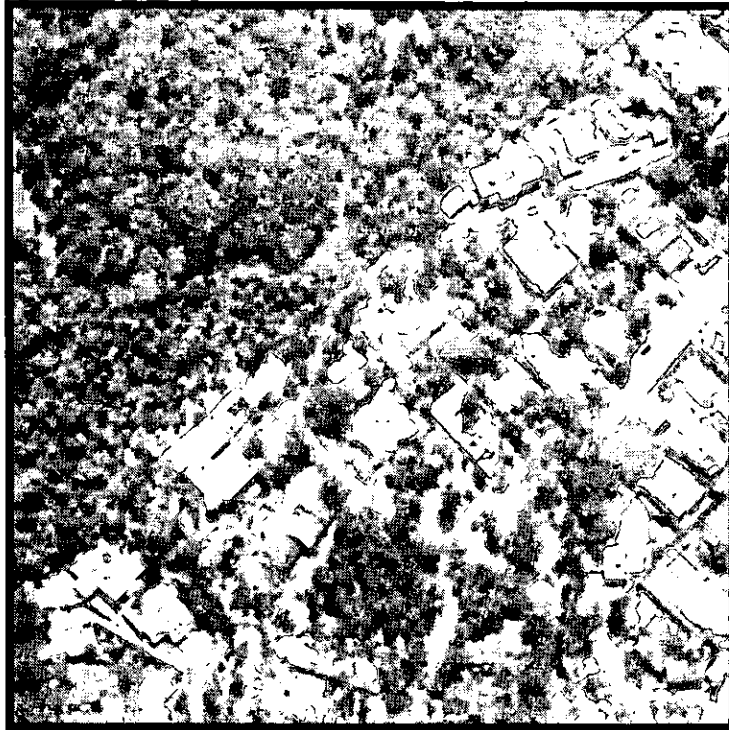


Figura 6.29

Fotografía aérea (ventana 2).

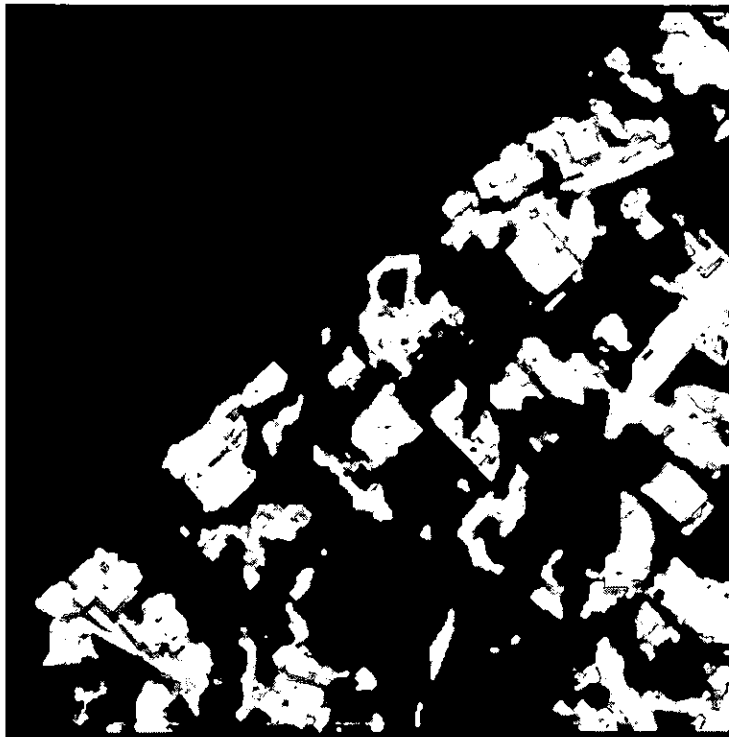


Figura 6.30

Resultado final: Máscara urbana de la fotografía aérea (ventana 2) con superposición de la zona urbana.

6.4.2 Imagen SPOT pancromática

Decidimos aplicar el método de extracción de zonas urbanas a imágenes de percepción remota que ofrecen una menor resolución, como lo son las imágenes SPOT y las imágenes SAR.

La ventana seleccionada de la imagen SPOT (figura 6.31) muestra una escena urbana de la ciudad de México que incluye la zona del aeropuerto y el bosque de Aragón, que en este caso representan las regiones más homogéneas de la imagen (región no urbana). La región urbana está representada por las colonias que se encuentran alrededor del aeropuerto.



Figura 6.31

Imagen SPOT pancromática con una resolución de 10 metros por pixel [512x512x256].

Obtención de la imagen binarizada: Calculamos el error global de las imágenes que surgen de algunos algoritmos de segmentación y binarización. Dichas imágenes son:

- Imagen segmentada en 2 clases con el método adaptivo de segmentación (figura 6.32).
- Imagen segmentada inicialmente en 4 clases con el método adaptivo de segmentación. Posteriormente el número de clases se reduce a 2: se fusionan las clases 1 y 2 en una clase, y las 3 y 4 en otra (figura 6.33).
- Imagen segmentada en 2 clases por el método de k-medias (figura 6.34).
- Imagen binarizada por correlación (figura 6.35).



Figura 6.32

Imagen SPOT segmentada en 2 clases mediante el método adaptivo de segmentación.



Figura 6.33

Imagen SPOT segmentada de 4 a 2 clases mediante el método adaptivo de segmentación.

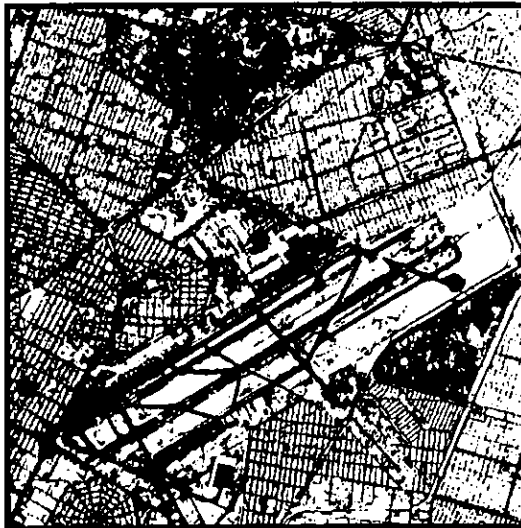


Figura 6.34

Imagen SPOT segmentada en 2 clases por el método de k-medias.



Figura 6.35

Imagen SPOT binarizada por correlación.

Los valores de error global son normalizados con respecto al menor de ellos (GE_n). Los resultados obtenidos se presentan en la tabla 6.3.

Imagen	GE_n
Imagen segmentada en 2 clases con el método adaptivo de segmentación	1.0277
Imagen segmentada de 4 a 2 clases con el método adaptivo de segmentación	1.1740
Imagen segmentada en 2 clases por el método de k-medias	1.0506
Imagen binarizada por correlación	1.0000

Tabla 6.3

Valores de error global para cada imagen considerada.

De estos valores podemos observar que la imagen con menor error global es aquella binarizada mediante el método de binarización por correlación, de tal manera que esa es la imagen considerada para la fusión.

Obtención de la imagen de textura: La imagen de textura imagen se obtiene a partir del parámetro de temperatura propuesto por X. Descombes [10] y descrito en el capítulo 5. La imagen de temperatura se presenta en la figura 6.36. Esta imagen es binarizada por correlación para hacer la clasificación de las regiones urbanas de la imagen. La imagen binarizada se presenta en la figura 6.37.

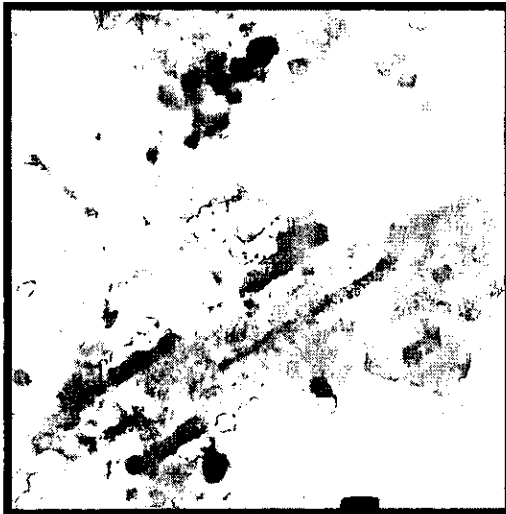


Figura 6.36

Imagen SPOT de temperatura.

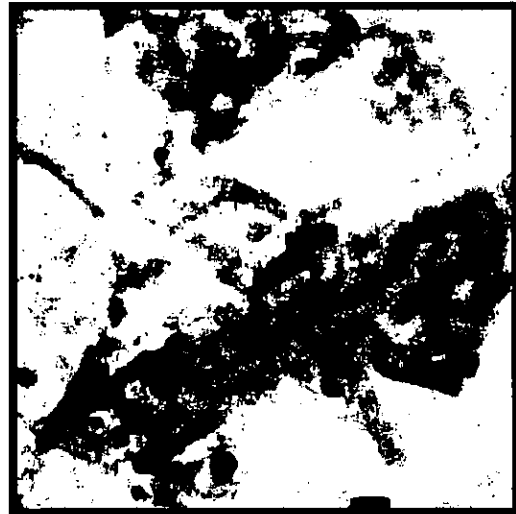


Figura 6.37

Imagen SPOT de temperatura binarizada por correlación.

Proceso de fusión: El proceso de fusión se lleva a cabo con las dos imágenes de entrada obtenidas anteriormente, es decir, la imagen original binarizada por correlación (figura 6.38) y la imagen de temperatura binarizada (imagen 6.39).



Figura 6.38

Imagen SPOT binarizada por correlación.

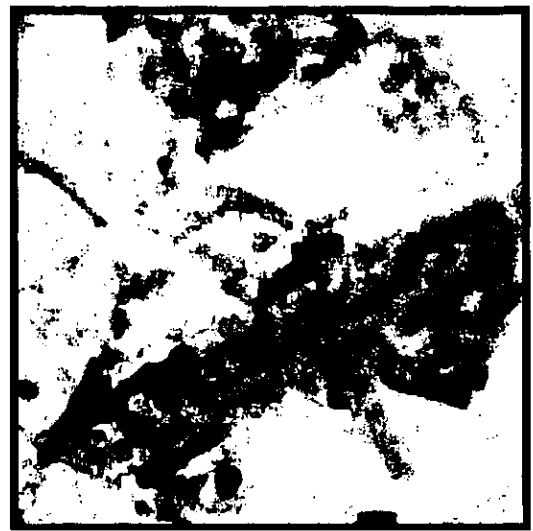


Figura 6.39

Imagen de temperatura binarizada por correlación de la imagen SPOT.

Utilizamos un mapa rudimentario de las regiones urbanas y no urbanas (figura 6.40) creado a partir de la imagen original.

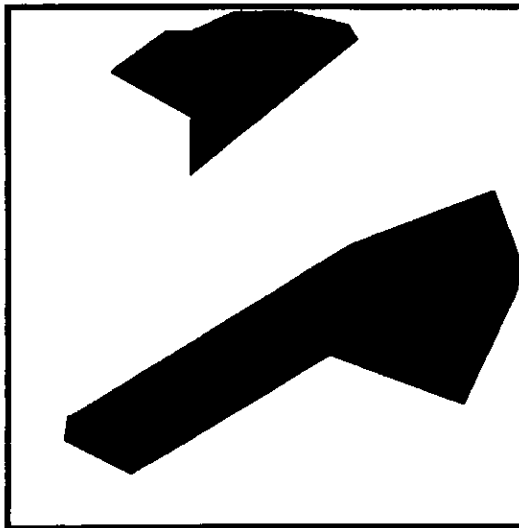


Figura 6.40

Mapa urbano de la imagen SPOT.



Figura 6.41

Máscara urbana de la imagen SPOT.

La máscara urbana obtenida después del proceso de fusión se presenta en la figura 6.41.



Figura 6.42

Imagen SPOT original.

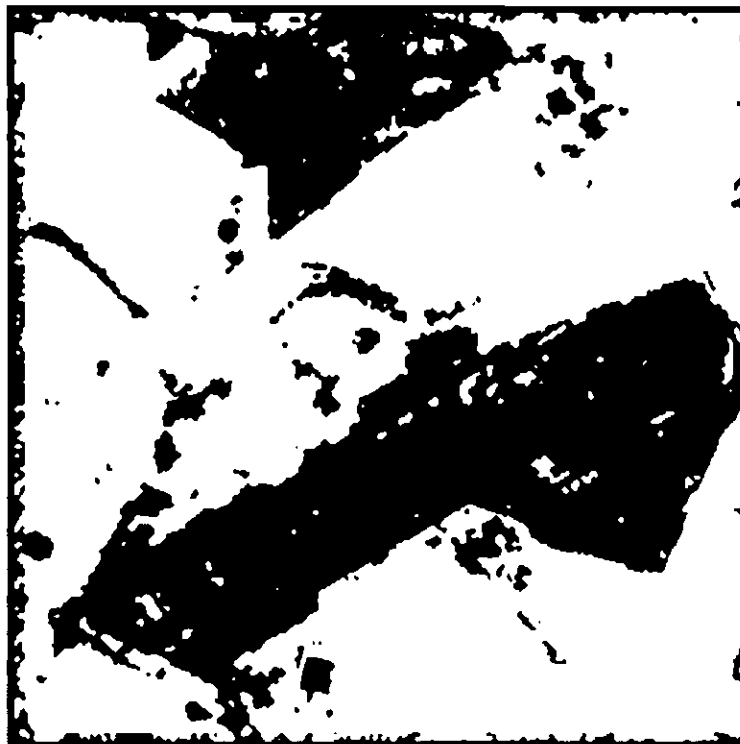


Figura 6.43

Máscara urbana.

En la página anterior mostramos la imagen SPOT (figura 6.42) con su máscara urbana (figura 6.43). El objetivo de mostrarlas en una misma página es para tener una vista general de la máscara urbana resultante.

En esta imagen podemos notar que, debido a la baja resolución de la imagen, nuestros resultados de extracción de zonas urbanas no fueron tan buenos como en la fotografía aérea. Sin embargo, sí tenemos una buena aproximación de las pistas del aeropuerto y del bosque de Aragón que, en este caso, son las zonas más significativas de regiones no urbanas. Debemos mencionar que, debido a la resolución de estas imágenes, las calles que están en las colonias del rededor del aeropuerto no son identificadas ya que el algoritmo de temperatura las analiza con una fuerte varianza.

En la página siguiente mostramos nuevamente la imagen SPOT (figura 6.44) con la superposición de la máscara urbana (figura 6.45) obtenida por nuestro algoritmo. El objetivo de mostrarlas en una misma página es para tener una vista general de la extracción de zonas urbanas.

Si comparamos la imagen original (figura 6.44) con la que obtuvimos con la superposición de las zonas urbanas, vemos que las zonas no urbanas tienen una buena clasificación de las regiones más grandes (las pistas del aeropuerto y el bosque de Aragón). Se observa también que algunas calles grandes también fueron identificadas, pero no del todo bien, ya que quedan fragmentadas.



Figura 6.44

Imagen SPOT original.



Figura 6.45

*Resultado final: Máscara urbana de imagen SPOT
con superposición de la zona urbana.*

6.4.3 Imagen SAR

Para calcular el parámetro de textura y la binarización en imágenes del tipo SAR, se utilizó la figura 6.46 con resolución de 30 metros por *pixel*. Esta imagen tuvo una etapa previa de filtrado basado en la Transformada Polinomial.



Figura 6.46

Imagen SAR adquirida por el satélite ERS-1 el 19 de septiembre de 1997, la escena corresponde a la zona urbana que rodea el aeropuerto internacional de la Ciudad de México [512x512 píxeles, 256 niveles de gris].

Obtención de la imagen binarizada: Calculamos el error global de las imágenes que surgen de algunos algoritmos de segmentación y binarización. Dichas imágenes son:

- Imagen segmentada en 2 clases con el método adaptivo de segmentación propuesto en el capítulo 4 (figura 6.47).
- Imagen segmentada inicialmente en 4 clases con el método adaptivo de segmentación. Posteriormente el número de clases se reduce a 2: se fusionan las clases 1 y 2 en una clase, y las 3 y 4 en otra (figura 6.48).
- Imagen segmentada en 2 clases por el método de k-medias (figura 6.49).
- Imagen binarizada por el método de binarización por correlación (figura 6.50).



Figura 6.47

Imagen SAR segmentada en 2 clases mediante el método adaptivo de segmentación.



Figura 6.48

Imagen SAR segmentada de 4 a 2 clases mediante el método adaptivo de segmentación.



Figura 6.49

Imagen SAR segmentada en 2 clases por el método de k-medias.



Figura 6.50

Imagen SAR binarizada por correlación.

Los valores de error global son normalizados con respecto al menor (GE_n). Los resultados obtenidos se presentan en la tabla 6.4.

Imagen	GE_n
Imagen segmentada en 2 clases con el método adaptivo de segmentación	1.0000
Imagen segmentada de 4 a 2 clases con el método adaptivo de segmentación	1.0945
Imagen segmentada en 2 clases por el método de k-medias	1.1290
Imagen binarizada por correlación	1.1093

Tabla 6.4

Valores de error global para cada imagen considerada.

De estos valores podemos observar que la imagen con menor error global es aquella binarizada mediante el método adaptivo de segmentación, de tal manera que esa es la imagen considerada para la fusión.

Obtención de la imagen de textura: Esta imagen se obtiene a partir del parámetro de temperatura propuesto por X. Descombes [10] y descrito en el capítulo 5. La imagen de textura se presenta en la figura 6.51. Esta imagen es binarizada por correlación para hacer la clasificación de las regiones urbanas de la imagen. La imagen binarizada de la imagen de temperatura se presenta en la figura 6.52.

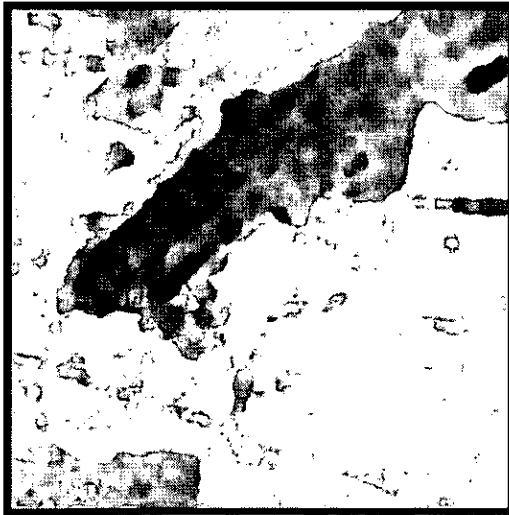


Figura 6.51

Imagen de temperatura de la imagen SAR.



Figura 6.52

Imagen de temperatura binarizada por correlación de la imagen SAR.

Proceso de fusión: El proceso de fusión lo llevamos a cabo con dos imágenes de entrada obtenidas anteriormente, es decir, la imagen original segmentada en dos clases por el método adaptivo de segmentación (figura 6.53) y la imagen de temperatura binarizada (imagen 6.54).

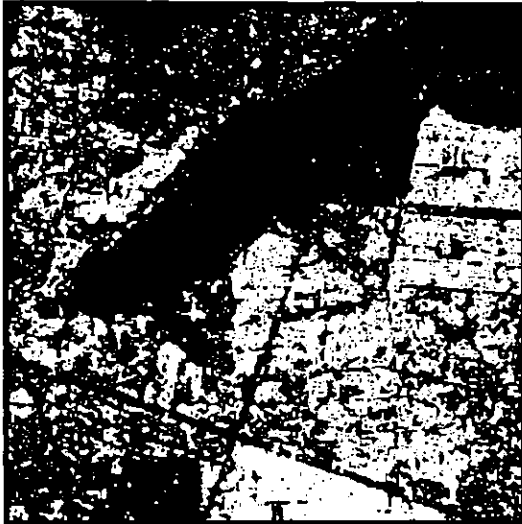


Figura 6.53

Imagen SAR segmentada en 2 clases mediante el método adaptivo de segmentación.



Figura 6.54

Imagen de temperatura binarizada por correlación de la imagen SAR.

Utilizamos un mapa rudimentario de las regiones urbanas y no urbanas (figura 6.55) creado a partir de la imagen original.

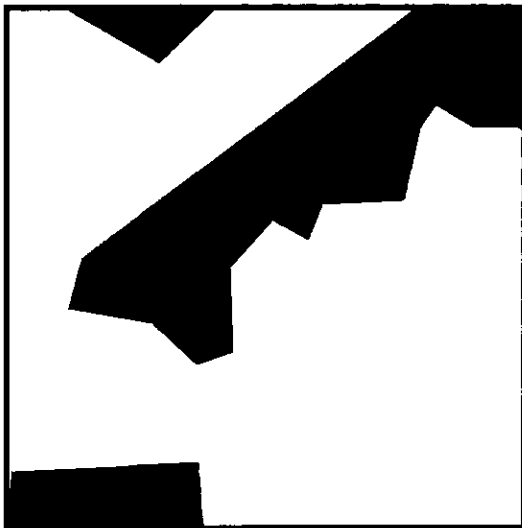


Figura 6.55

Mapa urbano de la imagen SAR.



Figura 6.56

Máscara urbana de la imagen SAR.

La máscara urbana obtenida después del proceso de fusión se presenta en la figura 6.56.



Figura 6.57

Imagen SAR original.



Figura 6.58

Imagen fusionada.

En la página anterior mostramos la imagen SAR (figura 6.57) con su máscara urbana (figura 6.58). El objetivo de mostrarlas en una misma página es para tener una vista general de la máscara urbana resultante.

En esta imagen podemos notar que, debido a la resolución de la imagen es aún más baja que en las imágenes SPOT, nuestros resultados de extracción de zonas urbanas no fueron tampoco tan buenos como en la fotografía aérea. A pesar de ello, las pistas del aeropuerto son perfectamente identificables, así como el fragmento del bosque de Aragón (parte superior de la imagen) y la región homogénea de la esquina inferior izquierda. En este tipo de imágenes sucede algo semejante que con las imágenes SPOT, es decir, que solamente las calles grandes pueden ser identificadas (las calles pequeñas que están en las colonias del rededor del aeropuerto se pierden) pero aparecen fragmentadas en la máscara urbana.

En la página siguiente mostramos nuevamente la imagen SAR (figura 6.59) con la superposición de la máscara urbana (figura 6.60) obtenida por nuestro algoritmo. El objetivo de mostrarlas en una misma página es para tener una vista general de la extracción de zonas urbanas.

Si comparamos la imagen original (figura 6.59) con la que obtuvimos con la superposición de las zonas urbanas, vemos que las zonas no urbanas tienen una buena clasificación de las regiones más grandes (las pistas del aeropuerto y el bosque de Aragón y la zona homogénea de la esquina inferior izquierda). Vemos también que algunas de las calles grandes también fueron identificadas, pero como mencionábamos dichas calles aparecen fragmentadas.



Figura 6.59
Imagen SAR original.



Figura 6.60
*Resultado final: Máscara urbana de imagen SAR
con superposición de la zona urbana.*

6.5 Conclusiones

En este capítulo extrajimos zonas urbanas de imágenes de fotografía aérea de alta resolución, SPOT y SAR. Como primer paso para la extracción de zonas urbanas obtuvimos una máscara urbana a partir del algoritmo de fusión de datos propuesto por Shan Yu [32]. La fusión se lleva a cabo con una imagen binarizada, que nos proporciona información muy aproximada acerca de los contornos de los elementos presentes en la escena en cuestión, y una imagen de temperatura binarizada, que nos dará información acerca de las texturas de los elementos presentes en la escena de interés, para disminuir el error que la segmentación genera al confundir zonas no urbanas con zonas urbanas por sus niveles de gris. Una vez que obtuvimos estas imágenes, procedimos a la fusión de estos datos para extraer una máscara urbana. Por último, con base a la máscara urbana, procederemos a extraer las regiones urbanas presentes en las escenas analizadas.

En los resultados presentados notamos que nuestro algoritmo presenta buenos resultados en el análisis de imágenes aéreas de alta resolución en comparación con las imágenes SPOT y SAR. Algo que es muy notorio es que nuestro algoritmo de extracción de zonas urbanas identifica perfectamente las calles en las fotografías aéreas, mientras que en las imágenes SPOT y SAR no se identifican, debido a su menor resolución. En estos dos tipos de imágenes sólo se identifican las avenidas de dimensiones considerables con respecto a la resolución de las imágenes.

Como vimos, el aporte principal de la imagen de temperatura es reducir el error que se presenta en la imagen segmentada al confundir las regiones no urbanas (como campos deportivos, bosques, copas de árboles) con las regiones urbanas cuando sólo se analizan los niveles de gris. Dado que la imagen de temperatura muestra un parámetro de textura, esto ayuda a tener resultados confiables.

7. CONCLUSIONES GENERALES

La percepción remota de recursos de la tierra mediante imágenes aéreas e imágenes de satélite es un área de gran actividad en investigación. Específicamente, la adquisición de datos remotos en regiones urbanas puede facilitar tareas en la actualización de cartografía y planeación urbana mediante la segmentación y extracción automática de regiones de interés. Debido a esto, el contar con algoritmos que faciliten estas tareas, mediante la generación de imágenes segmentadas de interpretación objetiva, es de especial importancia en el área del procesamiento digital de imágenes de percepción remota.

En este contexto, la extracción de zonas urbanas a partir de información proporcionada por los niveles de gris mediante una aproximación probabilística e información contextual vía un esquema estocástico es un campo de especial interés. Esto se debe a que la información que puede ofrecer una imagen, exclusivamente a partir de sus características espectrales (niveles de gris), resulta insuficiente.

El proceso completo para la extracción de regiones urbanas a partir de una imagen de percepción remota puede dividirse en tres etapas principales:

- ◆ Segmentación (esquema probabilístico y estocástico)
- ◆ Clasificación (esquema de fusión de datos)
- ◆ Extracción

Bajo esta perspectiva, hemos presentado aquí un esquema completo para la extracción de regiones urbanas a partir de fotografías aéreas de alta resolución, con el que se busca ofrecer resultados mostrando regiones homogéneas, al mismo tiempo que se preserven la mayor cantidad de detalles finos de la imagen.

Segmentación: Para el proceso de segmentación propusimos un algoritmo adaptivo que presenta dos fases principales: una no contextual y otra contextual. La idea de un algoritmo de este tipo surgió debido a que el tipo de imágenes consideradas para nuestro estudio (en especial las fotografías aéreas de alta resolución) presentan una gran cantidad de detalles, a los cuales nos referimos como 'estructuras finas'. Es

objetivo principal del algoritmo propuesto el preservar este tipo de estructuras, al tiempo que se buscó evitar una alta incidencia de *píxeles* aislados, generando así imágenes con regiones homogéneas.

El algoritmo adaptivo de segmentación se basa en una serie de conceptos y algoritmos de segmentación clásicos, que se listan a continuación:

- ◆ Mínima distancia
- ◆ Máxima probabilidad *a posteriori* (MAP)
- ◆ Campos aleatorios de Markov
- ◆ Modelo de Potts
- ◆ Recocido Simulado

Para decidir cuales píxeles de la imagen serían clasificados mediante un enfoque no contextual y cuales mediante uno contextual se decidió aproximar las funciones de probabilidad conjunta para cada una de las clases a partir de sus respectivos centros de clase y de su desviación estándar. Se pudo observar que esta decisión fue correcta, ya que al tomar en cuenta este concepto estadístico se obtuvo una diferente respuesta para cada una de las clases, lo cual representa una mayor adecuación a las características propias de cada clase.

Una vez definidas las funciones de probabilidad conjunta se definió un *umbral de rechazo* para decidir qué píxeles serían clasificados mediante un esquema no contextual (aquellos cuya probabilidad condicional máxima se encuentra por arriba del umbral son clasificados utilizando un estimador de máximo *a posteriori*), y cuales por el algoritmo contextual (los píxeles rechazados por el umbral se clasifican mediante un esquema de recocido simulado, basado en el modelo de Potts). Aquí una observación importante es el que este umbral de rechazo se debe considerar en base a la probabilidad conjunta, ya que si se considera la probabilidad *a posteriori* (la probabilidad conjunta por la probabilidad *a priori*) las clases no predominantes de la imagen –a las cuales por lo general pertenecen las estructuras finas que se quieren preservar– se ven afectadas, lo cual no conviene a nuestros intereses.

Los resultados obtenidos al aplicar el algoritmo propuesto sobre fotografías aéreas e imágenes SPOT muestran imágenes segmentadas con buena definición de contornos, preservación de las estructuras finas, y una baja incidencia de píxeles aislados, lo que resulta en imágenes homogéneas. Al hacer comparaciones visuales y numéricas del algoritmo adaptivo propuesto contra aquellos métodos clásicos en los que se basa, se hace evidente que el esquema propuesto ofrece imágenes segmentadas de calidad. Sin embargo, para el caso de imágenes SAR, es recomendable combinar nuestro método adaptivo de segmentación con un esquema de filtrado que también respete los contornos y los detalles finos de la escena analizada. Para este caso observamos que el utilizar un método adaptivo de reducción de ruido multiplicativo, basado en la transformada polinomial, representa un buen complemento para nuestro esquema de segmentación.

Las principales ventajas del algoritmo adaptivo de segmentación propuesto son:

- ◆ Generación de imágenes segmentadas con regiones homogéneas y respeto de los detalles finos de la imagen
- ◆ Facilidad de implementación
- ◆ Bajo tiempo de ejecución
- ◆ Cuenta con parámetros variables

Clasificación: En la etapa de clasificación, se obtuvo una máscara urbana (segmentación binaria) a través de la aplicación de un algoritmo de fusión de datos entre dos imágenes que nos proporcionaran información relevante. Las imágenes utilizadas fueron:

Imagen binarizada: Esta imagen nos dio información muy aproximada acerca de los contornos de los elementos urbanos, basada exclusivamente en los niveles de gris de la imagen original. Debido a que existen diversos métodos de segmentación y binarización, se hizo una evaluación de los resultados generados por algunos de estos métodos con el objetivo de determinar la imagen más adecuada para el proceso de fusión. El proceso de evaluación de resultados se basó en la estimación de un error global y considerando que, a menor error global, mayor confiabilidad de poder tomar este resultado como el mejor para el proceso de fusión. El método de segmentación propuesto mostró su efectividad al mostrar el menor error global en comparación con las medidas de error global de los métodos k-medias y binarización por correlación.

Imagen de temperatura binarizada: La imagen de temperatura nos dio información acerca de las texturas de los elementos urbanos y no urbanos (representados en la imagen de temperatura por niveles de gris altos y bajos, respectivamente) para disminuir el error que se genera con la segmentación al confundir zonas no urbanas con zonas urbanas, al considerar únicamente sus características espectrales. La forma en que se obtuvo la imagen de temperatura fue a través de la estimación de un parámetro característico de la textura urbana (parámetro de temperatura) basado en los campos aleatorios de Markov. La imagen de temperatura se binarizó a fin de que, a partir de un umbral, las regiones con mayor temperatura se asignen a una categoría urbana, mientras que aquellas con menor temperatura son asignados a una categoría no urbana. El método de binarización utilizado fue el de Correlación ya que éste evalúa la correlación entre la imagen de temperatura y la imagen de temperatura binarizada.

Extracción: En esta última etapa, se realizó una discriminación entre regiones urbanas y no urbanas. Esta discriminación se llevó a cabo superponiendo la máscara urbana (aquella que se obtuvo por la clasificación) sobre la imagen original, es decir, a las regiones no urbanas se les asignó un valor único de nivel de gris (color negro), mientras que para las regiones urbanas se consideró el nivel de gris de la imagen original.

En general, podemos considerar que el esquema de extracción de regiones urbanas presentado ofrece resultados de buena calidad, especialmente cuando se trabaja con fotografías aéreas de alta resolución.

En el esquema presentado resulta de especial relevancia el considerar para el proceso de fusión dos imágenes que ofrezcan información complementaria, como en este caso fueron la de temperatura (que ofrece una mejor discriminación entre regiones urbanas y no urbanas, pero sin proporcionar una máscara muy detallada) y la segmentada en dos clases mediante el algoritmo adaptivo propuesto (la cual si bien, al considerar únicamente las características espectrales de la imagen analizada, puede llevar a errores de clasificación; por otra parte ofrece una muy buena estimación de contornos, lo cual ayuda a preservar más detalles de la escena estudiada).

Como trabajo futuro, se propone agregar otros parámetros de textura al proceso de fusión con el objetivo de obtener una mejor estimación de las regiones urbanas.

GLOSARIO

Clasificación – Asignación de objetos contenidos en una imagen, a una clase específica, de acuerdo a sus propiedades.

Clasificador – Expresión matemática que define la pertenencia de un objeto a cierta clase, en una clasificación por medio de descriptores.

Conectividad – Relación geométrica que un *pixel* establece con sus vecinos inmediatos.

Contorno – Frontera de una región perteneciente a un objeto, dentro de la imagen, la cual se define por cambios abruptos en niveles de gris.

Convolución – Operación que consiste en sustituir a un *pixel* por la suma ponderada de los *pixeles* de su vecindad local. Los pesos correspondientes a cada *pixel* se almacenan en una matriz, que conforma un operador de convolución. Se utiliza para operaciones de filtrado en el *dominio espacial* (ver *filtros*).

Detector – Dispositivo utilizado para capturar la información de una escena, posibilitando la obtención de una imagen.

Dominio espacial – Se refiere a transformaciones sobre las coordenadas espaciales de la imagen.

Dominio de las frecuencias – Se refiere a transformaciones sobre la *frecuencia espacial* de una imagen.

Escena – Espacio físico del cual se obtiene la información de una imagen.

Especificación del Histograma – Consiste en aproximar el histograma de la imagen original a una función específica.

Filtro – En el *dominio espacial*, un filtro digital es una función sobre los *pixeles* de la *vecindad* local de cada *pixel* de una imagen.

Frecuencia espacial – Cambios de intensidad de los niveles de gris de los *pixeles* con respecto a las distancias espaciales.

Gradiente – Vector que apunta a la dirección de máximo cambio en la vecindad de un *pixel*. Se usa como *filtro* para resaltar los *contornos* de una región.

Histograma – Representa una estimación discreta de la función de densidad de probabilidad de los niveles de gris de una imagen digital.

Igualación del Histograma – Consiste en obtener una distribución uniforme de la función de densidad de probabilidad de los niveles de gris de una imagen.

Imagen – Una imagen digital es un arreglo bidimensional de enteros cuyos índices de filas y columnas denotan la posición de un punto, y el valor del punto corresponde a un nivel de gris.

Laplaciano – Operador diferencial que aplicado a una imagen, genera una buena definición de bordes en cualquier dirección. Las zonas con cambios abruptos de intensidad se resaltan mientras que las planas desaparecen. La desventaja de este filtro es que por ser un operador de 2^a. derivada, es muy sensible al ruido.

Mediana – Este filtro no lineal selecciona la mediana del conjunto de *pixeles* de una *vecindad*. Elimina el *ruido* aleatorio en forma de líneas y puntos de la imagen.

Modificación del Histograma – Consiste en aplicar cualquier función de transformación de niveles de gris a la imagen original según las características que se quieren resaltar en ella.

Normalización del Histograma – Consiste en asignar un peso igual a todos los valores de los *pixeles* de la imagen original, el resultado es un histograma plano cuyas barras tienen todas el mismo valor.

Patrón – Parámetro descriptivo de un objeto, que puede ser de tamaño o de forma.

Pixel – Elemento pictórico de la imagen.

Región – Conjunto de *pixeles* con propiedades similares, conectados entre sí.

Segmentación – Separación de una imagen en regiones disjuntas, cuyos *pixeles* mantienen un valor dentro de un rango.

Umbral – Valor de un nivel de gris utilizado como referencia para separar dos regiones.

REFERENCIAS

- [1] Aguilar, A; *et al.* **Procesamiento Digital de Imágenes.** *Notas de la Maestría en Ciencias de la Computación de la UACP y P del CCH con sede en el IIMAS-UNAM.* UNAM. México. 1994.
- [2] Azpiroz, J; Media, V. y Lerallut, J-F. **Procesamiento de imágenes biomédicas.** Universidad Tecnológica de Compiègne y UAM. México.1999.
- [3] Baraldi A. and Parmiggiani F. **Urban area classification by multispectral spot images.** IEEE Trans. On geoscience and Remote Sensing, vol. 28, n4. July,1990.
- [4] Besag, J. **On the Statistical Analysis of Dirty Pictures.** *Journal of the Royal Statistical Society.* Serie B 48. 1986.
- [5] Bow, Sing-Tze. **Pattern Recognition and Image Processing.** Marcel Dekker, Inc. USA. 1992.
- [6] Blink, A. **Grey-level thresholding of images using a correlation criterion.** *Pattern Recognition Letters.* North Holland. 1989.
- [7] Camarillo, P.; Varela, A. and Escalante, B. **Adaptive multiplicative-noise reduction in SAR images with polynomial transforms.** IEEE. 1998.
- [8] Chuvieco, E. **Fundamentos de Teledetección Espacial.** Ediciones Rialp. Madrid, España. 1996. Segunda edición.
- [9] Delves, L. M.; Wilkinsons, R.; Oliver, C. J. and White, R. G. **Comparing the Performance of SAR Image Segmentation Algorithms.** *International Journal of Remote Sensing.* Vol 13.
- [10] Descombes, X. **Champs markovien en analyse d'images.** Thèse de Doctorat, ENST. France. 1993.

- [11] Descombes, X. **Fission and Fusion Markovian approach for multi-channel segmentation.** *IGARSS'95.* Florence, Italy. 1995.
- [12] Descombes, X.; Moctezuma, M.; Maître, H. and Rudant, J.P. **Coastline Detection by a Markovian Segmentation on SAR Images.** *Signal Processing 55.* Netherlands. 1996.
- [13] Devroye, L.; Györfi, L. and Lugosi G. **A Probabilistic Theory of Pattern Recognition, Applications of Mathematics Stochastic Modelling and Applied Probability.** Springer-Verlay, 1996.
- [14] Fondeur, J.C. **Approche Markovienne Pour la Fusion de Données en Imagerie Satellitaire.** Département Image, Telecom. Paris, France. 1993.
- [15] Franco, J.A.; Moctezuma, M. y Parmiggiani, F. **Aproximación contextual a la segmentación de fotografía aérea.** *SIARP1999.* La Habana, Cuba. 1999.
- [16] Franco, J.A.; Moctezuma, M.; Parmiggiani, F. **Remote sensing images segmentation: A mixed approach adapted for fine structure detection.** *SIARP 2000.* Lisboa, Portugal. 2000.
- [17] Frost, V.S. *et al.* **A model for radar images and its applications to adaptive digital filtering of multiplicative noise.** *IEEE Transn. Pattern analysis and machine intelligence*, 4, 1982.
- [18] Gouinaud, D. **Traitment d'images satellitaires pour la detection d'agglomérations.** *Thèse de Doctorat.* ENST. France. 1996.
- [19] Haralick, R.M. **Statistical and structural approaches to texture.** *Proc. Of the IEEE.* Vol. 67. Número 5. Mayo, 1979.
- [20] Houzelle, S. y Giraudon, G. **Data fusion using SPOT and SAR images for bridge and urban area extraction.** *IGARSS'91.* Helsinki. Finlandia. 1991.
- [21] Kato, Z. **Modélisations markoviennes multirésolutions en vision par ordinateur. Application à la segmentation d'images SPOT.** *Thèse pour obtenir le titre de Docteur en Science.* L'Université de Nice Sophia Antipolis. France. 1994.
- [22] Leberl, F. **Radargrammetric image processing.** Artech House, Inc., 1990.
- [23] Li, S.Z. **Markov Random Field Modeling in Computer Vision.** Springer-Verlag. 1995.
- [24] Lira, J. **Introducción al tratamiento Digital de Imágenes.** Instituto de Geofísica, UNAM. México. 1995.
- [25] Lopes, A. *et al.* **Maximum a posteriori filtering and first order texture models in SAR images.** *IEEE IGARSS '90 proceedings.* Washington, D.C.

- [26] Lorette, A.; Descombes, X. et Zerubia, J. **Extraction des Zones Urbaines Fondée sur une Analyse de la Texture par Modélisation Markovienne.** *Rapport de Recherche.* INRIA. Francia. 1998.
- [27] Maitre, H. et Zinn-Justin, J., **Les Progrès du Traitement des Images.** Les Houches, Session LVIII. North-Holland, Elseiver. 1996.
- [28] Mendez, R. **Transformada Polinomial y Campos Aleatorios de Markov: una aplicación a la detección de líneas de costa en imágenes de radar de apertura sintética.** *Tesis para obtener el título de Ingeniero en Telecomunicaciones.* UNAM. México. 1997.
- [29] Tou, J. and González, R. **Pattern Recognition Principles.** Addison-Wesley. USA. 1981.
- [30] Vandeneede, J., Zhang, D., Wambacq, P. and Van Gool, L. **Supervised Classification of Remote Sensing Images With the Aid of Robust Statistics and the Gibbs Distribution.** ESAT-M12. Belgium. 1994.
- [31] Zhang, D.B., Van Gool L. and Oosterlinck, A. **Coastline Detection from SAR Images.** ESAT-M12. Belgium. 1994.
- [32] Yu, S. **Improving Satellite Image Analysis Quality by Data Fusion.** IGARSS'95. Florence, Italy. 1995.
- [33] Yu, S.; Berthod, M. and Giraudon, G. **Towards Robust Analysis of Satellite Images Using Map Information — Application to Urban Area Detection.** *Research Report.* INRIA. France. 1997.
- [34] **Agencia Espacial Europea**
<http://www.esrin.esa.it/export/esaCP/index.html>
- [35] **Bird's Eye Images**
<http://www.birdseyeimages.com/history.html>
- [36] **Centro de Percepción Remota y Sistemas de Información Geográfica**
<http://www.cprsig.puc.cl/>
- [37] **Instituto de Geografía UNAM**
<http://www.igeograf.unam.mx/>
- [38] **National Oceanic and Atmospheric Administration**
<http://www.noaa.gov/>
- [39] **Procesamiento de Imágenes de Radar**
http://sputnik.dpi.inpe.br/spring/usuario_spa/radar_in.htm
- [40] **Satélites SPOT**
http://ceos.cnes.fr:8100/cdrom98/ceos1/satellit/spotsys/english/ss_paylo.htm
- [41] **Space Imaging**
<http://www.spaceimaging.com/>

- [42] **SPOT**
<http://www.spot.com/>
- [43] **TELEDET**
<http://www.teledet.com.uy/>
- [44] **The French National Institute for Research in Computer Science and Control**
<http://www.inria.fr/index.en.html>

APÉNDICE A

En este apéndice se transcriben los artículos presentados en diferentes Simposia Internacionales, los cuales fueron aportes de esta tesis.

Los artículos que se presentan a continuación son los siguientes:

J. A. Franco, M. Moctezuma y F. Parmiggiani. "*Aproximación Contextual a la Segmentación de Fotografía Aérea*". SIARP '99. La Habana, Cuba. Marzo, 1999.

J. A. Franco, M. Moctezuma and F. Parmiggiani. "*Remote Sensing Images Segmentation: A mixed approach adapted for fine structure detection*". SIARP 2000. Lisboa, Portugal. Septiembre, 2000.

J. A. Franco, M. Moctezuma, M. E. Barilla, B. Escalante and F. Parmiggiani. "*SAR Images Filtering And Segmentation: A Multiresolution And Contextual Approach*". IEEE IGARSS 2001. Sydney, Australia. Julio, 2001.

Aproximación Contextual a la Segmentación de Fotografía Aérea

J. A. Franco¹, M. Moctezuma¹ y F. Parmiggiani²

¹ Universidad Nacional Autónoma de México

División de Postgrado, DEPMI-UNAM. México, D.F.

e-mail: josea2k@yahoo.com, miguelm@verona.fi-p.unam.mx

² ISAO-CNR. Via P. Gobetti 101-40129 Bologna, Italia.

Resumen

Basado en una formulación bayesiana, este artículo presenta un método de segmentación tendiente a eliminar efectos de sobresegmentación. La información requerida a priori es obtenida vía un clasificador clásico. Un algoritmo de relajación determinista es adaptado a fin de mejorar la calidad del resultado bayesiano. Para la aplicación de éstos se empleó una fotografía aérea pancromática de alta resolución de una región urbana de la Ciudad de México. Mediante el segmentador bayesiano se obtuvo una segmentación en cuatro categorías para su posterior procesamiento por un método de relajación determinista (algoritmo ICM). Los resultados obtenidos muestran una correcta funcionalidad de nuestro método, ya que con la fusión de *pixeles* aislados y de regiones de transición a regiones conexas se obtuvo una mejor estimación de contornos, así como regiones más homogéneas.

Introducción

La percepción remota de recursos de la tierra mediante imágenes de satélite y/o imágenes aéreas es un área de gran actividad en investigación. Específicamente, la adquisición de datos remotos en regiones urbanas puede facilitar tareas en la actualización de cartografía y planeación urbana mediante la segmentación y extracción automática de regiones de interés. Los métodos automáticos de procesamiento basados en funciones estadísticas presentan ventajas de consideración al incorporar conocimientos de teoría de la detección, de la decisión y, en general, de formulaciones estadísticas.

En este contexto, este artículo presenta un método de segmentación semiautomático aplicado a fotografía aérea pancromática de alta resolución. La escena de interés se localiza en la región urbana de la Ciudad de México. Nuestro algoritmo emplea una formulación de Bayes para el cálculo de probabilidades *a posteriori*. El resultado es una

imagen segmentada en diversas categorías. La siguiente etapa del proceso incorpora un algoritmo adaptivo de relajación determinista (vía un muestreador de Gibbs) a fin de obtener una mejor estimación de contornos.

Aproximación Estadística

El empleo de funciones estadísticas en procesos de segmentación/clasificación presenta ventajas interesantes al compararlas con métodos puramente supervisados; ejecuciones más automáticas y aplicables a un alto volumen de datos son algunas de ellas. En una formulación estadística se requiere de modelos matemáticos que establezcan ligas con los procesos del mundo real, que transforman a una matriz de *pixeles* en, por ejemplo, una imagen pancromática. Los modelos necesariamente deben incorporar información *a priori* sobre el proceso de formación de la imagen y de la imagen misma, tal y como se nos presenta.

Un algoritmo de segmentación debe dividir a una imagen en subregiones, cada una representando a objetos o regiones de interés. Dos problemas importantes surgen a partir de los requerimientos mencionados: la sobresegmentación (que produce demasiados segmentos) y la subsegmentación (que produce muy pocos segmentos) [3].

Ambos problemas se relacionan con el establecimiento del modelo de análisis y que bajo formulaciones estadísticas y estocásticas necesitan la incorporación de información (precisa) *a priori* y además, expresada por el menor número de parámetros posibles [2][4][5].

La intención del presente artículo es definir un algoritmo de segmentación estadístico funcional, basado en un número pequeño de parámetros y operando con información contextual.

Una función bayesiana de discriminación [2] incorpora la siguiente relación de probabilidades:

$$\rho(\omega_k | x) = \frac{\rho(x | \omega_k)\rho(\omega_k)}{\rho(x)} \quad \dots(1)$$

donde

$$\rho(x) = \sum_{k=0}^{M-1} \rho(x | \omega_k)\rho(\omega_k)$$

es la probabilidad total del evento/*pixel* x ,

$\rho(\omega_k)$ es la probabilidad *a priori* de ocurrencia de la clase ω_k ,

$\rho(x|\omega_k)$ es la probabilidad condicional del *pixel* x , dada la clase ω_k , y

M es el número de clases o regiones típicas.

En una aproximación clásica, el procedimiento de segmentación consiste en calcular los valores de las probabilidades *a posteriori* $\rho(\omega_k|x)$, asignando el *pixel* x a la clase ω_k que

maximiza $\rho(\omega_k|x)$. Ya que el denominador de la ecuación 1 es término común, puede ser eliminado.

La búsqueda de los términos máximos se realiza de acuerdo a:

$$\max_k \rho(\omega_k | x) = \rho(x | \omega_k) \rho(\omega_k) \quad \dots(2)$$

La función de verosimilitud de la clase o región ω_k con respecto al *pixel* x , $\rho(x|\omega_k)$ requiere de un modelo probabilístico de función de densidad. El cálculo de los términos probabilísticos requiere información tanto del proceso de segmentación como de los niveles de gris de la imagen original. Nuestra propuesta es emplear funciones $f_{x|\omega_k}$ similares a las utilizadas en teoría de lógica difusa. Ellas se esquematizan en la figura siguiente (figura 1):

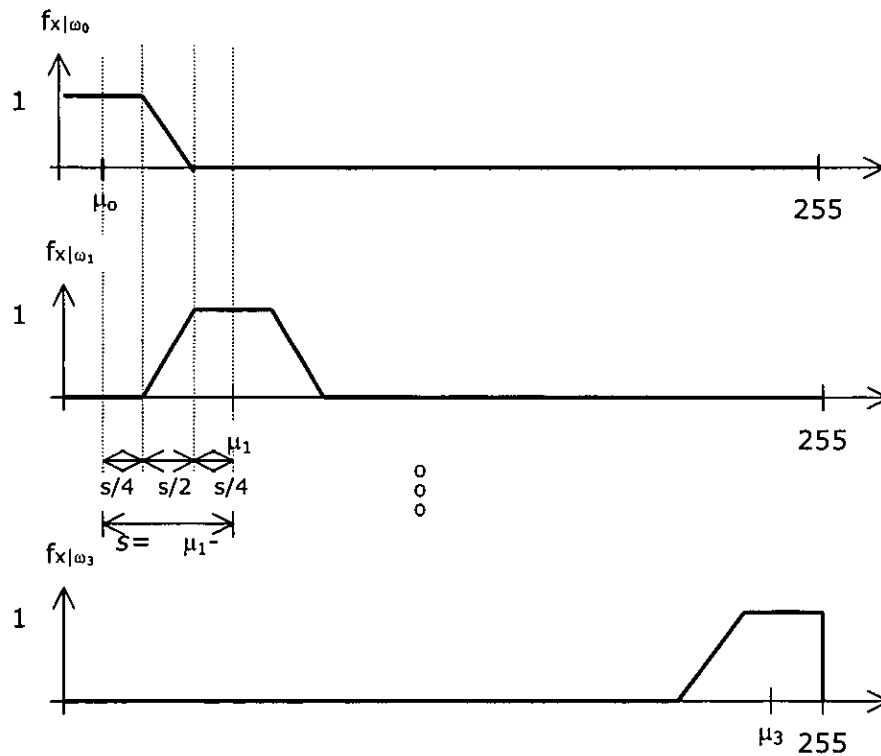
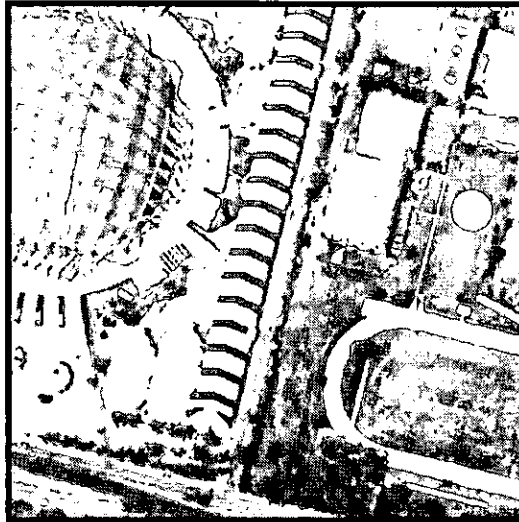


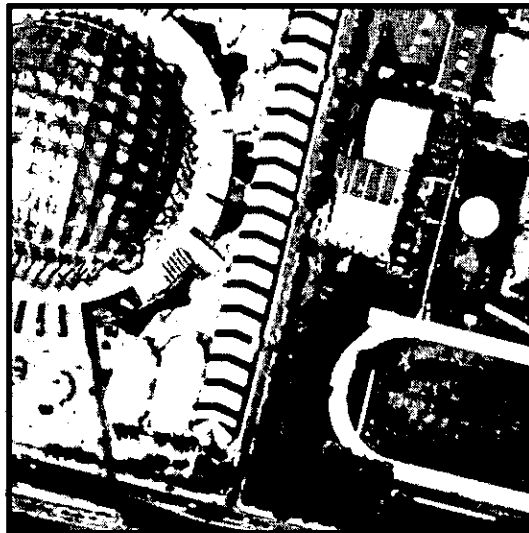
Figura 1

Funciones que aproximan a las funciones de probabilidad condicional $\rho(x | \omega_k)$.

Los parámetros $\rho(\omega_k)$ y μ_k pueden aproximarse por la aplicación de un algoritmo de segmentación, por ejemplo k-medias. La figura siguiente muestra una escena (foto aérea) de la Ciudad de México (figura 2), así como el resultado de la segmentación en 4 regiones (figura 3) empleando la ecuación 2.



a) Imagen original, con un tamaño de píxel de aproximadamente 20x20 cm.



c) Imagen segmentada.

Figura 2
Segmentación bayesiana.

Relajación Determinística

El resultado obtenido empleando el algoritmo de la sección anterior es bueno, en términos generales. Sin embargo, si el interés de la aplicación reside en la localización de regiones homogéneas del tipo: pasto, árboles, calles, banquetas, etc., los *pixeles* aislados y las regiones formadas por unos cuantos *pixeles* deben ser fusionados a alguna región conexas, cumpliendo criterios de optimización.

La siguiente etapa de nuestro esquema es, entonces, optimar el resultado previo mediante un algoritmo contextual. Para ello se empleó una relajación determinista. Los algoritmos de recocido simulado (*Simulated Annealing* - SA) permiten obtener (teóricamente) configuraciones globales mínimas, sin embargo, son computacionalmente costosos.

Para acelerar la convergencia, una alternativa es emplear algoritmos determinísticos. La esencia algorítmica de la relajación estocástica es que permite las transiciones de energía mediante incrementos y decrementos, mientras que en la relajación determinista solo las transiciones con decrementos son permitidas. Por tal razón, el algoritmo empleado, Modo Condicional Iterativo (*Iterated Conditional Mode* - ICM), debe partir de una configuración de segmentación "razonablemente" correcta [1].

Los pasos a seguir en el algoritmo ICM son los siguientes:

1. Comenzar en una "buena" configuración inicial ω^0 , con $k=0$, siendo k el número de iteración.
2. En cada iteración
 - a) escoger aleatoriamente un *pixel*
 - b) a partir de la configuración actual ω^k , determine posibles configuraciones de energía, $U(\eta)$, donde:

$$\eta \in N_{\omega^k}$$
 - c) seleccionar la configuración de energía mínima

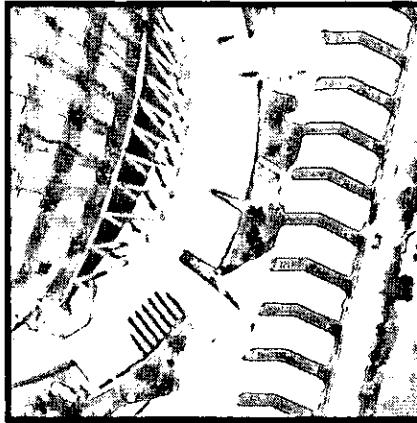
$$\omega^{k+1} = \arg \min_{\eta \in N_{\omega^k}} U(\eta)$$
 - d) regresar a 2a) hasta barrer toda la imagen
3. Repetir el punto 2, con $k=k+1$, hasta alcanzar el número fijado de iteraciones.

En nuestra propuesta las configuraciones de energía fueron calculadas empleando vecindades en 8-conexidad: $U_i(\eta) = -\sum \beta \delta_{ij}$, donde:

- i es la clase de *pixel* analizado,
- j es la clase de un *pixel* vecino y
- δ es la función impulso unitario.

En el cálculo de $U(\eta)$ debe considerarse únicamente a las clases de los *pixeles* contenidos en la vecindad de análisis.

En la figura 3 se muestra el resultado final de nuestro algoritmo. Se muestran tres ventanas: a) de la imagen original, b) de la imagen segmentada en 4 regiones mediante el criterio de Bayes, y c) el resultado del algoritmo ICM.



a) Imagen original.



b) Resultado empleando el esquema bayesiano.



c) Resultado de la relajación determinista.

Figura 3
Resultado final.

Conclusiones

El esquema presentado incorpora una relajación determinista a un segmentador bayesiano, a fin de eliminar efectos de sobresegmentación. En la figura 3c) se observa la fusión de los *pixeles* aislados y de las regiones de transición de la figura 3b) a regiones conexas. El resultado obtenido muestra regiones homogéneas con una detección precisa de contornos.

Por tratarse de material pancromático, una desventaja importante es la confusión de clases y la mezcla de los modos de la función de densidad de probabilidad. Actualmente se estudia un esquema de fusión, que incorpora datos de diversos sensores remotos a fin de determinar adecuadamente los elementos de escena.

Referencias

- [1]. Besag, J., "On the Statistical Analysis of Dirty Pictures", *Jl. Roy. Statis. Soc. B*, 1986.
- [2]. Bow, Sing-Tze, "Pattern Recognition and Image Processing", Marcel Dekker, Inc., 1992.
- [3]. Delves, L. M., Wilkinsons, R., Oliver, C. J. and White, R. G., "Comparing the Performance of SAR Image Segmentation Algorithms", *Int. J. Remote Sensing*, vol 13, Noll, pp. 2121-2149, 1992.
- [4]. Devroye, L., Györfi, L. and Lugosi G., "A Probabilistic Theory of Pattern Recognition, Applications of Mathematics Stochastic Modelling and Applied Probability", Springer-Verlay, 1996.
- [5]. Maitre, H. et Zinn-Justin, J., "Les Progrès du Traitement des Images. Les Houches 1992", Session LVIII, North-Holland, Elseiver, 1996.

Remote Sensing Images Segmentation: A Mixed Approach Adapted for Fine Structure Detection

J. A. Franco¹, M. Moctezuma¹ and F. Parmiggiani²

¹ National University of Mexico – Graduate Division, DEPMI-UNAM, Mexico

Mail: josea2k@yahoo.com , miguelm@verona.fi-p.unam.mx

² ISAO-CNR, Bologna, Italy.

Abstract

Remote sensing images segmentation can be approached in several ways: from the simplest and intuitive concept of minimum-distance classifiers to wide theory supported algorithms as Markov Random Fields and Gibbs' distribution based classifiers. The former approach and –generally speaking– non-contextual classifiers have an inherent drawback: they assume that spectral properties of each *pixel* are independent of its neighborhood, which may result in segmented images showing many isolated *pixels*. As the latter approach does take into account the neighborhood of each *pixel*, it generates labeled images showing homogeneous regions; nevertheless, a disadvantage of these algorithms is the eventual elimination of fine well-defined structures of the scene by fusing them to adjacent regions.

In this paper, we propose a mixed segmentation method focused on avoiding over-segmentation while preserving fine structures. Information of cluster centers is obtained using a non-contextual algorithm, then by means of a Bayes classifier a reject criterion is defined creating a 'reject class'. Finally, our proposal includes a Markovian model in order to reclass rejected *pixels*. Visual and numerical comparisons of our results with those generated by direct application of a classic Markovian classifier or by deterministic relaxation –among other methods– are encouraging.

Introduction

Scientific research in earth observation from remote sensing images allows an effective monitoring and management of its resources. In this domain image segmentation is an useful tool for obtaining thematic information, which allows easy actualization on cartography and urban distribution.

Several segmentation algorithms have been developed, ranging from the simplest concept of minimum-distance to classifiers supported by widely studied theories as

Markov Random Fields. Obviously these algorithms perform differently, showing each of them characteristic advantages and drawbacks.

Segmentation algorithms can be used in a mixed way to obtain better results. With this aim in mind we have applied in [5] the following scheme: cluster centers were defined using a simple non-contextual algorithm (K-means), then we approached the probability functions, based on Bayes' theorem, and classed the image via a maximum-likelihood classifier. Finally, we took this previous result as the initial configuration for Iterated Conditional Modes algorithm (ICM). Results obtained show homogeneous regions, but fine structures were fused to adjacent regions. Similar results may be obtained by contextual approaches [3,9].

Here, we propose another scheme focused on the preservation of fine structures but trying to avoid over-segmentation effects. Steps followed in this new scheme are briefly described below:

- Obtainment of cluster centers through a non-contextual algorithm.
- Approximation of probability functions based on both the standard deviation and the mean of each class.
- Image segmentation by means of a maximum-likelihood algorithm, but defining a reject criterion. Rejected *pixels* are labeled as 'undefined-class'.
- Finally, undefined-class *pixels* are reclassified by a Markov Random Fields (MRF) classifier defined in a non-homogeneous external field.

Results obtained with this approach show homogeneous regions and respect of representative fine structures, as expected.

Non-contextual segmentation

This approach to image segmentation does not take into account the neighborhood information of the *pixel* being classed, which may lead to over-segmentation problems. Two classical schemes used to perform non-contextual segmentation are the ones founded on the basic concepts of minimum-distance and maximum-likelihood.

Minimum-distance algorithms – The concept of minimum distance is obtained intuitively from the fact that similarity between elements in a given space can be expressed in terms of their proximity to each of cluster centers. These classifiers are specially effective when pattern classes show a 'reasonably small' degree of variability, that is to say, when the elements of each class cluster tightly about its mean [7].

Minimum-distance algorithms are very simple and offer fast executions, but as they only consider the properties of the *pixel* being analyzed, images so classed may show many isolated *pixels* resulting in a snow-like segmentation.

Maximum-likelihood algorithms – They consider that spectral properties of *pixels* belonging to a class follow a probability law, consequently each class may be described by means of a probability function. Briefly, it signifies that these functions resemble the real distribution of each class, which allows us to determine the probability of appartenance to a given class for each one of *pixels* composing the scene.

Computation is performed for all the categories considered in the segmentation process and *pixels* are assigned to the one maximizing the probability function [2].

Maximum-likelihood algorithms generally are more complex and require a greater amount of calculation than distance classifiers and, although they offer better results, over-segmentation effects are not completely avoided.

Contextual segmentation

Contextual classifiers commonly require a major amount of processing time than non-contextual methods and are applied in an iterative way in order to minimize a non-convex energy function. Two well-known contextual classifiers are the ICM (a deterministic relaxation algorithm) and the Markovian model.

Iterated Conditional Modes (ICM) – Proposed by Besag [1], it is the most popular deterministic algorithm and, as it does not include a temperature parameter, it is not annealing. ICM algorithm requires a 'good' initial configuration (the quality of the final result strongly depends on it) and it guarantees a fast convergence –in approximately 10 iterations– towards the local minimum.

It is not known how to ensure a 'good' initial configuration but, for image segmentation, the maximum-likelihood classifier (MLC) is usually adopted as the initialization method for the ICM algorithm.

Markovian model – Using MRF and Maximum A Posteriori (MAP) estimation, minimization of energy functions can be reached by Simulated Annealing (SA). SA algorithms reach (theoretically) a global minimum, but they require a great amount of computation and hence they are very time-consuming, which is a great disadvantage. Simulated annealing is controlled by a parameter called 'temperature' (T), which decreases sequentially. At high T , large energy increases may be allowed; as T decreases, only small energy increases may be accepted; and around the freezing T , no energy increases are admitted. This criterion allows the algorithm to avoid local minima and to reach the global minimum [6].

As the process to reach the global optimum is too slow –temperature decreases minimally after each iteration–, approximations of the SA algorithm are usually implemented and, as a consequence, the convergence towards the global optimum is no longer guaranteed. The regularization process usually adopts the following schedule for decreasing T : $T^{(t)} = kT^{(t-1)}$, where k takes values between 0.8 and 0.99, allowing a faster convergence.

Images classed via contextual segmentation show homogeneous regions and isolated *pixels* are hardly encountered but, under this scheme, fine well-defined structures of the scene are eventually eliminated by merging them to contiguous regions.

A mixed approach to image segmentation

We have briefly described above the main advantages and drawbacks of some of the principal contextual and non-contextual classifiers. Now our aim is to define a mixed algorithm focused on the obtainment of labeled images showing homogeneous regions but preserving fine well-defined structures.

When we use the term 'fine well-defined structures' we refer to small structures whose spectral properties are really close to any of cluster centers, which means that they are representative of the class they belong to, hence they must be preserved.

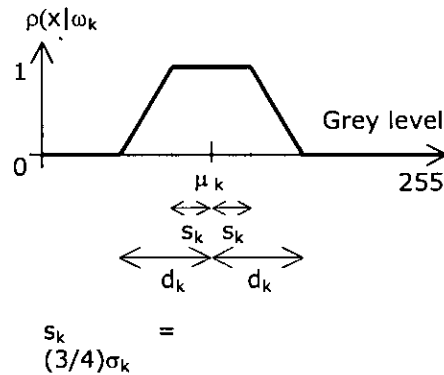


Figure 1

Model used to approximate the conditional probability function of each class.

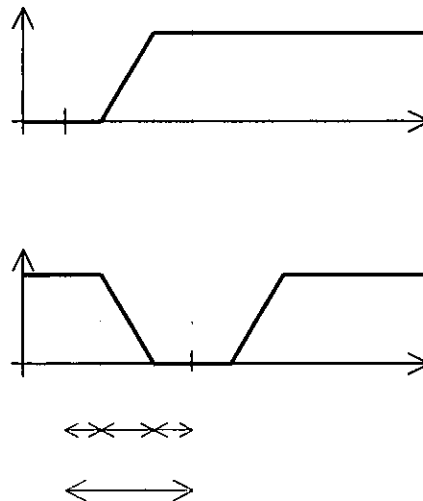


Figure 2

The energy functions of the imaging process (statistics of the data) for the MRF classifier.

The algorithm we propose is easy to implement and it is composed of four primary steps which are described below.

- 1) Obtainment of cluster centers through a non-contextual algorithm – In this case we decided to employ the widely known K-means algorithm, which offers fast executions and does not require a training stage, to obtain a *K-class* image.
- 2) Approximation of the probability functions based on the standard deviation of each class – The standard deviation (a measure of the dispersion or variation in a distribution) was used to create the conditional probability functions describing the normal distribution of each class. These functions were produced following the piecewise linear function shown in figure 1.
- 3) Image segmentation by means of a maximum-likelihood algorithm, but defining a reject criterion – Once we define the probability functions for each class, the next step was to perform a new segmentation based on the concept of maximum-likelihood. Here we define a parameter called the *reject threshold*, in such a way that *pixels* whose maximum probability of appartenance to a given class is smaller than this threshold are assigned to a new class *K+1* (reject or non-defined class), otherwise they are definitely assigned to the class maximizing its probability function.
- 4) Reassignment of *pixels* labeled as non-defined class using a MRF classifier – The last step was to perform a reclassification, but affecting only the *pixels* assigned to the reject class. For this step we decided to employ a Markovian Potts model, considering a second-order neighborhood system (η_2) –which considers the eight nearest neighbors of the central *pixel*. Piecewise linear functions defining the statistics of the data were defined following the scheme shown in figure 2. A simulated annealing process is applied in order to optimize the energy function. The algorithm is described next.

I.- Initialization

Random initialization of the solution $Y^{(0)}$, using a generator of uniformly distributed random numbers belonging to the finite set of classes $\Omega = \{0, 1, \dots, K-1\}$.
 Selection of a high enough initial temperature $T^{(0)}$ ($T^{(0)} = 2$ in this model).
 $\beta[\text{actual}] = \beta[\text{election}] = 0.3$

II.- Classification

For $n = 0, n = 150$

In each stage (n):

For each site $r(i,j)$:

Its value (actual_class) is $Y_r^{(n)}$ and its neighborhood configuration is $\eta_r^{(n)}$.

Draw a uniformly distributed random variable (election_class) belonging to Ω .

II.1.- Analysis in 8-conexity

prob[actual] = 0

prob[election] = 0

For each neighbor (labeled as class x) of r :

If actual_class = x :

prob[actual] = prob[actual] - $\beta[\text{actual}]$

End if

If election_class = x :

prob[election] = prob[election] - $\beta[\text{election}]$

End if

End For

II.2.- Local energy variation associated to site r

prob[actual] = prob[actual] + $U[\text{actual}]$

prob[election] = prob[election] + $U[\text{election}]$

If prob[election] \leq prob[actual]:

$r(i,j) = \text{election_class}$

Else:

$p = \exp -[(\text{prob}[\text{election}] - \text{prob}[\text{actual}])/T^{(n)}]$

$\xi_2 =$ random variable chosen from a uniform distribution on the interval $[0,1]$.

If $\xi_2 < p$:

$r(i,j) = \text{election_class}$

End if

End if

End for

$T^{(n+1)} = 0.95 * T^{(n)}$

End for

Results

To evaluate our algorithm, we applied it to an aerial photograph (in 256 grey levels) showing an urban scene of Mexico City valley (figure 3). In this image, transition areas –that must be rejected in step 3– are easily identified, as well as different fine structures –which must be preserved.

For the segmentation process we defined four typical classes, based on the spectral properties of representative structures. K-means algorithm showed a fast convergence and hence, a minimal time requirement. The value of cluster centers and their respective standard deviations are registered in table 1.

	μ_0	μ_1	μ_2	μ_3
Grey level	66	107	164	210
Standard deviation	11	14	15	14

Table 1

Cluster centers and their standard deviations.

Based on these data we approximated the conditional probability function for each class and performed the reclassification of the image via a MLC (bayesian classifier), *pixels* whose maximum probability was minor than 0.5 (the *reject threshold*) were labeled as class 5. These *pixels* are shown –in a light tone– in figure 4. As it can be observed, all transition zones were assigned to the reject class, and clearly differentiated fine structures were not altered. It is also important to point out that the reject class does not only lodge isolated *pixels* but also homogeneous groups of a few hundred *pixels*, so even relatively large zones may be reassigned depending on their surroundings. *Pixels* rejected represent the 20.15% of the image.

The definitive 4-class image is displayed in figure 5. It is evident that transition zones became homogeneous while fine structures were almost completely preserved. An special observation in this stage is the low-time requirement for image processing, which was of only 21% of the time required for a classical Markovian approach and simulated annealing minimization, as only rejected *pixels* were considered in this step.

A series of magnified windows are used to evidence the performance of our proposal as well as the differences between this result and those obtained via two widely known algorithms.

In the original subimage (figure 6) we can easily identify an important transition zone at the right-bottom corner. In this window it is also of special importance the region placed at the center, where the concept of fine well-define structure is clearly exemplified. These structures were almost completely preserved (only some *pixels* on the edges were rejected) as seen in figure 7.

Figures 8 to 10 make evident the differences between labeled images obtained through the proposed scheme, a classic Bayes classifier and a classic Markovian model.

The classic Bayesian classifier (figure 9) offers fast executions and preserves fine structures, but it allows the existence of many isolated *pixels*; consequently, transition zones show a snow-like appearance and therefore, edges are not clearly defined.

A classic Markovian classifier provides labeled images showing homogeneous regions, but it requires a great amount of calculation for reaching the global minimum, and fine structures may be eventually merged to adjacent regions. Figure 10 shows how fine structures are dramatically reduced.

On the contrary, the definitive 4-class image obtained with our proposal –shown in figures 5 and 8– exhibits homogeneous regions, clearly defined edges, and fine

structures are almost completely preserved. Although this proposal may also allow the presence of isolated *pixels* –but in a much lower rate than the Bayesian classifier and K-means– it shows a very good performance, overcoming the results provided by the classical approaches previously analyzed.

As a last comment, in table 2 we can realize how the ICM algorithm causes considerable changes in label assignments: the number of *pixels* assigned to class 2 – the one fine structures at the center of the window belong to– decreases 4% (around 10,500 *pixels*) with reference to our proposal, as a consequence of an almost absolute elimination of these structures.



Figure 3

*Original image
[512x512x256].*



Figure 4

Result of the maximum-likelihood algorithm with a reject threshold (after step 3). Rejected pixels are shown in white.



Figure 5

Result of the mixed segmentation algorithm. Image is segmented in 4 classes. Rejected pixels were reclassified using a MRF model.

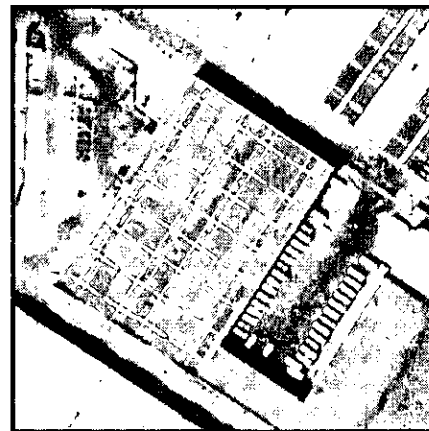


Figure 6

*Original subimage
[256x256x256].*



Figure 7

Subimage with the result of the maximum-likelihood algorithm with a reject threshold. The reject class is represented for the clearest pixels.



Figure 8

Subimage showing the definitive 4-class segmentation using the proposed algorithm.



Figure 9

A classic bayesian segmentation result.



Figure 10

A classic Markovian classifier result.

	ω_0 [%]	ω_1 [%]	ω_2 [%]	ω_3 [%]
K-means	32.62	23.83	20.33	23.22
Classic Bayes	34.05	22.72	19.46	23.77
ICM	37.94	21.75	16.74	23.56
Classic MRF	33.60	23.19	20.01	23.20
Our proposal	31.83	24.84	20.75	22.98

Table 2. *Percentage of pixels belonging to each class, according to the employed segmentation algorithm.*

Comments and concluding remarks

The proposed approach –as clearly exemplified by the labeled images presented–overcomes results supplied by the direct implementation of the methods it is based on (K-means, Bayes, MRF model).

Results provided by the proposed segmentation scheme, evidence its good performance in generating labeled images showing homogeneous regions, clearly defined edges, respect of fine structures, as well as a minimal presence of isolated *pixels*. In addition, it is a simple easy-to-implement algorithm offering fast executions and variable parameters.

Initial information of cluster centers is obtained via a non-contextual algorithm, then a bayesian classifier is adapted for generating an additional class (reject or non-defined class) finally, rejected *pixels* are reclassified through a MRF classifier.

The presence of isolated *pixels* may be avoided by preprocessing the original image using a filtering algorithm, improving the final result.

References

- [1] Besag, J., "On the Statistical Analysis of Dirty Pictures", Journal of the Royal Statistical Society, Series B 48 (1986), pp. 259-302.
- [2] Chuvieco, E., "Fundamentos de Teledetección Espacial", Madrid: RIALP, 1996.
- [3] Descombes, X., Moctezuma, M., Maître, H. and Rudant, J.P., "Coastline Detection by a Markovian Segmentation on SAR Images", Signal Processing 55 (1996), Netherlands. pp. 123-132.
- [4] Fondeur, J.C., "Approche Markovienne Pour la Fusion de Données en Imagerie Satellitaire", Département Image, Télécom Paris, 1993.
- [5] Franco, J.A., Moctezuma, M. and Parmiggiani, F., "Aproximación Contextual a la Segmentación de Fotografía Aérea", SIARP '99, La Habana, pp. 583-590.
- [6] Li, S.Z., "Markov Random Field Modeling in Computer Vision", Springer-Verlag, 1995.
- [7] Tou, J. and González, R., "Pattern Recognition Principles", USA: Addison-Wesley, 1981.
- [8] Vandeneede, J., Zhang, D., Wambacq, P. and Van Gool, L., "Supervised Classification of Remote Sensing Images With the Aid of Robust Statistics and the Gibbs Distribution", ESAT-M12, Belgium, 1994.
- [9] Zhang, D.B., Van Gool L. and Oosterlinck, A., "Coastline Detection from SAR Images", ESAT-M12, Belgium, 1994.

SAR Images Filtering and Segmentation: A Multiresolution and Contextual Approach

J. A. Franco¹, M. Moctezuma¹, M. E. Barilla¹, B. Escalante¹ and F. Parmiggiani²

¹ National University of Mexico – Graduate Division, DEPMI-UNAM, Mexico

² ISAO-CNR, Bologna, Italy.

Abstract

In this work we present a mixed contextual algorithm for segmenting SAR-ERS1 images (© ESA). The first step was to pre-process the original SAR image by means of a polynomial transform based filter in order to decrease the effect of speckle. Segmentation stage was performed as follows: cluster centers were obtained using a non-contextual algorithm; then, based on a Bayes classifier we achieved a first segmentation step, but defining a 'reject class'; finally, rejected *pixels* were reclassified via a Markovian model.

Results obtained show segmented images exhibiting homogeneous regions and a minimal presence of isolated *pixels*. As well, they evidence that combined use of polynomial transform and Markov Random Field theory does not introduce noticeable degradation of edges in segmented regions.

Introduction

Synthetic Aperture Radar (SAR) images are of special interest for the obtainment of thematic information over extensive regions. SAR images have several well-known advantages [1], however the presence of speckle and, generally speaking, the poor quality of SAR images represent a serious drawback when applying classical segmentation algorithms: over-segmented images showing excessive isolated *pixels* are obtained.

It is evident that a preprocessing stage must be included in order to take advantage of the high spatial resolution of SAR images. In this context a number of filters have been proposed [2][3]. However, in most cases speckle filtering algorithms smooth speckle as well as relevant information.

In this work we make use of an adaptive noise-reduction algorithm based on polynomial transforms, which provides smoothed SAR images but retaining subtle details.

Once the filtering stage is performed, next step is to define a segmentation algorithm intended to minimize the presence of isolated *pixels* while preserving fine structures, i.e. small structures whose spectral properties are really close to any of cluster centers, which means they are representative of the class they belong to, hence they must be preserved. We achieved this task by means of a combined segmentation scheme based on both, a classical Bayesian model and a Markov Random Fields (MRF) classifier defined in a non-homogeneous external field [4].

Adaptive Noise-Reduction by Polynomial Transforms

In order to analyze an image on a local basis, the image is multiplied by a window function. This windowing takes place at several positions over the entire input image comprising a sampling lattice S . Within every window, the image is described by a weighted sum of polynomials, which are orthogonal with respect to the window function. The mapping from the input image to the weights of polynomials, henceforth referred to as the polynomial coefficients, is called a forward polynomial transform. By interpolating the polynomial coefficients with specific synthesis filters the original image can be recovered. This process is called an inverse polynomial transform.

Polynomial coefficients $L_{m,n-m}(p,q)$ belonging to polynomial $G_{m,n-m}(x-p,y-q)$ for all positions (p,q) in S of the window function $V(x-p,y-q)$, are derived by the convolution of the input image $L(x,y)$ with a filter

$$D_{m,n-m}(x,y) = G_{m,n-m}(-x,-y)V^2(-x,-y) \quad (1)$$

and selecting the outputs

$$L_{m,n-m}(p,q) = \int_{-\infty-\infty}^{+\infty+\infty} \int L(x,y)D_{m,n-m}(p-x,q-y)dx dy \quad (2)$$

at positions (p,q) in the sampling lattice S , for $m = 0, \dots, n$ and $n = 0, \dots, N$. The polynomial $G_{m,n-m}(x,y)$ has a degree m in x and $n-m$ in y . The maximum order of the polynomial expansion is denoted by N .

The resynthesized image $L'(x,y)$ is obtained by the interpolation process of the inverse transform, i.e.

$$L'(x,y) = \sum_{n=0}^N \sum_{m=0}^n \sum_{(p,q) \in S} L_{m,n-m}(p,q)P_{m,n-m}(x-p,y-q) \quad (3)$$

Interpolation functions are defined by

$$P_{m,n-m}(x,y) = G_{m,n-m}(x,y)V(x,y)/W(x,y) \quad (4)$$

for $m = 0, \dots, n$ and $n = 0, \dots, N$, where

$$W(x, y) = \sum_{(p, q) \in S} V(x - p, y - q) \quad (5)$$

is called the weighting function. The only condition for the existence of the polynomial transform is that this weighting function must be different from zero for all coordinates (x, y) .

The aim of the adaptive noise-reduction algorithm is to decrease the presence of multiplicative noise on SAR images while retaining subtle details. The algorithm locates edges of the image and builds a binary image called "energy mask", which is intended to determine those parts of the image that must be considered as noise, and therefore eliminated. In the binary energy mask, one values represent the regions of the image to be reconstructed at full resolution, while zero values represent those regions to be reconstructed at the lowest resolution.

Discrimination between edges and noise is performed by means of a threshold technique. In order to determine the optimal threshold, we take into account two basic characteristics of SAR images:

- One-look amplitude SAR images have a Rayleigh distribution and the signal-noise relation (SNR) is about 1.9131.
- In general, the SNR of multi-look SAR images does not change over the whole image.

As the threshold depends on the local mean of the image, the algorithm applies a higher threshold on bright regions than on dark regions of the image [5].

Contextual Segmentation

Image segmentation may be approached in several ways: from the basic concept of minimum-distance to wide theory supported algorithms as MRF based classifiers. These algorithms show characteristic advantages and drawbacks, and may be used in a mixed way to obtain better results.

The segmentation scheme we implement in this work is a blend of three well known segmentation algorithms based on the following concepts:

- a) Minimum-distance - This concept is obtained intuitively from the fact that similarity between elements in a given space can be expressed in terms of their proximity to each of cluster centers.
- b) Maximum-likelihood - Considers that spectral properties of *pixels* belonging to a class follow a probability law, i.e. each class may be described by means of a probability function. These functions resemble the real distribution of each class, allowing to determine the probability of appartenance to a given class for each one of the *pixels* composing the scene.

Markovian model – Using MRF theory and Maximum a posteriori (MAP) estimation, minimization of energy functions can be reached by Simulated Annealing (SA). Simulated annealing is controlled by a parameter called 'temperature' (T), which decreases sequentially. At high T , large energy increases may be allowed; as T decreases, only small energy increases may be accepted; and around the freezing T , no energy increases are admitted. This criterion allows the algorithm to avoid local minima and to reach the global minimum.

The algorithm we propose is easy to implement and it is composed of four primary steps which are described below.

- 1) Obtainment of cluster centers through a non-contextual algorithm – We employed the widely known K-means algorithm, which offers fast executions and does not require a training stage, to obtain K cluster centers.
- 2) Approximation of the probability functions based on the standard deviation and the mean of each class – The standard deviation (a measure of the dispersion or variation in a distribution) and the mean of each class were used to create the conditional probability functions $p(x|w_k)$ describing the normal distribution for each class w_k , where x is the gray level of the original image.
- 3) Image segmentation by means of a maximum-likelihood algorithm, but defining a reject criterion – A first segmentation step based on the concept of maximum-likelihood was performed. We define a parameter called the reject threshold, in such a way that *pixels* whose maximum probability of appartenance to a given class is smaller than this threshold are assigned to a new class $K+1$ (reject or non-defined class), otherwise they are definitely assigned to the class maximizing its probability function.
- 4) Reassignment of rejected *pixels* by means of a MRF classifier – Finally, we perform a reclassification stage, but affecting only *pixels* assigned to the reject class. For this step the energy function is defined by a Markovian Potts model and piecewise linear functions defining the statistics of data [6]. Potts model considers a second-order neighborhood system, i.e. the eight nearest neighbors to the central *pixel*. A simulated annealing process was applied in order to optimize the energy function.

Results

To evaluate our algorithm, we applied it to a SAR-ERS1 image (in 256 gray levels) showing an urban scene of Mexico City valley (Fig. 1).

The result of the adaptive noise-reduction algorithm is shown in Fig. 2. It is clear how brighter regions of the image were smoothed in a higher degree than dark regions, according to the way the threshold is defined.

For the segmentation step we defined four typical classes. Fig. 3. shows the reject class, which not only lodges isolated *pixels*, but also homogeneous groups of a few

hundred *pixels*, so even relatively large zones may be reassigned depending on their surroundings. *Pixels* rejected represent the 34.19% of the image.

The definitive 4-class image is displayed in Fig. 4. An special observation in this stage is the low-time requirement for image processing, which was of only 35% of the time required for a classical Markovian approach and simulated annealing minimization, as only rejected *pixels* were considered in this step.

Conclusions

Results obtained with this scheme show segmented SAR-ERS1 images exhibiting homogeneous regions and a low incidence of isolated *pixels*. As well, results show that the combined use of polynomial transform and Markov Random Field theory does not introduce noticeable degradation of edges in segmented regions.



Figure 1

Original image in 256 gray levels.



Figure 2

Result of the adaptive noise-reduction algorithm.

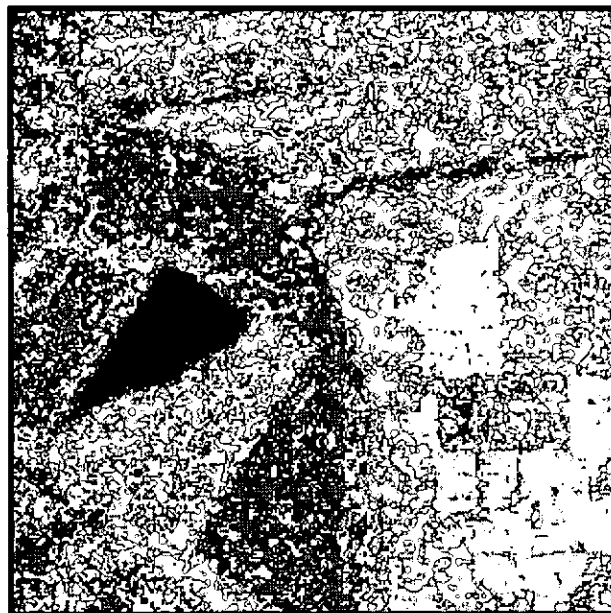


Figure 3

*Result of the maximum-likelihood algorithm
with a reject threshold (after step 3).
Rejected pixels are shown in white.*

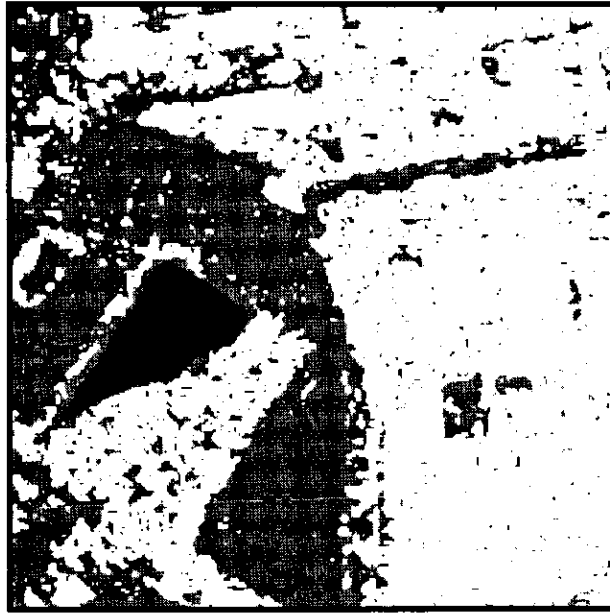


Figure 4

*Result of the mixed segmentation algorithm.
Image is segmented in 4 classes.
Rejected pixels were reclassified using a MRF model.*

References

- [1]. F. Leberl, "Radargrammetric image processing", Artech House, Inc., 1990.
- [2]. V. S. Frost, J. A. Stiles, K. S. Shanmugan and J. C. Holtzman, "A model for radar images and its applications to adaptive digital filtering of multiplicative noise", IEEE Trans. Pattern analysis and machine intelligence, 4, 1982, pp. 157-166.
- [3]. A. Lopes, E. Nezry, R. Touzi and H. Laur, "Maximun a posteriori filtering and first order texture models in SAR images", IEEE IGARSS '90 proceedings, Washington, D.C. pp. 2409-2412.
- [4]. S. Z. Li, "Markov random field modeling in computer vision", Springer-Verlag, 1995.
- [5]. P. Camarillo, A. Varela and B. Escalante, "Adaptive multiplicative-noise reduction in SAR images with polynomial transforms", IEEE, 1998, pp. 1171-1173.
- [6]. J. A. Franco, M. Moctezuma and F. Parmiggiani, "Remote sensing images segmentation: a mixed approach adapted for fine structure detection", SIARP 2000 proceedings, Lisbon. pp. 393-404.
- [7]. J. Vandeneede, D. Zhang, P. Wambacq and L. Van Gool, "Supervised classification of remote sensing images with the aid of robust statistics and the Gibbs distribution", ESAT-M12, Belgium, 1994.

APÉNDICE B

Características de Cámaras Fotográficas

Cámaras Analógicas	Detector	Emulsión fotosensible compuesto de sales de plata	
	Rango De Sensibilidad	Ortocrómico	0,35 a 0,6 μm (Azul - Verde)
		Pancromático	0,35 a 0,7 μm (Azul - Rojo)
		Infrarrojo B/N	0,7 a 0,85 μm (Infrarrojo Cercano)
		Color	0,4 a 0,5 μm (Azul); 0,5 a 0,6 μm (Verde) y de 0,6 a 0,7 μm (rojo)
Infrarrojo	0,5 a 0,6 μm (Verde); de 0,6 a 0,7 μm (rojo) y de 0,7 a 0,8 μm (IrC)		
Cámaras Digitales	Detector	Fotocelda	CCD, Couple Charge Device
	Rango De Sensibilidad	Del visible al infrarrojo cercano, en anchos de bandas que varían entre 0.012 a 0.3 μm .	
	Tamaño de la Focelda	9 x 9 μm	
	Array	De 512 x 512 a 2048 x 2048 foceldas.	

Características de Sistemas de Percepción Remota Espacial

Satélite	Operador	Lanzamiento	Óptico o Radar	Pancromático o Multiespectral	Resolución
SPOT 1/2/3	CNES/SPOT	1986/90/93	Óptico	Pancromático Multiespectral	10 20
SPOT 4	CNES/SPOT	1998	Óptico	Pancromático Multiespectral Multiespectral	10 20 100
Landsat 5	Space Imaging	1984	Óptico	Multiespectral Multiespectral	30 80
Landsat 7	U.S. Government	1999	Óptico	Pancromático Multiespectral	15 30
RADARSAT	Canadian Space Agency	1995	Radar	N/A	8-100
ERS-1/2	European Space Agency	91/94	Radar	N/A	30-50
JERS-1	Japan	1992	Radar	N/A	18
IKONOS	Space Imaging	1999	Óptico	Pancromático Multiespectral	1 4
QuickBird	EarthWatch	1999	Óptico	Pancromático Multiespectral	0.82 3.28
OrbView 2	ORBIMAGE	1997	Óptico	Multiespectral	1.1 Km
OrbView 3	ORBIMAGE	1999	Óptico	Pancromático Multiespectral	1 4