

01170

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE INGENIERIA

DIVISION DE ESTUDIOS DE POSGRADO

TESIS

**METODO DE ESTRUCTURAS DE ARBOL (QUADTREE) PARA
LA SEGMENTACION DE IMAGENES DE PERCEPCION
REMOTA.**

PRESENTADA POR:

OLIMPIA ARELLANO NERI

PARA OBTENER EL GRADO DE:

**MAESTRO EN INGENIERIA
(ELECTRICA)**

DIRIGIDA POR:

DR. MIGUEL MOCTEZUMA FLORES

TESIS CON Cd. Universitaria, Noviembre de 1997.
FALLA DE ORIGEN



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

DEDICO ESTE TRABAJO A:

- A mis padres Antonio y Elena:

Agrediendo así su ayuda y comprensión, y el haberme permitido realizarlo lejos de ellos, aunque a pesar de la distancia siempre sentí el calor de sus manos junto a mí, ofreciéndome el amor y el cariño de siempre en los momentos tristes y alegres.

- A Araceli y Luis:

Para corresponder a la confianza depositada y para complementar la alegría de lo que recientemente han recibido de las manos de Dios. La presencia de Brandon y Brenda ha sido una bendición en su hogar y en la familia, y es la prueba de que la gente buena siempre está en gracia con Dios y ustedes han sido muy buenos con nosotros. A mi manera esta es una pequeña forma de darles las gracias. Todos mis logros siempre serán una forma de decirles "gracias Ara y Luis". Gracias por acompañar mi alegría con sus sonrisas y abrazos, y gracias por acompañar mi tristeza con sus lágrimas y con sus palabras de aliento y esperanza.

AGRADECIMIENTOS

- Dios mio, te doy gracias por haberme creado, hecho cristiano y conservado durante estos días. Perdóname el mal que he cometido y acepta el bien que haya podido hacer.

***- Gracias a mis hermanas:
Sunny, Rosalía, Gaby, Araceli y Adela,***

***y a mis hermanos:
Antonio, Raúl, Armando, Arturo y Luis Moisés,***

por haberme apoyado en mis decisiones, no sin antes ofrecerme un sincero y buen consejo. Les agradezco el haberme permitido tener el cariño de todos y cada uno de sus hijos.

- Gracias a mi siempre amigo M. en I. Salvador Ayala Raggi, por su ayuda incondicional y por su sincera amistad.

Agradezco a la Universidad Nacional Autónoma de México, al Posgrado de Ingeniería, a los investigadores del departamento de Eléctrica, compañeros y amigos, y muy en especial a mi director de tesis Dr. Miguel Moctezuma Flores, el apoyo que me brindaron, así como sus valiosos conocimientos y su sincera amistad.

De la misma manera quiero ofrecer mi agradecimiento al Dr. Xavier Descombes del INRIA (Institut National pour la Recherche en Informatique et Automatique) en Francia, por su ayuda y sus aportaciones en las técnicas de Segmentación Markoviana.

A mis inolvidables y queridas amigas Maricela Torres Lugo y Olga Puntos Carbajal, por su cariño y su compañía.

ÍNDICE

INTRODUCCIÓN	1
CAPÍTULO 1	
Segmentación de Imágenes	3
1.1 Clasificación estadística	5
1.2 Detección de bordes	6
1.3 Crecimiento de regiones	6
1.4 Algunos métodos de segmentación	7
1.4.1 Clasificación de una imagen multiespectral usando redes neuronales	7
1.4.2 Método de segmentación híbrido	9
1.4.3 Clasificación de imágenes de satélite multifuente basada en un modelo de campos aleatorios de Markov (Markov Random Fields MRFs)	12
1.4.4 Comportamiento de una clasificación no-supervisada a partir de una clasificación supervisada	14
1.4.5 Clasificación de terrenos integrando imágenes SPOT y ERS-1 SAR	16

CAPÍTULO 2

Sistemas de percepción remota.	17
2.1 Componentes de un sistema de percepción remota ideal	18
2.2 Características de un sistema de percepción remota real	19
2.3 Características de los instrumentos de imágenes de percepción remota	20
2.3.1 Resolución espacial	20
2.3.2 Resolución espectral	22
2.3.3 Resolución radiométrica	23
2.4 Principales sistemas sensores	23
2.4.1 Sistemas Landsat	23
2.4.2 Sistemas SPOT	26
2.4.3 Sistemas SAR	29

CAPÍTULO 3

Campos Aleatorios de Markov	31
3.1 Nociones topológicas	31
3.1.1 Vecinos	31
3.1.2 Sistema de vecindad	32
3.1.3 Clique	32
3.2 Teorema de Hammersley-Clifford	33
3.2.1 Distribución de Gibbs	33
3.2.2 Distribución de Gibbs con temperatura	34
3.3 Segmentación markoviana (análisis bayesiano)	35
3.4 Recocido simulado	37
3.5 Modelo de Ising y regularización de imágenes binarias	38
3.6 Restauración de imágenes multiclase y modelo de Potts	40

CAPÍTULO 4	
Técnicas de segmentación markoviana.	43
4.1 Modelo de Potts clásico	43
4.2 Modelo de Potts matricial	48
CAPÍTULO 5	
Método de estructuras de árbol (quadtree)	53
5.1 Suavizado por quadtree	54
5.2 Agrupamiento de centróide local	57
5.3 Estimación de frontera	62
5.3.1 Filtrado y clasificación por umbral de la región de incertidumbre	64
5.3.2 Clasificación por distancia de Mahalanobis	68
5.3.3 Clasificación usando doble decisión	70
5.3.4 Clasificación usando el modelo de Potts para el último nivel de resolución	72
CAPÍTULO 6	
Comparaciones y conclusiones	74
6.1 Comparaciones	74
6.2 Conclusiones	98
REFERENCIAS	100

INTRODUCCIÓN

La detección de frontera es crucial para la búsqueda y descripción de las diferentes regiones de la tierra. La cartografía, los estudios de evolución geográfica y la navegación marítima automática requieren información exacta de dicha detección. Desafortunadamente la no muy buena calidad de ciertas imágenes hace más difícil la tarea de extraer información y de garantizar una detección de fronteras confiable. Es por esto que surge la necesidad de un proceso de segmentación.

La variedad de técnicas de segmentación y los diferentes tipos de imágenes sobre las que se aplican dichas técnicas, hacen difícil el llegar a una definición general aceptable del problema. Sin embargo, es posible hacer algunas afirmaciones de naturaleza general acerca de la segmentación de imágenes.

La mayoría de los autores coinciden en que la segmentación representa un proceso de modelización de la imagen, pero no todos consideran este proceso como una transformación de los datos de la imagen original a una descripción simbólica de tales datos. La segmentación puede ser vista como un proceso de simbolización. Para apreciar este punto de vista, es conveniente definir el objetivo de la segmentación de imágenes como: *la división de una imagen en un grupo de regiones disjuntas, cada una de propiedades constantes, separadas por fronteras bien definidas*. Sin hacer ningún compromiso, como el de la naturaleza de sus propiedades relevantes en una tarea de segmentación particular ya sea nivel de gris, color o textura, es posible ver que la descripción simbólica resultante contiene dos elementos: un componente de clase, indicando el valor de la propiedad relevante en una región y un componente de posición, indicando dónde se encuentra la región en la imagen. Así los símbolos asociados con la segmentación tienen una estructura particularmente simple, lo cual probablemente explica el porqué este aspecto del problema había recibido insuficiente atención en el pasado.

Por otra parte, la descripción simbólica de una imagen debe capturar aquellas características relevantes de la imagen, e ignorar las variaciones irrelevantes de las que la imagen es objeto. Esas variaciones podrían incluir la posición relativa y el movimiento de la cámara o transductor, las fluctuaciones debidas al ruido o a la iluminación del ambiente, etc. La distinción esencial entre los símbolos y las señales (imagen) de la cual se derivan los símbolos, es que cada símbolo representa una clase completa de las señales.

En este trabajo se ha desarrollado un método de segmentación, el cual combina un clasificador no-paramétrico, basado en un algoritmo de agrupamiento de centróides, con una representación de estructura de árbol (quadtree) de la imagen. La operación de suavizado por quadtree se realiza hasta llegar a un nivel de resolución óptimo de baja resolución, donde se efectúa la clasificación estadística. Una vez encontrada la frontera, ésta se proyecta a los siguientes niveles del árbol hasta llegar al nivel de máxima resolución (imagen original). La segmentación, aunque es estadística, no requiere un conocimiento a priori de las clases. Esto es, porque el agrupamiento se basa en operaciones locales. Para la estimación de frontera se siguieron diferentes técnicas tales como:

- decisión por umbral,
- decisión por distancia de Mahalanobis,
- un esquema de doble decisión y
- el modelo de Potts clásico.

Ante la necesidad de comparar los resultados obtenidos, con algún esquema de referencia el cual fuera confiable y óptimo, se eligieron dos esquemas basados en la teoría de los campos aleatorios de Markov (Markov Random Fields MRFs). Ya es conocido que la gran ventaja de usar los MRFs en segmentación de imágenes, es que las propiedades globales de la imagen son controladas por las propiedades locales. Se desarrollaron dos modelos, uno de ellos es el modelo de Potts clásico, donde el coeficiente de homogeneidad tiene un valor único y el modelo de Potts matricial, el cual sustituye el valor único del coeficiente de homogeneidad por una matriz que se compone por valores diferentes para cada par de clases.

La distribución de este trabajo se realiza en 6 capítulos. El capítulo 1 hace una introducción al concepto de segmentación de imágenes y presenta algunas técnicas que se han desarrollado para tal objetivo. Ya que el propósito aquí es el de segmentar imágenes de percepción remota, se consideró conveniente hacer una descripción en el capítulo 2 de los diferentes sistemas que se utilizan para la obtención de éstas imágenes. En el capítulo 3 se desarrolla la teoría matemática de los campos aleatorios de Markov. Los modelos de Potts clásico y Potts matricial se desarrollan en el capítulo 4. El capítulo 5 contiene la teoría, desarrollo y resultados del método de estructuras de árbol (quadtree). El capítulo 6 está dedicado a la comparación de resultados de los diferentes esquemas de segmentación, así como a las conclusiones de tales comparaciones.

CAPÍTULO 1

SEGMENTACIÓN DE IMÁGENES

La figura 1.1 revela una idea fundamental de la segmentación de imágenes. Obviamente en esta imagen hay dos regiones, cada una de ellas tiene distinta brillantez o nivel de gris. Además, las dos regiones son espacialmente distintas, ya que se encuentran en lados opuestos a partir de la línea de frontera. La segmentación de imagen es simplemente el proceso por el cual, mediante un cálculo, se convierte la descripción de imagen original (un arreglo de niveles de gris), en una descripción de dos regiones: las propiedades representativas y el conjunto de direcciones de elementos dentro de cada región.



Figura 1.1: Imagen que contiene dos regiones.

Al igual que muchos otros problemas en visión, mientras una definición de "sentido común" se puede dar fácilmente, no es así de fácil crear algoritmos computacionales cuya función es comparable con el sistema visual humano. Uno de los mayores impedimentos para un desarrollo exitoso de teorías de visión, ha sido una tendencia a subestimar la complejidad del problema, precisamente porque la función humana es prodigiosa y mediada por procesos en su mayoría subconscientes. A esto podría sumarse que, en común con otras tareas perceptuales, la visión en general y la segmentación en particular, son afectadas por varios tipos de incertidumbre, cuyo efecto combinado da como resultado técnicas perceptualmente ineficaces más simples.

Regresando a la figura 1.1, observamos en primer lugar, que la imagen está segmentada puramente en base a niveles de gris en dos regiones, una inspección a cada una de ellas revela que los píxeles deben ser colocados en distinta región de acuerdo a sus niveles de gris: pueden ser muy brillantes o muy oscuros para la región a la cual pertenecen virtualmente. Esto nos hace pensar que la segmentación desempeñada por un humano se realiza mediante el promedio de nivel de gris de la imagen y que no es simplemente una clasificación pixel por pixel.



Figura 1.2 Imagen con varias regiones de diferentes texturas.

El proceso de promedio espacial mediante el cual se calculan los valores representativos de las regiones, inevitablemente introduce cierta confusión en la localización de la frontera. Por lo tanto hay dos tipos de incertidumbre en este caso: la incertidumbre con respecto al valor representativo de la propiedad relevante en cada región debido a las fluctuaciones y la incertidumbre con respecto a dónde se encuentra la frontera. Si tales dificultades ocurren en un caso simple como escala de gris, es muy probable que se multipliquen cuando se usan propiedades de regiones más complejas, tal es el caso de la textura (figura 1.2).

Los primeros trabajos en segmentación se centraron en el problema aparentemente simple de binarizar imágenes en escala de gris, de esta manera podrían identificarse por ejemplo objetos industriales en un transfondo uniforme. Pronto llegó a ser muy claro, que sólo en los casos más simples donde la combinación de iluminación y transfondo en objetos reflejantes, llevaba a una marcada separación de niveles de gris, puede utilizarse un simple cálculo de umbral con alguna esperanza de éxito. Esto condujo a un número de aproximaciones a problemas más generales. De entre las técnicas que han surgido, es posible distinguir tres grupos principales: clasificación estadística, detección de bordes y crecimiento de regiones.

1.1 Clasificación estadística.

Cuando un simple cálculo de umbral falla, es natural tomar bases estadísticas más rigurosas para la segmentación. Una manera obvia de mejorar el funcionamiento, es tratar el problema como un problema de decisión, en el cual, las estadísticas de la población total de píxeles son consideradas como una mezcla de aquellas regiones componentes en las que la imagen será segmentada. Estos métodos primero estiman los parámetros más relevantes, tales como media y varianza, para las distribuciones componentes de la mezcla y después seleccionan umbrales óptimos, para separar los píxeles pertenecientes a las diferentes poblaciones, basados en algún criterio teórico de decisión.

Las distribuciones componentes pueden estimarse a su vez usando un modelo paramétrico, por ejemplo, asumiendo que son Gaussianas y aplicando máxima probabilidad o técnicas de estimación Bayesiana o bien usando un método derivado heurísticamente como es el agrupamiento (clustering). Dichos métodos son tradicionalmente usados en reconocimiento de patrones (por ejemplo [2],[3],[4]). La aplicación directa de segmentación de imágenes es posible en casos limitados ya sea por métodos teóricos de decisión [3] o por un algoritmo de agrupamiento [5],[6]. Frecuentemente, estos métodos se reemplazan por procedimientos derivados heurísticamente, como lo es dividir la imagen en un número de áreas superpuestas e interpolar los umbrales resultantes, para superar grandes variaciones de escala en el nivel de gris [7].

La diversidad de las modificaciones a la clasificación estadística, indica que se tienen algunas serias deficiencias cuando se aplican a segmentación de imágenes. La deficiencia principal, es la incapacidad de métodos estadísticos de primer orden para tomar en cuenta la relación espacial entre los datos. Por ejemplo, un método basado solamente en estadísticas de píxel, así como el histograma de la escala de gris, da resultados, los cuales son invariantes a permutaciones arbitrarias de la imagen. Aún así, tales permutaciones destruyen completamente la coherencia espacial, la cual es generalmente considerada como una propiedad esencial de las regiones, si dicha coherencia está expresada en términos de conectividad, relación de perímetro y área o algún criterio similar. Entre los más notables intentos están aquellos basados en los métodos de relajación o iterativo, en los cuales, una clasificación de píxel preliminar es hecha probabilísticamente, con las probabilidades

actualizándose sucesivamente, referenciándose en la clase de vecindarios de píxeles [8],[9],[10].

1.2 Detección de bordes.

Estos métodos, están basados en la convolución con la imagen de uno o más funciones *kernel*, escogidas para dar respuesta máxima a áreas de la imagen donde el nivel de gris cambia muy rápidamente.

Por la demanda computacional de la convolución, los primeros trabajos se centraron en el uso de *kernels* definidos en vecindarios pequeños, típicamente de sólo 3 x 3 píxeles en extensión (ej. [12],[13],[14]). A finales de los setentas, con la llegada de computadoras más potentes, se eliminó en gran parte esta limitación y se empezó a trabajar con algoritmos basados en aproximaciones matemáticas más rigurosas, involucrando *kernels* de tamaños de ventanas de arriba de 15 x 15 píxeles. Con el empleo de filtros de cuadratura, fue posible aplicar las mismas técnicas, no solamente para el análisis de características simples de la imagen como líneas y bordes, sino también análisis de textura [15]. Este método es adoptado y adaptado para la solución al problema de segmentación por textura.

En el caso de detección de bordes, si éste es considerado como un problema de filtro de empate o como un problema de estimación estadística, el dilema que surge al realizar el algoritmo, es la necesidad tanto de la localización espacial, como de la inmunidad al ruido. Estos dos requisitos, contradicen una determinada extensión del principio de incertidumbre de análisis de señal [16], al incrementarse la inmunidad al ruido, implica un promedio espacial más grande y por lo tanto la localización espacial se reducirá. Así, aunque el problema de incertidumbre asume una apariencia insignificamente diferente en los métodos de detección de bordes, éste aún existe.

En otros trabajos se utilizaron los filtros, cubriendo varias escalas diferentes en un intento por superar las limitaciones de la incertidumbre. La idea era usar la consistencia espacial de las salidas de varias escalas diferentes de filtros, para producir una estimación más confiable de la posición del borde, más que la que puede obtenerse mediante una simple operación de filtrado.

1.3 Crecimiento de regiones.

Entre los métodos estadísticos clásicos y la detección de bordes, se encuentra una diversidad de métodos, los cuales caen más o menos en la categoría de crecimiento de regiones. La idea principal detrás de estas técnicas es simple: dado uno o más puntos de inicio (starting points) en la imagen, los píxeles adyacentes se examinan uno por uno y se prueban para ver si son lo "suficientemente cercanos" a las propiedades de la región examinada en ese momento, para ser aceptados. Si fallan la prueba para cada región propuesta, éstos deben desecharse y por último asignarse a una nueva región.

La motivación de esta aproximación es muy clara: es tratar de tomar en cuenta tanto distancia en espacio, como similaridad de propiedades. Ya que el problema no se resuelve fácilmente en términos teóricos de decisión, se han encontrado una gran variedad de técnicas [17],[18], que van desde un simple agrupamiento, hasta estimación de frontera estocástica, la cual es una mezcla de crecimiento de regiones y detección de bordes.

El inconveniente con estos métodos, es que como en muchos problemas de optimización no-lineal, tanto la solución final, como el número de iteraciones requeridas para lograrla, son altamente dependientes de las condiciones iniciales: el número y la localización de los puntos de inicio. En general, no hay garantía de que una imagen dada podría ser segmentada de la misma manera por el mismo algoritmo y dos grupos de puntos de inicio. Otra objeción, es la naturaleza esencialmente secuencial de la búsqueda a través de la imagen.

Para resolver estos problemas, se han propuesto varios métodos [19],[20],[11], en los que se adopta una representación piramidal de la imagen. Primero se usó un tipo de "relajación" heurísticamente derivado del proceso de crecimiento de regiones, después estuvo firmemente basado en la teoría de decisión estadística. Ambos métodos, aprovechan la descripción piramidal de la imagen, para combinar la información global y local guiando el proceso de segmentación.

1.4 Algunos métodos de segmentación.

1.4.1 Clasificación de una imagen multispectral usando redes neuronales.

Este método fusiona las ventajas de las estrategias paramétricas y no-paramétricas, utilizando una red neuronal Kohonen y una red neuronal multicapa con un algoritmo de retro-propagación BP (Back Propagation) [21].

Las técnicas recientes de inteligencia artificial de redes neuronales, ofrecen métodos potencialmente robustos para la distinción de las diferentes áreas de la superficie de la tierra. Debido a que una red neuronal es relativamente tolerante a pérdidas y datos ruidosos, ésta tiene un alto potencial en la clasificación de imágenes de satélite. Esta investigación compara el resultado de un clasificador de red neuronal artificial ANN (Artificial Neural Network) con un algoritmo de entrenamiento BP, con un clasificador de máxima probabilidad MLC (Maximum Likelihood Classifier) supervisado [22].

La metodología ANN se resume en dos fases: la fase de aprendizaje y la fase de clasificación. El primer paso de la fase de aprendizaje es seleccionar el entrenador del área. Después el entrenador de área seleccionado se agrupa por una red neuronal Kohonen de tal manera que el número de regiones clasificadas sea mucho más grande que el número de clases actualmente deseado. Las regiones pequeñas con muy pocos píxeles, cuyo número

está abajo del valor del umbral, se fusionan con las regiones vecinas tomando en consideración los factores espaciales y geométricos.

En este método la red Kohonen agrupa la imagen en 64 clases para obtener valores promedios de cada una de las clases. Entonces las características de cada clase se calculan y se accesan a la red neuronal para entrenamiento y clasificación. La información utilizada, son los valores de las bandas 1,2,3 de 20 x 20 m² de una imagen SPOT¹, sus índices de vegetación, tales como verdosidad y brillantez de las áreas de la superficie de la tierra. Después del entrenamiento, el dato muestra utiliza el algoritmo BP, se procede igual para obtener las características del dato de prueba. La red se aplica sobre el área de prueba y la clasificación se obtiene por medio de un Perceptron Multicapa MLP (MultiLayer Perceptron) [23],[24],[25].

El patrón de clasificación adoptó 4 clases: agua, campo, bosque y zona urbana. La red neuronal triplecapa (Threelayered), tiene 5 neuronas en la capa de entrada y 4 neuronas en la capa de salida. El algoritmo BP es un gradiente iterativo para minimizar el error cuadrático medio entre la salida actual de una percepción multicapa retroalimentada y la salida deseada. Los datos SPOT de la banda 1 y 3, su brillantez y su verdosidad se aplican a cada neurona como dato de entrada. Tanto el modelo MLC, como el modelo ANN se entrenaron y se aplicaron al sitio de prueba y los resultados se evaluaron de acuerdo a la exactitud de la clasificación.

El rango de clasificación rechazada para cada categoría fue del 0.05 % y los pixeles desechados se calificaron como no-clasificados. Las tablas 1.1 y 1.2 muestran los resultados.

Clase	Precisión de clasificación
Agua	92.0 %
Campo	67.0 %
Bosque	78.3 %
Zona Urbana	54.1 %

Precisión de clasificación total 78.1 %

Clase	Precisión de clasificación
Agua	95.7 %
Campo	62.7 %
Bosque	90.5 %
Zona Urbana	57.9 %

Precisión de clasificación total 86.1 %

Tabla 1.1 Método de Máxima Probabilidad MLC

Tabla 1.2 Método de Red Neuronal Artificial ANN

¹¹ Las características de las imágenes SPOT, se verán a detalle en el siguiente capítulo: "Sistemas de Percepcion Remota"

El Método ANN mejoró la precisión de clasificación total de un 78.1 % a un 86.1 % (mejora de 8 %).

La comparación cuantitativa, indica que el clasificador estadístico convencional de máxima probabilidad, fue menos capaz de distinguir las clases de las áreas de la superficie de la tierra, comparado con el clasificador ANN. El MLC está restringido por la distribución normal, por lo que las redes neuronales BP son más tolerantes con el dato de entrada y pueden clasificar datos espaciales multifuente con distribuciones diferentes. La preclasificación con red Kohonen redujo significativamente el número de datos de entrada al MLP.

1.4.2 Método de segmentación híbrido.

El método integra detección de bordes y crecimiento de regiones para superar las propias debilidades de cada uno de éstos y consiste en los siguientes estados [26]:

- i) Filtrado.
- ii) Detección y seguimiento de bordes.
- iii) Enlace de fragmentos de borde.
- iv) Crecimiento de regiones.

En ii) los valores de la magnitud del borde resultante, se comparan con un umbral y se les aplica una operación de adelgazamiento (thinning) para crear bordes de un pixel de ancho. En iii) los fragmentos de borde resultantes, se enlazan donde se detectan huecos de un pixel de ancho entre dichos fragmentos de borde. Debido al enlace se forman polígonos cerrados, los cuales dividen la imagen en un grupo de subimágenes. Los fragmentos de borde que no pertenecen a un polígono cerrado, se eliminan. En iv) el crecimiento de regiones se lleva a cabo en cada polígono. Las regiones no pueden crecer fuera de los polígonos. El método de crecimiento de regiones es la mejor fusión (best merge) iterativa, el cual fusiona en cada exploración sobre la imagen, el par pixel/región al más bajo costo. Para fusionar pixeles aislados se emplea la regla de contexto. El método se aplicó a una escena Landsat - TM².

Los objetivos del método son: asegurar que las regiones no van a sobrecrecer bordes y permitir que los bordes detectados como incompletos puedan detectarse por crecimiento de regiones.

Filtrado. Es aconsejable primeramente filtrar las imágenes a segmentar, para remover el ruido en el dominio espectral. El método usado en esta prueba, está definido en [27] y se basa en la regla pulso/desertor (pulse/dropout) [28].

Detección de bordes: Se utiliza el método desarrollado por Lacroix [29], el cual es una versión aumentada del detector definido por Frei & Chen [30]. Utiliza el producto escalar

² Las imágenes Landsat - TM se verán a detalle en el siguiente capítulo: "Sistemas de Percepción Remota".

de 4 máscaras bases ortonormales con una ventana de imagen de 3 x 3. El cálculo de la magnitud del borde, depende de los valores espectrales. Dadas dos ventanas de imagen con una significativa discontinuidad de intensidad, las magnitudes de bordes serían diferentes por el método de Lacroix. No obstante, en percepción remota, las diferencias absolutas entre vecindarios de píxeles son significantes e indican la presencia de bordes importantes, sin tomar en cuenta la intensidad total de la imagen en la localidad de interés. Por lo tanto, es importante ajustar el cálculo de la magnitud de borde, para compensar los efectos del nivel de intensidad local. Se espera que la diferencia absoluta entre los valores de los píxeles, sea la misma que la magnitud. Esta dependencia se puede superar sustrayendo para cada ventana de tamaño 3 x 3 su valor mínimo menos 1. Las direcciones de borde y las magnitudes de borde, se calculan para cada canal, resultando vectores de dirección y magnitud de borde de 6 componentes. Por el procesamiento subsiguiente, la dirección del borde se toma del canal que tiene el valor de magnitud de borde más alto. Los valores de la magnitud de borde, se comparan con un umbral, por el procedimiento automático definido por Kapur *et al* [31], el cual utiliza la entropía en la magnitud de la imagen. Una vez teniendo la magnitud y la dirección del borde, se crean los fragmentos de borde de un píxel de amplitud, explotando la probabilidad de que sea una imagen borde LBE (Likelihood of Being Edge) [29]. La imagen LBE, se usa como entrada a un algoritmo de seguimiento de bordes, el cual selecciona los píxeles con más altos valores LBE. Así se crean los bordes de un píxel de amplitud. Si la resta del mínimo no se aplica mientras se calcula la magnitud del borde, el resultado puede ser otro muy diferente.

Enlazado: Se basa en el siguiente principio: Donde dos fragmentos de bordes terminan y son separados por un píxel solamente, se adhiere entre ellos un píxel, para juntarlos en conectividad 8, siguiendo un conjunto de reglas simples. Para implementar este procedimiento, la imagen se explora y se crea una estructura de datos para describir la red de fragmentos de borde en la imagen entera. El proceso comienza así: cuando un píxel de borde no-marcado se detecta, este píxel se marca. Si tiene algún píxel de borde vecino no-marcado y en conectividad 8, uno de ellos también se marca y se suma como un hijo del primer píxel de borde. El proceso de seguimiento de borde continúa a partir del hijo. Si el primer píxel, tuvo más de un píxel de borde vecino no-marcado y en conectividad 8, éstos se almacenan recursivamente en una pila y son visitados después, durante un procedimiento reversivo. Por consiguiente, se crea un grupo de datos con una estructura de árbol.

La mejor fusión para crecimiento de regiones: Los métodos de crecimiento de regiones son procesos secuenciales y están sujetos a problemas de dependencia de orden¹. Así, los primeros puntos de la imagen que serán procesados, pueden afectar la elección de los puntos a ser procesados después. El orden en el cual, los puntos de la imagen se procesan, puede afectar el resultado final. Para resolver el problema de dependencia de orden, una solución podría ser, fusionar solamente el par más similar de regiones espacialmente adyacentes, en la imagen entera en cada iteración [33], [34]. Un problema es cuando más pares tienen el mismo grado de similaridad o tienen un vecino en común. Es importante la

¹Zucker [32] define dependencia de orden como: "La información adquirida en los primeros estados del procesamiento, influirán en el procesamiento de estados posteriores".

selección del par correcto en caso de ambigüedad, porque puede afectar el resultado de segmentación final.

Una posibilidad de resolver este problema, es aplicar fusión simultáneamente a todos los pares que tengan el grado de similaridad más alto. La condición aquí, es que el orden completo de independencia, se logra si y sólo si, los pares que tienen el mismo grado de similaridad se fusionan simultáneamente. A este procedimiento se le llama Fusión Iterativa de Orden Independiente Simultáneo SOIM (Simultaneous Order Independent Iterative Merging).

Integración de los dos métodos de segmentación: Usando los polígonos cerrados derivados de la detección de bordes, la imagen se divide en un grupo de subimágenes independientes. Cada polígono cerrado cubre uno o más objetos. Asumiendo que no hay correlación espacial entre dos campos cubiertos con el mismo tipo de cultivo en diferentes polígonos, entonces se puede reducir el problema de la siguiente manera: en lugar de aplicar la mejor fusión de la imagen, se aplica la mejor fusión en un conjunto independiente de subimágenes, donde cada subimagen tiene un tamaño mucho menor que $N \times M^4$.

Para procesar píxeles en un polígono solamente, los píxeles externos al polígono se etiquetan como "FUERA" del polígono. Solamente los píxeles que están dentro del polígono se consideran para fusionarse. Los píxeles "BORDE", los cuales comprenden la orilla del polígono son también rechazados al aplicar la fusión. Una desventaja de este método, es que los píxeles de la orilla no se consideran en el proceso de crecimiento de región y así se mantienen aislados. Un método que adhiere esas orillas es [35].

Este procedimiento híbrido de crecimiento de regiones está hecho solamente para píxeles que no son "BORDE", de este modo, el proceso de crecimiento nunca puede fusionar píxeles fuera del polígono cerrado, al mismo polígono. La detección de borde y el crecimiento de región, se combinan de tal forma que se toma la ventaja de sus respectivas funcionalidades. La funcionalidad de la detección de borde, es que las fronteras se colocan de una manera espacialmente precisa. Su debilidad es que no todos los bordes se conectan dentro de contornos cerrados y por lo tanto pueden resultar muchas regiones grandes. La funcionalidad del algoritmo de agrupamiento (crecimiento de región), radica en su habilidad para colocar fronteras en áreas de gradientes débiles.

Costo computacional: El proceso se aplicó a una imagen de 1953 x 1801 píxeles. Se usó la configuración SUN/SPARC-10 con 64 Mb de RAM con el disco de datos montado via NFS. Para el procedimiento de la mejor fusión se dan dos tiempos. El primero es el tiempo requerido cuando se consideraron todos los polígonos. El segundo es el tiempo requerido cuando los cuatro polígonos más grandes (los que cubren áreas de agua), se excluyeron del proceso de fusión.

⁴N = número de renglones, M = número de columnas en la imagen completa.

	Tiempo transcurrido HH:MM:SS.S	Tiempo-CPU HH:MM:SS.S
Detección de borde	7:50.7	5:56.6
Enlace de borde paralelo	16:56.4	50.4
Determinación de polígono	32.7	9.6
Mejor fusión con polígonos (1)	58:16:14.8	55:20:06.7
Mejor fusión con polígonos (2)	18:40.0	10:50.0

Tabla 1.3: (1) La mejor fusión con todos los polígonos. (2) La mejor fusión excluyendo los cuatro polígonos más grandes (75000 píxeles).

1.4.3 Clasificación de imágenes de satélite multifuente basada en un modelo de campos aleatorios de Markov (Markov Random Fields MRFs).

Es un modelo general para clasificación multifuente de datos de percepción remota [36]. Se desarrolla y se prueba un modelo específico para fusión de imágenes ópticas, imágenes de apertura de radar sintéticas SAR⁵ (Synthetic Aperture Radar) y datos del Sistema de Información Geográfica GIS (Geographic Information System). El modelo MRF, explota el contexto de dependencia de clase espacial entre píxeles vecinos en una imagen y el contexto de dependencia de clase temporal entre las diferentes imágenes. El desarrollo del modelo específico se investigó fusionando imágenes Landsat TM, imágenes multitemporal ERS-1 SAR⁶ y mapas de la superficie de la tierra GIS.

Para una aplicación particular de percepción remota, los datos de la imagen deberían estar disponibles desde varias fuentes. Un número incremental de sensores basados en satélites están recolectando datos geográficos, oceánicos, geofísicos, meteorológicos y estudios de vigilancia. Además de las imágenes de satélite, los datos espaciales de GIS podrían proporcionar información acerca de un fenómeno en particular. Para muchas aplicaciones, la información proporcionada por cada sensor, podría ser incompleta, inconsistente o imprecisa. Las fuentes adicionales pueden proporcionar datos complementarios y la fusión de datos multifuente pueden crear una interpretación más consistente de la escena, en la cual la incertidumbre asociada disminuye. Así, en la

⁵Las imágenes SAR se verán a detalle en el Capítulo 2

⁶Las imágenes ERS-1 SAR se verán a detalle en el Capítulo 2.

interpretación de una escena, el uso de datos multifuente puede proporcionar la base de una interpretación más exacta que los métodos de simple-fuente tradicionales.

La fusión del MRF permite un mejor acoplamiento entre los datos de las fuentes.

Primero se presenta el modelo de clasificación multifuente general y el modelo para fusionar las imágenes Landsat TM, ERS-1 SAR y los datos de GIS, después se divide el modelo de fusión en dos partes. La primera: el módulo de tiempo fijo, describe la fusión de los datos recolectados simultáneamente, o con un intervalo de tiempo muy corto entre ellos, donde no han ocurrido cambios. La segunda: el módulo multitemporal, describe la fusión de las imágenes capturadas en diferentes fechas donde ocurrieron cambios en las clases. Para el módulo de tiempo fijo, la tarea es crear una interpretación de la escena asignando cada pixel a la clase que maximiza la probabilidad posterior, tomando en cuenta el factor confiabilidad de cada sensor. Se asume que las medidas de los diferentes sensores son independientes.

Incorporar el aspecto temporal, representa las mediciones del sensor s a un tiempo t .

El clasificador multisensor puede extenderse incorporando datos de GIS.

Para especificar el modelo de fusión particular para la tarea de prueba, es necesario definir el modelo de ruido, para imágenes Landsat TM y ERS-1 SAR.

Para el modelo de ruido de la imagen Landsat TM, se usaron las seis bandas no-térmicas originales de dicha imagen. La imagen no se degradó severamente por el ruido y fue bien modelada por la distribución normal multivariada (basada en la inspección de los histogramas univariados).

Para el modelo de ruido de la imagen ERS-1 SAR, se puede usar ya sea la función de densidad de probabilidad o un modelo más complejo, describiendo las estadísticas y la función de correlación. Aquí se modeló usando un campo aleatorio multiplicativo autoregresivo MAR (Multiplicative Autoregressive Random field). Los parámetros del modelo se han usado como descriptores de textura [37].

Para probar el método, se utilizó un conjunto de datos consistente en una imagen Landsat TM y tres imágenes ERS-1 SAR. Este conjunto de datos, se adquirió en un proyecto de monitoreo del arado en las áreas agrícolas en el otoño. Debido a la normalmente alta cubierta de niebla y las condiciones de iluminación, la posibilidad de obtener una escena Landsat TM libre de niebla en el periodo óptimo, después de que las actividades de cultivo han terminado y se cubra la tierra de nieve, es muy baja.

Se consideró un problema de clasificación de cinco clases, con las más generales áreas de la superficie de la tierra: agua, zona urbana, bosque y dos clases de áreas agrícolas: aradas y sin-arar. Entre las adquisiciones de las diferentes imágenes una gran porción de los

campos agrícolas han sido cultivados. Las muestras de control de la tierra de las áreas agrícolas, estuvieron disponibles para todas las fechas de adquisición de imágenes. Para checar el efecto del dato del mapa incluido en la clasificación, se digitalizó un mapa topográfico y se transformó a la proyección UTM.

El desarrollo del nuevo modelo de fusión MRF se comparó con el modelo de [38]. El modelo de referencia más simple, realiza fusión en un nivel pixel por pixel, fusionando las probabilidades posteriores del clasificador sensor-específico, tomando en cuenta los factores de confiabilidad y la probabilidad de cambios de clase. La exactitud de clasificación de simple-fuente para la escena fue del 95 % para Landsat y 65-80 % para SAR. Las grandes variaciones para las exactitudes de SAR, fueron debido a las condiciones del tiempo. La habilidad del sensor SAR para distinguir entre áreas agrícolas aradas y sin arar de la temperatura y de la humedad de la tierra.

Para fusión de imágenes adquiridas al mismo tiempo, el modelo de fusión MRF resultó en un rango de error de $3.1 \pm 0.1\%$, comparado a $6.3 \pm 0.3\%$ del modelo de referencia. Incluyendo los datos GIS, el rango de error correspondiente se redujo a $1.4 \pm 0.1\%$ para el modelo MRF y $2.9 \pm 0.2\%$ para el modelo de referencia. Para el modelo de fusión temporal, el rango promedio de error para el modelo MRF fue 5.6%, comparado con el 7.2% para el modelo de referencia. Los rangos de error promedio con datos GIS incluidos, fueron 2.7% para el modelo MRF y 4.4% para el modelo de referencia.

Para fusión de ciertas imágenes, el método MRF mejoró la exactitud de clasificación para áreas de agua y zonas urbanas comparado con el método de referencia. Para el modelo de fusión temporal, la diferencia principal entre al método de fusión MRF y el de referencia, fue la exactitud de clasificación para zonas urbanas y áreas agrícolas (aradas y sin arar).

En promedio, la exactitud al detectar los cambios actuales en arado fue 75.1% para el modelo MRF, comparado al 62.2% para el modelo de referencia.

1.4.4 Comportamiento de una clasificación no-supervisada a partir de una clasificación supervisada.

El método de clasificación está basado en un proceso de Poisson no-estacionario, para la distribución de puntos en el espacio radiométrico. Las funciones de intensidad se estiman no-paramétricamente usando el *kernel* uniforme. Para estimar las probabilidades *a priori*, se usa un algoritmo EM, el cual mantiene las densidades constantes. Después, a la imagen clasificada, se le aplica un filtrado especial: un clasificador contextual ICM. La distribución de puntos en la lattice geométrica-2 se modeliza con campos aleatorios de Markov. El clasificador ICM también toma en cuenta las densidades radiométricas.

Se trata de adaptar esta clasificación supervisada a una no supervisada. La filosofía, es encontrar automáticamente zonas homogéneas, las cuales podrían ser usadas como entrenadores de datos en la clasificación supervisada.

El proceso de Poisson no-homogéneo, la estimación de intensidades, la estimación a *priori* y el clasificador contextual se describen en [39].

Enseguida se describe el procedimiento de selección automática de puntos de entrenamiento: La filosofía consiste en detectar regiones homogéneas en la imagen, incluyendo no solamente los valores de los píxeles, sino que también la textura de cada una de las regiones: esta técnica es justamente la contraparte de encontrar puntos de cambio. Stephens y Smith [40], propusieron un modelo paramétrico para realizar esto, en el cual se usa en cada renglón y columna de la imagen, una prueba de hipótesis para detectar si hay o no un punto de cambio.

Otra técnica es la propuesta por Wichern, Miller y Hsu [41], se basa en encontrar grupos de píxeles contiguos que no contengan puntos de cambio. Este método dió no muy buenos resultados, por lo que se modificó: la prueba no se aplica a cambios en los parámetros, pero si a cambios en la distribución en cualquier dirección. Los resultados de la modificación, no fueron mejores, esto debido a que cada coordenada de la distribución fue tratada separadamente, la distribución completa multivariada no se tomó en cuenta. Devroye y Wise [42], dieron una solución para trabajar en el espacio completo, éste consiste de una comparación de apoyos de densidades. Por tal motivo se hizo lo siguiente:

- Primero se estiman los apoyos en un rango del mismo tamaño en la parte del lado derecho que del lado izquierdo del píxel (para el paso horizontal), y se hace lo mismo verticalmente y a lo largo de las dos diagonales;
- después se verifica en el número de puntos en un lado, estando en el soporte de los puntos del otro lado; si el número es suficientemente grande en las cuatro direcciones.

La elección del ancho de banda, genera una dificultad en el procedimiento. Usando la misma estrategia de comparación de soportes, las partes homogéneas encontradas de la imagen, se fusionan para formar los grupos de entrenamientos de clases. Si el grupo contiene muchos puntos, se toma aleatoriamente una muestra de cada uno de ellos.

Los resultados obtenidos de la segmentación en varias imágenes de diferente naturaleza (datos de percepción remota, fotografías aéreas), fueron muy satisfactorios, están muy cercanos a los resultados de los métodos supervisados. Esta técnica es además computacionalmente muy eficiente.

1.4.5 Clasificación de terrenos integrando imágenes SPOT y ERS-1 SAR.

Se enfoca en la integración de una imagen visible de alta resolución multispectral SPOT y datos de imagen ERS-1 banda-C. SAR, para la clasificación de terrenos. Se identificaron 7 tipos diferentes de terreno que van desde agua hasta vegetación. Como primera parte del estudio, se aplicó un modelo para compensar el efecto de elevación en la imagen ERS-1 SAR. La imagen SAR geoméricamente rectificadas, se reviste con la ortoimagen SPOT, produciéndose un error rms de este revestimiento, el cual es aceptable para este estudio. La segunda parte, trata con la clasificación de áreas de la superficie de la tierra. En particular, se suma la imagen fractal que es extraída de la SAR. Entonces ambas, la información espectral y la espacial se utilizan simultáneamente para la clasificación. Se siguió la segmentación supervisada para diferenciar las áreas de la superficie de la tierra de las imágenes combinadas [43].

Una red neuronal de aprendizaje dinámico DL (Dynamic Learning), se adaptó como un clasificador. Los detalles de la red y los métodos para su entrenamiento se describen en [44]. El entrenamiento de la red, procede presentándola con el dato de entrenamiento obtenido de la intensidad y de las imágenes fractales. El número de grupos de entrenamiento depende del área de la clase y del grado de variaciones espaciales en esa clase. Como un resultado, se obtuvieron cuatro canales de patrones de entrenamiento. Una clara mejora, se observa en la exactitud de la clasificación sobre áreas heterogéneas espacialmente y espectralmente. También se observa una mejora en el área de agua de ríos, utilizando información fractal adicional. Sin la ayuda de imágenes fractales, esta área sería clasificada como un estanque de pesca. Esto es porque ambas áreas son espacialmente homogéneas pero espectralmente heterogéneas, lo cual puede resolverse midiendo la textura. Las orillas de los ríos también se realzan. La tasa de exactitud de clasificación se basó en la matriz de confusión y el coeficiente de Kappa, los cuales son cerca de 0.95 y 0.94, respectivamente.

En este capítulo se mencionaron ya algunos métodos para segmentar y clasificar imágenes de varios tipos. También se mencionaron algunas técnicas de post-procesamiento para mejorar la clasificación. Los métodos de Estructuras de árbol (quadtree), así como el modelo de Potts se describen a detalle en capítulos posteriores. Existen otros métodos, todos ellos de no menos importancia, pero que no fue posible incluir aquí.

CAPÍTULO 2

SISTEMAS DE PERCEPCIÓN REMOTA.

La percepción remota de la superficie de la tierra, ha sido una historia muy larga que data desde el uso de globos en los siglos XVIII y XIX, pero en su connotación moderna, la percepción remota se inicializa con los sistemas montados en naves aéreas, que fueron desarrollados en el presente siglo, inicialmente para propósitos militares.

El nombre de percepción remota, aparece durante las primeras misiones interplanetarias de sondas no tripuladas. A bordo de las sondas, se instalaron cámaras de televisión que permitían capturar y enviar a la tierra imágenes de regiones previamente seleccionadas de la superficie de otros planetas. Por este hecho, la percepción remota, se define como la ciencia de captar y coleccionar información de un objeto a través de medidas hechas a distancia, sin estar en contacto físico con él [43]. Esta definición involucra todos aquellos campos posiblemente asociados a un objeto.

La medición de las intensidades de los campos asociados a un objeto, permite deducir algunas características del cuerpo que los produce o los altera. Algunos de los principales campos que se miden son el electromagnético, el eléctrico, el magnético, el gravitacional y el acústico.

Actualmente, el concepto está restringido a los métodos que emplean la energía electromagnética como medio de detección y medición de las características de un objeto [43]. Con base en esto, se puede redefinir percepción remota, como un conjunto de actividades que buscan la caracterización de las propiedades de los cuerpos naturales, a través de la detección, registro y análisis del flujo de energía radiante, por ellos reflejado o emitido [44].

2.1 Componentes de un sistema de percepción remota ideal.

Los componentes básicos de un sistema de percepción remota ideal, se muestran en la figura 2.1. Incluye los siguientes [45]:

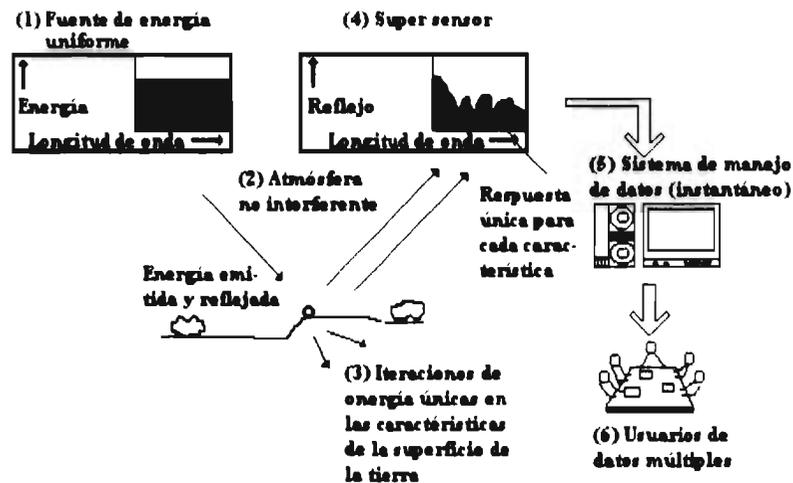


Figura 2.1 Componentes de un Sistema de Percepción Remota ideal [45].

1. *Fuente de energía uniforme*: Esta fuente proporcionaría energía sobre todas las longitudes de onda, a un nivel de salida alto, conocido y constante, sin tomar en cuenta tiempo y lugar.
2. *Atmósfera no interferente*: Sería una atmósfera que no modificaría la energía de la fuente, de ninguna manera, aún si la energía estuviera en camino a la superficie de la tierra o viniera de ésta, la atmósfera la mantendría sin tomar en cuenta la longitud de onda, tiempo, lugar y la altitud de percepción involucrada.
3. *Serie de iteraciones únicas de materia de energía en la superficie de la tierra*: Esas iteraciones, generarían señales reflejadas y/o emitidas que no solamente son seleccionadas por su longitud de onda, sino que también son conocidas, invariantes y únicas a todos y cada uno de los tipos y subtipos de interés de las características de la superficie de la tierra.
4. *Super sensor*: Sería un sensor, altamente sensible a todas las longitudes de onda, proporcionando datos espacialmente detallados en la brillantez absoluta (o resplandor) de una escena, como una función de longitud de onda, a través del espectro. Este sensor sería simple y confiable, no requiere virtualmente potencia o espacio y sería exacto y económico al operar.
5. *Sistema de manejo de datos en tiempo real*: En este sistema, fueron generados los instantes de resplandor *versus* la respuesta de longitud de onda sobre un elemento de terreno, esto sería procesado en un formato interpretable y reconocido como el

único, para el elemento del terreno particular del cual viene. Este procedimiento sería realizado casi instantáneamente ("tiempo real"), proporcionando información oportuna. Por la naturaleza consistente de las iteraciones de materia de energía, no habría necesidad de datos de referencia en el desarrollo del análisis. Los datos derivados proporcionarían una idea de los estados físico, químico y biológico de cada una de las características de interés.

6. *Usuarios de datos múltiples:* Esta gente tendría profundos conocimientos tanto de sus respectivas disciplinas como de técnicas de adquisición y análisis de datos de percepción remota. Esta información estaría disponible para ellos rápidamente y a un menor costo. Con esta información, diferentes usuarios tomarían decisiones en cuanto a la mejor manera para dirigir los recursos de la tierra bajo escrutinio.

Desafortunadamente, un sistema de percepción remota ideal como el descrito no existe. Los sistemas de percepción remota reales están alejados con respecto a cada uno de los puntos de un sistema ideal.

2.2 Características de un sistema de percepción remota real.

Considerando los elementos del sistema ideal, se han desarrollado los siguientes defectos generales, que podrían ser reconocidos en cualquier sistema de percepción remota real.

1. *Fuente de energía:* Todos los sistemas de percepción remota pasivos, dependen de la energía, ya sea reflejada y/o emitida por las características de la superficie de la tierra. La distribución espectral de la luz del sol reflejada y la energía auto-emitida, está muy lejos de ser uniforme. Los niveles de la energía solar, obviamente varían con respecto al tiempo y el lugar, y los diferentes materiales de la superficie de la tierra, emiten energía variando los grados de eficiencia. Mientras se tiene algún control sobre la naturaleza de las fuentes de energía para sistemas activos, las fuentes de energía usadas en todos los sistemas reales, son generalmente no-uniformes con respecto a la longitud de onda y sus propiedades varían con el tiempo y el lugar. Consecuentemente, se debe calibrar las características de la fuente misión por misión, o tratar con unidades de energía *relativa*, sensadas en un lugar y tiempo dado.
2. *Atmósfera:* La atmósfera normalmente incrementa los problemas provocados por la variación de la fuente de energía. La atmósfera siempre modifica la intensidad y la distribución espectral de la energía recibida por un sensor. Esto limita "donde podemos ver" espectralmente y sus efectos varían con la longitud de onda, tiempo y lugar. La eliminación o compensación de los efectos atmosféricos, vía una forma de calibración, es particularmente importante en aquellas aplicaciones donde se involucran observaciones repetitivas de la misma área geográfica.
3. *Las iteraciones de materia de energía en la superficie de la tierra:* La percepción remota podría ser simple, si cada material reflejara y/o emitiera energía en una forma conocida y única. A pesar de que los patrones de respuesta espectral, juegan un papel principal en la detección, identificación y análisis de los materiales de la superficie de la tierra, el mundo espectral está lleno de ambigüedad. Radicalmente, los diferentes tipos de materiales pueden tener grandes similitudes espectrales, haciendo difícil

una diferenciación. Además, la comprensión general de las iteraciones de materia de energía para cada característica de la superficie de la tierra, está en un nivel elemental para algunos materiales y virtualmente no existe para otros.

4. *Sensor*: Un sensor simple no es sensible a todas las longitudes de onda. Todos los sensores reales, han fijado límites de *sensibilidad espectral*. También tienen un límite en cuanto a qué pequeños en un objeto de la superficie de la tierra, pueden ser "vistos" todavía por un sensor. Este límite, se conoce como *resolución espacial* de un sensor, es una indicación de qué también puede un sensor grabar detalles espaciales.
5. *Un sistema de manejo de datos*: La capacidad de los actuales sensores para generar datos, no excede la capacidad actual del manejo de los mismos. El procesamiento de datos en un formato interpretable, puede ser y es un logro que implica algunas consideraciones, instrumentación, tiempo, experiencia y datos de referencia. Mientras se tenga el manejo de más datos por medio de máquinas (computadoras y otros dispositivos mecánicos y electrónicos), la intervención humana en el procesamiento de datos es y seguirá siendo esencial para la aplicación productiva de los datos de percepción remota.
6. *Usuarios de datos múltiples*: La parte medular de la aplicación exitosa de cualquier sistema de percepción remota, es la persona (o personas) que usan los datos del sensor remoto del sistema. Los "datos" generados por un sistema de percepción remota llegan a ser "información", solamente si alguien entiende su generación, sabe cómo interpretarlos y conoce la mejor manera de usarlos.

2.3 Características de los instrumentos de imágenes de percepción remota.

Las principales características de un instrumento de imagen de percepción remota, operando en las bandas infrarroja y visible, son descritas en términos de su resolución espacial, espectral y radiométrica [46]. Otras características importantes son: la manera de operar el dispositivo de rastreo (electromecánico o electrónico) y sus propiedades geométricas.

2.3.1 Resolución espacial.

Este no es un concepto fácil de definir. Esta resolución puede medirse de muchas maneras, dependiendo de los propósitos del usuario. Townshend [47] utiliza cuatro diferentes criterios en los cuales se basa para definir la resolución espacial. Esos criterios son las propiedades geométricas del sistema de imágenes, la capacidad para distinguir los objetivos puntuales, la capacidad para medir la periodicidad de los objetivos repetitivos y la capacidad de medir las propiedades espectrales de los objetivos pequeños. La medida más comúnmente usada, basada en las propiedades geométricas de la imagen, es el campo instantáneo de vista CIV (Instantaneous Field of View IFOV) de un sensor. Está definido como el área en la tierra, que en teoría, es visto por el instrumento desde una altitud dada y en un instante de tiempo dado. El CIV puede medirse de dos maneras: como el ángulo α o como la distancia equivalente XY en la tierra (Figura 2.3, note que la sección de cruce XY, es de hecho, el diámetro de un círculo).

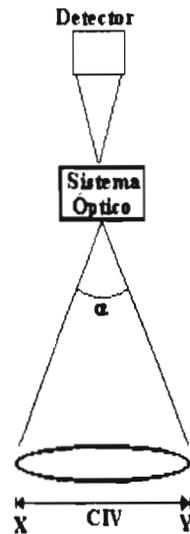


Figura 2.2 Campo instantáneo de vista angular [46].

El campo instantáneo de vista, depende de un gran número de factores. Los satélites no tienen una órbita perfectamente estable; por lo que su altura variará en decenas de kilómetros. El CIV llegará a ser más pequeño a altitudes más bajas e incrementará con el aumento de altitud. El CIV es la medida de resolución más frecuentemente citada, aunque no es necesariamente la más útil. Para explicar el porqué, se considera la manera en la cual la reflexión de una fuente puntual se expresa en una imagen. Una fuente puntual altamente reflectiva en la tierra, no produce un simple punto de brillo en una imagen capturada por una nave aérea o satélite. La fuente puntual es vista en una imagen como una región circular difusa, debido a las propiedades ópticas involucradas en la imagen. Una sección de cruce de lo grabado o de la distribución de intensidad de la imagen de una fuente puntual, puede verse como una distribución de tipo Gaussiano. La forma dependerá de las propiedades de los componentes ópticos del sistema y de la brillantez relativa de la fuente puntual. La definición de potencia de resolución espacial, basada en el CIV, no es completamente satisfactoria, ya que como una definición geométrica, no toma en cuenta las propiedades espectrales del objetivo. Si la percepción remota está basada en la detección y grabación de la reflexión de los objetivos, la reflexión estaría siendo medida en un número de puntos discretos, entonces podría ser razonable una definición de resolución espacial, que tomara en cuenta, la manera en la cual se genera la reflexión. Esto es la base de la definición del elemento de resolución efectiva ERE (*effective resolution element*), el cual es definido por Colvocoresses, como "el tamaño de una área para la cual, el valor de la simple reflexión puede asignarse con razonable certeza, que la respuesta está en 5% del valor representando la reflexión relativa actual".

Otros métodos para medir la potencia de resolución espacial de un dispositivo de imagen, están basados en la capacidad del dispositivo para distinguir objetos específicos. Hay dos medidas en uso y ambas son quizás más fácilmente definidas por sensores fotográficos. El primer método está basado en el hecho de que la función de difusión puntual PSF (Point Spread Function) de una fuente puntual, es un disco central brillante con anillos brillantes y oscuros alrededor de éste. El criterio Rayleigh, asume dos fuentes puntuales igualmente brillantes y especifica que las dos fuentes serán distinguibles en una imagen, si el disco central brillante de una, cae en el primer anillo oscuro de la PSF de la otra. La separación mínima entre las fuentes puntuales para lograr este grado de separación en la imagen, es una medida de la potencia de resolución espacial del sistema. El segundo método asume que los objetivos no son puntos, sino objetos lineales y paralelos con una separación conocida, la cual está relacionada con su frecuencia espacial. Si los objetos contrastan muy fuertemente con su transfondo, entonces podrían moverse para juntarlos hasta alcanzar el punto en donde no sean muy distinguibles. La frecuencia espacial de los objetos, en la cual ellos sean lo menos distinguibles, es una medida de la potencia de resolución espacial.

2.3.2 Resolución Espectral.

La segunda propiedad importante de un sistema óptico de imágenes, es la resolución espectral. Las imágenes digitales capturadas por sensores a bordo de satélites, han sido multibanda o multispectrales; lo cual significa que las imágenes individuales han sido separadamente grabadas en bandas espectrales discretas. El término resolución espectral se refiere al ancho de esas bandas espectrales. Enseguida se nombran dos importantes puntos: i) La posición en el espectro, el ancho y el número de bandas espectrales, determinarán el grado del cual los objetivos individuales (especies de vegetación, tipos de cosecha o rocas), pueden ser discriminados en las imágenes multispectrales y ii) el uso de imágenes multispectrales, pueden llevar a un grado más alto de potencia discriminatoria que una simple banda tomada de ellas mismas. Las diferencias entre el espectro reflejado de varias rocas, por ejemplo, podría ser muy sutil y los tipos de rocas podrían por lo tanto ser separables, solamente si el dispositivo de grabación fuera capaz de detectar la reflexión espectral del objetivo en una banda de ondas muy estrecha. Para proporcionar una identificación más confiable de objetivos en particular, en una imagen sensada remotamente, la resolución espectral del sensor debe coincidir tanto como sea posible, con la curva de la reflexión espectral del objetivo proyectado.

Desafortunadamente, no es posible incrementar la resolución espectral de un sensor simplemente para satisfacer las necesidades del usuario; hay un precio a pagar. La resolución espectral muy alta, reduce la relación señal-a-ruido de la salida del sensor [46].

Debe encontrarse un compromiso entre los requerimientos del ancho de banda estrecho (resolución espectral alta) y una relación señal-a-ruido baja.

2.3.3 Resolución radiométrica.

La resolución radiométrica o sensibilidad radiométrica, se refiere al número de niveles digitales usados para expresar las datos capturados por el sensor. En general, entre más grande sea el número de niveles, mayor será el detalle en la información. Podría considerarse una imagen digital compuesta de sólo dos niveles en la que el nivel 0 se muestra como negro y el nivel 1 como blanco. El ojo no es tan sensitivo a cambios en la intensidad, como a variaciones en el color. El número de niveles es comúnmente expresado en términos del número de dígitos binarios, necesarios para almacenar el valor del nivel máximo. El número de niveles, no puede incrementarse para satisfacer las preferencias del usuario. Debe tenerse en cuenta la relación señal-a-ruido del sensor. El tamaño del escalón de un nivel de reflexión al siguiente, no puede ser menor que el nivel del ruido o sería imposible decir si un cambio en el nivel, se debió a un cambio real en la reflexión del objetivo o a un cambio en la magnitud del ruido superimpuesto. Un instrumento de baja calidad con un nivel de ruido alto, tendría por lo tanto una resolución radiométrica menor, comparado con uno de alta calidad y de alta relación señal-a- ruido. Tucker [48], investigó la relación entre la resolución radiométrica y la capacidad para distinguir diferentes tipos de vegetación; en relación a una imagen de 64 niveles, encontró solamente una mejora de 2 a 3% al emplear una de 256 niveles. *Bernstein et al.* [49], usó una medida de entropía, para comparar la cantidad de información (en bits por pixel) para datos de 8 y 6 bits.

2.4 Principales sistemas sensores.

2.4.1 Sistemas Landsat.

El programa Landsat, llamado en sus inicio ERTS (Earth Resources Technology Satellite), se desarrolló con el objetivo de permitir la adquisición de datos espaciales, espectrales y temporales de la superficie de la tierra de una forma global y repetitiva. Este programa fue operado por la NASA/NOAA (National Oceanic and Atmospheric Administration) hasta 1985, cuando fue transferido a EOSAT (Earth Observation Satellite Company). La serie Landsat consta hasta el momento de 6 satélites y se contempla extenderla hasta 7. La tabla 2.1 muestra los satélites componentes de esta serie, el periodo de operación y los sensores que contienen. Como se observa en la tabla, la serie Landsat ha producido imágenes por medio de tres sensores diferentes MSS (imágenes en 4 bandas con un pixel rectangular de resolución 79 m x 56 m, el Landsat 3 contó con una quinta banda en la porción térmica del espectro, que tenía una resolución de 120 m por lado), RBV (sistema de televisión de 3 canales, el Landsat 3 tenía un solo canal) y TM (barredor multiespectral de 7 bandas y resolución de 30 m). Sin embargo, se tiene contemplado que el último satélite que se tiene proyectado lanzar, cuente con dos sensores nuevos, que son el ETM, instrumento formador de imagen semejante al TM pero con 8 bandas y una resolución de 15 m y el MLA, barredor multiespectral de captura a lo largo de la dirección de desplazamiento del satélite y que contará con 8 bandas y una resolución máxima de 10 metros en la porción visible.

Los satélites Landsat se desplazan en una órbita quasi polar y heliosincrona (el ángulo sol-tierra permanece constante), esto permitió a los Landsat 1, 2 y 3 obtener datos de una misma región de la tierra cada 18 días y actualmente a los Landsat 4 y 5 cada 16, en condiciones semejantes de iluminación.

Satélites	Periodo	Sensor
Landsat 1	23/07/72 al 05/01/78	MSS/RBV
Landsat 2	22/01/75 al 27/07/83	MSS/RBV
Landsat 3	05/03/78 al 07/09/83	MSS/RBV
Landsat 4	16/07/82 operando	MSS/TM
Landsat 5	01/03/84 operando	MSS/TM
Landsat 6	10/93 falló lanzamiento	ETM
Landsat 7	En planes	ETM/MLA

Tabla 2.1 Serie de satélites Landsat. MSS (Multispectral scanner), RBV (Return beam vidicom), TM (Thematic mapper), ETM (Enhanced thematic mapper) y MLA (Multilinear array).

Los datos colectados son transmitidos vía microondas a las estaciones terrenas de recepción, donde son grabados y procesados. Los datos son ofrecidos a los usuarios como imágenes multispectrales en formato de fotografías, en cintas magnéticas compatibles para computadora (CTT, ½ pulgada ó 8 mm), en discos flexibles y actualmente en CD-ROM.

El MSS es un sistema formador de imágenes, cuyo barrido es perpendicular a la trayectoria y es obtenido a través de un espejo oscilante, con un campo de vista de 11.6°. El sistema permite la obtención de líneas de barrido de aproximadamente 185 km. La radiación proveniente de la escena, después de la reflexión en el espejo oscilante, es focalizada por el sistema óptico sobre una matriz de 24 terminales de fibras ópticas de manera que cada terminal recibe el flujo de radiación proveniente de una área de 79 m x 79 m en la superficie (ERT).

Conducido por las fibras ópticas, el flujo de radiación es descompuesto por medio de filtros en 4 conjuntos de las siguientes bandas:

Banda 4: de 0.5 a 0.6 μm .

Banda 5: de 0.6 a 0.7 μm .

Banda 6: de 0.7 a 0.8 μm .

Banda 7: de 0.8 a 1.1 μm .

Banda 8: de 10.4 a 12.6 μm .(únicamente Landsat 3).

Al alcanzar los detectores (6 por banda), la radiación es convertida en una señal eléctrica proporcional a la radiada de la superficie incluida en el elemento de resolución del terreno correspondiente. La señal que puede variar de 0 a 4 V, es entonces cuantizada en 63 niveles posibles, siendo en seguida transmitida a la tierra o, eventualmente grabada para transmisiones posteriores. El arreglo de la matriz de detectores, permite que durante la oscilación del espejo, 6 líneas sean barridas, del oeste hacia el este, una vez cada 9.95 μ s. Como durante este intervalo el ERT, se desplaza apenas 56 m en el sentido del barrido; a cada elemento de resolución de la imagen o pixel, se debe asociar una área de 56 m x 79 m en la superficie.

Generalmente, una imagen MSS está formada por 2983 líneas y 3548 pixeles, cubriendo una área de 185 km x 185 km, aproximadamente, sobre la superficie.

El mapeador temático (Thematic Mapper), es un instrumento sensor (captador de imágenes), proyectado para conseguir mejor resolución (30 m x 30 m máxima sobre la superficie), separación espectral (7 bandas) y fidelidad geométrica precisa, en comparación con el MSS. Es un sistema formador de imagen de barrido perpendicular a la trayectoria y la captación de la imagen es hecha a través del espejo oscilante.

Los datos del TM son adquiridos simultáneamente en 7 bandas espectrales que son:

- Banda 1: de 0.45 a 0.52 μ m.
- Banda 2: de 0.52 a 0.60 μ m.
- Banda 3: de 0.63 a 0.69 μ m.
- Banda 4: de 0.76 a 0.90 μ m.
- Banda 5: de 1.55 a 1.75 μ m.
- Banda 6: de 10.4 a 12.5 μ m.
- Banda 7: de 2.08 a 2.35 μ m.

La imagen está formada por bandas de 6120 x 6120 pixeles, cubriendo una área de 185 km x 185 km aproximadamente. El tamaño del elemento de resolución del terreno, es de 30 m para las bandas 1,2,3,4,5 y 7 y de 120 m para la banda 6. Para las bandas 1 a 5 y la 7, los sensores están dispuestos en forma de una estructura de líneas y columnas de 16 elementos, mientras que la banda 6 cuenta únicamente con 4 detectores. La energía captada por el sensor, es convertida en señales eléctricas de bajo nivel y registrada como uno de los 256 niveles de grises posibles. Los satélites Landsat se desplazan de norte a sur en una órbita geocéntrica, circular quasi polar y heliosíncrona, lo que garantiza condiciones semejantes de iluminación a lo largo del año en el área observada.

El sistema mundial de referencia WRS (World Reference System), clave de localización global para las imágenes Landsat, está basado en un sistema de columnas y líneas cuyos cruces determinan la localización nominal del centro de cada escena. Este sistema proporciona a los usuarios la correlación con las latitudes y longitudes y la facilidad de identificar cualquier posición sobre la superficie terrestre. Refleja la naturaleza repetitiva de las órbitas Landsat y la manera cómo las imágenes son encuadradas después de la adquisición. Aquí, los números de las órbitas están

correlacionadas con los centros nominales de cada órbita. El número de líneas, identifica la posición de la escena en el sentido vertical y son numerados del 1 al 24 en el sentido descendente del satélite y de 125 a 248 en el sentido ascendente. La hora solar media de cada línea individual en una órbita se mantiene fija, por lo tanto todos los centros para una línea de una determinada latitud tienen una misma hora solar. Los cambios en el ángulo de elevación solar, provocan variaciones en las condiciones de iluminación sobre las cuales una imagen es obtenida. La elevación solar fluctúa mucho a lo largo del año. Sus efectos dependen de la región a ser observada, las regiones de alta reflectancia son más sensibles a la elevación del sol.

2.4.2 Sistemas SPOT

El programa de satélites SPOT (Système Pour l'Observation de la Terre), es una co-participación de Francia, Suecia y Bélgica y consta de 3 satélites que poseen las mismas características. El primer satélite SPOT fue puesto en órbita el 22 de febrero de 1986, el segundo en febrero de 1990 y el tercero fue lanzado en septiembre de 1993. Los satélites SPOT han sido colocados a una altitud media de 830 km, en una órbita circular quasi polar inclinada 8° con respecto al norte geográfico. Éstos dan la vuelta a la tierra en 101 minutos, lo cual combinado con el movimiento de rotación de la tierra hace que la traza en el suelo de 2 órbitas consecutivas, aparezca desplazada en el ecuador 2823 km. Cada uno de los satélites SPOT vuelve a pasar en la vertical de un mismo punto al cabo de 369 revoluciones, o sea, cada 26 días.

Los satélites SPOT están dotados de 2 instrumentos idénticos, el HRV1 y HRV2 (Haute Résolution Visible), capaces de funcionar de forma independiente. Cada instrumento barre una banda de dimensión, en sentido este-oeste de 60 km, en mira vertical y puede alcanzar los 80 km en visión oblicua.

A lo largo de la traza del satélite, se dividen las escenas cada 60 km. Así pues, la escena SPOT cubre una superficie de 60 km x 60 km, en mira vertical y hasta 60 km x 80 km en visión oblicua.

Ambos instrumentos cubren áreas distintas cuando realizan una mira vertical. Sin embargo, cuando los dos sensores funcionan al mismo tiempo en una toma de mira vertical, las escenas adquiridas por ambos instrumentos sufren un traslape de 3 km, pudiendo integrarse a una sola franja de 117 km. La orientación de las escenas con respecto al norte, varía en función de la latitud y del ángulo de toma de vista. Los instrumentos poseen dos arreglos de detectores, el primero de ellos consta de 3000 sensores y se emplea para la captura de imágenes en el modo multispectral, mientras el segundo tiene 6000 sensores y permite la adquisición de imágenes pancromáticas. Por tanto cada línea de barrido comprende 3000 ERT's o píxeles para cada banda del modo multispectral y 6000 en el modo pancromático.

Una escena SPOT consta de 3000 líneas de 3000 ERT's de 20 m en modo multibanda y de 6000 líneas de 6000 píxeles de 10 m en modo pancromático. En mira oblicua, hasta 27 grados hacia el este o hacia el oeste, la dimensión en el suelo de una línea varía entre 60 km (mira vertical) y 80 km (mira oblicua externa), siendo constante el número de detectores. El paso de muestreo varía en consecuencia entre 10 y 13.5 m o

entre 20 y 27 m según el modo espectral. A continuación las líneas se someten a un remuestreo al paso fijo de 10 o 20 m, con lo cual aumenta el número de píxeles en la línea. El descentrado observado entre cada línea, se debe, entre otros, a la rotación de la tierra entre los registros de dos líneas consecutivas. Dada su órbita quasi polar y heliosincrona, SPOT puede adquirir imágenes hasta una latitud de 84° norte-sur.

La elección de la órbita, se hizo con el objeto de asegurar una sincronización perfecta con el sol (heliosincrona). Las condiciones de iluminación, son por lo tanto idénticas para todas las escenas de una misma latitud en determinada época. Si a lo largo del año, varían el azimut y la elevación del sol, el ángulo entre la dirección solar, el centro de la tierra y el plano de la órbita permanecen constantes.

Los satélites SPOT, capturan la información mediante dos modos distintos de observación: el modo pancromático "P" y el modo multibanda o multiespectral "XS", ambos modos funcionan indiferentemente en uno u otro de los instrumentos y de manera simultánea o individual. Sin embargo, debido a que el satélite posee únicamente dos canales para transmisión de datos, no pueden funcionar las 4 posibilidades a un mismo tiempo. Las combinaciones posibles se muestran en la tabla 2.2.

HRV1	XS	+		+		+	
	P		+		+	+	
HRV2	XS	+			+		+
	P		+	+			+

Tabla 2.2 Posibles modos de funcionamiento simultáneo de los instrumentos sensores HRV1 y HRV2.

Cuando un instrumento captura una escena al mismo tiempo, tanto en modo pancromático como multiespectral, las dos imágenes generadas (XS y P), no se superponen perfectamente, debido a la disposición de los sensores dentro del instrumento.

Las longitudes de onda para los diferentes modos son:

Banda pancromática: de 0.51 a 0.73 μm .

Banda XS1: de 0.50 a 0.59 μm .

Banda XS2: de 0.61 a 0.68 μm .

Banda XS3: de 0.79 a 0.89 μm .

Como se puede apreciar, para el modo pancromático, la observación se realiza en una banda espectral única, correspondiente a la parte visible del espectro sin el azul. Dicha toma de vista, efectuada en un solo canal, proporciona imágenes en blanco y negro. El tamaño del píxel es de 10 m. En el modo multibanda, se obtienen imágenes para las porciones verde, rojo e infrarojo cercano del espectro. La combinación de las 3 bandas posibilita la consecución de composiciones a color. El tamaño del píxel es de 20 m.

Los dos instrumentos HRV del SPOT, están dotados de espejos orientables que permiten la visión oblicua, hacia el este o hacia el oeste de la traza, con un límite de $\pm 27^\circ$ de la vertical. Para poder tener una mira oblicua, los espejos orientables se desplazan de 0.6° en 0.6° , lo que permite captar hasta 91 escenas o posiciones diferentes. Tal inédita característica posibilita, por un lado, la visión estereoscópica y por el otro, la observación repetida de una misma zona.

Un desplazamiento del ángulo de vista de 24° al oeste y de 24° al este, permite obtener una relación base/altura (b/h) de 1. Para un par estereoscópico, compuesto por una vista vertical y otra segunda realizada desde un ángulo de 27° , el valor de la relación b/h es de 0.5.

Gracias a la mira oblicua, resulta posible adquirir escenas dentro de una banda de 950 km, paralela a la dirección de desplazamiento del satélite y centrada en su traza. Dicha técnica permite aumentar la frecuencia de observación de un mismo sitio durante un ciclo. Tal frecuencia varía en función de la latitud: en el ecuador se puede observar la misma región 7 veces durante los 26 días del ciclo orbital, lo que da una media de 3.7 días entre dos observaciones. Para una latitud de 45° , una misma región se puede observar 11 veces durante el ciclo orbital, dando una media de 2.4 días.

La serie SPOT presenta también su propio sistema de referencia mundial, para la ubicación geográfica de sus imágenes, denominado GRS (Grade de Reference Spot). La GRS es un sistema mediante el cual se atribuye a cada escena dos marcas de referencia K y J, correspondientes a columnas y líneas. Las columnas quedan orientadas paralelamente a la dirección de desplazamiento del satélite y las líneas son paralelas a las latitudes. La intersección de una columna con una línea determina un nodo. El centro de cada escena SPOT, está relacionado con un nodo GRS. Una escena en la GRS, cubre una área de 60 km x 60 km. Si una escena cubre una área mayor y su centro se encuentra desplazado con respecto a cualquier nodo de la GRS, debido a que fue tomada mediante una mira oblicua, es incorporada al nodo más cercano. Una escena puede estar desplazada hasta 30 km (para la derecha o la izquierda de la K nominal), pero no existirá desplazamiento en el sentido norte-sur, esto es, el centro de la imagen siempre caerá en una J del sistema GRS. La figura 2.3 muestra una imagen SPOT.



Figura 2.3 Imagen SPOT. © CNES.
(Tamaño de la escena: 256 x 256 píxeles. Tamaño del píxel: 10 mts.)

2.4.3 Sistemas SAR

El satélite marino (Seasat), fue uno de los satélites experimentales lanzados durante los setentas. Seasat fue lanzado en 1978 en una órbita casi polar (Tabla 2.3) y fue operado por sólo 106 días, entre el 26 de junio y el 9 de octubre de 1978. Sin embargo, fue de gran importancia en la percepción remota, mediante éste se llevó el primer sensor de imágenes de microonda activo conducido en el espacio, un Radar de Apertura Sintética SAR (Synthetic Aperture Radar). Muy diferente a los sensores de imágenes ópticas tales como MSS y TM, los cuales son dispositivos pasivos que detectan la energía solar reflejada o la energía térmica emitida, el Seasat SAR generó su propia energía de microonda y detectó la cantidad de esta energía que fue retornada al sensor desde la superficie de la tierra. Se usaron microondas con una longitud de onda de 23.5 cm. Un sensor de microonda activo, pudo utilizarse independientemente de las condiciones de iluminación solar y así el Seasat SAR pudo ser operado tanto de día como de noche. Por otro lado, la radiación de microonda a esas longitudes de onda, es capaz de introducirse entre las nubes y de proporcionar imágenes de la superficie de la tierra bajo cualquier condición del clima.

SAR(Synthetic Aperture Radar)	Satelite Seasat
Frecuencia: 1274.8 GHz	Órbita: circular, quasi polar
Longitud de onda: 23.5 cm (banda L)	Altitud: 790 km
Polarización: HH	Inclinación: 108 °.
Ángulo de depresión: 20 °	Periodo: 101 minutos
Resolución espacial: 25 m	Ciclo de repetición: 152 días

Tabla 2.3 Satélite Seasat y sensor SAR.

Un radar se usa normalmente en el modo mira-lateral y el ángulo de depresión del instrumento es uno de los parámetros más importantes. El Seasat SAR usó un ángulo de 20 ° (17.5 ° - 22.5 ° de la vertical). Es principalmente la relación entre el ángulo de depresión, la longitud de onda del radar y la rugosidad del objetivo, que determina la magnitud del retorno, el cual se expresa en términos de un valor de la escala de gris en la imagen. La principal diferencia entre un radar a bordo de una nave aérea y uno de mira-lateral (SLAR o Side Looking Airborne Radar) y el Seasat SAR, es el hecho de que para alcanzar una resolución espacial razonable de altitudes orbitales, sería necesaria una antena muy grande. El Seasat SAR produjo imágenes con una resolución espacial de 25 m; para esto se necesitaría una longitud de antena del orden de varios kilómetros, si se hubiera usado un radar convencional "fuerza bruta". La antena cargada por Seasat, medía solamente 12 m de longitud, pero se usó el procesamiento matemático de retornos de radar, para simular el efecto de una antena muy grande, tomando en cuenta el movimiento hacia adelante del satélite. Así, las imágenes fueron generadas matemáticamente de una antena sintética o simulada y de aquí precisamente viene su nombre.

Desde Octubre de 1957, la humanidad explora el espacio poniendo en órbita terrestre lo que ya son muchos centenares de satélites. Cada uno de estos vehículos cósmicos han revelado una pequeña parte de los secretos de nuestro planeta. Los satélites lanzados hasta ahora nos han ayudado a entender las características físicas del entorno próximo y las remotas regiones que se hallan entre las estrellas. Han incrementado también los conocimientos sobre nuestro planeta escudriñando los recursos naturales escondidos en los océanos y el subsuelo y revelando el complejo mecanismo de la atmósfera. Ciertos satélites se han convertido además en valiosas herramientas cotidianas, tal es el caso de los satélites de percepción remota como los mencionados aquí.

La percepción remota se considera hoy en día una desarrollada forma tecnológica para la adecuada explotación de los recursos naturales, es una actividad que comprende la administración de terrenos urbanos, la producción y control sanitario de cosechas y la identificación de depósitos minerales. La percepción remota como las telecomunicaciones es una de las actividades espaciales más importantes.

CAPÍTULO 3

CAMPOS ALEATORIOS DE MARKOV.

Un Campo Aleatorio de Markov MRF (Markov Random Field) es un proceso estocástico discreto cuyas propiedades globales se controlan por medio de las propiedades locales. Éstas están definidas por las probabilidades condicionales locales. Su existencia como procesos estocásticos discretos en enrejados infinitos, se mostró en 1968 [50]. En la última década su uso se ha proliferado en las ciencias de la ingeniería [51], especialmente en procesamiento de imágenes [52][53].

La manera más común de definir los MRFs relacionados con modelos de imágenes es definirlos en un enrejado. Sin embargo, aquí se definirán más generalmente como gráficos. Considérese un gráfico $\mathcal{G} = (S, E)$ donde $S = \{s_1, s_2, \dots, s_N\}$ es un grupo de vértices (o sitios) y E es el grupo de bordes.

3.1 Nociones Topológicas.

3.1.1 Vecinos.

Dos puntos s_i y s_j son vecinos si hay un borde $e_{i,j} \in E$ conectándolos. El grupo de puntos vecinos de un sitio s se denota por γ_s .

3.1.2 Sistema de vecindad.

$\gamma = \{\gamma_s, /s \in S\}$ es un sistema de vecindad para φ si

- (i) $s \notin \gamma_s$,
- (ii) $s \in \gamma_r \Leftrightarrow r \in \gamma_s$,

Los sistemas de vecindad más usuales para dimensión 2D se muestran en la figura 3.1.

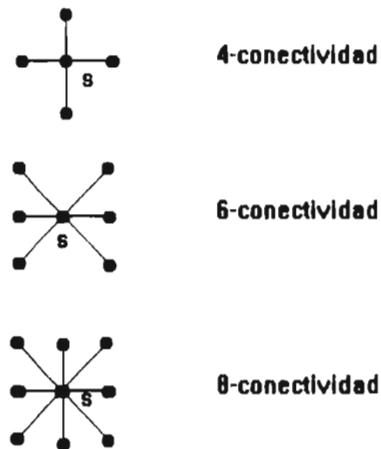


Figura 3.1 Sistemas de vecindad en dimensión $d=2$ [55].

En los modelos 3D se manejan 6-,18- y 26-conectividad.

3.1.3 Clique.

Un subgrupo $C \subseteq S$ es un clique si cada par de sitios distintos en C son vecinos. \mathbf{C} denota el grupo de cliques y $deg(\mathbf{C}) = \max_{c \in C} C_i$. La figura 3.2 muestra los cliques asociados a diferentes sistemas de vecindad en dimensión $d=2$.

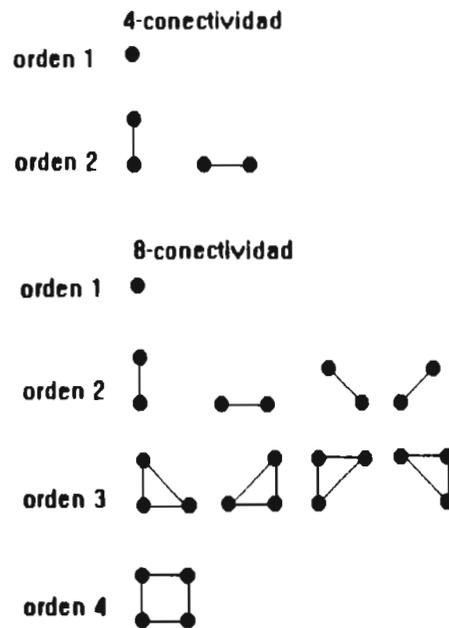


Figura 3.2 Cliques asociados a diferentes sistemas de vecindad en dimensión $d=2$ [55].

3.2 Teorema de Hammersley-Clifford.

3.2.1 Distribución de Gibbs.

El interés del teorema fundamental de Hammersley-Clifford es el de relacionar las formas globales y locales de un Campo de Markov. Este teorema es el siguiente [51]:

Suponiendo el conjunto de sitios S finito o enumerado. Todo Campo de Markov limitado (tamaño del vecindario limitado) tal que el espacio de los niveles de gris Ω es discreto y la probabilidad estrictamente positiva en el espacio de las configuraciones E es una campo de Gibbs:

$$P(X = x) = \frac{1}{Z} \exp - U(x) \quad (3.1)$$

con:

$$U(x) = \sum_{c \in \mathcal{C}} U_c(x) \quad (3.2)$$

donde cada término $U_c(x)$ es la restricción del clique c de una función llamada potencial del clique.

Así para los modelos en 2D y en 4-conectividad (cliques de orden 1 y 2), todo potencial se escribe:

$$U(x) = \sum_{c=(s) \in \mathcal{C}_1} U_c(x_s) + \sum_{c=(r,s) \in \mathcal{C}_2} U_c(x_r, x_s) \quad (3.3)$$

La función multi-variables $U(x)$ se desarrolló pues de manera muy específica y "limitada". Sin embargo, la expresión de los potenciales del clique U_c puede depender de la localización del clique considerado, lo que permite incluir naturalmente el aspecto no-estacionario de las funciones de energía en el formalismo markoviano [54]. Un caso común consiste por ejemplo en adoptar una dependencia paramétrica lineal de los potenciales del clique de orden 2, así:

$$U_c(r, s, x_r, x_s) = \alpha(r, s)W(x_r, x_s) \quad (3.4)$$

Se tomará por la serie $\sum_{(r,s)} U_c(x_r, x_s)$ la suma de los términos del clique de orden 2, a fin de disminuir las notaciones.

3.2.2 Distribución de Gibbs con temperatura.

Se introduce un parámetro T de temperatura en la expresión de la distribución de Gibbs:

$$P(X = x) = \frac{1}{Z(T)} \exp - \frac{U(x)}{T}, \quad (3.5)$$

la función de normalización se escribe en tal caso:

$$Z(T) = \sum_{x \in \mathcal{E}} \exp - \frac{U(x)}{T} \quad (3.6)$$

El comportamiento de la distribución de Gibbs cuando $T \rightarrow 0$ y $T \rightarrow \infty$ se basa en lo siguiente:

$$\forall x_1, x_2 \in \mathcal{E}, \frac{P(X = x_2)}{P(X = x_1)} = \exp - \frac{U(x_2) - U(x_1)}{T} \quad (3.7)$$

La distribución P converge en tal caso en una temperatura alta, hacia la ley equiparable en E :

$$T \rightarrow \infty, \text{ i.e. } T \gg \sup_{x_1, x_2 \in E} (|U(x_2) - U(x_1)|) \Rightarrow \lim P(X = x) = \frac{1}{\text{Card}(E)} \quad \forall x \in E.$$

mientras que una temperatura base, se "concentra" en las configuraciones de baja energía:

$$T \rightarrow 0, \text{ i.e. } T \ll \inf_{x_1, x_2 \in E} (|U(x_2) - U(x_1)|) \Rightarrow \lim P(X = x) = \frac{1}{\text{Card}(E^*)} \quad \forall x \in E^*, = 0 \quad \text{si no}$$

donde E^* es el espacio-bajo de las configuraciones x alcanzando el mínimo global de la energía.

Estas propiedades son la base del método de "recocido simulado" [56].

3.3 Segmentación markoviana (análisis bayesiano).

El razonamiento utilizado para la segmentación es el de asignar una etiqueta a partir del nivel de gris de un pixel conocido:

$$X \rightarrow L \tag{3.8}$$

Se define la probabilidad a posteriori de dicho proceso:

$$P(L = l | X = x) = \frac{P(X = x | L = l)P(L = l)}{P(X = x)} \tag{3.9}$$

y analizando el producto

$$P(X = x | L = l)P(L = l) \tag{3.10}$$

el cual puede ser interpretado como la distribución conjunta $P(X = x, L = l)$. El primer término de este producto es la probabilidad de realización de una configuración conociendo su etiquetado. En la hipótesis de independencia de los sitios para las probabilidades condicionales se escribe:

$$P(X = x, L = l) = \prod_{\lambda} P(X_{\lambda} = x_{\lambda} / L_{\lambda} = l_{\lambda}) \quad (3.11)$$

Este término describe el conocimiento del histograma condicional de cada zona de interés en la imagen. Suponiendo a cada región con media de intensidad uniforme a además degradada por un ruido blanco gaussiano no correlacionado y cuya varianza depende de cada región:

$$P(X_{\lambda} = x_{\lambda} / L_{\lambda} = l_{\lambda}) = \frac{1}{\sqrt{2\pi\sigma_{l_{\lambda}}}} \exp\left(-\frac{(x_{\lambda} - \mu_{l_{\lambda}})^2}{2\sigma_{l_{\lambda}}^2}\right) \quad (3.12)$$

Incorporando la hipótesis markoviana en la probabilidad a priori de clases $P(L)$, ésta puede ser escrita mediante la distribución de Gibbs.

$$P(L = l) = \frac{1}{Z} \exp(-U(l)) \quad (3.13)$$

Un caso frecuente es adoptar un modelo de Potts o de Ising para definir la energía $U(l)$. Similarmente a la ec. 3.13, $P(X = x / L = l)$ puede expresarse por:

$$P(X = x / L = l) = \prod_{\lambda} P(X_{\lambda} = x_{\lambda} / L_{\lambda} = l_{\lambda}) = \prod_{\lambda} \exp(-U(x_{\lambda} / l_{\lambda})) \quad (3.14)$$

La función de energía total asociada es:

$$U(x/l) = \sum_{\lambda} U(x_{\lambda} / l_{\lambda}) \quad (3.15)$$

de la ec. 3.12, la energía se expresa:

$$a^x = \exp(x \ln a)$$

$$\left(\sqrt{2\pi\sigma_{l_{\lambda}}}\right)^{-1} = \exp\left(-\ln \sqrt{2\pi\sigma_{l_{\lambda}}}\right)$$

$$P(X_{\lambda} = x_{\lambda} / L_{\lambda} = l_{\lambda}) = \exp\left(-\ln \sqrt{2\pi\sigma_{l_{\lambda}}}\right) \exp\left(-\frac{(x_{\lambda} - \mu_{l_{\lambda}})^2}{2\sigma_{l_{\lambda}}^2}\right)$$

$$= \exp\left(-\left(\frac{(x_{\lambda} - \mu_{l_{\lambda}})^2}{2\sigma_{l_{\lambda}}^2} + \ln \sqrt{2\pi\sigma_{l_{\lambda}}}\right)\right)$$

entonces

$$U(x|l) = \sum_i \left(\frac{(x_i - \mu_{l_i})^2}{2\sigma_{l_i}^2} + \ln \sqrt{2\pi\sigma_{l_i}} \right) \quad (3.16)$$

$$U(l|x) = U(x|l) + U(l) \quad (3.17)$$

la segmentación es pues la optimización de:

$$\arg \min_l U(l|x) = U(x|l) + U(l)$$

3.4 El recocido simulado (simulated annealing).

La teoría de campos de Markov no es únicamente una manera de analizar y representar imágenes, si no que está también vinculada a un método específico de optimización: el recocido simulado [56]. Este algoritmo nos permite resolver el problema de estimación en sentido del estimador *Máximo a Posteriori* MAP (*Maximum A Posteriori*). Además nos asegura, si se obedece a sus condiciones, el alcanzar un global óptimo. Sabemos que un campo de Markov se caracteriza por una energía $U(X)$, y puesto que la ley del campo se caracteriza por una ley de Gibbs:

$$P(X) = \frac{e^{-U(X)}}{Z} \quad (3.18)$$

Ahora se introduce un parámetro de temperatura $T(n)$, entonces la probabilidad de Gibbs se escribiría:

$$P(X) = \frac{e^{-\frac{U(X)}{T(n)}}}{Z(T(n))} \quad (3.19)$$

El objetivo al optimizar una imagen es el de encontrar un elemento de Ω_0 donde:

$$\Omega_0 = \{\omega \in \Omega\}. \quad (3.20)$$

la medida de la probabilidad en este conjunto es una medida π_0 constante en Ω_0 y nula además. Se denota U^* , la energía máxima en todas las configuraciones y U_* , la energía mínima, entonces $\Delta = U^* - U_*$ será la diferencia entre estas energías. Esto conlleva a un muestreador de Gibbs, con una temperatura $T(n)$ que varía después de cada barrido de la imagen. La variación de la temperatura deberá seguir el esquema siguiente:

(a) $T(n) \rightarrow 0$ cuando $n \rightarrow \infty$

(b) $T(n) \geq N\Delta / \log(n) \quad \forall n \geq 2$

en tal caso, para toda configuración (imagen) inicial η y para toda configuración (imagen) ω :

$$\lim_{n \rightarrow \infty} P(X(n) = \omega / X(0) = \eta) = \pi_0 \quad (3.21)$$

Esto significa que aplicando un muestreador de Gibbs, al problema de estimación, se está variando la temperatura después de cada barrido de la imagen, el esquema asegura la convergencia a una verdadera configuración óptima y no sólo a un óptimo local. Nótese que en este método de optimización, se permite la elección de las configuraciones presentes en la energía, pero que el decremento en la temperatura concentra la medida de la probabilidad poco a poco en el conjunto de configuraciones óptimas. Todos los métodos clásicos de optimización, operan aún por el decremento de la energía y todos llegan invariablemente a un óptimo local.

Con el método de recocido simulado, se tomará en cuenta la lentitud del esquema y la necesidad de manejar la diferencia entre la energía máxima y la energía mínima, Δ .

3.5 Modelo de Ising y regularización de imágenes binarias.

El modelo de Ising se desarrolló en 1925 en el marco de la teoría del ferromagnetismo para analizar las propiedades de transición de fase de ciertos sólidos. En el modelo de enrejado isotrópico y en 4-conectividad, un spin de un sitio s se integra con sus vecinos en 4-conectividad (denotados por r), por una constante de acoplamiento J y con un campo magnético B , de modo que la energía del sitio s condicionado a sus vecinos se escribirá [55]:

$$W_s = -J \sum_r \sigma_r \sigma_s - B \sigma_s \quad (\sigma_s = \pm 1) \quad (3.22)$$

La energía total E del enrejado es la suma de los términos de las iteraciones sitio-sitio (cliques de orden 2) y de sus términos de iteración sitio-campo magnético (cliques de orden 1). Una forma equivalente de la función de energía es la forma de Heisenberg:

$$E = \frac{J}{2} \sum_{r,s} (\sigma_r - \sigma_s)^2 + \frac{B}{2} \sum_r (\sigma_r - h)^2 \quad (3.23)$$

también:

$$E = \frac{J}{2} \sum_{(r,s)} \|\bar{u}_r - \bar{u}_s\|^2 + \frac{B}{2} \sum_r \|\bar{u}_r - \bar{b}\|^2 \quad (3.24)$$

la primera sumatoria es la de los cliques de orden 2 y la segunda es la de los cliques de orden 1. donde \bar{u}_r , son los vectores unitarios asociados con las direcciones de los spins en cada sitio y \bar{b} , es el vector unitario asociado al sentido del campo magnético.

Recordando ahora la demostración en [56], se establece un vínculo entre la regularización de una imagen binaria y el modelo de Ising. Es la fuente de toda extensión que se efectúa para las imágenes multiclases. Considerando una imagen, la cual se quiere regularizar $f^0(x, y)$ como una derivada continua en un dominio \mathbf{R}^2 , se busca la solución al problema.

$$\min E = E_1 + E_2$$

donde

$$E_1 = \lambda \|f - f^0\|^2 = \lambda \iint_{\text{imagen}} (f(x, y) - f^0(x, y))^2 \quad dx dy \quad (3.25)$$

es el término de la diferencia en la imagen original

$$E_2 = \mu \|\nabla f\|^2 = \mu \iint_{\text{imagen}} \left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2 \quad dx dy \quad (3.26)$$

es el término de la regularización de la imagen. E es la energía total de la imagen. Discretizando la imagen,

$$E_1 = \lambda \sum_{i,j} \|f_{i,j} - f^0_{i,j}\|^2 \quad (3.27)$$

$$E_2 = \mu \left(\sum_{i,j} \|f_{i,j} - f_{i+1,j}\|^2 + \|f_{i,j} - f_{i,j+1}\|^2 \right) \quad (3.28)$$

Cuando la imagen es binaria: $f_{i,j} = 0$ ó $1 \forall i, j$ asociando a cada pixel su "variable de spin" $\sigma_{i,j}$.

$$f_{i,j} = \frac{1 + \sigma_{i,j}}{2} \quad (\sigma_{i,j} = \pm 1) \quad (3.29)$$

Retomando las ecuaciones (3.25-3.26),

$$E = E_1 + E_2 \quad (3.30)$$

$$E_1 = \frac{\lambda}{4} \sum_{i,j} \|\sigma_{i,j} - \sigma^0_{i,j}\|^2 \quad (3.31)$$

$$E_2 = \frac{\mu}{4} \left(\sum_{i,j} \|\sigma_{i,j} - \sigma_{i+1,j}\|^2 + \|\sigma_{i,j} - \sigma_{i,j+1}\|^2 \right) \quad (3.32)$$

Comparando con ec. 3.24, se observa que la energía total E es la de un modelo de Ising 2D isotrópico de constante de acoplamiento entre sitios $J = \frac{\mu}{2}$ en presencia de un campo magnético de módulo constante $B = \frac{\lambda}{2}$ pero no-homogéneo y cuyo sentido en cada pixel depende de su valor binario original, es decir, un campo magnético casi "homogéneo por fragmentos" si la imagen de salida no es demasiado ruidosa.

3.6 Restauración de imágenes multiclase y modelo de Potts.

El potencial de cliques de orden 2 actualmente más utilizado [57], [58], [59], [60], [61], en los métodos de regularización de imágenes es:

$$V(r, s) = \begin{cases} -\beta & \text{si } l_r = l_s \\ +\beta & \text{si } l_r \neq l_s \end{cases} \quad (3.33)$$

(suponiendo 4-conectividad).

Se discute una constante cercana al modelo estandar de Potts en física estadística:

$$V(r, s) = -K\delta(l_r, l_s) \quad (K = 2\beta) \quad (3.34)$$

δ es el símbolo de Kronecker definido por:

$$\delta(x_1, x_2) = \begin{cases} 1. & x_1 = x_2 \\ 0. & \text{de otro modo} \end{cases}$$

Este modelo ha sido ampliamente estudiado por sus propiedades de transición de fase [62], las cuales son ahora conocidas y que dependen del número de clases q . El modelo de Potts coincide con el modelo de Ising cuando $q = 2$, de acuerdo a la fórmula:

$$\delta(I, J) = \frac{1 + \sigma_I \sigma_J}{2}.$$

Entonces $K_{Potts} = 2 J_{Ising}$, los valores σ_i son las variables de spin ± 1 asociados a los valores binarios de los píxeles. Es entonces una extensión cuando q es cualquier otro número. Partiendo de la siguiente ecuación:

$$\delta(m, n) = \frac{1}{q} (1 + (q-1) \bar{u}_m \cdot \bar{u}_n) \quad (3.35)$$

donde m y n son dos enteros dentro del intervalo $[0, \dots, q-1]$ y \bar{u}_i son los q vectores unitarios "uniformemente repartidos" en el vértice de un hipertetraedro de \mathbf{R}^{q-1} (figura 3.3), es decir, tales que:

$$\sum_{k=0}^{q-1} \bar{u}_k = \bar{0} \quad \text{y} \quad \bar{u}_m \cdot \bar{u}_n = -\frac{1}{q-1} \quad \text{si} \quad m \neq n \quad (3.36)$$

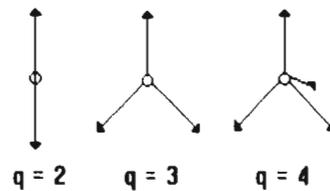


Figura 3.3 Modelo de Potts para diferentes números de clases [55].

Cuando $q = 2$, es el caso del modelo de Ising, se obtienen dos valores de spin de direcciones opuestas. Los vectores \bar{u}_i son unitarios, la función de energía obtenida en las ecuaciones 3.30, 3.31, 3.32, se extiende a la siguiente forma:

$$E = E_1 + E_2 \quad (3.37)$$

$$E_1 = \lambda \sum_{\gamma, \gamma'} \left\| \bar{u}_{i, \gamma} - \bar{u}_{i, \gamma'} \right\|^2 \quad (3.38)$$

$$E_s = \mu \left(\sum_{i,j} [\bar{u}_{i,j} - \bar{u}_{i,j+1}]^2 + [\bar{u}_{i,j} - \bar{u}_{i+1,j}]^2 \right) \quad (3.39)$$

El contenido escalar de la imagen $f(x,y)$ ha sido reemplazado por una expresión vectorial en \mathbf{R}^{q-1} , y la fórmula de la energía obtenida es una generalización de $q-1$ dimensiones del modelo de Heisenberg. La energía condicional en un sitio s tiene la forma:

$$W_s = -K \sum_r \delta(l_r, l_s) - B \delta(l_s, l_s^0) \quad (3.40)$$

$$\text{donde } \lambda = \frac{q-1}{q} \frac{B}{2} \quad \text{y} \quad \mu = \frac{q-1}{q} \frac{K}{2} \quad (3.41)$$

las notaciones están dadas en las ecuaciones 3.25 y 3.26.

La fórmula 3.40 que es probablemente el modelo más simple asociado a la restauración de imágenes de clases, es independiente a toda "relación de orden" entre distintas etiquetas o de todo conocimiento *a priori* en las regiones.

Se ha definido aquí una serie de conceptos imprescindibles para comprender la utilidad y la aplicación de los campos aleatorios de Markov MRF en el procesamiento de imágenes. En el siguiente capítulo se describen dos esquemas de segmentación de imágenes usando MRF, por lo que se consideró pertinente dedicar el presente espacio al desarrollo matemático de conceptos tales como: distribución de Gibbs, análisis bayesiano de la segmentación, recocido simulado, modelo de Ising y modelo de Potts.

CAPÍTULO 4

TÉCNICAS DE SEGMENTACIÓN MARKOVIANA.

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

- La segmentación debe ser completa, es decir, todos los píxeles deben pertenecer a una clase.
- Los píxeles en dentro de una región deben estar conectados.
- Los regiones deben ser disjuntas.

Nótese que el segundo término significa que los píxeles cercanos deben pertenecer a la misma región. Esta restricción puede expresarse en términos de un campo aleatorio de Markov MRF. Para ser más precisos, lo que se intenta es asignar una etiqueta a cada pixel.

En este capítulo se describen dos métodos de segmentación basados en la teoría de los campos aleatorios de Markov MRF, de los cuales se han obtenido resultados muy confiables para la detección de frontera en diferentes imágenes de percepción remota.

4.1 Modelo de Potts clásico.

Al utilizar MRF, las propiedades globales de las imágenes se controlan solamente por las propiedades locales de las mismas. El objetivo del proceso de segmentación es asociar a cada pixel de los datos (imagen original) una etiqueta de un grupo finito [63]. En el sentido probabilístico este método consiste en definir un modelo de un MRF por medio de potenciales de cliques y seleccionar la etiqueta más probable usando un estimador *Máximo A Posteriori* (*Maximum A Posteriori* MAP).

Se denota X la imagen correspondiente a los datos (imagen original) y Y la imagen segmentada. La segmentación consiste en maximizar la probabilidad condicional $P(Y|X)$, la cual, del teorema de Bayes, es proporcional a $P(X|Y)P(Y)$. $P(Y)$ está referido al modelo a *priori* mientras que $P(X|Y)$ está referido al término de dato-direccionado. En estudios prácticos, el modelo probabilístico depende del análisis de las energías de los cliques. Esto da como resultado, funciones de energía relacionadas a la función de potencial (ecuación 4.1).

$$P(\Sigma) = \frac{1}{Z} \exp[-U(\Sigma)] = \frac{1}{Z} \exp\left[-\sum_{c \in C} V_c(\sigma_i, i \in c)\right], \quad (4.1)$$

donde:

C es el grupo de cliques relacionados a un sitio dado.

V_c es la función de potencial asociada con el clique c .

Z es una constante de normalización llamada función de partición.

El primer término probabilístico (el modelo a *priori*) es frecuentemente compuesto por un modelo de Potts [64]. La función de energía asociada con este modelo se escribe:

$$U(Y) = \sum_{c=\{i,j\}} \beta \delta_{y_i, y_j} \quad (4.2)$$

Los cliques tienen dos elementos. El coeficiente β define las propiedades de homogeneidad de la solución, esto es, entre más grande sea este término mayor será la probabilidad de que dos píxeles adyacentes tengan la misma etiqueta.

El término de dato-direccionado se define mediante funciones de costo que dependen de la etiqueta l y se denotan por f^l . Los parámetros de entrada se extraen directamente de los datos (imagen original). El potencial asociado, aplicado a los cliques de primer orden se escribe como:

$$U(X|Y) = \sum_{c=\{i\}} \sum_l f^l(x_i) \delta_{y_i, l} \quad (4.3)$$

donde x_i y y_i son los valores de el dato y la etiqueta respectivamente en el sitio i .

El problema de la segmentación consiste en minimizar la energía global:

$$U(X|Y) + U(Y) \quad (4.4)$$

En el marco del estimador MAP, el proceso de minimización se realiza usando una técnica estocástica tal como el recocido simulado [65].

La función de costo lineal se define para cada clase (figura 4.1).

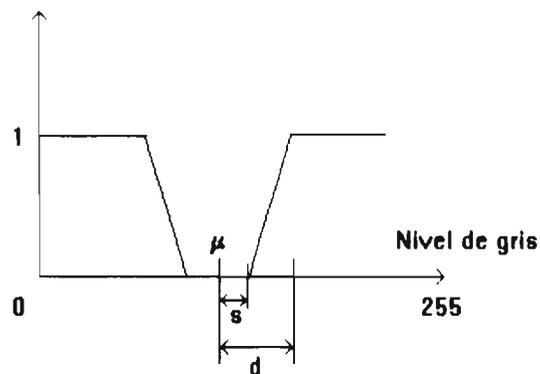


Figura 4.1 Función de costo definiendo el término de dato-direccionado para una clase típica [63].

Los diferentes parámetros asociados reflejan el histograma para cada clase. Para definir cada función de costo, se tiene que estimar la media μ_l de la clase l y dos parámetros de desviación, s_l y d_l . Si la distancia entre un pixel del dato dado y la media μ_l es más pequeña que s_l , este pixel pertenece muy probablemente a la clase l . En cambio, si la distancia entre el pixel del dato y la media de la clase l es más grande que d_l , la probabilidad de que el pixel pertenezca a la clase l es muy baja.

De acuerdo a la función de energía de la ecuación 4.4, la energía asociada se escribe:

$$U(Y) + U(X|Y) = \sum_{c \in \{1,2\}} \beta \delta_{1,c} + \sum_{c \in \{1,2\}} \sum_l f'(x_l) \delta_{1,c,l} \quad (4.5)$$

El mínimo global de esta energía se obtiene mediante un proceso de recocido simulado. Considerando el enfriamiento, la temperatura final $\lim_{k \rightarrow \infty} T_k = 0$, es aproximada por la regla del incremento geométrico $T_{k+1} = \tau T_k$, donde k denota la k -ésima transición y τ es la razón de decremento. Como se señala en [66], para estimar el mínimo global de las funciones de energía τ debe ser menor que 1. En este trabajo $\tau = 0.95$.

Diferentes tipos de parámetros están involucrados en este método. Primero, los términos de dato-direccionado manejan parámetros que dependen de la escena misma. Su estimación puede realizarse por medio de áreas de entrenamiento o analizando el histograma cuando las diferentes clases pueden separarse fácilmente. Aquí, se utilizó el modelo de k -medias [67] (se procesó en khoros), para la estimación de los centros de agrupamiento de los términos de dato-direccionado. Los parámetros del modelo *a priori* se fijan por estudios empíricos pero se mantienen robustos para aplicaciones dadas.

En la figura 4.2 se muestra una fotografía aérea que será clasificada por este esquema. Visualmente es posible distinguir las diversas regiones que componen la escena, tales como árboles, calles, pasto, casas, etc.

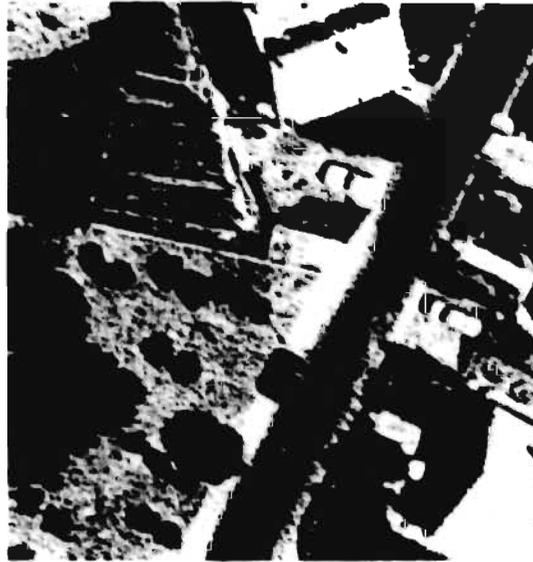


Figura 4.2 Foto aérea.
(Tamaño 256 x 256 pixeles¹. Tamaño del pixel: 15 cm.)

Los parámetros de entrada para el algoritmo de segmentación son:

Número de clases:	3 (se obtiene del método de k -medias).
El coeficiente β :	0.3 (se fijó después de un estudio empírico).
Las medias de las clases:	57,120,185 (se obtienen del método de k -medias).
Temperatura inicial:	2.0 (se fijó después de un estudio empírico).
Razón de decremento de T:	0.95 (debe ser menor que 1).
Número de iteraciones:	150 (se fijó después de un estudio empírico).

¹Esta imagen se ha ampliado para poder apreciar los cambios al efectuar la segmentación.

Inseguida se muestra una serie de imágenes (figura 4.3a-d) que ilustran como evoluciona la imagen segmentada, la cual es inicialmente una imagen que se conforma de píxeles aleatorios con valores relacionados con las diferentes clases.

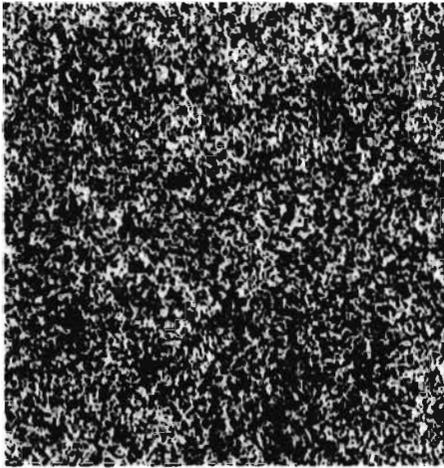


Figura 4.3a Imagen segmentada.
(Iteración no. 1).
(Tamaño: 256 x 256 píxeles)

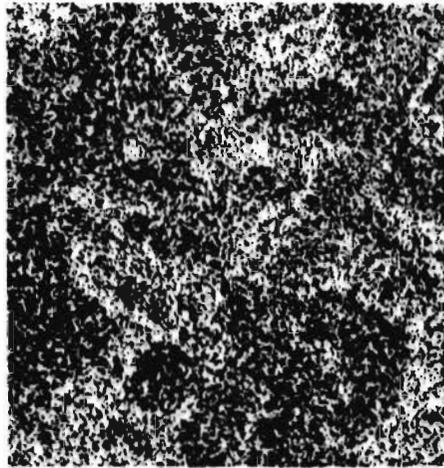


Figura 4.3b Imagen segmentada
(Iteración no. 5).
(Tamaño: 256 x 256 píxeles)



Figura 4.3c Imagen segmentada.
(Iteración no. 10).
(Tamaño: 256 x 256 píxeles)



Figura 4.3d Imagen segmentada
(Iteración no. 15).
(Tamaño: 256 x 256 píxeles)



Figura 4.3e Imagen segmentada.
(Iteración no. 50).
(Tamaño: 256 x 256 píxeles)



Figura 4.3f Imagen segmentada
(Iteración no. 150).
(Tamaño: 256 x 256 píxeles)

4.1 Modelo de Potts matricial.

En el modelo anterior, se mencionó que el esquema de k -medias fue utilizado para conocer el número de clases y las medias de estas clases, sin embargo, la segmentación partía de la imagen X , la cual contiene los datos de la imagen original. También se describió el coeficiente de homogeneidad β como un solo valor para todas las clases. En el modelo de Potts matricial lo que se intenta es implementar un método post-clasificación del esquema de k -medias, es decir, tomar el resultado de esta clasificación (figura 4.4) y mejorar la definición en las regiones homogéneas, así como minimizar la presencia de píxeles mal clasificados.



Figura 4.4 Segmentación por k -medias.
(Tamaño 256 x 256 píxeles¹).

Se crea entonces una matriz M , que para la imagen de la figura 4.4, la cual está segmentada por k -medias para tres clases, quedaría de la siguiente forma:

$$M = \begin{pmatrix} \beta_{1,1} & \beta_{1,2} & \beta_{1,3} \\ \beta_{2,1} & \beta_{2,2} & \beta_{2,3} \\ \beta_{3,1} & \beta_{3,2} & \beta_{3,3} \end{pmatrix}.$$

asumiendo que $\beta(i,j) = \beta(j,i)$, se tienen ahora B valores de β . Donde B se define como:

$$B = \frac{n(n-1)}{2} + n, \quad (4.6)$$

y n es el número de clases presentes en la imagen.

Para nuestro ejemplo:

$$\beta_{11} = \beta_{12} = \beta_{21}$$

$$\beta_{22} = \beta_{23} = \beta_{32}$$

$$\beta_{33} = \beta_{31} = \beta_{13}$$

¹Esta imagen se ha ampliado para poder apreciar los cambios al efectuar la segmentación.

Por lo que la matriz ahora es:

$$M = \begin{pmatrix} \beta_{1,1} & \beta_{1,2} & \beta_{1,3} \\ \beta_{2,1} & \beta_{2,2} & \beta_{2,3} \\ \beta_{3,1} & \beta_{3,2} & \beta_{3,3} \end{pmatrix}$$

Se asume también que $\beta(1,1) = \beta(2,2) = \dots = \beta(n,n)$, entonces:

$$\beta = \beta_{1,1} = \beta_{2,2} = \beta_{3,3}$$

la matriz M finalmente es:

$$M = \begin{pmatrix} \beta & \beta_{1,2} & \beta_{1,3} \\ \beta_{2,1} & \beta & \beta_{2,3} \\ \beta_{3,1} & \beta_{3,2} & \beta \end{pmatrix}$$

El criterio para asignar los diferentes valores de β , es el siguiente:

Primeramente $\beta < 1$, de aquí se asignará un valor grande de β al par de clases cuya probabilidad de que una evolucione a la otra es máxima, en el caso contrario, se asigna un valor de β pequeño al par de clases cuya probabilidad de que una evolucione a la otra es mínima.

Se muestran aquí tres diferentes resultados de la segmentación, para cada uno de éstos se ha dado como parámetro de entrada una matriz M diferente para cada caso. En el primer caso (figura 4.5), se ha favorecido la adyacencia entre las tres clases, asignando a todas las β valores muy cercanos a 1. En el caso 2 (figura 4.6), se han asignado a las β valores de cero, de esta manera, se impide que las clases evoluciones entre sí, por lo que se tiene en la imagen segmentada resultante, la presencia de muchos pixeles aislados. El caso 3 (figura 4.7), es en el que, de acuerdo al conocimiento *a priori* que se obtiene de la segmentación por k -medias y después de un estudio empírico, se introduce una matriz M cuyos valores favorecen la evolución de unas clases a otras de una manera óptima, obteniéndose así una segmentación más certera.



Figura 4.5 Post-segmentación del algoritmo de k -medias.

$$M = \begin{pmatrix} 0.8 & 0.4 & 0.1 \\ 0.4 & 0.8 & 0.4 \\ 0.1 & 0.4 & 0.8 \end{pmatrix}$$

(Tamaño: 256 x 256 píxeles)



Figura 4.6 Post-segmentación del algoritmo de k -medias.

$$M = \begin{pmatrix} 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 \end{pmatrix}$$

(Tamaño: 256 x 256 píxeles)



Figura 4.5 Post-segmentación del algoritmo de k -medias.

$$M = \begin{pmatrix} 0.3 & 0.1 & 0.0 \\ 0.1 & 0.3 & 0.1 \\ 0.0 & 0.1 & 0.3 \end{pmatrix}$$

En este capítulo se describió el modelo de Potts clásico y el modelo de Potts matricial, con el objetivo de analizar y remarcar que la segmentación markoviana es una de las herramientas más utilizadas dentro del procesamiento de imágenes para la detección de frontera en imágenes de percepción remota. Además los resultados obtenidos se utilizan como referencia para compararlos con los resultados obtenidos de la segmentación por medio de estructuras de árbol (quadtree), que se describirá en el capítulo 5. Las comparaciones y conclusiones se discutirán en el capítulo 6.

CAPÍTULO 5

MÉTODO DE ESTRUCTURAS DE ÁRBOL (QUADTREE).

Para resolver el problema de resolución espacial alta y resolución espacial baja, se adopta un método multi-resolución. La información obtenida en cada resolución (nivel de quadtree), se usa para controlar el procesamiento en el nivel más bajo del quadtree (resolución máxima).

Específicamente, usando una estructura de quadtree [68], se aplica un algoritmo de agrupamiento (clustering) no paramétrico a la resolución espacial óptima mas baja (nivel óptimo más alto del quadtree). El algoritmo seleccionado no requiere de conocimiento *a priori* del número de clases presentes, por lo tanto aumenta la generalidad del esquema de segmentación.

Utilizando la información de clases obtenida del agrupamiento, se realiza un proceso de estimación de frontera. En cada nivel del quadtree se lleva a cabo una segmentación. La segmentación en el nivel actual depende de la información de la frontera obtenida del nivel previo.

El razonamiento detrás de este método multi-resolución es que la frontera entre un par de regiones homogéneas debe aparecer en cada nivel de quadtree, dado que cada región es significativamente más grande que el operador-promedio usado para crear el quadtree. Así, si una frontera puede localizarse en la resolución espacial óptima más baja, esto facilita la búsqueda de frontera en el siguiente nivel de resolución.

La figura 5.1 representa el procedimiento de segmentación y las siguientes secciones describen cada etapa a detalle.

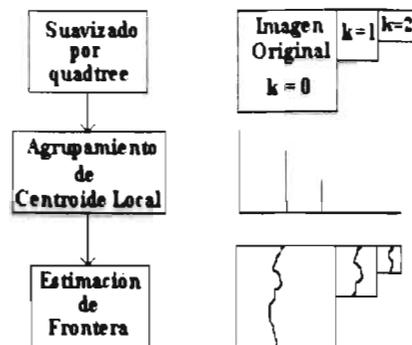


Figura 5.1 Etapas de la Segmentación [69].

5.1 Suavizado por quadtree.

Si se considera una imagen de $N \times N$ píxeles, se define $d(i, j)$ para $0 \leq (i, j) < N$ y $N = 2^m$ (m es el número de niveles del quadtree). El quadtree de esta imagen para el nivel k está definido como:

$$q(i, j, k) = \frac{1}{4} \left[q(2i, 2j, k-1) + q(2i+1, 2j, k-1) + q(2i, 2j+1, k-1) + q(2i+1, 2j+1, k-1) \right] \quad (5.1)$$

donde:

$$0 < k \leq m, \quad 0 \leq (i, j) < 2^{m-k}$$

y

$$q(i, j, 0) = d(i, j)$$

Se observa que un quadtree está basado en el promedio de un bloque de 2×2 . El nivel inmediato al nivel base (imagen original), consiste de nodos que representan bloques de píxeles de 2×2 de la imagen original no-translapados, entonces el tamaño de este nivel es $2^{m-1} \times 2^{m-1}$. Este proceso puede repetirse hasta alcanzar el nodo raíz cuyo valor es la media del nivel de gris de la imagen original.

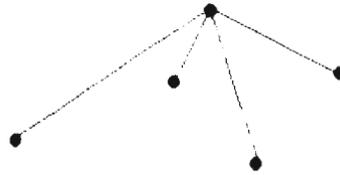


Figura 5.2 Un nodo padre con sus cuatro hijos.

La figura 5.2 representa un nodo padre enlazado a sus cuatro hijos. La figura 5.3 muestra una imagen con distintas regiones, cada una con diferente media de nivel de gris y su histograma. En 5.4a-c se muestran los 3 niveles siguientes del quadtree de la misma imagen con sus respectivos histogramas, puede observarse en la secuencia de histogramas un incremento en la resolución de las clases.

Las estructuras de quadtree son útiles en segmentación de imágenes porque el proceso de promediación, el cual genera el siguiente nivel de quadtree, reduce la varianza de la señal en una región homogénea. Sin embargo, el proceso de suavizado también introduce una distorsión debido a la fusión de datos de diferentes regiones.

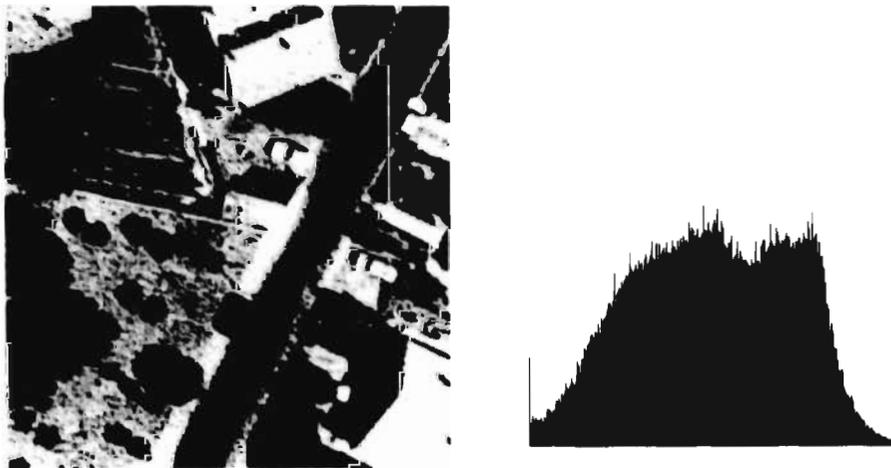
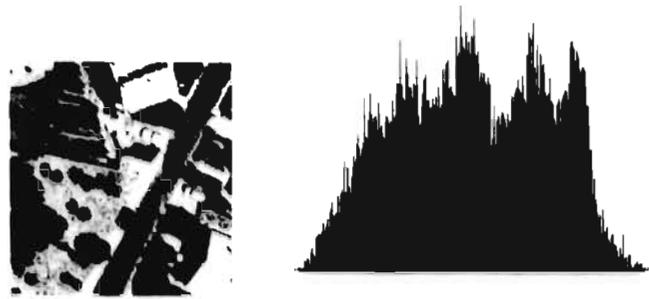
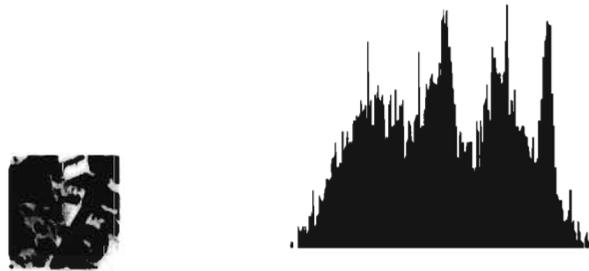


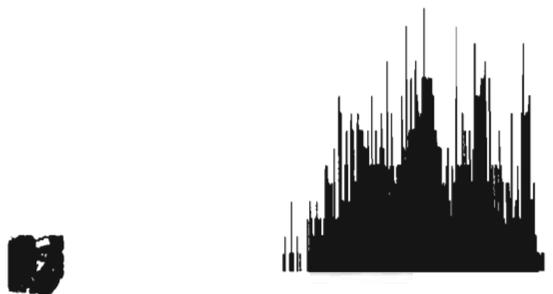
Figura 5.3 Foto aérea con distintas regiones y su histograma.
(Tamaño: 256 x 256 píxeles. Tamaño del píxel: 15 cm.)



(a) Nivel de suavizado $k=1$.
(Tamaño: 128 x 128 píxeles)



(b) Nivel de suavizado $k=2$.
(Tamaño: 64 x 64 píxeles)



(c) Nivel de suavizado $k=3$.
(Tamaño: 32 x 32 píxeles)

Figura 5.4a-c Efectos de suavizado por quadtree.

Ya que el último nivel de quadtree sería únicamente un píxel, es necesario truncar el proceso de suavizado. En [69] Spann y Wilson proponen un método para encontrar el nivel óptimo de quadtree, mediante una ganancia de suavizado. La elección de dicho nivel óptimo es de suma importancia para el proceso de segmentación, pues en este nivel es donde se llevará a cabo la primera clasificación, es por eso que en el presente trabajo se realiza la elección del nivel óptimo subjetivamente, visualizando la secuencia de histogramas del proceso de suavizado completo. Se elige aquel nivel cuyo histograma muestre una reducción de la varianza de la señal en regiones homogéneas, pero también debe tomarse en cuenta que la distorsión debida a la fusión de datos de distintas regiones sea mínima. Por citar un ejemplo, para la imagen de la figura 5.3, el nivel óptimo de suavizado es la imagen de la figura 5.4 b.

5.2 Agrupamiento de centróide local.

Para definir un centróide local, considere una densidad de probabilidad $p(x)$. El centróide local definido en cada punto x en el espacio de clase, está dado por:

$$\mu(x) = x + \frac{\int_{-w}^w x' p(x+x') dx'}{\int_{-w}^w p(x+x') dx'} \quad (5.2)$$

Esta ecuación establece que el centróide local del punto x es justamente el centro de masa de la distribución de probabilidad calculado sobre una ventana $2w$ y centrado en el punto x como se ilustra en la figura 5.5.

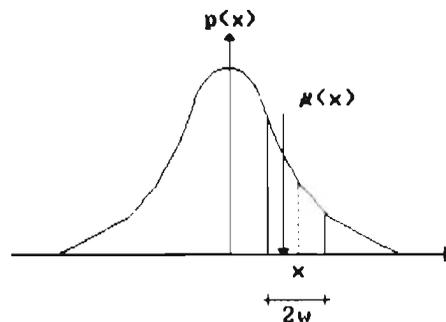


Figura 5.5 Cálculo del centróide local [69].

La ecuación 5.2 puede extenderse al espacio discreto, reemplazando las integrales por sumatorias. también puede ser vista como una operación de un filtro generalizado en la distribución de probabilidad, la respuesta impulso del filtro $w(x)$, sería una función "diente de sierra".

$$w(x) = x, \quad x < w$$

$$= 0, \quad \text{de otro modo} \tag{5.3}$$

Debido a que la distribución de probabilidad de una imagen entera no puede ser escrita como la suma de un grupo de distribuciones locales no traslapadas, se utiliza un esquema iterativo para las distribuciones de clase en general [70].

Considérese $h(x)$ el valor del histograma para la posición x en el espacio de clase. El algoritmo trabaja continuamente actualizando el histograma moviendo las masas de probabilidad a la posición de su centróide local, hasta que el histograma no presente cambio alguno con respecto al histograma de la iteración anterior. Así, si $h^n(x)$ es el histograma actualizado en la n -ésima iteración, el algoritmo procede como sigue:

$$n = 1$$

$$\downarrow$$

$$h^0(x) = h(x)$$

$$\downarrow$$

$$h^n(x) = \sum_{y \in B(x)} h^{n-1}(y)$$

donde $y \in B(x)$ si y solo si $\mu^n(y) = x$

$$\text{donde } \mu^n(y) = y + \frac{\sum_{y'=-m}^m y' h^{n-1}(y+y')}{\sum_{y'=-m}^m h^{n-1}(y+y')}$$

↓

?

$$h^n(x) = h^{n-1}(x) \quad \forall x$$

↓

↓

$n = n + 1$ ←← no ←← ↓ →→ si →→ alto

En este caso los centróides locales se calculan en una ventana de ancho $2m+1$. Un histograma final de la forma:

$$h(x) = \sum_{i=1}^N n_i \delta_{ix} \quad (5.4)$$

con

$$\min_{i,j} |x_i - x_j| > 2m+1 \quad (5.5)$$

ya no presentará cambio alguno en el histograma al ser sometida a un lazo de iteraciones. Esta distribución terminará el algoritmo.

En la ecuación 5.4 N_i es el número de clases encontrado y x_i es aproximadamente:

$$x_i \approx \frac{\sum_{x \in c_i} x h(x)}{\sum_{x \in c_i} h(x)} \quad (5.6)$$

donde $x \in c_i$ si x está clasificada en la i -ésima clase. Por lo tanto el procedimiento iterativo preserva las medias de las clases.

Ya que las propiedades de convergencia del algoritmo son difíciles de determinar, en este trabajo se ha observado que éste siempre converge en un número pequeño de iteraciones (generalmente entre 5 y 15). El número de clases encontradas depende tanto del tamaño de la ventana como del número de "picos" presentes en el histograma. Para el ejemplo citado en la sección anterior, después de haber elegido el nivel óptimo de suavizado ($k=2$, tamaño: 64×64), se somete esta imagen al proceso de clasificación por medio de agrupamiento de centróides locales. Las figuras 5.6a-f ilustran un proceso completo de la clasificación, el número de iteraciones es 5. El tamaño de la ventana es de 20, por lo que las clases resultantes son 3.

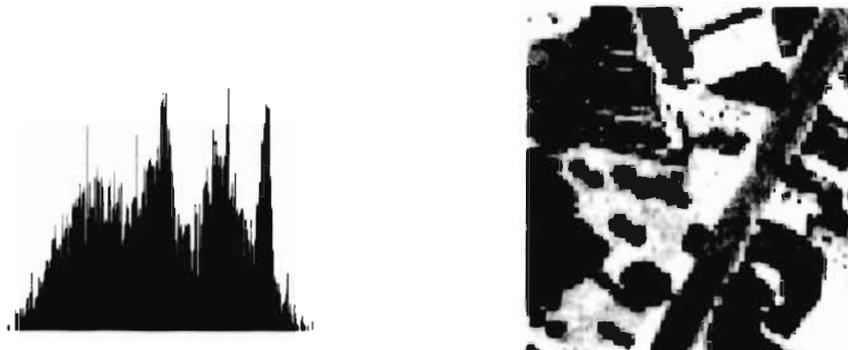


Figura 5.6a Nivel de suavizado óptimo más alto $k=2$.
(Tamaño: 64×64 píxeles¹)

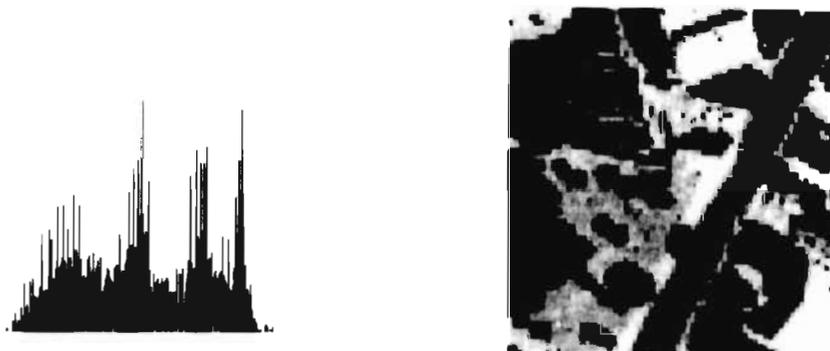


Figura 5.6b Resultado de la primera iteración.
(Tamaño: 64×64 píxeles¹)

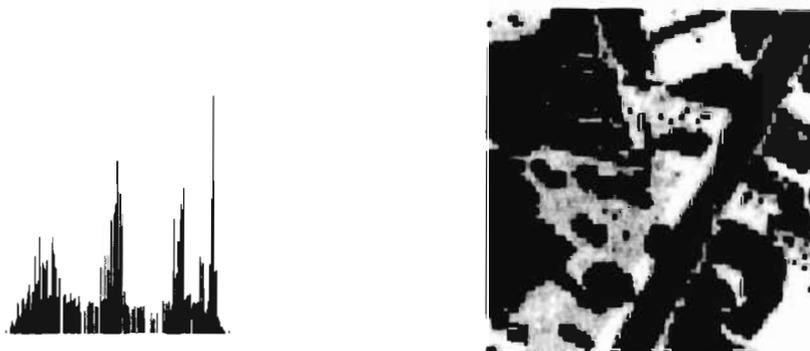


Figura 5.6c Resultado de la segunda iteración.
(Tamaño: 64×64 píxeles¹)

¹Se realizó una ampliación a la imagen para poder apreciar los cambios en cada iteración



Figura 5.6d Resultado de la tercera iteración.
(Tamaño: 64 x 64 píxeles¹)

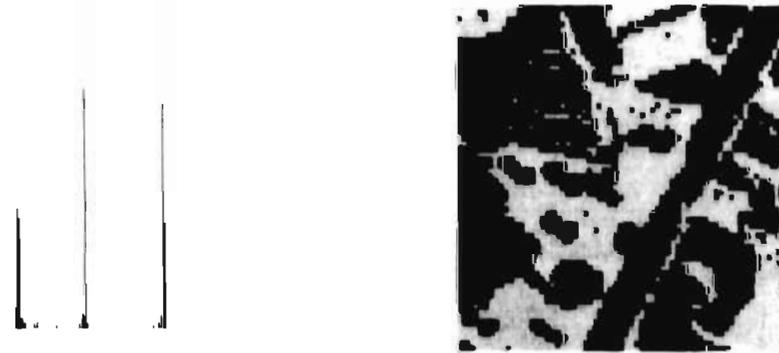


Figura 5.6e Resultado de la cuarta iteración.
(Tamaño: 64 x 64 píxeles¹)

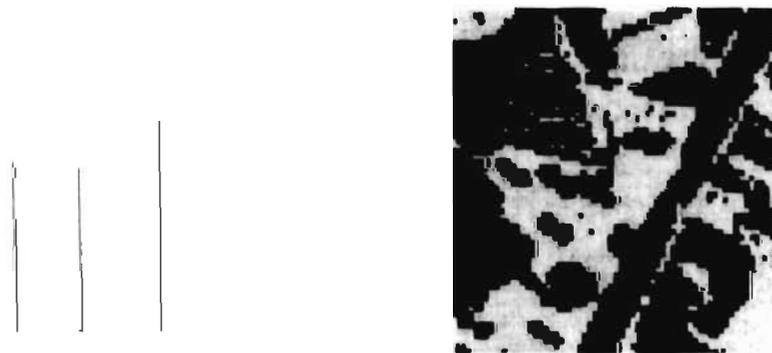


Figura 5.6f Resultado de la quinta iteración.
(Tamaño: 64 x 64 píxeles¹)

¹ Se realizó una ampliación a la imagen para poder apreciar los cambios en cada iteración.

5.3 Estimación de frontera.

Se ha mencionado anteriormente que el suavizado por quadtree es un medio de sobrellevar la resolución en espacio de clase y la resolución espacial. Por lo tanto, siguiendo el procedimiento de agrupamiento de centróides en el nivel de quadtree óptimo más alto, cada nodo frontera en este nivel, define un bloque $L \times L$ de píxeles del nivel de quadtree más bajo (resolución espacial más alta), con $L = 2^k$, siendo k el nivel de quadtree. El problema ahora es cómo restaurar la imagen original, a partir de la imagen correspondiente a este nivel óptimo de quadtree.

Para la solución a este problema, se hace la siguiente suposición: La clasificación realizada en el nivel óptimo más alto del quadtree es válida para los niveles bajos. Así, una región frontera está definida; los nodos no-frontera son asignados a la misma clase a la que pertenecen sus padres; los nodos frontera se clasifican de tal manera que el ancho de la región frontera se reduce por un factor de 2 conforme el nivel de quadtree k disminuye. El resultado es una frontera entre píxeles en el nivel más bajo del árbol, es decir, en la resolución espacial más alta. Como se estableció anteriormente, la suposición clave que se consideró al obtener la frontera es que la segmentación es invariante a la escala sobre la cual ésta es vista [71]. Las restricciones resultantes del nivel previo, controlan el procesamiento del nivel actual.

Una descripción más precisa de la estimación de frontera es la siguiente: asumiendo que la clasificación del nivel de quadtree óptimo mediante el algoritmo de agrupamiento de centróides, es bastante precisa, se establece que esta clasificación es la de un nivel $k+1$ determinado y lo que procede es hacer la clasificación del nivel k . Siguiendo con el ejemplo que se ha estado manejando, la clasificación del nivel $k+1$ sería la clasificación del nivel óptimo de quadtree o sea $k=2$ (tamaño: 64×64) y la clasificación que se hace posteriormente es la del nivel $k=1$ (tamaño: 128×128).

Se define $q(i, j, k)$ como el (i, j) -ésimo nodo en el nivel k y $c(q(i, j, k))$ como la clase de este nodo. Inicialmente cada nodo tiene la misma clase que su padre.

$$c(q(i, j, k)) = c\left(q\left(\frac{i}{2}, \frac{j}{2}, k+1\right)\right) \quad (5.7)$$

La figura 5.7 ilustra la imagen resultante de este proceso.



Figura 5.7 Se asignaron a los hijos las clases de los padres.
(Tamaño: 128 x 128 píxeles¹)

De esta clasificación, se define la región frontera $\Lambda_b(k)$ como:

$$(i, j, k) \in \Lambda_b(k) \Leftrightarrow c(q(i, j, k)) \neq c(q(i', j', k))$$

donde $(i', j') \in N_8(i, j)$, el grupo de píxeles (i, j) en 8-conectividad (figura 5.8).

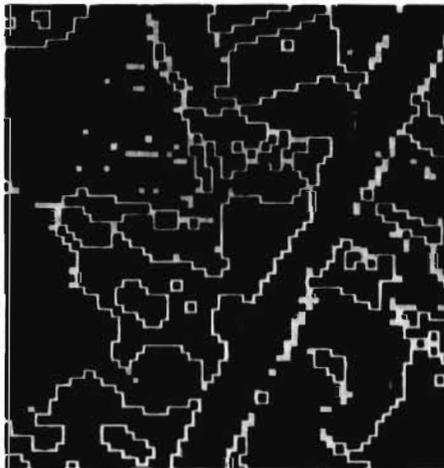


Figura 5.8 Región frontera $\Lambda_b(k)$.
(Tamaño: 128 x 128 píxeles¹)



Figura 5.9 Región de incertidumbre
 $\Lambda_u(k) = \Lambda_b(k) + \Lambda_1(k)$.
(Tamaño: 128 x 128 píxeles¹)

¹ Se realizó una ampliación a la imagen para poder apreciar los cambios en cada iteración

Una vez que se ha determinado $\Lambda_b(k)$, esta región se aumenta con un grupo de píxeles $\Lambda_i(k)$, el cual tiene 8-conectividad en $\Lambda_b(k)$.

$$(i, j) \in \Lambda_i(k) \Leftrightarrow (i', j') \in \Lambda_b(k)$$

$$\text{y } (i', j') \in N_8(i, j).$$

Esto genera una región de incertidumbre definida por $\Lambda_c(k) = \Lambda_b(k) + \Lambda_i(k)$. En la figura 5.9 se muestra la región de incertidumbre correspondiente a nuestro ejemplo.

Debido a que esta región de incertidumbre se convertirá en la frontera entre clases después de ser clasificada, resulta de vital importancia que dicha clasificación se realice lo más certeramente posible, ya que esta frontera se proyectará a los niveles bajos del quadtree hasta llegar al nivel de resolución más alto (nivel de quadtree $k=0$).

En este trabajo la clasificación de la región de incertidumbre se realizó mediante 4 esquemas diferentes:

- Filtrado de la región y clasificación por umbral.
- Clasificación por distancia de Mahalanobis.
- Clasificación usando doble decisión.
- Clasificación usando modelo de Potts.

Enseguida se analiza cada esquema y se presentan los resultados obtenidos.

5.3.1 Filtrado y clasificación por umbral de la región de incertidumbre.

La región de incertidumbre es pues una línea de frontera de un ancho de 4 píxeles. Esta región es suavizada con un filtro lineal cuyo ancho espacial depende de la relación señal-a-ruido ρ_k estimada para el nivel k .

$$\Lambda_c(k) = h(i, j, \rho_k) * \Lambda_c(k), \quad (5.8)$$

donde * denota convolución y

$$\rho_k = \frac{(\mu_1 - \mu_2)}{\sigma_k}. \quad (5.9)$$

donde μ_1 y μ_2 son las medias encontradas con el algoritmo de centroide local.

$$\sigma_k^2 = \frac{1}{N(N'(k))} \sum_{(i,j,k) \in N'(k)} [q(i,j,k) - \mu(i,j,k)]^2 \quad (5.10)$$

donde $N'(k)$ es el grupo de pixeles no-frontera y $\mu(i,j,k)$ es la media asociada con el punto (i,j,k) .

$$\mu(i,j,k) = \mu_{[q(i,j,k)]}$$

y $N(A)$ es el número de puntos dentro de la región A . El filtro $h(i,j,\rho)$ se forma mediante una convolución iterada usando un filtro (3×3) .

$$\begin{aligned} h_1(i,j,\rho) &= \lambda(\rho), \quad (i,j) = 0, \\ &= \frac{1 - \lambda(\rho)}{8}, \quad -1 \leq (i,j) \leq 1 \quad (i,j) \neq (0,0). \end{aligned} \quad (5.11)$$

La función $\lambda(\rho)$, la cual maneja en ancho relativo del filtro resultante fue seleccionado después de algunos experimentos a ser una función lineal de ρ , dando un máximo suavizado para $\rho < 2$ y no suavizado para $\rho > 8$. La figura 5.10 muestra el resultado del suavizado.



Figura 5.10 Imagen suavizada por el template de 3×3 .
(Tamaño: 128×128 pixeles)

Se realizó una ampliación a la imagen para poder apreciar los cambios en cada iteración.

Después del suavizado, se clasifican los píxeles de la región de incertidumbre usando un criterio de umbral.

Se representa como i , el número de clases presentes y μ_i son las medias de las clases. Para cada par de clases se encuentra un umbral $\Theta_{i,i+1}$, se tendrán entonces $i-1$ umbrales.

$$\Theta_{i,i+1} = \frac{|\mu_i - \mu_{i+1}|}{2} \quad (5.12)$$

Posteriormente, se les asigna una clase $\mu_{p(x)}$ a los píxeles de la región de incertidumbre que se encuentran en $\mu_{i+1} \geq p(x) \geq \mu_i$.

$$\mu_{p(x)} = \begin{cases} \mu_i & \mu_i \leq p(x) \leq \mu_i + \Theta_{i,i+1} \\ \mu_{i+1} & \mu_i + \Theta_{i,i+1} < p(x) \leq \mu_{i+1} \end{cases} \quad (5.13)$$

La figura 5.11 y 5.12 ilustran la asignación de clases a los píxeles $p(x)$ y la imagen resultante, respectivamente. En la figura 5.13 se han agregado a la imagen 5.12 los píxeles no-frontera.

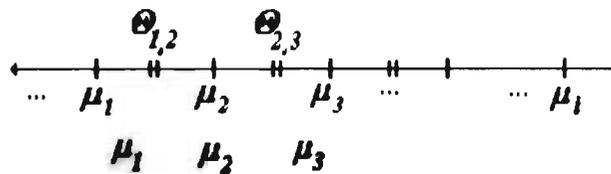


Figura 5.11 Asignación de clases a los píxeles $p(x)$.



Figura 5.12 Región de incertidumbre clasificada usando criterio de umbral. (Tamaño 128 x 128 píxeles¹)



Figura 5.13 Imagen clasificada. (Tamaño 128 x 128 píxeles¹)

Después de remover los píxeles aislados, el proceso se repite en el nivel $k-1$ hasta llegar al nivel de máxima resolución (figura 5.14).



Figura 5.14 Clasificación final por decisión de umbral. (Tamaño: 256 x 256 píxeles¹)

¹ Se realizó una ampliación a la imagen para poder apreciar los cambios en cada iteración.

La segmentación en este esquema depende de la forma de la frontera y de la propagación de errores a los niveles siguientes hacia abajo del quadtree. La proporción de errores se debe a que el suavizado para los píxeles adyacentes a la frontera introduce un efecto indeseado, ya que un 50% de sus vecinos pertenecen a diferente clase. Por otra parte, si al hacer la clasificación en el nivel óptimo más alto se tiene un error, al proyectar la frontera a los niveles siguientes bajos, este error se incrementará debido al ancho de la región de incertidumbre.

5.3.2 Clasificación por distancia de Mahalanobis.

Con el objetivo de hacer una decisión más certera al clasificar la región de incertidumbre, se aplicó el esquema de distancia de Mahalanobis [72], la cual se define de la siguiente manera:

$$r_i^2 = \frac{(p(x) - \mu_i)^2}{\sigma_i^2} \quad \forall p(x) \in \Lambda_i(k) \quad (5.14)$$

donde:

i es el número de clases presentes,

$p(x)$ son los píxeles que conforman la región de incertidumbre (con sus niveles de gris).

μ_i son las medias de las clases,

σ_i^2 es la varianza de las clases y se obtiene de la clasificación obtenida por el algoritmo de agrupamiento de centróides:

$$\sigma_i^2 = \frac{1}{N(p(\mu_i))} \sum [p(k) - \mu_i]^2 \quad \forall p(k) \text{ asociado con } \mu_i \quad (5.15)$$

$p(\mu_i)$ es el grupo de píxeles etiquetados con la clase i .

$p(k)$ es el grupo de píxeles (con su nivel de gris) asociados con la media μ_i .

$N(A)$ es el número de elementos del conjunto A .

La clase $\mu_{p(x)}$ asignada a los píxeles $p(x)$ de la región de incertidumbre:

$$\mu_{p(x)} = \min_{\mu} \left[\frac{(p(x) - \mu_i)^2}{\sigma_i^2} \right] \quad (5.16)$$

Retomando la imagen de la figura 5.9, la cual muestra la región de incertidumbre, ahora la figura 5.15 muestra su clasificación utilizando la distancia de Mahalanobis. En 5.16 se han agregado los pixeles no-frontera a la imagen.

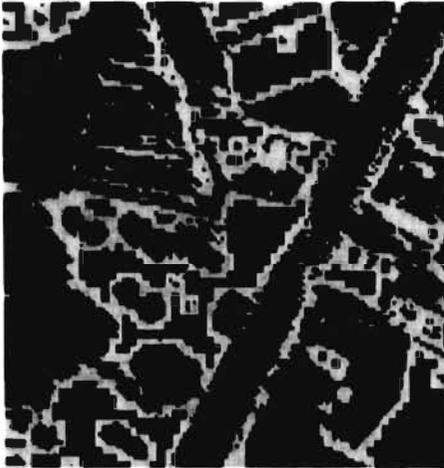


Figura 5.15 Frontera clasificada por distancia de Mahalanobis.
(Tamaño: 128 x 128 pixeles¹)



Figura 5.16 Imagen completa clasificada.
(Tamaño: 128 x 128 pixeles¹)

Este proceso se repite en cada nivel de quadtree hasta llegar al de máxima resolución, tal es el caso de la figura 5.17.

¹ Se realizó una ampliación a la imagen para poder apreciar los cambios en cada iteración



Figura 5.17 Imagen clasificada por distancia de Mahalanobis.
(Tamaño: 256 x 256 píxeles¹)

5.3.3 Clasificación usando doble decisión.

Analizando los resultados de los dos esquemas anteriores, y ante la necesidad de clasificar lo más certero posible la región de incertidumbre, se ha sometido dicha región a un proceso de doble decisión, de esta manera se busca asegurar la precisión al asignar a los píxeles $p(x)$ la clase más conveniente. Es así como se minimiza la presencia de errores desde la primera clasificación de esta región, por lo tanto, al proyectar la región frontera a los siguientes niveles, disminuye la probabilidad de arrastrar errores desde los niveles de quadtree más altos.

Una vez que se ha encontrado la región de incertidumbre (Figura 5.9), ésta es clasificada por medio de la distancia de Mahalanobis (figura 5.15), posteriormente se somete al proceso de suavizado usando el template de 3 x 3 descrito en la sección 5.3.1. Este resultado se muestra en la figura 5.18. Una vez suavizada la imagen, la región de incertidumbre es clasificada nuevamente utilizando el criterio de umbral (figura 5.19).



Figura 5.18 Imagen que fue clasificada por distancia de Mahalanobis sometida a un proceso de suavizado. (Tamaño: 128×128)



Figura 5.19 Imagen suavizada es nuevamente clasificada utilizando el criterio de umbral. (Tamaño: 128×128)

Este procedimiento de doble decisión se realiza en cada nivel de quadtree hasta llegar al de máxima resolución (figura 5.20).



Figura 5.20 Imagen de máxima resolución clasificada mediante un proceso de doble decisión. (Tamaño: 256×256)

Se realizó una ampliación a la imagen para poder apreciar los cambios en cada iteración

5.3.4 Clasificación usando el modelo de Potts para el último nivel de resolución.

En el capítulo 3 se analizó la teoría e implementación del modelo de Potts, este método también fue utilizado para clasificar los píxeles $p(x)$ de la región de incertidumbre únicamente para el nivel $k=0$, es decir para la región de incertidumbre de la imagen original, para dicha región en el nivel $k+1$ se utilizó el modelo de doble decisión descrito en 5.3.3.

Una vez que se ha realizado la clasificación del nivel $k+1$ (tamaño: 128×128), por el método de doble decisión (figura 5.19), esta frontera se proyecta al nivel $k=0$ (tamaño: 256×256), creándose una nueva región de incertidumbre (figura 5.21), ahora esta región es clasificada usando el modelo de Potts. La figura 5.22 muestra la clasificación final.



Figura 5.21 Región de Incertidumbre con sus niveles de gris.
(Tamaño 256×256 píxeles¹)

¹ Se realizó una ampliación a la imagen para poder apreciar los cambios en cada iteración.



Figura 5.22 Imagen clasificada usando el modelo de Potts.
(Tamaño 256 x 256 pixeles¹)

El método reportado aquí combina una operación de suavizado por estructuras de árbol (quadtree) con una clasificación estadística realizada en el nivel óptimo más alto del quadtree, seguida por una estimación de frontera basada en los segmentos obtenidos de dicho nivel del árbol. La segmentación es estadística, pero está basada en un algoritmo de centróide local, el cual no requiere información *a priori* de las clases presentes.

Se ha mostrado paso a paso la clasificación por medio de estructuras de árbol aplicada a una sola imagen, se presentaron también los resultados obtenidos usando las diferentes opciones que se tienen al hacer la estimación de frontera (clasificación de la región de incertidumbre). En el siguiente capítulo se hace una comparación de los diferentes resultados así como la discusión de los mismos.

¹ Se realizó una ampliación a la imagen para poder apreciar los cambios en cada iteración.

CAPÍTULO 6

COMPARACIONES Y CONCLUSIONES.

Los resultados obtenidos en la segmentación de imágenes de cada método utilizado, fueron mostrados en sus capítulos respectivos, sin embargo, estos esquemas fueron aplicados a una sola imagen. En este capítulo, tales métodos se aplican a imágenes provenientes de otras fuentes (aerea, SPOT y SAR) y los resultados se confrontan para llegar a una evaluación de la funcionalidad. En la última sección del capítulo se presentan los puntos concluyentes del trabajo, así como futuras líneas de investigación.

6.1 Comparaciones.

Las figuras 6.1, 6.8 y 6.15 muestran las imágenes que se sometieron al proceso de segmentación utilizando las 6 técnicas desarrolladas aquí. Después de cada imagen, se muestran los resultados obtenidos con las 6 diferentes técnicas. En la tabla 6.1 se analizan y comparan dichos resultados.

La manera en que se presentan las imágenes resultantes es:

- (a) Imagen segmentada.
- (b) El contorno de la imagen segmentada superpuesto en la imagen original.

Todas las imágenes han sido ampliadas a un 120 % para apreciar mejor los detalles.



Figura 6.1 Foto aérea. (PARIS)*
 (Tamaño de la escena: 256 x 256 píxeles, tamaño del píxel: 1 x 1m.)
 (Coordenadas (metros): sup. izq. X-6200, sup. izq. Y-13100, inf. der. X-7200, inf. der. Y-12100)

*AREA: PARIS, FRANCE	
Dataset: Ortho image dataset	Source: GeoInformation International
Date collected: June 1995	Media: CD-ROM
Area covered (km ²): 1	Flying height: 4270 meters (14,000 feet)
Focal length camera: 304 mm (12 inches)	Nominal scale: 1:14,000
Image control: GPS	Projection: Clarke 1880 IGN
Pixel resolution (cm): 100	Scanning: 24-bit colour
Georeference: Ville de Paris	Total file size (KB): 2,930



(a)

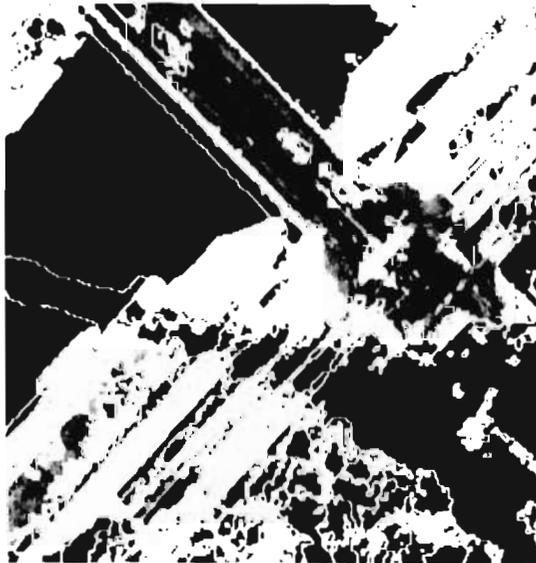


(b)

Figura 6.2 Imagen Segmentada en 6 clases por estructuras de árbol (quadtree) con estimación de frontera basada en decisión por umbral.



(a)

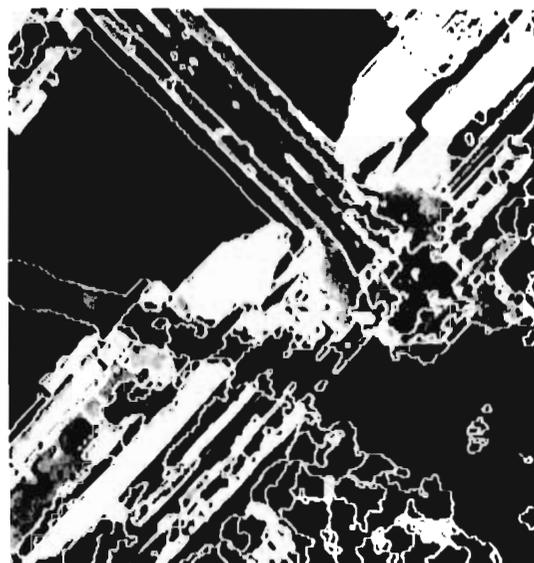


(b)

Figura 6.3 Imagen segmentada en 6 clases por estructuras de árbol (quadtree) con estimación de frontera basada en decisión por distancia de Mahalanobis.



(a)



(b)

Figura 6.4 Imagen segmentada en 6 clases por modelo de Potts matricial.

ESTA TESIS NO DEBE
SALIR DE LA BIBLIOTECA



(a)

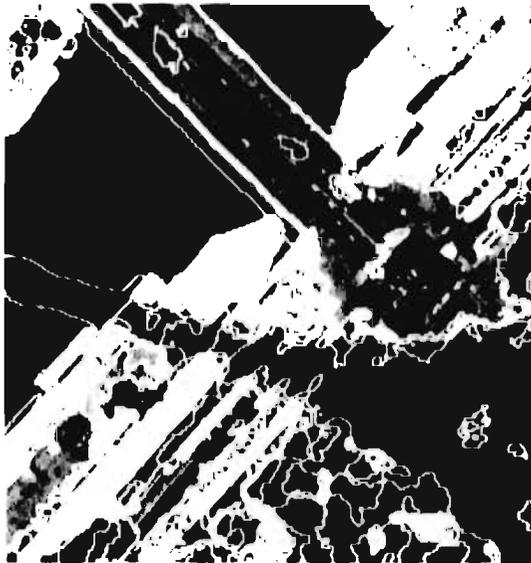


(b)

Figura 6.5 Imagen segmentada en 6 clases por estructuras de árbol con estimación de frontera basada en modelo de Potts clásico.



(a)

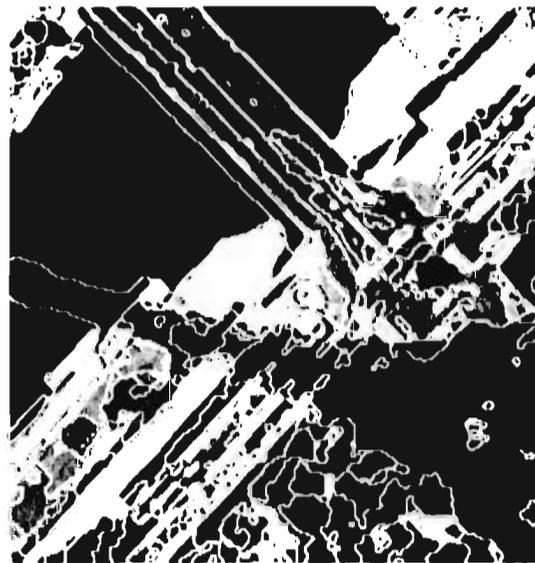


(b)

Figura 6.6 Imagen segmentada en 6 clases por modelo de Potts clásico.



(a)



(b)

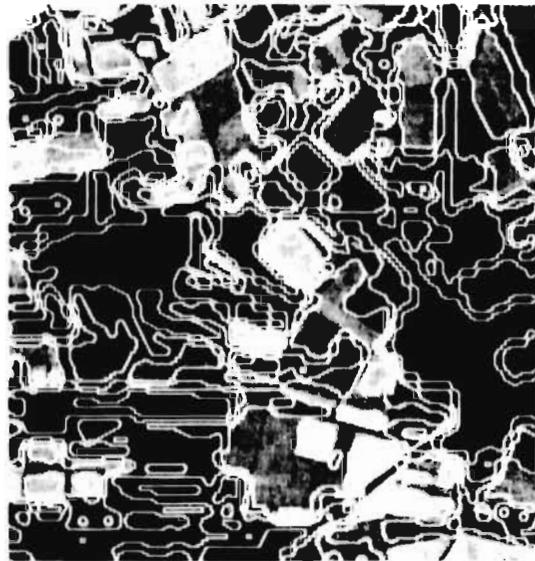
Figura 6.7 Imagen segmentada en 6 clases por estructuras de árbol (quadtree) con estimación de frontera basada en una doble decisión.



Figura 6.8 Imagen SPOT pancromática. (SCENE)
(Tamaño de la escena: 256 x 256 píxeles, tamaño del píxel: 10 x 10m.)



(a)

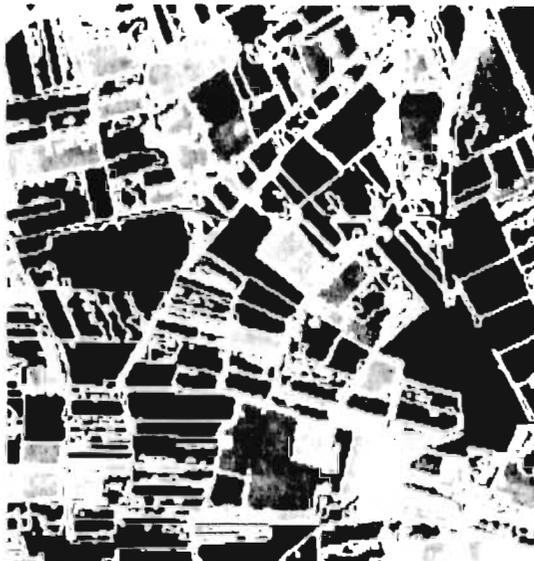


(b)

Figura 6.9 Imagen segmentada en 7 clases por estructuras de árbol (quadtree) con estimación de frontera basada en decisión por umbral.



(a)



(b)

Figura 6.10 Imagen segmentada en 7 clases por estructuras de árbol (quadtree) con decisión de frontera basada en la decisión por distancia de Mahalanobis.



(a)

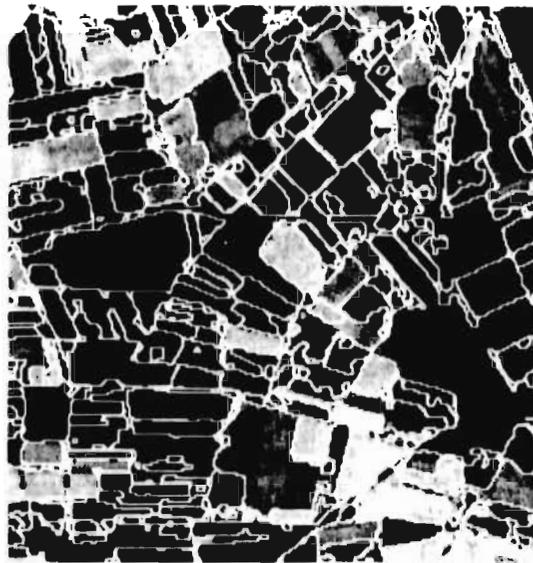


(b)

Figura 6.11 Imagen segmentada en 7 clases por modelo de Potts matricial.



(a)

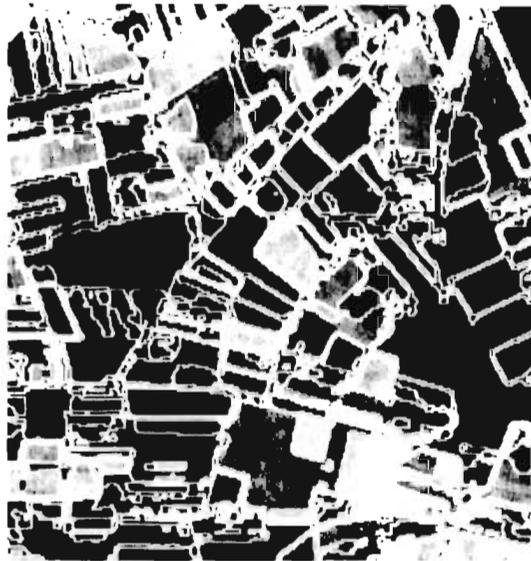


(b)

Figura 6.12 Imagen segmentada en 7 clases por estructuras de árbol (quadtree) con estimación de frontera basada en el modelo de Potts clásico.



(a)



(b)

Figura 6.13 Imagen segmentada en 7 clases por el modelo de Potts clásico.



(a)



(b)

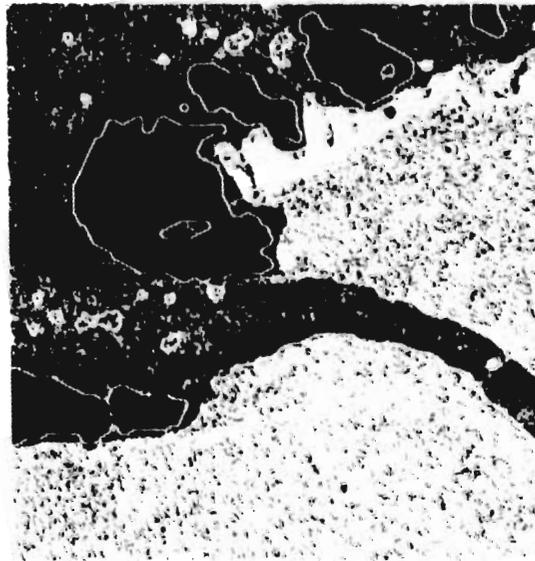
Figura 6.14 Imagen segmentada en 7 clases por estructuras de árbol (quadtree) con estimación de frontera basada en una doble decisión.



Figura 6.15 Imagen ERS-1, Precision Image-PR1. (WS52)
(Tamaño de la escena: 256 x 256 píxeles, tamaño del píxel: 12.5 x 12.5 m.)



(a)

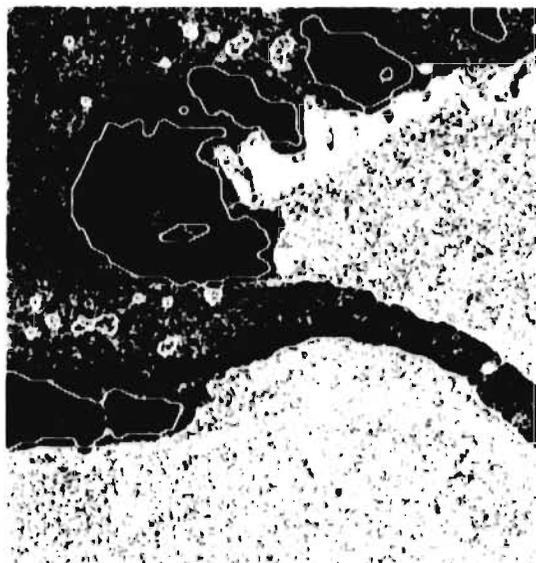


(b)

Figura 6.16 Imagen segmentada en 3 clases por estructuras de árbol (quadtree) con estimación de frontera basada en decisión por umbral.



(a)

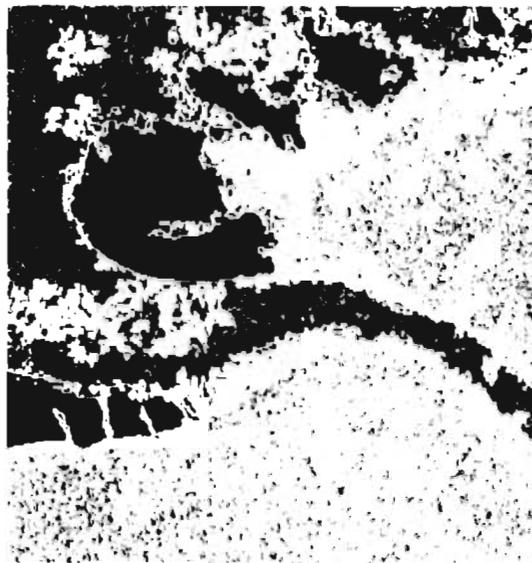


(b)

Figura 6.16 Imagen segmentada en 3 clases por estructuras de árbol (quadtree) con estimación de frontera basada en decisión por umbral.

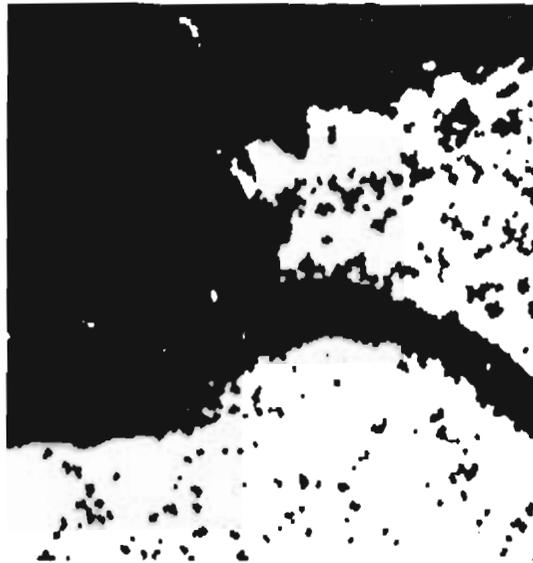


(a)

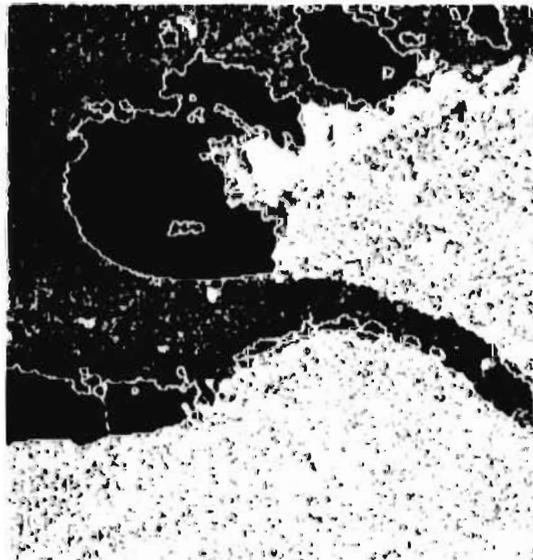


(b)

Figura 6.17 Imagen segmentada en 3 clases por estructuras de árbol (quadtree) con estimación de frontera basada en decisión por distancia de Mahalanobis.



(a)

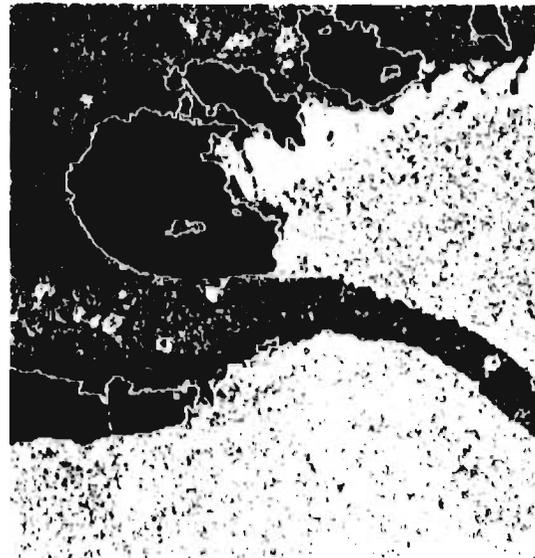


(b)

Figura 6.18 Imagen segmentada en 3 clases por model de Potts matricial.



(a)

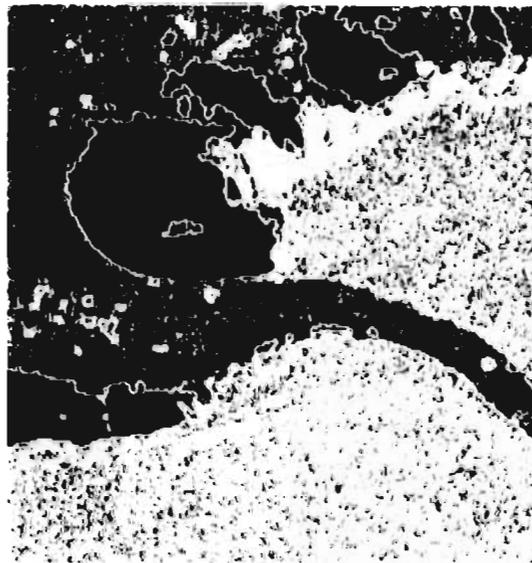


(b)

Figura 6.19 Imagen segmentada en 3 clases por estructuras de árbol (quadtree) con estimación de frontera basada en el modelo de Potts clásico.



(a)

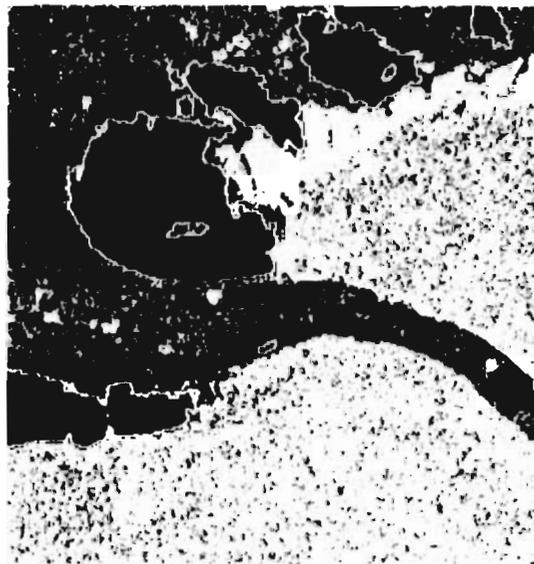


(b)

Figura 6. 20 Imagen clasificada en 3 clases por modelo de Potts clásico.



(a)



(b)

Figura 6.21 Imagen segmentada en 3 clases por estructuras de árbol (quadtree) con estimación de frontera basada en una doble decisión.

Imagen	Quadtree/Umbra	Quadtree/ Distancia de Mahalanobis	Potts Matricial	Quadtree/ Potts	Potts Clásico	Quadtree/ Doble Decisión
PARIS	Presenta distorsión máxima en los bordes y fusiona erróneamente las clases. Pierde detalles pequeños presentes en la escena.	Contiene pixeles mal clasificados, no define las fronteras con exactitud y no presenta homogeneidad en las regiones ya definidas. Conserva detalles pequeños en la escena.	Mejora el resultado de la segmentación por K-means, el cual es muy parecido al de Quadtree/distancia de Mahalanobis. No distorsiona los bordes y presenta regiones homogéneas. Si conserva detalles pequeños.	Presenta pixeles bien clasificados, no hay distorsión en los bordes de las regiones, pero tiende a perder pequeños detalles en la escena.	Pixeles bien clasificados. presenta homogeneidad en las regiones definidas, ligera distorsión en los bordes y pierde los pequeños detalles presentes en la escena	Pixeles bien clasificados. homogeneidad en las regiones definidas. no presenta distorsión en los bordes y conserva un poco más los pequeños detalles de la escena.
SCENE	Presenta distorsión en los bordes y fusiona erróneamente las clases.	Presenta pocos pixeles mal clasificados, fronteras bien definidas y homogeneidad dentro de las regiones.	Fusiona demasiado las clases y distorsiona los bordes.	Pixeles bien clasificados, fronteras bien definidas y regiones homogéneas. Ligera distorsión en los bordes.	Pixeles bien clasificados, bordes ligeramente distorsionados, presenta homogeneidad en las regiones ya definidas.	Pixeles bien clasificados, fronteras bien definidas y regiones homogéneas.

Imagen	Quadtree/Umbra	Quadtree/ Distancia de Mahalanobis	Potts Matricial	Quadtree/ Potts	Potts Clásico	Quadtree/ Doble Decisión
WS52	Máxima distorsión en los bordes, regiones homogéneas.	Presencia de muchos píxeles mal clasificados, no define bien los bordes y las regiones presentes no presentan homogeneidad.	Presencia de píxeles mal clasificados, bordes bien definidos, la homogeneidad de las clases no es muy óptima.	Presencia de muy pocos píxeles mal clasificados, bordes bien definidos, homogeneidad en las regiones presentes.	Píxeles bien clasificados, bordes bien definidos, homogeneidad en las regiones presentes.	Píxeles bien clasificados, bordes bien definidos, homogeneidad en las regiones presentes.

Tabla 6.1 Evaluación de los diferentes esquemas de segmentación.

6.2 Conclusiones.

Se desarrolló un método de segmentación de imágenes, con una visión a la solución del problema de la multiresolución. Este esquema consta de tres etapas: suavizado por estructuras de árbol (quadtree), clasificación por agrupamiento de centróides y finalmente la estimación de frontera. A lo largo de este estudio, se implementaron distintos modelos para llevar a cabo la etapa de estimación de frontera, que representa la más importante, ya que al proyectar la frontera hasta el nivel de máxima resolución, un mínimo de errores deberá ser arrastrado hasta dicho nivel. De aquí la importancia de tener una clasificación certera desde el primer nivel del árbol. Los modelos desarrollados fueron:

- * *Decisión por umbral*: Los resultados muestran una vez más como la segmentación es un proceso más complicado que un simple análisis pixel por pixel de la imagen. En general presenta distorsión máxima de los bordes y una errónea fusión de clases.
- * *Decisión por distancia de Mahalanobis*: Los resultados superan a los obtenidos en el esquema anterior, sin embargo, las regiones no se presentan homogéneas debido a la presencia de píxeles mal clasificados, los resultados son muy parecidos a los que se obtienen al aplicar el algoritmo de *k*-medias.
- * *Clasificación por doble decisión (decisión por umbral y decisión por distancia de Mahalanobis)*: Es sin duda uno de los esquemas que ha proporcionado los mejores resultados, no obstante la presencia de ruido en la imagen a clasificar. Es un método robusto y muy general, que no requiere información *a priori* del número de clases presentes en la imagen, ni de las medias de dichas clases, simplemente se fija una ventana que es la que va a recorrer el histograma de la imagen suavizada, creándose así una serie de nuevos histogramas que van agrupando los centróides de los histogramas, dando como resultado final una imagen clasificada. Pese a que la frontera original se proyecta desde los niveles de baja resolución hasta los de alta resolución, ésta no sufre distorsión alguna, ya que se va regenerando en cada nivel. Los resultados muestran además como la presencia de los píxeles mal clasificados y de los píxeles aislados es mínima, por lo que las regiones bien definidas se presentan como homogéneas.
- * *Clasificación de frontera por modelo de Potts*: La combinación de estos dos esquemas resulta bastante satisfactoria, debido a que la frontera en el segundo nivel del árbol, se clasifica por el modelo de doble decisión. En el primer nivel o nivel de máxima resolución, la frontera se clasifica mediante el modelo de Potts clásico. Sin embargo es evidente en los resultados, una ligera distorsión en los bordes de frontera.

- * *Modelo de Potts*: Es el otro esquema del que se obtuvieron los mejores resultados. Ya es bien sabido el éxito de las técnicas markovianas en segmentación de imágenes de percepción remota, es por eso que se incluyeron dos esquemas basados en tales técnicas. El modelo de Potts clásico nos asegura una imagen bien clasificada, con regiones homogéneas y con bordes de frontera bien definidos. Sin embargo este modelo requiere de información *a priori* del número de clases y de las medias de las clases. Esta información puede bien obtenerse del algoritmo de *k*-medias.
- * *Modelo de Potts matricial*: En la búsqueda de encontrar un método de segmentación y post-segmentación con resultados satisfactorios, se desarrolló el modelo de Potts matricial, el cual, a partir de una imagen segmentada (por *k*-medias, en este estudio), realiza una reetiquetación de los píxeles. Esto es, a partir de de una matriz de coeficientes de homogeneidad cuyos valores favorecen o desfavorecen la evolución de una clase a otra. Los resultados indican una mejor fusión de clases, pero se presenta distorsión en los bordes de frontera en el caso de ciertas imágenes, tal como SCENE.

Es resumen, los esquemas más certeros al segmentar las diferentes imágenes fueron el de estructuras de árbol (quadtree) con estimación de frontera basada en una doble decisión y el modelo de Potts clásico.

Con el objetivo de seguir con la investigación, la búsqueda y la obtención de mejores resultados, se sugieren algunas perspectivas para la continuación del presente trabajo:

- En el caso del método de quadtree con distancia de Mahalanobis, el resultado no es muy óptimo por la presencia de píxeles mal clasificados, este resultado podría ser sometido a un proceso de relajación probabilística con el objeto de mejorar la clasificación de los píxeles aislados.
- Debido a los resultados tan satisfactorios que presenta el esquema de quadtree con doble decisión, éste podría extenderse para ser usado en segmentación de textura.
- Una mejora al esquema, en el problema de la clasificación de frontera, sería la incorporación del modelo de Ising, basado en la teoría de los campos aleatorios de Markov.
- Finalmente sería conveniente tratar de segmentar la región de frontera con un modelo de filtro adaptivo mayoritario.

Referencias.

- [1] Marr, D., "Vision", San Francisco, Freeman, 1982
- [2] Duda, R.O., Hart, P.E., "Pattern Classification and Scene Analysis", New York, Wiley, 1973.
- [3] Fukunaga, K., "Introduction to Statistical Pattern Recognition", New York, Academic Pr., 1979.
- [4] Devijver, P.A., Kittler, J., "Pattern Recognition: A Statistical Approach", London, Prentice Hall, 1983.
- [5] Narendra, P.M., Goldberg, M., "A Non-Parametric Clustering Scheme for LANDSAT", *Patt. Rec.* 9, 207-215, 1977.
- [6] Wilson, R., "A Class of Local Centroid Algorithms for Classification and Quantisation", *Proc. IEEE Conf. Comput. Sys. Sig. Proc.*, Bangalore, India, 1984.
- [7] Chow, C.K., Kaneko, T., "Automatic Boundary Detection of the Left Ventricle form Cineangiograms", *Comput. Biomed. Res.* 5, 388-410, 1972.
- [8] Davis, L.S., Rosenfeld, A., "Co-operating Processes for Low-Level Vision", *Art. Int.* 17, 245-263, 1981.
- [9] Peleg, S., "A New Probabilistic Relaxation Scheme", *IEEE Trans. Patt. Anal. Mach. Int.*, PAMI-2, 362-369, 1980.
- [10] Kittler, J., Föglein, J., "A General Contextual Classification Method for Image Segmentation", *Proc. 3rd Scand. Conf. on Image Anal.*, 90-95, Copenhage, 1983.

- [11] Chen, P.C., Pavlidis, T., "Image Segmentation as an Estimation Problem", in "Image Modeling" (A. Rosenfeld ed.), New York, Academic Pr., 9-28, 1981.
- [12] Roberts, L.G., "Machine Perception of 3-d Solids", in *Optical and Electro-Optical Information Processing* (J. T. Tippett ed.), Cambridge, MA, MIT Pr., 159-197, 1965.
- [13] Prewitt, J. M.S., "Object Enhancement and Extraction", in "Picture Processing and Psychopictories" (B.S. Lipkin, A. Rosenfeld eds.), New York, Academic Pr., 75-149, 1970.
- [14] Frei, W., Chen, C.C., "Fast Boundary Detection: a Generalisation and a New Algorithm", *IEEE Trans. Comput.*, C-26, 988-998, 1977.
- [15] Knutsson, H., Grandlund, G.H., "Texture Analysis Using 2-d Quadrature Filters", *Proc. IEEE CAPADIM Workshop*, 1983.
- [16] Papoulis, A., "Signal Analysis", New York, McGraw Hill, 1977.
- [17] Brice, C., Fennema, C., "Scene Analysis Using Regions", *Art. Int. J.*, 205-226, 1970.
- [18] Wilson, R., Spann, M., "Finite Prolate Spheroidal Sequences and their Applications. pts I y II". *Accepted for publication IEEE Trans. Patt. Anal. Mach. Int.*
- [19] Burt, P.J., Hong, T.H., Rosenfeld, A., "Segmentation and Estimation of Region Properties Through Co-operative Hierarchical Computation", *IEEE Trans. Sys. Man Cyber. SMC-11*, 802-809, 1981.
- [20] Hong, T.H., Rosenfeld, A., "Compact Region Extraction Using Weighted Pixel Linking in a Pyramid", *IEEE Trans. Patt. Anal. Mach. Int. PAMI-6*, 222-229, 1984.
- [21] Kyoung-Ok Kim, Young-Kyu Yang, Jong-Hoon Lee, Kyung-Ho Choi, Tae-Kyun Kim, "Classification of Multispectral Image Using Neural Network", *IEEE International Geoscience and Remote Sensing Symposium: IGARSS'95, vol. 1*, pp.446-448, Firenze Italy, 10-14 July 1995.
- [22] G. Pasquariello and P. Blonda, "Multitemporal Remote Sensing Data Classification Using Neural Network", *Proceeding of ISPRS, vol. 29*, pp. 922-929, Washington D.C., 1992.
- [23] Yeqiao Wang, Daniel L. Civco, "Artificial Neural Networks in High Dimensional Spatial Data Classification: A Performance Evaluation", *ACSM 55th Annual Convention and ASPRS 61st Annual Convention*, pp. 662-671, Feb 27, 1995.

- [24] Jing Wu, Hong Yan and Andrew N. Chalmers, "Color Image Segmentation Using Fuzzy Clustering and Supervised Learning", *Journal of Electronic Imaging*, vol. 3, No. 4 pp. 397-403, October 1994.
- [25] P. Blonda, V. la Forgia, Pasquariello, Santalino, "Multispectral Classification by a Modular Neural Network Architecture", *IGARSS'94, California Institute of Technology, Pasadena*, vol. 4, pp. 1873-1876, 1994.
- [26] Ronald P.H.M. Schoenmakers, Graeme G. Wilkinson, Theo E. Schouten, "Results of Hybrid Segmentation Method", *Proceedings Europto Series, Image and Signal Processing for Remote Sensing*, vol. 2315 pp. 113-126, 26-30 September 1994, Rome, Italy.
- [27] R.P.H.M. Schoenmakers and J.H.T. Stakenborg, "Adaptive Filtering Using Gradient Morphology", *Proceeding of the Eleventh International Conference on Pattern Recognition (IAPR_11)*, pp. C240-C243, The Hague, The Netherlands, August 30-september 3 1992.
- [28] J.H.T. Stakenborg, "Per-Field Classification of a Segmented SPOT Simulated Image", *Proceeding of the 7th International Symposium on Remote Sensing for Resources Development and Enviroment Management, ISPRS Commission VII*, pp. 73-78, Enschede, The Netherlands, August 25-29 1986.
- [29] V. Lacroix, "A Three Module Strategy for Edge-Detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10 (6):803-810, 1988.
- [30] W. Frei and C-C. Chen, "Fast Boundary Detection: a Generalization and a New Algorithm", *IEEE Transactions on Computers*, C-26(10):988-998, 1977.
- [31] N. Kapur, P.K. Shao and A.K.C. Wong, "A New Method for Gray-Level Picture Thresholding Using the Entropy of the Histogram", *Computer Vision, Graphics and Image processing*, 29:273-285, 1985.
- [32] S.W. Zucker, "Region Growing: Childhood and Adolescence", *Computer Graphics and Image Processing*, 5(3):382-399, 1976.
- [33] J-M. Beaulieu and M. Goldberg, "Hierarchy in Picture Segmentation: A Stepwise Optimization Approach", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(2):150-163, 1989.
- [34] J.C. Tilton, "Image Segmentation by Iterative Parallel Region Growing and Splitting", *Proceeding of the International Geoscience and Remote Sensing Symposium (IGARSS89)*, pp. 2235-2238, Vancouver 1989.
- [35] Th. E. Schouten, "Adding Edges to Regions", *Private Communication, Katholieke Universiteit Nijmegen, Toernooiveld 1*, 6525 DE Nijmegen, 1993.

- [36] Anne H. Schistad Solberg and Torfinn Taxt. "Classification of Multisource Satellite Imagery Based on a Markov Random Field Model", *Proceedings Europto Series. Image and Signal Processing for Remote Sensing*, vol. 2315 pp. 322-331, 26-30 September 1994, Rome, Italy.
- [37] A.H. Schistad Solberg and T. Taxt. "A Markov Random Field Model for Classification of Multisource Satellite Imagery", *Submitted to IEEE Trans. Geosc. Rem. Sens.*, 1994.
- [38] A. H. Schistad Solberg, A. K. Jain and T. Taxt. "Multisource Classification of Remotely Sensed Data: Fusion of Landsat TM and SAR Images", *To appear in IEEE Trans. Geosc. Rem. Sens.*, 1994.
- [39] J.P.Rasson, F. Montaigne, F. Orban-Ferauge. "From the Supervised to the Unsupervised Behavioral Classification", *IEEE International Geoscience and Remote Sensing Symposium: IGARSS'95*, vol. 2, pp. 1159-1161, Firenze Italy, 10-14 July 1995.
- [40] Stephens D.A. and Smith A.F.M., "Bayesian Edge Detection in Images Via Changepoint Methods", *Computer Intensive Methods in Statistics*, Härdle W. and Simar De., Physica-Verlag-Heidelberg, pp. 1-18, 1993.
- [41] Wichern D.W., Miller R.B. and Hsu D.A., "Changes of Variance in First-Order Autoregressive Time Series models", *Appl. Statist.* 25, no. 3, pp. 248-256, 1976.
- [42] Devroye L. and Wise G.L., "Detection of Abnormal Behavior Via Non Parametric Estimation of the Support", *SIAM J.Appl. math.*, vol. 28, no. 3, pp. 480-488, 1980.
- [43] Sabin F.F., "Remote Sensing: Principles and Interpretation", USA. *W.H. Freeman and Company*, 1a de. 426 pp, 1978.
- [44] Steffen, C.A., Lorenzetti, J.A., Stech, J.L. e Souza, R.C., "Sensoramiento Remoto: Principios Físicos: Sensores e Productos e Sistema Landsat", *INPE*, 1a de. 81 pp, Brasil 1981.
- [45] Thomas M. Lillesand and Ralph W. Kiefer, "Remote Sensing and Image Interpretation" *John Wiley & Sons, Inc* New York 1987.
- [46] Paul M. Mather, "Computer Processing of Remotely-sensed Images" *John Wiley & Sons, Inc* New York 1987.
- [47] Townshend, J.R.G., "The Spatial Resolving Power of Earth Resources Satellites: a Review", *NASA Technical Memorandum 82020*, Goddard Spaceflight Center, Greenbelt, Maryland 1980.
- [48] Tucjer, C.J., "Red and Photographic Infrared Linear Combinations for Monitoring Vegetation", *Remote Sensing of Enviroment*, 10 127-150, 1979.

- [49] Bernstein, R., Lotpiech, J.B., Myers, J., Kolsky, H.G. and Less, R.D., " Analysis and Processing of Landsat-4 Sensor Data Using Advanced Image Processing Techniques and Technologies. *IEEE Trans. on Geoscience and Remote Sens.*, GE-22, 192-221, 1984.
- [50] R. Dobrushin, "The description of a random field by means of conditional probabilities and conditions of its regularity", *Theory Probab. Appl.*, Vol. 13, No. 2, 1968, pp. 197-224.
- [51] J. Besag, "Spatial interaction and statistical analysis of lattice systems", *Acad. Roy. Statist. Soc. Ser. B*, Vol. 36, 1974, pp. 721-741.
- [52] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images", *IEEE Trans. Pattern. Anal. Machine Intell.*, Vol. 6, No. 6, 1984, pp. 721-741.
- [53] R. Chellappa and A. Jain, "Markov Random Fields: Theory and Application. Academic Press, New York, 1993. Collective work.
- [54] X. Descombes, "Champs Markoviens en Analyse d'Images", TELECOM Paris. These de doctorat, dcembre 1993.
- [55] M. Sigelle, "Champs Markoviens en Traitement des Images", TELECOM Paris. Support de Cours, avril 1992.
- [56] P. Carnevali, L. Coletti and S. Patarnello, "Image Processing by Simulated Annealing", *IBM Journal of Research and Development* Vol. 29, nov. 1985.
- [57] J. Besag, "On the Statistical Analysis of Dirty Pictures", *Journal of the Royal Statistical Society* Vol. B-48 (1986) pp. 259-302.
- [58] H. Derin and H. Elliot, "Modeling and Segmentation of Noisy and Textured Images using Gibbs Random Fields", *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 9 No. 1., Jan 1987 pp. 39-55.
- [59] P. A. Kelly, H. Derin and K.D. Hartt, "Adaptative Segmentation of Speckled Images using a Hierarchical Random Field Model", *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 26 No. 10, Oct. 1988 pp. 1628-1641.
- [60] N. Karssemeijer, "A Relaxation Method for Image Segmentation using a Spatially Dependant Stochastic Model", *Pattern Recognition Letters* Vol. 11, N. 1, 1990 pp. 13-23.
- [61] N. Karssemeijer, "A Statistical Method for Automatic Labeling of Tissues in Medical Images", *Machine Vision and Applications* Vol. 3 N. 2 Spring 1990 pp. 75-86.

- [62] F. Y. Wu. "Potts Model". Review of Modern Physics Vol. 54 No. 1, Jan 1982.
- [63] X. Descombes, M. Moctezuma, H. Maitre, J. Rudant. "Coastline detection by a Markov segmentation on SAR images". Signal Processing No. 55, 1996, pp. 123-132.
- [64] R. Chellappa and A. Jain. "Markov Random Fields: Theory and Application". Academic Press, New York, 1993. Collective work.
- [65] S. Geman and D. Geman. "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images". IEEE Trans. Pattern Anal. Machine Intell. Vol. 6, No. 6, 1984, pp. 721-741.
- [66] S. Kirkpatrick, C. Gellatt and M. Vecchi. "Optimization by simulated annealing". Science, No. 220, 1983, pp. 671-680.
- [67] R.O.Duda and P.E. Hart. "Pattern Recognition and Scene Analysis". Wiley, New York, 1973.
- [68] Tanimoto, S. L., Pavlidis, T., "A Hierarchical Data Structure for Image Processing". Comput. Graph. and Image Proc. 4, 104-119, 1975.
- [69] Spann M., Wilson R., "A quadtree Approach to Image Segmentation that Combines Statistical and Spatial Information", Patt. Rec. 18, 257-269, 1985.
- [70] Wilson R., Spann M., "Image Segmentation and Uncertainty". Research Studies Press LTD, England, 1988.
- [71] Wilson R., "From Signals to Symbols- the Inferential Structure of Perception". Proc. IEEE COMPINT-85, 221-225, Montreal, 1985.
- [72] Bow Sing-Tze. "Pattern Recognition and Image Preprocessing". Marcel Dekker, Inc., New York, 1992.