



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

SEGMENTACIÓN ESTADÍSTICA NO-SUPERVISADA DE IMÁGENES DE PERCEPCIÓN REMOTA:
ESQUEMA EM.

T E S I S
Que para obtener el título de :
I N G E N I E R O E N
TELECOMUNICACIONES
P r e s e n t a :
SALVADOR GARCIA TAVERA



Director de tesis:
DR. MIGUEL MOCTEZUMA FLORES

MÉXICO, D. F.

TESIS CON
FALLA DE ORIGEN

2002.



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

Dedicado a:

Mi s Padres Raúl y Ernestina , que me apoyaron y confiaron en mi, durante todo el tiempo que duró esta aventura.

Mi Abuelita Luisa y mi abuelito José María, espero que me estén viendo desde el cielo.

Abue Gaudencia, te lo prometí, y hasta hoy puedo cumplirlo.

A Roberto, Raúl, Marisol y Uriel.

Mi sobrino Alexis, el más pequeño y latoso.

Agradecimientos:

Al Dr. Miguel Moctezuma , por acceder a ser mi director de tesis, gracias por su apoyo, confianza y consejos para terminar a bien este trabajo.

A la Universidad Nacional Autónoma de México por brindarme la oportunidad de recibir mi educación en sus aulas.

A los profesores : Ing. Enrique Arenas, Ing. Joel Gómez, M.C. Carlos Hernández Garciadiego.

Al bloque 6, por demostrar un verdadero sentimiento de amistad, lo siento, ya no nos veremos en las islas.

A Gabriela, Irene y Sofía, por demostrarme su gran amistad no sólo con palabras si no con hechos.

Al CONACYT por haberme entregado una beca para la realización de esta tesis.

1. INTRODUCCIÓN	1
1.1 RESUMEN.	1
1.2 CONTENIDO DE LA TESIS	3
2.FUNDAMENTOS DE PERCEPCIÓN REMOTA.	5
2.1. INTRODUCCIÓN.	5
2.2. ORÍGENES.	7
2.3. FACTORES QUE INTERACTÚAN EN LA PERCEPCIÓN REMOTA.	8
2.3.1 ESPECTRO ELECTROMAGNÉTICO.	8
2.3.2. INTERACCIONES DE LA ENERGÍA CON LA ATMÓSFERA.	11
2.3.2.1 DISPERSIÓN.	11
2.3.2.1.1 Dispersión de Rayleigh.	11
2.3.2.1.2 Dispersión de Mie.	12
2.3.2.2 ABSORCIÓN.	12
2.4 SENSORES.	13
2.4.1 TIPOS DE SENSORES.	14
2.5 TIPOS DE IMÁGENES.	15
2.5.1 FOTOGRAFÍA PANCROMÁTICA.	16
2.5.2 FOTOGRAFÍA INFRARROJA (A BLANCO Y NEGRO).	16
2.5.3 FOTOGRAFÍA DE COLOR NATURAL.	17
2.5.4 IMÁGENES MULTIESPECTRALES.	17
2.6. SATÉLITES: UNA FUENTE DE ADQUISICIÓN DE IMÁGENES.	19
2.6.1.1 SISTEMA LANDSAT.	19
2.6.1.1.1 Sistemas generadores de imágenes a bordo de los satélites LANDSAT.	20
2.6.1.2 SISTEMA SPOT.	21
2.6.1.3 SISTEMA ERS-1.	22
2.6.1.4 SATÉLITE IKONOS.	24
3. MÉTODOS DE SEGMENTACIÓN.	25
3.1 INTRODUCCIÓN.	25
3.2 DEFINICIÓN.	26
3.3 SEGMENTACIÓN DE IMÁGENES (PROCESOS BASADOS EN PÍXELES).	27
3.3.1 DETECCIÓN DE DISCONTINUIDADES.	28
3.3.2 DETECCIÓN DE PUNTOS.	28
3.3.3 DETECCIÓN DE BORDES.	29
3.3.3.1 Operador Gradiente (Sobel).	30
3.3.3.2 Operador Laplaciano o Pseudo Laplace.	31

3.3.3.3 Detección de bordes por filtrado Gaussiano.	32
3.3.3.4 Operador de Frei y Chen.	33
3.3.4 DETECCIÓN DE LÍMITES.	34
3.3.4.1 Procesamiento local.	34
3.3.4.2 Procesamiento global(Transformada de Hough).	35
3.3.5 Umbralización.	37
3.3.5.1 Umbralización: Método de White y Rohrer.	38
3.4 SEGMENTACIÓN DE IMÁGENES (PROCESOS BASADOS EN REGIONES).	38
3.4.1 MÉTODOS BASADOS EN CRECIMIENTO DE REGIONES.	38
3.4.1.1 Algoritmo de Brice y Fenema.	39
3.4.1.2 Algoritmo DN.	40
3.4.2 MÉTODOS BASADOS EN DIVISIÓN Y FUSIÓN DE REGIONES.	41
EJEMPLOS .	43
CONCLUSIONES .	47
<u>4. TEORÍA DE LOS CAMPOS ALEATORIOS DE MARKOV (MRF).</u>	<u>48</u>
4.1 PROBABILIDAD Y VARIABLES ALEATORIAS.	48
4.1.1 DEFINICIÓN.	48
4.1.2 FUNCIÓN DE DISTRIBUCIÓN ACUMULADA.	48
4.1.2.1 Momentos de variables aleatorias.	49
4.1.3 FUNCIONES DE DISTRIBUCIÓN DE VA DISCRETAS.	50
4.1.3.1 Distribución Binomial.	50
4.1.3.2 La distribución de Poisson.	50
4.1.4 FUNCIONES DE DISTRIBUCIÓN DE VA CONTINUAS.	51
4.1.4.1 Función de densidad exponencial.	51
4.1.4.2 Función de densidad normal.	51
4.2 TEORÍA BAYESIANA.	52
4.3 ANTECEDENTES EN LOS MRF.	53
4.3.1 SISTEMAS DE VECINDAD Y CLIQUES.	53
4.3.2 CAMPOS ALEATORIOS DE MARKOV.	54
4.3.3 CAMPOS ALEATORIOS DE GIBBS.	55
4.3.4 EQUIVALENCIA ENTRE MRF Y GRF.	56
4.4 ESTIMACIÓN BAYESIANA.	56
4.4.1 ESTIMADOR MÁXIMO A POSTERIORI.	57
4.4.2 MODOS MARGINALES A POSTERIORI.	57
4.4.3 MEAN FIELD (MF).	58
4.4.4 FUNCIÓN DISCRIMINANTE DE BAYES.	58
4.4.4.1 Función discriminante óptima para modelos distribuidos normalmente.	60
4.4.4.1.1 Función discriminante óptima.	60
4.5 EJEMPLOS DE MODELOS MARKOVIANOS.	62
4.5.1 MODELO DE ISING.	62

4.5.2 MODELO DE POTTS.	63
4.6 MÉTODOS LOCALES Y GLOBALES DE MINIMIZACIÓN.	63
4.6.1 MÉTODOS LOCALES.	63
4.6.1.1 Minimización con etiquetas discretas.	64
4.6.1.1.1 Modos condicionales iterados(ICM).	64
4.6.2 MÉTODOS DE MINIMIZACIÓN GLOBALES.	65
4.6.2.1 Muestreador de Gibbs.	65
4.6.2.2 Algoritmo de Metrópolis (a temperatura constante).	66
4.6.2.3 Algoritmo de Recocido Simulado.	67
EJEMPLOS .	70
CONCLUSIONES	76
<hr/>	
5.ESTIMACIÓN DE PARÁMETROS: ALGORITMOS EM Y SEM.	77
<hr/>	
5.1 INTRODUCCIÓN.	77
5.2 MODELANDO A LA IMAGEN.	77
5.2.1 MODELADO POR PÍXEL.	78
5.3 ESTIMACIÓN DE PARÁMETROS.	79
5.3.1 ALGORITMO EM.	79
5.3.2 ALGORITMO SEM.	83
5.4 CÁLCULO DE LOS PARÁMETROS DE UNA IMAGEN POR MEDIO DEL ALGORITMO EM.	84
CONCLUSIONES .	90
<hr/>	
6.SEGMENTACIÓN CONTEXTUAL NO SUPERVISADA: ESQUEMA EM.	91
<hr/>	
6.1 INTRODUCCIÓN.	91
6.2 MODELADO DE CONTEXTO.	91
6.3 ALGORITMO EM Y SEGMENTACIÓN.	93
6.3.1 PROCEDIMIENTO PARA LA SEGMENTACIÓN.	95
6.3.2 RESULTADOS	97
CONCLUSIONES	122
<hr/>	
CONCLUSIONES GENERALES	123
<hr/>	
GLOSARIO	126
<hr/>	
BIBLIOGRAFÍA	128
<hr/>	
ANEXO 1	132

Introducción

1.1 Resumen.

La percepción remota es una ciencia y/o un arte, es el conocer a la tierra desde un lugar donde tal vez algunos nunca soñamos llegar a estar: el cielo. Pero toda observación no es gratis, esta ciencia nos da imágenes muy bellas a las cuales no solo debemos de admirar si no también debemos de investigar y analizar. Es decir les debemos de preguntar que es lo que le pasa a nuestro planeta.

Y es ahí en donde entra el análisis de imágenes, para tener un resultado o una explicación debemos de preguntarle a la imagen, y una forma de preguntarle es segmentándola y clasificándola. Estas son dos técnicas de interrogatorio muy interesantes, prueba de ello son los trabajos que se desarrollan diariamente para dar procesos de segmentación o formas de clasificación.

Al terminar de leer dichos trabajos, a uno ya no le quedan dudas que esta ciencia es muy atractiva y a la vez es un campo muy fértil para el desarrollo de investigaciones.

Es por eso que decidí iniciar esta tesis sobre un tema bastante interesante, la segmentación no – supervisada. Este tipo de segmentación es muy socorrida hoy día por el hecho de que nos presenta resultados sólo con el hecho de entregarle una imagen. Pero no sólo en el campo de la percepción remota, si no también en el campo de las imágenes biomédicas y de la visión de robots.

La presente tesis esta enfocada a definir una herramienta en el tema de la segmentación de imágenes y concretamente en el área de la percepción remota. El esquema propuesto incorpora una etapa de estimación estadística, a fin de realizar procesos de segmentación más automáticos en imágenes pancromáticas.

Mi trabajo se puede resumir de una forma sencilla de la siguiente manera:

1. Obtuve los parámetros característicos de una imagen (número de clases y de estas su media y variancia), de una forma "automática" o no supervisada, al modificar el algoritmo EM en su forma más pura y sencilla para que fuese aplicable a los problemas de segmentación de imágenes. El algoritmo EM es una alternativa para estimar la media y variancia de datos estadísticos dispersos. El cambio no es muy complicado, pero al modificarlo se previó que esto no alterara su convergencia y si aumentara su velocidad de procesamiento, ambos requerimientos se lograron.
2. Una vez obtenidos los datos anteriores sólo se necesita un algoritmo de segmentación para presentar un resultado para nuestros ojos de tal forma que seamos capaces de analizarlos y dar una conclusión. Para realizar esto se proponen dos procedimientos:
 - a) Un algoritmo de umbral: Donde se propone que las medias obtenidas por el algoritmo EM sean consideradas como umbrales para realizar la segmentación.
 - b) Un algoritmo bayesiano: El resultado final de segmentación se obtiene por la aplicación de un algoritmo óptimo bayesiano, ya que el algoritmo EM, tiene como premisa que los datos son de tipo Gaussiano y toma muy en cuenta a la teoría Bayesiana para obtener los parámetros correspondientes de la imagen analizada.

Para evaluar los resultados obtenidos, estos se compararan con un algoritmo clásico de segmentación: el algoritmo de k - medias, de esta forma se observará si la clasificación obtenida por medio del algoritmo EM, siendo este una propuesta de segmentación no - supervisada es de buena calidad visual.

Hay que aclarar, que al referirme a calidad, hago la referencia a que la segmentación debe de cumplir con que se pueda distinguir zonas de interés tales como: construcciones, carreteras, árboles y sombras, al igual que observar una buena contornalización y diferenciar entre zonas claras, oscuras u opacas.

Esto se pide, ya que los resultados de este trabajo se pueden aplicar a la rama de monitoreo vía satélite, pues con este procedimiento se pueden detectar cambios tales como: la disminución gradual de las selvas en el planeta, el monitoreo de los huracanes, etc. Ya que nos presenta la gran ventaja de generar una imagen segmentada de forma no - supervisada, y si se aplica a una secuencia del mismo lugar, se pueden detectar desde cambios muy obvios o difíciles de observar, dependiendo de que se esta buscando.

Por ejemplo: Si se esta monitoreando un lugar del Pacífico, es decir se mandan secuencias temporales de fotografías al centro de búsqueda, para prevenir a la población más cercana contra un posible huracán; se puede generar una alarma tal que nos indique que las imágenes han cambiado y nos de la localización de en donde ha ocurrido tal hecho y observar si en efecto se esta dando el nacimiento de un huracán.

Otra aplicación es en el campo de la visión de robots, pues este algoritmo se puede aplicar para genera las imágenes que "ve" el robot, y por medio de este avisarle al mismo que se ha generado un cambio en lo que esta observando, por ejemplo variaciones en un camino o que se le ha quitado el objeto que estaba observando.

1.2 Contenido de la tesis

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

Capítulo 1.

Presenta un panorama general del problema que se aborda, la motivación para resolverlo y las técnicas que se utilizaron como propuesta de solución.

Capítulo 2.

En este capitulo se explica la teoría básica de la percepción remota y muestra información sobre los principales generadores de imágenes y sus diferentes formas de obtenerlas.

Capítulo 3.

Se presentan las técnicas más básicas para realizar una segmentación, se exponen las diferentes formas de obtener un mapa de contornos o los procesos para realizar un cerrado de contornos. Al igual que explicar como se lleva a cabo una segmentación por medio de las técnicas de crecimiento de regiones.

Capítulo 4.

En este apartado se estudian y desarrollan conceptos básicos de la teoría de los Campos Aleatorios de Markov y de la Teoría Bayesiana, para aplicarlos más adelante en la segmentación de imágenes.

Capítulo 5.

Aquí se explica la estimación de los parámetros estadísticos básicos de una imagen (media, variancia y números de clases). Se exponen dos algoritmos para llegar a dichos resultados, estos son el Algoritmo EM y el Algoritmo SEM. Yo me enfoqué a desarrollar la teoría del algoritmo EM para comprobar su efectividad.

Capítulo 6.

Es este capítulo, se propone una forma de segmentación óptima en el sentido bayesiano, para aplicar los resultados obtenidos del Algoritmo EM. Aquí tengo que ser muy específico, pues los resultados arrojados por el Algoritmo EM nos dan la gran ventaja de saber cual es el número de clases en que estadísticamente se puede segmentar la imagen y de cada una de estas clases, su media y desviación estándar, partiendo de aquí se pueden aplicar otros algoritmos de segmentación. Los resultados que arroja esta tesis permiten que sean utilizados en diferentes algoritmos de segmentación.

Capítulo 7.

Se presentan las conclusiones del presente trabajo de tesis, así como futuras aplicaciones y extensiones de investigación.

Fundamentos de percepción remota.

2.1. Introducción.

La percepción remota es el arte y la ciencia de obtener información acerca de un objeto, área o fenómeno a través del análisis de datos adquiridos por un mecanismo que no esta en contacto con el objeto, área o fenómeno bajo investigación.

Usando varios sensores se pueden adquirir datos remotamente, los cuales se deben de analizar para obtener información acerca de los objetos, área o fenómeno que se esta investigando. La colección remota de los datos puede darse en distintas formas, incluyendo variaciones en la distribución de fuerzas, en las ondas acústicas, de los campos electromagnéticos, etc. Un ejemplo de esto es el sonar, este se comporta como un sistema de navegación de un murciélago, obtiene información de la variación de la distribución de las ondas acústicas.

Pero en muchos aspectos, la percepción remota puede ser explicada (a grandes rasgos) por el proceso de lectura, es decir, nuestros ojos actúan como sensores que responden a la luz reflejada de la página que se está leyendo. Los "datos" que los ojos adquieren son impulsos correspondientes a la cantidad reflejada de luz por la hoja. Estos datos son analizados e interpretados en el cerebro; lo que permite interpretar que las partes más oscuras de la página son letras que forman palabras [12].

A continuación se presenta los principales elementos de un sistema de percepción remota, de acuerdo al esquema de la figura 2.1.

1. Fuente de energía o iluminación: (A) - La Fuente de energía que supone el origen del flujo energético detectado por el sensor. Puede tratarse de un foco externo al sensor, en cuyo caso se habla de teledetección pasiva, o de un haz energético emitido por este, conocido como teledetección activa. La fuente de energía más importante es la energía solar.

2. Radiación y la atmósfera (B) - Es el proceso en el que la energía electromagnética es propagada a través del espacio libre.

3. Interacción con la superficie terrestre (C) – La cubierta terrestre, formada por distintas masas de vegetación, suelo, agua y construcciones, que reciben la señal energética procedente de la fuente de energía, y la reflejan o emiten de acuerdo a sus características físicas.

4. El sistema Sensor (D) - Compuesto por el sensor propiamente dicho y la plataforma que lo sustenta. Tiene como misión captar la energía procedente de las cubiertas terrestres, codificarla y enviarla directamente al sistema de recepción.

5. Sistema de transmisión, Recepción, y proceso (E) - El Sistema de recepción es en donde se recibe la información transmitida por el sensor. Esta información es grabada en un formato apropiado y es distribuida a los intérpretes para su tratamiento visual o digital.

6. Interpretación y análisis (F) - En esta etapa el intérprete, analiza la información, en donde es recibida después de ser transmitida por el sensor. Los datos de interés son grabados en un formato apropiado y distribuidos a los intérpretes para su tratamiento visual o digital; esta información llega al usuario final, quien es el encargado de analizar el documento fruto de la interpretación, así como dictaminar el impacto que de él se derive.

7. Aplicación (G) - El último proceso es el de extraer la información necesaria para resolver un problema específico. El usuario final es el encargado de analizar el documento fruto de la interpretación, así como dictaminar el impacto que de él se derive.

En la figura 2.1 se ejemplifica los 5 primeros elementos de un sistema de percepción remota.

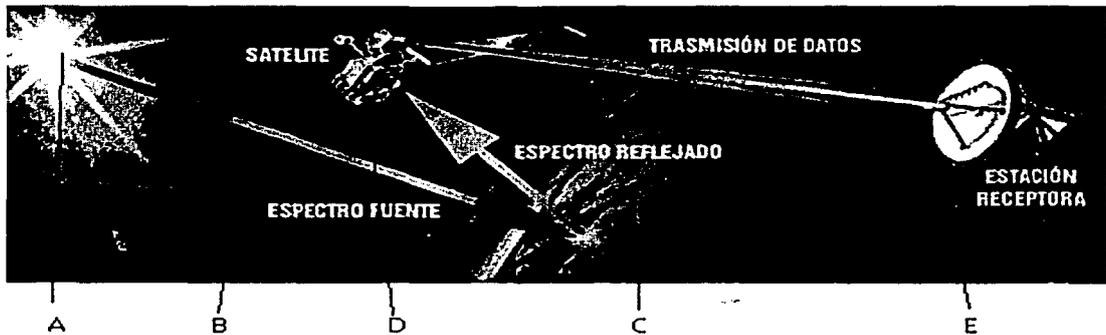


Figura 2.1 Elementos de un sistema de percepción remota.

2.2. Orígenes.

El origen de la percepción remota o teledetección está ligada a los experimentos de Newton(1822), quien constató que un rayo luminoso(luz blanca), al atravesar un prisma, se desdobra en un haz de colores.

Desde entonces los científicos fueron ampliando sus estudios al respecto de tan fascinante materia. Verificaron que la luz blanca es una síntesis de diferentes tipo de luz, es decir un haz compuesto por la mezcla de muchos haces.

Posteriormente, descubrieron que cada color descompuesto en el espectro correspondía a una temperatura diferente y que la luz roja incidiendo sobre un cuerpo, lo calentaba más que la luz violeta. Se descubrió que además del rojo visible, existen radiaciones invisibles para los ojos, llamadas infrarrojas. Después, un experimento de Titter reveló otro tipo de radiación: la ultravioleta.

Sin embargo, algunos autores colocan el origen de la percepción remota ligada al desarrollo de sensores fotográficos, cuando las fotografías aéreas eran obtenidas desde globos.

**TESIS CON
FALLA DE ORIGEN**

2.3. Factores que interactúan en la percepción remota.

2.3.1 Espectro electromagnético.

La luz visible es solamente uno de las pocas formas de la energía electromagnética. Las ondas de radio, el calor, los rayos ultravioleta y los rayos X son otras de sus formas.

Así, al estudiar la electricidad y el magnetismo uno pronto se entera del hecho de que hay un número de relaciones que se describen por productos vectoriales, es decir, un suceso de un tipo produce una respuesta afín perpendicularmente dirigida. De interés inmediato es el hecho de que un campo E variable en el tiempo, genera un campo M que en todas partes es perpendicular a la dirección en la que E cambia. En la misma forma, un campo M variable en el tiempo genera un campo E perpendicular en todas partes a la dirección en la que M cambia[11].

Ahora, consideremos una carga que de alguna manera se acelera desde el reposo. Cuando la carga esta sin movimiento, tiene asociada a ella un campo eléctrico uniforme radial que se extiende hasta el infinito. En el instante en que la carga se mueve, el campo E se altera en la vecindad de la carga y esta alteración se propaga hacia el espacio con velocidad finita. El campo eléctrico variable en el tiempo induce un campo magnético M , de la misma forma este campo M variable en el tiempo genera otro campo eléctrico E , y el proceso continua con E y M acoplados uno a otro en la forma de un pulso. A medida que un campo cambia, genera un nuevo campo que se extiende un poco más allá, y así el punto se mueve a través del espacio.

Los campos E y M pueden, más apropiadamente, considerarse como dos aspectos de un solo fenómeno físico: el campo electromagnético, cuya fuente es una carga en movimiento. La perturbación, una vez que ha sido generada en el campo electromagnético, es una onda sin atadura que se mueve más allá de su fuente e independientemente de ella. Ligados uno a otro como una sola unidad, los campos magnéticos y eléctricos variables en el tiempo se regeneran uno a otro en un ciclo sin fin. Las ondas electromagnéticas que llegan a nosotros del centro de nuestra galaxia, llevan viajando cerca de treinta mil años.

Pero, ¿qué trayectoria describen las ondas electromagnéticas al viajar por el espacio?

Bueno, la teoría electromagnética nos dice que la energía electromagnética viaja como una onda armónica sinusoidal a la velocidad de la luz "c". Y como ya se menciona, el campo eléctrico y el campo magnético son perpendiculares entre sí[34], *observe figura no2.1*

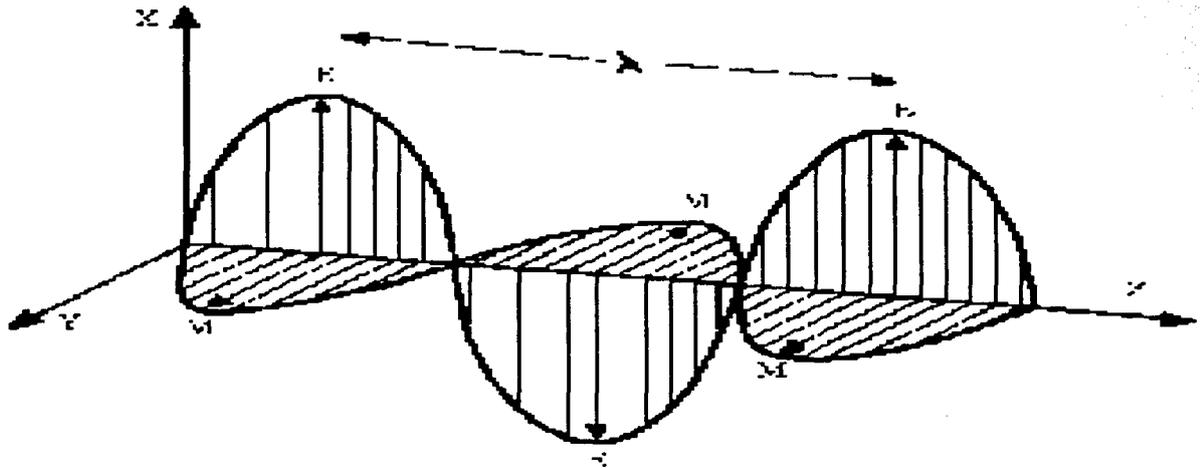


Figura 2.2 Una onda electromagnética. Esta compuesta por un campo eléctrico(E) y un campo magnético(M), donde estos son perpendiculares a la dirección de propagación.

La distancia de un pico de la onda al otro recibe el nombre de longitud de onda λ , y el número de picos por segundo que pasan en un punto determinado, es la frecuencia ν de la onda.

De este pequeño párrafo podemos desprender la siguiente ecuación:

$$C = \lambda \nu \quad 2.3.1-1$$

Dado que C es una constante (por convención se adopta el valor de 3×10^8 m/s), la frecuencia ν y la longitud de onda λ para una onda determinada se relacionan inversamente.

En percepción remota, la clasificación más común para una onda electromagnética es su longitud de onda(λ), la cuál se localiza en el esquema del espectro electromagnético. La unidad que más se utiliza para medir una longitud de onda es el micrómetro (μm). Aunque también se utilizan los nombre (tales como ultravioleta o microondas) con los que generalmente se denomina a las regiones del campo electromagnético. A continuación se enuncian algunos de ellos[34]:

Ondas de radio: Bajas frecuencias y grandes longitudes de onda. Las ondas electromagnéticas en esta faja son utilizadas para la comunicación a larga distancia, ya que, además de ser poco atenuadas por la atmósfera, son reflejadas por la ionosfera, propiciando una propagación de largo alcance.

Microondas: Se sitúan en la faja de 1 mm a 30 cm o 3×10^{11} a 3×10^9 Hz. En esta faja de longitudes de onda se puede construir dispositivos capaces de producir haces de radiación electromagnética altamente concentrados, llamados radares. La poca atenuación por la atmósfera o por las nubes, hace propicio un excelente medio para uso de sensores de microondas en cualquier condición climática.

Infrarrojo: Engloba la radiación con longitudes de onda de $0,75 \mu\text{m}$ a $1,0 \text{ mm}$. La radiación infrarroja es fácilmente absorbida por la mayoría de las sustancias (efecto de calentamiento).

Visible: Es definida como la radiación capaz de producir la sensación de visión para el ojo humano normal. Presentan una pequeña variación de longitud de onda (380 a 750 nm). Importante para la Percepción Remota, ya que las imágenes obtenidas en esta faja, generalmente presentan una excelente correlación con la experiencia visual del intérprete.

Ultravioleta: Extensa faja del espectro (10 nm a 400 nm). Las películas fotográficas son más sensibles a la radiación ultravioleta, que a la luz visible. Esta faja es utilizada para la detección de minerales por luminiscencia y polución marina. Uno de los grandes obstáculos para la utilización de esta región del espectro, es la fuerte atenuación atmosférica.

Rayos X: Faja de 1 \AA a 10 nm ($1 \text{ \AA} = 10^{-10} \text{ m}$). Son generados predominantemente, por la parada o frenada de electrones de alta energía. Por estar constituido por fotones de alta energía, los rayos X son altamente penetrantes, siendo una poderosa herramienta en la investigación sobre la estructura de la materia.

Rayos GAMA: Son los rayos más penetrantes de las emisiones de sustancias radioactivas. No existe, en principio, límite superior para la frecuencia de las radiaciones gama, aunque se encuentre aún una faja superior de frecuencias para la radiación, conocida como rayos cósmicos.

Esta clasificación se ve englobada en la *figura no. 2.3*.

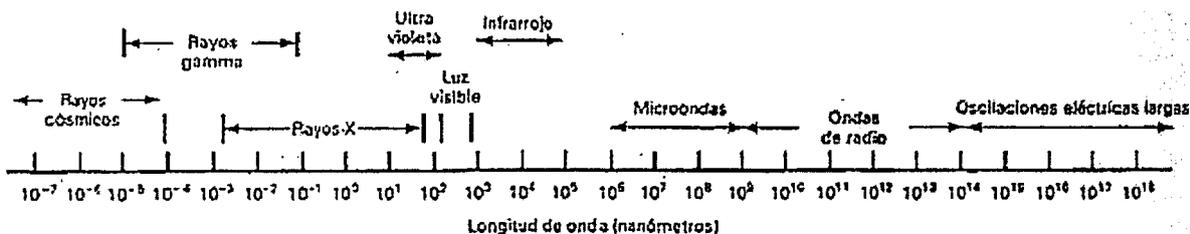


Figura 2.3 Clasificación del espectro electromagnético.

2.3.2. Interacciones de la energía con la atmósfera.

La detección de los objetos por percepción remota esta basada en el procesamiento (tratamiento) de la energía que es emitida o reflejada por la fuente, hecho debido a una agitación térmica. Las longitudes de onda de la energía reflejada usada en las técnicas de percepción remota abarcan de 0.3 a 0.4 μm (llamado fajo óptico), aunque la emisión de energía generalmente oscila en los 3 μm . La energía recibida por un mecanismo de percepción remota como una cámara fotográfica o scanner multiespectral es modificada por la atmósfera antes de llegar al censor.

Si la energía se moviera en el espacio libre, esta no sufriría modificaciones ni pérdidas, pero esto no ocurre, ya que esta pasa por la atmósfera, es decir este cambio de medio (espacio libre – atmósfera) es la que origina las pérdidas en la onda electromagnética, siendo más marcados a grandes alturas. Este fenómeno es lo que provoca que el proceso de estudio por medio de la percepción remota debe ser analizado con detalle, puesto que mucha de la energía emitida por la fuente podría perderse por la interacción con la atmósfera[24].

2.3.2.1 Dispersión.

Una de las principales causas de la modificación de la energía a la llegada de los sensores es la llamada dispersión atmosférica, que es la difusión no predecible de las partículas radiadas en la atmósfera.

2.3.2.1.1 Dispersión de Rayleigh.

La dispersión de Rayleigh es común cuando la radiación interactúa con moléculas atmosféricas y otras partículas diminutas que son más pequeñas en diámetro que la longitud de onda de la radiación. El efecto de la dispersión de Rayleigh es inversamente proporcional a la cuarta potencia de la longitud de onda. Este tipo de dispersión se presenta de forma más marcada en radiaciones que tiene pequeñas longitudes de onda.

Un cielo azul es una manifestación de la dispersión de Rayleigh, ya que si no se presentara, el cielo se vería de color negro. Por que cuando la luz blanca del sol es dispersada, el color azul se dispersa más. No parece ser violeta, aún que el color violeta se dispersa más que el color azul, no hay mucha luz violeta en la luz solar. El lado rojo del espectro puede viajar a través de la atmósfera más lejos porque no se dispersa mucho(Figura no. 2.3). Pero el cielo no siempre es azul. Por momentos, durante el amanecer o el ocaso, el cielo parece ser rojizo. Esto se debe a que los rayos del sol se aproximan a la tierra en ángulos más cercanos a la atmósfera y tienen más moléculas por las que penetrar. La luz azul se dispersa muy lejos, y las luces roja y naranja son las únicas que pueden pasar a través de la atmósfera.

La dispersión de Rayleigh es una de las principales causas de “neblina” en las imágenes. Visualmente este fenómeno disminuye el “encrespado”, o contraste de una imagen. En fotografía de color, los efectos de la dispersión hacen que se observe un tono azul – gris en toda la fotografía, particularmente cuando se toma desde grandes alturas[33].

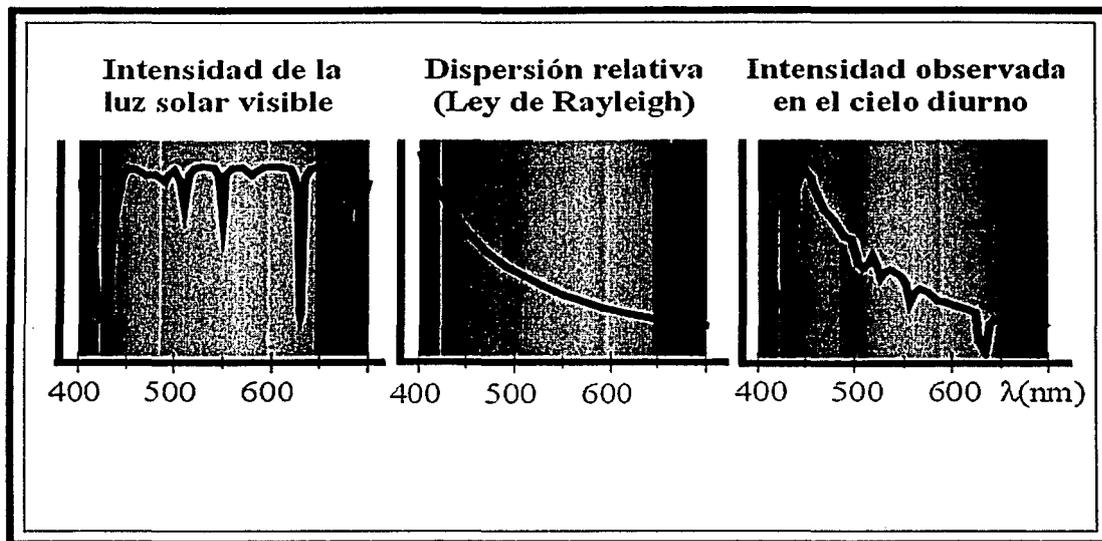


Figura 2.4 Dispersión de Rayleigh.

2.3.2.1.2 Dispersión de Mie.

Otro tipo de dispersión es la de Mie, esta se presenta cuando el diámetro de las partículas atmosféricas son relativamente iguales a la longitud de onda de energía del objeto examinado. El vapor de agua y el polvo son los principales causantes de la dispersión de Mie.

2.3.2.2 Absorción.

En contraste con la dispersión, la absorción atmosférica resulta de la pérdida efectiva de energía en la atmósfera. Esto provoca una absorción de energía en una longitud de onda dada. Los mejores absorbentes de radiación solar son el vapor de agua, el dióxido de carbono y el ozono. Ya que estos gases tienden a absorber la energía electromagnética en bandas específicas de longitud de onda (de hecho, la atmósfera es transparente tan sólo a la luz visible y a ciertas regiones del infrarrojo y de las ondas de radio, siendo opaca al resto

de las frecuencias), esto está íntimamente ligado con el tipo de sensor espectral que se tiene para hacer la percepción remota ya que si lo que se pretende es estudiar la atmósfera (satélites meteorológicos), los sensores se diseñarán para captar la radiación precisamente en aquellos lugares donde la absorción atmosférica sea más alta.

2.4 Sensores.

Como ya fue mencionado, todos los materiales y fenómenos naturales absorben, transmiten, reflejan y emiten selectivamente radiación electromagnética. Con el desarrollo actual es posible medir con una precisión razonable y a distancia, las propiedades espectrales de aquellos materiales y fenómenos[38].

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

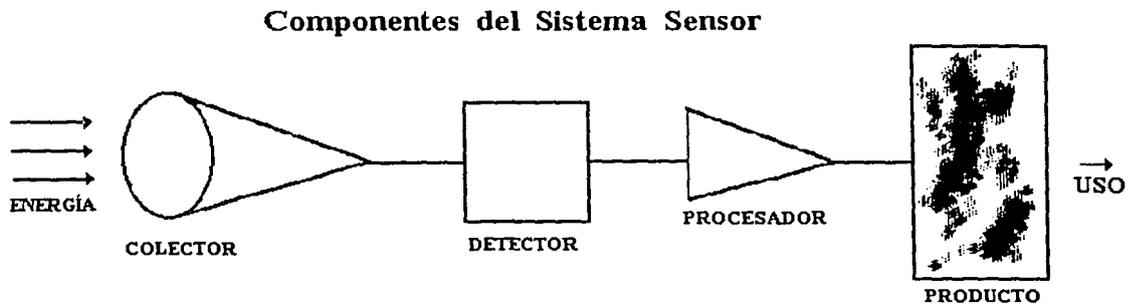


Figura 2.5 Componentes de un sensor.

donde :

Colector. Recibe la energía a través de una lente, espejo, o antena.

Detector. Capta la energía colectada de una determinada faja del espectro;

Procesador. La señal registrada es sometida a un procesamiento (revelado, ampliación) a través del cual se obtiene el producto;

Producto. Contiene la información necesaria para el usuario.

El ejemplo más sencillo de un sistema sensor es una cámara fotográfica, puesto que cuando la cámara es accionada, el flash se activa y emite radiación. Dicha radiación fluye hacia el objeto y es reflejada por este, hacia el sistema óptico de la cámara. La radiación es registrada por la película de la cámara (siendo esta un detector fotoquímico de radiación). Para que después de ser tratada químicamente, de cómo resultado una imagen en un pedazo de papel fotográfico.

2.4.1 Tipos de Sensores.

Una de las formas más habituales para clasificar a los sensores remotos es mediante el procedimiento para recibir la energía procedente del objeto que se está estudiando.

En este sentido, se habla de dos tipos de sensores:

1. *Sensores pasivos*, cuando se limitan a recibir la energía proveniente de un foco exterior a ellos es decir que este tipo de sensores recogen la energía electromagnética procedente de la cubierta terrestre, ya sea reflejada de los rayos solares o emitida debido a su propia temperatura *figura no 2.6.*

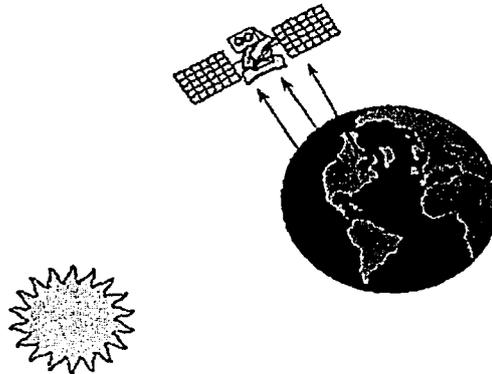


Figura 2.6 Un ejemplo de un receptor pasivo.

En función del procedimiento que emplean para recibir la radiación procedente de los objetos, los sensores pasivos se dividen en:

- a) sensores fotográficos.
- b) exploradores de barrido y empuje, cámaras de vidicon, cámaras de video
- c) sensores de antena o radiómetros de microondas.

TESIS CON
FALLA DE ORIGEN

2. *Sensores activos*, cuándo son capaces de emitir su propio haz de energía, que posteriormente recogen por reflexión sobre la superficie que se pretende observar.

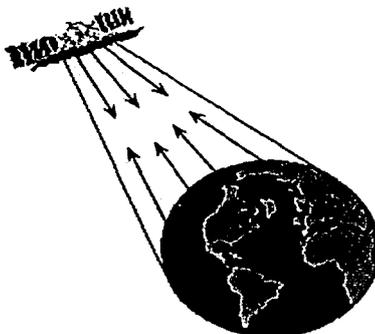


Figura 2.7 Ejemplo de un sensor remoto activo.

El sistema más conocido es el radar (radiómetro activo de microondas), que trabaja en una banda comprendida entre 0.1cm y 1m. Debido a su flexibilidad, puede trabajar en cualquier tipo de condición atmosférica. Por ello, el radar ha acaparado la atención de la comunidad científica internacional, especialmente en países ecuatoriales, donde son persistente las masa nubosas[36].

Una ulterior clasificación entre los sensores pasivos, considera su procedimiento para grabar la energía recibida. De acuerdo a ello, puede hablarse de sensores fotográficos, opto electrónicos y de antena. Los primeros resultan muy similares, puesto que se basan en las mismas técnicas aplicadas en la fotografía convencional. El resto son menos conocidos por público(ejemplos de ellos son las cámaras de video y scanner).

2.5 Tipos de imágenes.

Entre los sistemas aéreos disponibles para poder obtener una imagen se destacan las cámaras aéreas, los scanners electrónicos multispectrales, scanners electrónicos térmicos infrarrojos (TIR), radiómetros pasivos para producir imágenes por microondas y radares aéreos de vista lateral (SLAR).

Cada uno de estos da una presentación diferente de una imagen o fotografía, siendo las principales las siguientes[24]:

1. Pancromática
2. Fotografía infrarroja
3. Multiespectral
4. Color Natural
5. Falso color

2.5.1 Fotografía Pancromática.

Es una película simple que registra variaciones en el rango visible del espectro electromagnético (0.4 a 0.7 μm) en blanco y negro y en escala de grises. La imagen que se obtiene de resultado es la fotografía pancromática.

Las ventajas que presenta este tipo de película son las siguientes:

- No necesita de procesamiento sofisticado para su revelado.
- La mayoría de los países cuentan con este tipo de película, y tienen registros de sus territorios de años atrás, permitiendo una fácil comparación entre fotografías recientes y pasadas.

Las desventajas principales son:

- Por ser fotografías en blanco y negro la identificación de caminos, puentes y construcciones es hasta cierto punto posible, pero otros objetos no son distinguibles.
- La diferenciación entre diferentes superficies, como por ejemplo distintas cosechas, es casi imposible.
- Además, el tono de algún rasgo en particular se debe a los efectos combinados de la reflectancia a través del espectro visible.

2.5.2 Fotografía infrarroja (a blanco y negro).

La sensibilidad de esta película al contrario de la pancromática, es que se extiende al infrarrojo cercano. La sensibilidad de esta película esta dentro del rango de los 0.78 a los 0.88 μm .

Ventajas:

- La dispersión en una imagen de tipo infrarrojo es menor que en una pancromática.
- Las imágenes infrarrojas son ideales para delimitar las fronteras entre el agua y la tierra.
- Es ideal para el estudio de vegetación. La reflectancia por la grasa y la hierba son mucho mayores en el rango infrarrojo.

Desventajas:

- Tiene muy poco detalle en zonas oscuras.
- Como no penetra el agua, los objetos que están bajo su superficie no son observables.

2.5.3 Fotografía de color natural.

Diferentes colores pueden ser producidos por una combinación aditiva de los colores primarios: rojo, verde y azul. Al sobreponerlos en orden de dos, podemos obtener otros colores, a los cuales se les denomina complementarios. Cuando los tres colores primarios son sobrepuestos en cantidades diferentes, se produce una gran gama de colores.

Ventajas de la fotografía de color natural:

- Superficies que son indistinguibles en una fotografía en blanco y negro son fácilmente diferenciables en una fotografía a color, por que el sistema óptico humano puede diferenciar fácilmente millares de colores, pero muy pocos niveles de gris.
- Los colores producidos están de acuerdo con nuestra percepción diaria.

Desventajas:

- Se necesita un equipo de procesamiento para la película muy caro.
- Este tipo de fotografía contiene mucha información de distracción.

2.5.4 Imágenes multiespectrales.

La fotografía multiespectral envuelve simultáneamente a imágenes obtenidas de la misma escena en diferentes longitudes de onda. El arreglo más común de la adquisición de una imagen multiespectral es la obtención de cuatro placas, una en azul, rojo, verde y el cercano infrarrojo.

Las imágenes multiespectrales permiten obtener imágenes de banda simple, de color natural y falso color.

Una imagen de falso color puede ser formada los datos del verde sobre el azul, el rojo sobre el verde y el infrarrojo sobre el rojo.

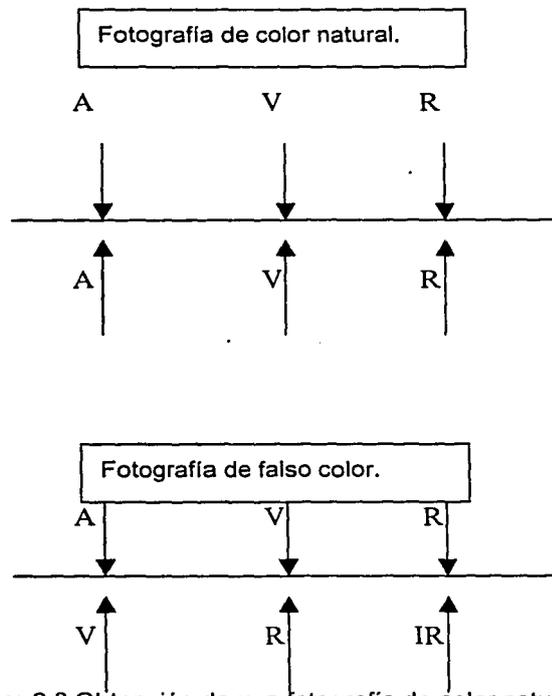


Figura 2.8 Obtención de una fotografía de color natural y falso color..

Donde

A = Azul

R = Rojo

V = Verde

IR = Infrarrojo.

La disponibilidad de imágenes de percepción remota aérea varía según el tipo de datos requeridos. La fotografía aérea está disponible para muchas áreas de estudio en la mayor parte del mundo, aunque en algunas ocasiones deberá ser autorizada para uso no militar por el gobierno del país objeto del estudio. Las imágenes de radar suelen tener también carácter reservado.

2.6. Satélites: Una fuente de adquisición de imágenes.

En la actualidad, el sistema Landsat representa la mayor fuente de datos de percepción remota y es el que tiene mayor potencial de continuidad a lo largo del tiempo.

Antes de comentar sobre cada uno de los sistemas, se presentan sus características orbitales, las cuales se basan en las siguientes premisas:

- la órbita debe ser circular, para garantizar que las imágenes tomadas en diferentes regiones de la Tierra tuvieran la misma resolución y escala.
- la órbita debería permitir el imageamiento (generación de una imagen de la superficie observada) cíclico de la superficie, para garantizar la observación periódica y repetitiva de los mismos lugares;
- la órbita debe ser síncrona con el sol (heliosíncrona), para que las condiciones de iluminación de la superficie terrestre se mantuvieran constantes;
- el horario de pasada del satélite debería atender las solicitudes de diferentes áreas de aplicación (geología, geomorfología, agricultura, etc., ...).

La tabla siguiente presenta las características de los sistemas Landsat, SPOT y ERS-1.

	Landsat 4 y 5	SPOT 1 y 2	ERS-1
Órbita	circular 98,2 grados Heliosíncrona	Circular 98,7 grados heliosíncrona	Circular 98,5 grados Heliosíncrona
Período	99 minutos	97 minutos	100,467 minutos
Altitud	705 Km	832 Km	785 Km
Cruce	9:45 horas	10:39 horas	10:30 horas (desc.)
Ciclo	16 días	26 días	35 días (SAR)
Órbita adyacente	172 Km	108 Km	100 Km
Órbita sucesiva	2.750 Km	2.700 Km	-

2.6.1.1 Sistema Landsat.

El sistema LANDSAT está compuesto hasta el momento por 5 satélites, que fueron desarrollados por la NASA (National Aeronautics and Space Administration), inicialmente recibieron el nombre Earth Resources Technology Satellite-1 (ERST-1) y en enero de 1975 pasaron a ser llamados LANDSAT.

El Landsat 1 y 2 llevaban a bordo dos sistemas sensores con la misma resolución espacial, pero con diferentes concepciones de imageamiento: el sistema RBV, con imageamiento instantáneo de toda la escena y el sistema MSS, con imageamiento del terreno por barrido de líneas (line-scanner).

Ambos sistemas tenían la finalidad de adquirir datos multispectrales, pero el desempeño del sistema MSS, en términos de fidelidad radiométrica, hizo que el tercer satélite de la serie tuviera un sistema RBV modificado, de modo de proveer datos con mejor resolución espacial en una única faja del espectro. Por otro lado, se adicionó una banda espectral al sistema MSS, para operar en la región del infrarrojo termal. Observe la figura 2.9.



Figura 2.9 Imagen del estado de Baja California Norte tomada por el sistema Landsat.

2.6.1.1.1 Sistemas generadores de imágenes a bordo de los satélites LANDSAT.

IMAGEADOR RBV (Return Beam Vidicon): es un sistema semejante a una cámara de televisión y permite el registro instantáneo de un área del terreno. La energía proveniente de toda la escena sensibiliza la superficie fotosensible del tubo de la cámara y durante un determinado tiempo, la entrada de energía es interrumpida por un obturador, para que la imagen del terreno sea barrida por un haz de electrones. La señal de vídeo puede ser transmitida teleméricamente.

IMAGEADOR MSS (Multispectral Scanner): es un sistema sensor que permite el imageamiento de líneas del terreno en una faja de 185 Km, perpendicularmente a la órbita del satélite. El barrido del terreno es realizado con ayuda de un espejo que oscila perpendicularmente al desplazamiento del satélite. Durante la oscilación del espejo, la imagen del terreno, a lo largo de la faja, es captada por una matriz de detectores. La dimensión de cada detector que compone la matriz de detectores, es responsable por el campo de visión instantáneo (área de la superficie de la Tierra observada por cada detector). La energía registrada por cada detector es transformada en una señal eléctrica y posteriormente es transmitida para las estaciones en la Tierra.

A cada oscilación del espejo, el satélite se desliza a lo largo de la órbita, para proporcionar un imageamiento continuo del terreno. El movimiento de rotación de la Tierra ocasiona un pequeño desplazamiento del punto inicial del barrido para el oeste, a cada oscilación del espejo, o sea, a cada seis líneas imageadas. Si se considera el desplazamiento de 185 Km a lo largo de la órbita del satélite, hay un desplazamiento de 12.5 cm entre la primera y la última columna de píxeles.

IMAGEADOR TM (Thematic Mapper): es un sistema avanzado de barrido multispectral concebido para proporcionar: resolución espacial más fina, mejor discriminación espectral entre objetos de la superficie terrestre, mayor fidelidad geométrica y mejor precisión radiométrica con relación al sensor MSS.

La energía proveniente de la escena alcanza el espejo de barrido que oscila perpendicularmente a la dirección de desplazamiento del satélite en sentido este-oeste y oeste-este. La señal atraviesa un telescopio y un conjunto de espejos, cuya función principal es corregir la señal colectada por el espejo de barrido. De esta manera, la señal detectada en cada matriz de detectores de cada banda es transferida para un amplificador y convertida en una señal digital a través de un sistema A/D (analógico/digital). Los datos de salida son entonces transmitidos vía telemetría.

2.6.1.2 Sistema SPOT.

El sistema SPOT es un programa espacial francés semejante al programa Landsat. En 1986, Francia lanzó el primer satélite de este sistema desde Guinea Francesa en África, que fue concebido por el Centre National d'Etudes Spatiales (CNES) y lleva a bordo dos sensores de alta resolución (HRV - HAUT Resolution Visible). Estos sensores fueron concebidos para operar en dos modos diferentes. El modo multispectral permite la adquisición de datos en tres fajas del espectro electromagnético con una resolución espacial de 20 metros y el modo pancromático con una banda de resolución espacial de 10 metros. *Vea figura 2.9*

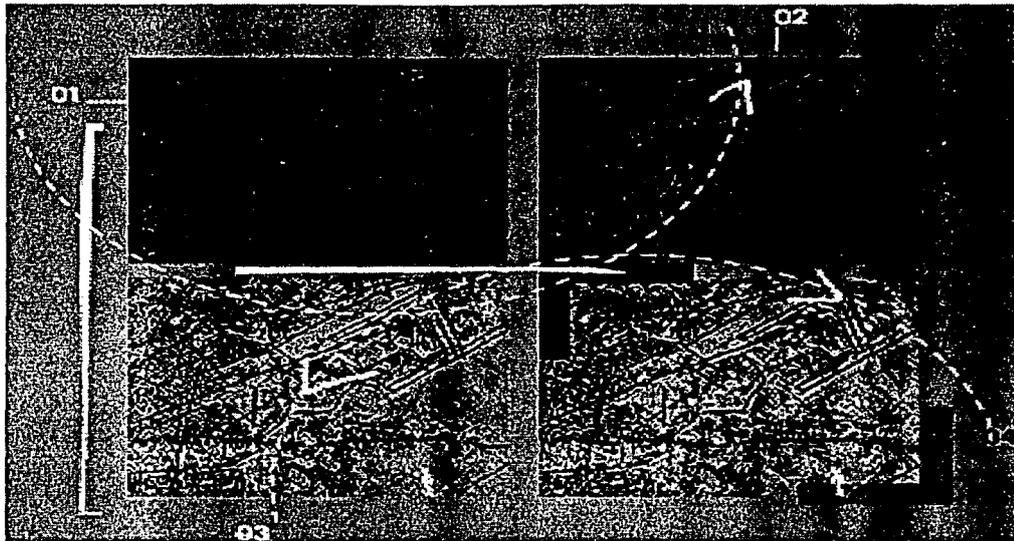


Figura 2.9. Imagen tomada por el sistema Spot.

Una de las características principales de los instrumentos a bordo del SPOT es la posibilidad de observación "off-nadir". El sensor podrá ser direccionado de modo de observar escenas laterales a la órbita en la que se encuentra el satélite en un momento determinado. Esta posibilidad de observación "off-nadir" incrementa los medios de obtener un aumento en el recubrimiento repetitivo de determinadas áreas. Otra ventaja de la vista "off-nadir" es la posibilidad de obtener pares estereoscópicos de determinadas áreas. La luz proveniente de la escena alcanza un espejo plano, que puede ser controlado a partir de las estaciones terrestres variando en ángulos de $\pm 0,6$ hasta 27° con relación al eje vertical[37].

La energía que alcanza el espejo plano es captada por una matriz lineal de detectores del tipo CCD (Charge-Coupled Device). Cada matriz consiste en 6000 detectores ordenados linealmente, formando lo que se denomina "push-broom scanner" o sistema de barrido electrónico. Este sistema permite el imageamiento instantáneo de una línea completa en el terreno, perpendicularmente a la dirección de desplazamiento del satélite en su órbita.

2.6.1.3 Sistema ERS-1.

Fue construido por la Agencia Espacial Europea (ESA), el ERS-1 fue lanzado del centro espacial de Guyana Francesa por el cohete Ariane 4 el 16 de julio de 1991. Originalmente con una misión de dos años, los objetivos estaban orientados principalmente a estudios oceánicos y de heladas, en varias áreas de ciencias naturales.

Entre los diferentes sensores a bordo del satélite, se tiene el AMI (Active Microwave Instruments), constituido por un radar de Abertura Sintética (SAR) y un escaterómetro (para medición de vientos).

SAR (Radar de apertura sintética): Este tipo de radar difiere del Radar de Apertura Real principalmente en el hecho de que usa una antena de longitud efectiva, sólo que se sintetiza al usar procesamiento de señales en vez de usar una antena físicamente grande. Esto se logra por que un sistema SAR usa el principio de Doppler para generar imágenes con una gran resolución de azimuth.

Las imágenes adquiridas por el SAR, suministran datos de una faja de 100 x 100 Km, con una resolución espacial de 30 metros (*figura 2.10*).

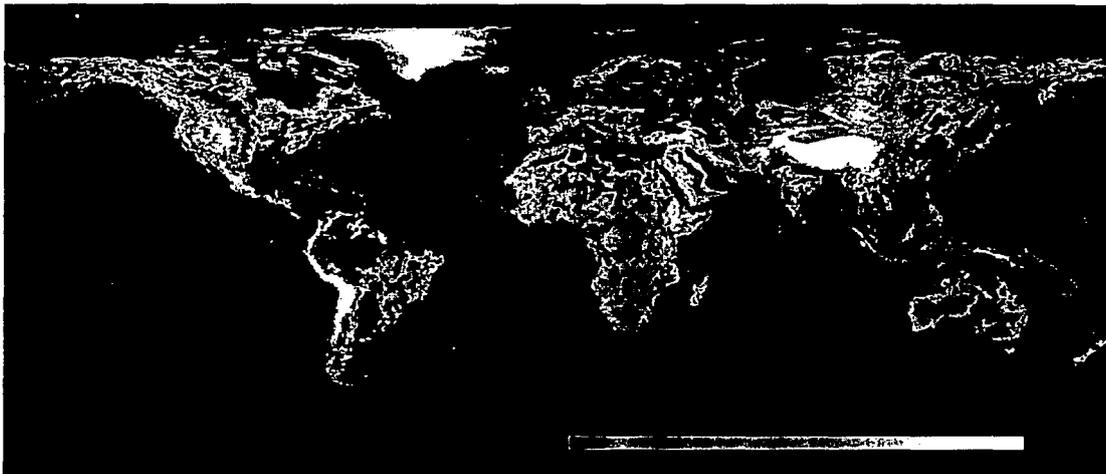


Figura 2.10. Imagen tomada por el sistema IRS, utilizando el modelo de elevación global de la tierra, la escala de colores muestra la altura sobre el nivel del mar.

Una antena de 10 x 1 metros emite y recibe un haz de microondas en la región de 5,3 Ghz (banda C), con polarización VV y un ángulo de incidencia de 23 grados.

La operación del SAR en modo Imagen produce una tasa de datos muy alta (105 Mbps), haciendo que las imágenes sólo puedan ser generadas en zonas equipadas con estaciones receptoras. La superficie terrestre podrá estar enteramente cubierta e imageada en ciclos de 35 días.

TESIS CON
FALLA DE ORIGEN

2.6.1.4 Satélite Ikonos.

Este satélite fue lanzado en 1999. Tiene la posibilidad de producir imágenes de 1 metro de resolución pancromática (0.45 a 0.9 μm), 4 metros en imágenes multispectrales, y un metro en imágenes a color (Banda Azul: 0.45 – 0.53 μm , Banda Verde: 0.52 – 0.61 μm , Banda Rojo: 0.64 – 0.72 μm , Banda IR de 0.77 – 0.88 μm).

Los datos son coleccionados en 11 bits por píxel, para su presentación puede realizarse en 11 bits por píxel o en ocho bits por píxel.

Este satélite envía las imágenes capturadas hacia las estaciones terrenas a una tasa comparable a ver 50 canales de televisión simultáneamente.

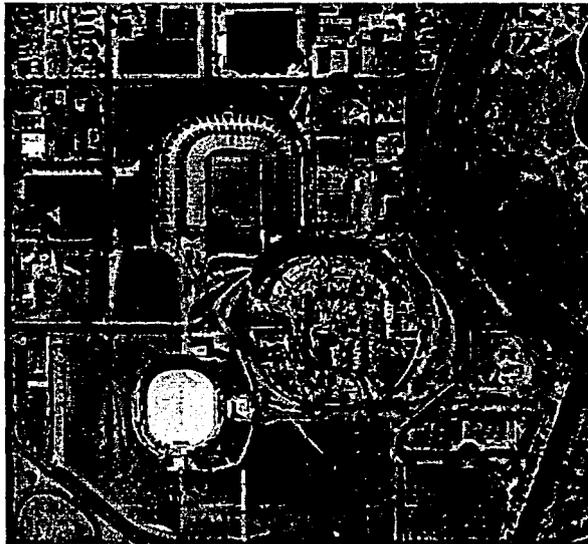


Figura 2.11 Imagen del satélite Ikonos.

Métodos de Segmentación

3.1 Introducción.

El principal objetivo del proceso de segmentación de una imagen es la partición de la imagen en regiones muy significativas que corresponden a una parte o un conjunto de objetos dentro de la misma. Esto se puede hacer al dividir sistemáticamente a la imagen en regiones o áreas.

Los procesos de subdivisión deben cesar cuando todas las regiones de interés han sido identificadas y ya no se debe hacer otra subdivisión más. Esto puede ser fácil de alcanzar en aplicaciones con cierto control donde el resultado de la segmentación es bien definida (inspección visual) pero presenta muchas dificultades en aplicaciones donde el resultado de la segmentación no se conoce (guía de robots)[10]. Pero, en todos los casos el nivel de segmentación dependerá de que problema se quiera solucionar.

Existen dos principales formas de realizar una segmentación:

- a) Métodos locales: Basado en píxeles.
- b) Métodos globales: Basado en regiones.

Estos procedimientos son complementarios y podrían generar resultados iguales, pero en la práctica, rara vez se da ese caso. Los procedimientos basados en píxeles buscan, detectan y aumentan un contorno, después conectan todos los elementos encontrados para formar una frontera que encierra la región de uniformidad.

Los métodos basados en regiones tratan de crear directamente regiones agrupando los píxeles que comparten características comunes dentro de áreas y regiones de uniformidad.

3.2 Definición.

El fin de la segmentación es dividir a la imagen en regiones que tengan una cierta uniformidad. Pero para poder dar la definición, primero se explicara que es un “predicado lógico”.

Considérese una imagen de m columnas por n renglones (*figura 3.1 a*), la letra R denota el arreglo completo de píxeles, esto es todos los pares $\{i,j\}$ donde $i = 0,1,2,3,\dots(m-1)$, y $j = 0, 1, 2,3,\dots (n-1)$.

Sea R_a un subconjunto no vacío de R de los píxeles contiguos de la imagen. Un “predicado lógico”[23], $P(R_a)$ es un enunciado lógico que asigna un valor: Verdadero o Falso a R_a , dependiendo solamente de la propiedades relacionadas a la intensidad de la matriz $f(i,j)$ para los de R_a . Además P tiene la propiedad de que si R_b no es un subconjunto vacío de R_a , entonces:

$P(R_a) = \text{Verdadero}$ implica que $P(R_b)$ es verdadero.

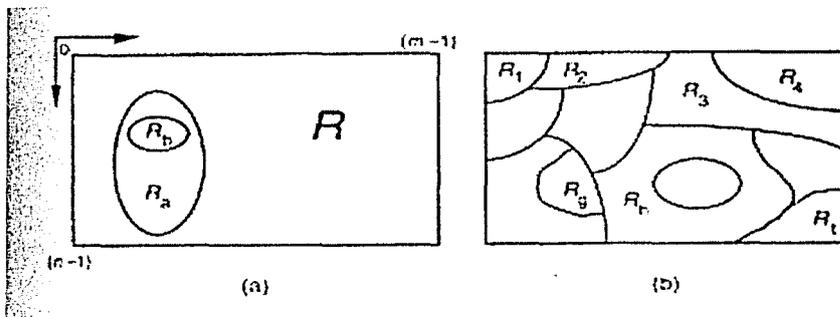


Figura 3.1 Ejemplo de un predicado lógico para el proceso de segmentación

Una segmentación del arreglo R , vea la *figura 3.1 b*, para un predicado de uniformidad P es una partición de R con subconjuntos no vacíos R_1, R_2, R_3,\dots,R_t de tal forma que:

- a) $\bigcup R_g = R$ para $g = 1,2,3,\dots,t$.
- b) R_g es una región conectada para $g = 1,2,3,\dots, t$.
- c) $R_g \cap R_h = \phi$ para todo g y h ; donde $g \neq h$.
- d) $P(R_g) = \text{Verdadero}$ para $g = 1,2,3,\dots, t$.
- e) $P(R_g \cup R_h) = \text{Falso}$ para $g \neq h$.

Donde $P(R_g)$ es la probabilidad de alguno de los subconjuntos de R y ϕ es el conjunto vacío.

Hay una gran variedad de rasgos o características de una imagen a las que se les puede aplicar un proceso de segmentación, estas incluyen intensidad (valores de niveles de gris), parámetros de color (RGB, color y saturación), rango y límite de información, textura y movimiento.

3.3 Segmentación de Imágenes (procesos basados en píxeles).

Los métodos basados en píxeles cuentan con la habilidad de detectar la presencia de cambios que se den en cualquiera de sus puntos (píxeles), también detectan los elementos de contorno, segmentos de contorno (líneas), o segmentos de líneas conectados (límites) que aparezcan en la imagen[3].

La segmentación basada en detección de contornos puede resumirse como una serie de pasos (*figura 3.2*). El primer paso la imagen es filtrada para superar las irregularidades producidas por el ruido. Después se aplica un detector de contornos , la imagen obtenida puede tener información que no nos interesa (provocadas por el ruido) y deberá ser removida por un método de umbralización, así se espera que en la imagen sólo aparezcan contornos (mapa binario de contornos). Finalmente los contornos individuales son agrupados respetando los límites encontrados en este proceso.

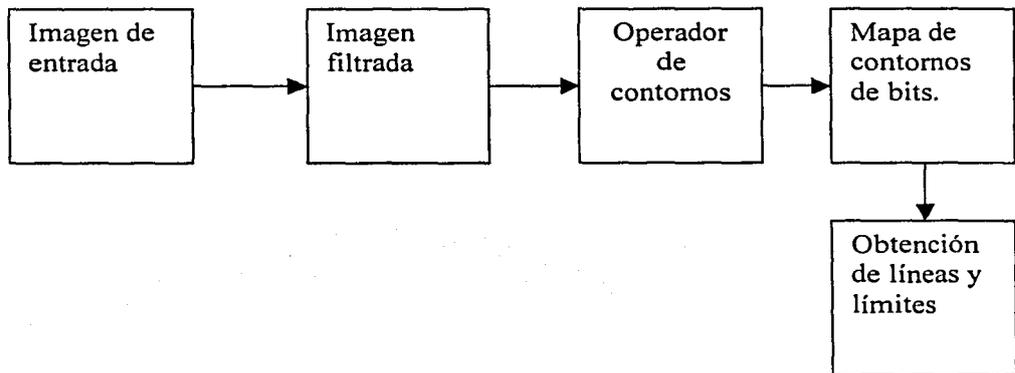


Figura 3.2 Procedimiento global de segmentación para un proceso de detección de contornos.

A continuación se enunciarán algunos de los procedimientos básicos que se conocen para hacer detecciones de discontinuidades y de puntos, ya que estos procedimientos son de gran importancia, puesto que de ellos depende si se prepara a la imagen directamente para la segmentación o se realiza un proceso de realce o filtrado.

3.3.1 Detección de discontinuidades.

En la práctica, la forma más común de ver las discontinuidades es la de pasar una máscara a través de la imagen. Así por ejemplo para una máscara de 3X3 como se muestra en la figura 3.3

W1	W2	W3
W4	W5	W6
W7	W8	W9

Figura3.3. Una máscara general de 3X3

Este procedimiento implica calcular la suma de los productos de los coeficientes por los niveles de gris contenidos en la región encerrada por la máscara. Así entonces, la respuesta de la máscara en un punto cualquiera de la imagen es:

$$R = w1z1 + w2z2 + \dots + w9z9 \quad (3.3.1 - 1)$$

Donde $z1$ es el nivel de gris asociado con el coeficiente de la máscara $w1$. Se observa que la respuesta de la máscara esta definida en el centro de la misma.

3.3.2 Detección de puntos.

La detección de los puntos aislados de una imagen es directa. Utilizando la máscara que se muestra en la figura 3.2, se sabe que se ha detectado un punto en la posición en la que está centrada la máscara si

$$|R| > T \quad (3.3.2 - 1)$$

donde T es un umbral no negativo, y R está dado por la ecuación(3.3.1 -1). Básicamente, lo que se hace es medir las diferencias ponderadas entre el punto central y sus vecinos. Ya que se parte de la idea de que si hay un punto aislado será bastante diferente del de sus vecinos.

La máscara de la figura 3.4 es la misma que la utilizada para el filtrado de alta frecuencia espacial. Sin embargo, aquí se hace hincapié estrictamente en la detección de puntos. Esto es, solamente son de interés las diferencias lo suficientemente grandes(como las determinadas por T) para que se consideren puntos aislados de una imagen.

-1	-1	-1
-1	8	-1
-1	-1	-1

Figura 3.4 Una mascara utilizada para detectar diferentes puntos aislados.

Después de aplicar estas dos pequeña máscaras, se podrá decidir si se hace un filtrado o realce de la imagen, puesto que si los resultados salen mucho más bajos que el umbral que se escogió, querrá decir que la imagen esta (hasta cierto punto) libre de ruido, y que la imagen está lista para ser segmentada.

3.3.3 Detección de bordes.

Aunque la detección de puntos y líneas son evidentemente elementos de cualquier presentación de la segmentación, la detección de bordes es con mucho el método más común para detectar discontinuidades significativas en el nivel de gris. La razón es que los puntos aislados y las líneas delgadas no son de frecuente aparición en la mayor parte de las aplicaciones prácticas.

Formulación Básica:

Un borde es la frontera entre dos regiones con propiedades de nivel de gris relativamente distintas. En la siguiente presentación se supone que las regiones en cuestión son lo suficientemente homogéneas para que la transición entre dos de ellas se pueda determinar sobre la base de las discontinuidades de nivel de gris solamente.

Básicamente la idea en que subyacen la mayor parte de las técnicas de detección de bordes es el cálculo de un operador local de derivación. En donde el módulo de la primera derivada se puede utilizar para detectar la presencia de un borde en la imagen y el signo de la segunda derivada se puede utilizar para determinar si un píxel está situado en el lado oscuro o claro del mismo.

La derivada primera en un punto de una imagen se obtiene utilizando el módulo del gradiente en ese punto. La segunda derivada se obtiene de forma similar utilizando el laplaciano.

Sin embargo, se presentarán algunas simplificaciones matemáticas, para poder obtener tanto el gradiente como el laplaciano de una imagen, estas simplificaciones reciben el nombre de operadores.

Estos operadores tienen en común una generalidad, son llamados "mapas de bordes" de una imagen. Un "mapa de bordes" contiene todas las líneas de contorno de un objeto o de una región homogénea de una imagen (por ejemplo líneas brillantes o fondo negro). En el caso ideal, el resultado de aplicar alguno de estos operadores es la presencia de sólo líneas de contorno en la imagen de resultados.

3.3.3.1 Operador Gradiente (Sobel).

El gradiente de una imagen $f(x,y)$ en la posición (x,y) es el vector:

$$\nabla f = \begin{pmatrix} G_x \\ G_y \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} \quad (3.3.3.1 -1)$$

Se sabe de antemano que el gradiente de un vector indica la dirección de la máxima desviación de f en (x,y) .

En procesamiento de imágenes para obtener el valor del gradiente se utiliza la siguiente simplificación:

$$\nabla f = |G_x| + |G_y| \quad (3.3.3.1 -2)$$

Pero para simplificar el cálculo, no se obtiene el valor del gradiente en la imagen, si no, lo que se hace es pasar dos máscaras de 3X3 por la imagen, dichas máscaras reciben el nombre de *operadores de Sobel* y son las siguientes:

-1	0	1
-2	0	2
-1	0	1

Figura 3.5 Operador de Sobel Y

-1	-2	-1
0	0	0
1	2	1

Operador de Sobel X

Estas máscaras se aplican de la siguiente manera:

El operador de Sobel X se aplica a la dirección horizontal de la imagen y el operador de Sobel Y se aplica a la dirección vertical de la imagen. Estas máscaras se aplican a cada píxel de la imagen. El píxel de estudio se coloca en el centro de la máscara y los píxeles alrededor de él son multiplicados por los factores del núcleo.

Para obtener el valor del gradiente en el píxel de estudio, lo único que se hace es aplicar la siguiente ecuación

$$G_{xy}(W_s) = \sqrt{(G_x)^2 + (G_y)^2} \quad (3.3.3.1 -3)$$

Para obtener el valor del siguiente nivel, sólo se desplaza la máscara y se repite el proceso. Así, el resultado del operador de Sobel es el resultado de la suma de dos convoluciones

3.3.3.2 Operador Laplaciano o Pseudo Laplace.

El laplaciano de una función bidimensional $f(x,y)$ es una derivada de segundo orden definida por:

$$\nabla^2 f = \frac{\partial^2 f}{\partial^2 x} + \frac{\partial^2 f}{\partial^2 y} \quad (3.3.3.2 -1)$$

Como se hizo con el gradiente, el Laplaciano puede ser implementado por una máscara de la siguiente forma:

0	-1	0
-1	4	-1
0	-1	0

Figura 3.6 Mascara del Laplaciano.

Para una región de 3X3, el operador Laplaciano se encuentra de la siguiente manera:

$$\nabla^2 f = 4W5 - (W2 + W4 + W6 + W8) \quad (3.5.2.2 -1)$$

donde el valor de W ya fue definido.

El requisito básico para la definición del Laplaciano es que el coeficiente asociado con el píxel central sea positivo y los coeficientes asociados a los píxeles exteriores sean negativos. Como el Laplaciano es una derivada, la suma de los coeficientes debe de ser cero. En consecuencia la respuesta es cero siempre que el punto en cuestión y sus vecinos tengan el mismo valor.

Pero este operador rara vez es ocupado para detectar bordes, pues es un operador muy sensible al ruido, tiene una respuesta de bordes dobles y no detecta la dirección de los bordes. Por estas razones, el Laplaciano se usa para detectar si un píxel esta en la zona clara u oscura de un borde.

3.3.3.3 Detección de bordes por filtrado Gaussiano.

La detección de bordes por filtro Gaussiano es conocida también como: filtro Log, u operador de sombrero mexicano.

Este procedimiento es la combinación de una segunda derivada (a nivel bidimensional) de los valores de gris de la imagen y el suavizado de ruido gracias a la aplicación de un filtro paso bajas Gaussiano. Los bordes de son detectados cuando hay un cruce por cero al final de la operación de convolución. Esto se explica en la siguiente ecuación:

Filtro paso bajas Gaussiano:

$$h(x,y) = \exp\left(\frac{-x^2 + y^2}{2\sigma^2}\right) \quad (3.3.3.3 -1)$$

donde σ es la desviación estándar. Si se propone que $r^2 = x^2 + y^2$, entonces de la ecuación 3.3.3.2 -1 el laplaciano de h con respecto a r es:

$$\nabla^2 h = \exp\left(\frac{-r^2 - \sigma^2}{\sigma^4}\right) \exp\left(\frac{-r^2}{2\sigma^2}\right) \quad (3.3.3.3. -2)$$

La figura 3.7 muestra una sección transversal de esta función. Se observa su suavizado, los pasos por cero son en $r = \pm\sigma$, con el centro positivo y los extremos negativos. Este es el modelo en que se basa la ecuación 3.3.3.2 -1 y la máscara[3] de la figura 3.6.

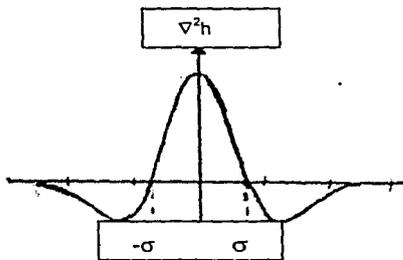


Figura 3.7 Respuesta de r al aplicarle el laplaciano $\nabla^2 h$

Pero, este operador también se puede expresar con la ayuda de un operador, el cuál recibe el nombre de operador del sombrero mexicano[23].

0	0	-1	-1	-1	0	0
0	-1	-3	-3	-3	-1	0
0	-3	0	7	0	-3	0
-1	-3	7	24	7	-3	-1
-1	-3	0	7	0	-3	-1
0	-1	-3	-3	-3	-1	0
0	0	-1	0	-1	0	0

Figura 3.8 Operador del filtro Gaussiano.

3.3.3.4 Operador de Frei y Chen.

Este operador fue propuesto por Frei y Chen en 1977[23]. Es de orden de 3X3 y tiene características semejantes al operador de Sobel, ya que también encuentra a los contornos por dos direcciones, vertical y horizontal, a continuación se presentan estos operadores:

-1	$\sqrt{2}$	-1		-1	0	1
0	0	0		$-\sqrt{2}$	0	$-\sqrt{2}$
1	$\sqrt{2}$	1		-1	0	1

Figura3.9 Operador vertical y Operador Horizontal de Frei Chen

Los resultados de aplicar cada operador a cada píxel son sumados para obtener nuestro estimador. El coseno de la raíz cuadrada de este valor es el vector de proyección de la información de vecindad en la dirección de "contornalización" y es asignado a la locación del píxel estudiado, pero en la imagen de salida.

Las ventajas comparadas con los más convencionales detectores de contornos (como el operador de Sobel) es la sensibilidad a la configuración relativa de los valores de los píxeles independientemente de la magnitud de la brillantez, que puede variar en algunos lugares de la imagen.

3.3.4 Detección de límites.

Los operadores antes descritos son aplicados a imágenes para detectar elementos de contornos. Sin embargo una imagen con bordes no - uniformes necesita un proceso adicional para poder agruparlos dentro de los límites encontrados.

La meta de la formación de límites o de la unión de los bordes, es hacer un borde global coherente de varios elementos de bordes locales. Las técnicas de formación de límites tienen que manejar instancias donde los elementos de borde no corresponden a algún límite significativo, y recíprocamente deben de formar un límite donde "posiblemente" deba de existir alguno.

Este tipo de detección se divide en dos formas de realizar el procesamiento.

- Procesamiento local
- Procesamiento global.

3.3.4.1 Procesamiento local.

Uno de los procedimientos más sencillos para enlazar puntos con un borde consiste en hacer un análisis de las características de los píxeles en una pequeña vecindad (esta puede ser de 3X3 o 5X5) respecto a cada uno de los puntos (x,y) de la imagen que ha sufrido una detección de bordes. Así todos los puntos comunes se enlazan formando un límite, los cuales tienen algunas características similares.

Las principales características que se usan para detectar si un píxel esta dentro de un borde son los siguientes:

- 1) La intensidad de la respuesta del operador que se utilizo para producir el píxel de borde.
- 2) La dirección del gradiente.

De este modo, la primera propiedad nos dice que si un píxel del borde de coordenadas (x',y') y en la vecindad de predefinida, es similar al modulo de gradiente del píxel si:

$$| \nabla f(x,y) - f(x',y') | \leq T \quad (3.3.4.1. -1)$$

donde T es un umbral no negativo[].

Para poder usar la segunda propiedad, se toman en cuenta las mismas condiciones anteriores, pero ahora se aplica la siguiente ecuación:

$$| \alpha(x,y) - \alpha(x',y') | \leq A \quad (3.3.4.1. -2)$$

donde A es un umbral angular.

Así entonces un punto de la vecindad predefinida de (x,y) está enlazada con un píxel de $f(x,y)$ si los criterios de ángulo y módulo se cumplen. Este proceso se debe de repetir para cada punto de la imagen.

3.3.4.2 Procesamiento global(Transformada de Hough).

En este apartado se considerarán puntos enlazados para determinar si están situados en una curva de límite.

Supóngase que para n puntos de una imagen se desea encontrar subconjuntos de puntos que se encuentran en líneas rectas. Una solución tentativa es encontrar en primer lugar todas las líneas determinadas por cada par de puntos y después encontrar los subconjuntos de puntos que estén cerca de ellas. Este procedimiento se llevaría mucho tiempo de cálculo, por lo que se propone una solución alternativa: La transformada de Hough[3].

Este método opera de la siguiente manera:

Consideres un punto (x_i,y_i) y la ecuación general de una recta en forma explícita $y_i = mx_i + b$. A través de (x_i,y_i) pasa un número infinito de líneas, pero todas ellas satisfacen a la ecuación $y_i = mx_i + b$ para diverso valores de m y b.

Pero si se rescribe la ecuación de la siguiente forma: $b = -mx_i + y_i$ y considerando el plano mb, se obtiene la ecuación de una única línea que pasa por un par determinado (x_i,y_i) . Además un segundo plano (x_j,y_j) también tiene una línea asociada a él, y esta línea corta a la línea asociada a (x_i,y_i) en (b',m') , donde m' es la pendiente y b' la ordenada en el origen de la línea que contiene a (x_i,y_i) y (x_j,y_j) en el plano xy. De hecho todos los puntos contenidos en esta línea tienen líneas en el espacio de parámetros que corta a (a',b') . La figura 3.8 explica lo anterior.

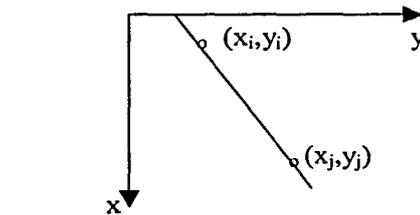
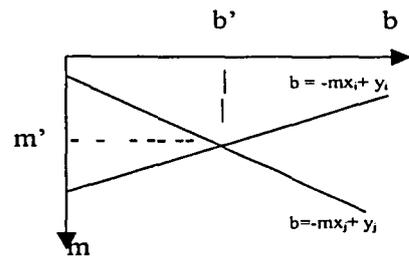


Figura 3.8 a) Plano XY



b) Espacio parámetro.

La ventaja de calcular la transformada de Hough es que el espacio parámetro es dividido en bloques llamadas células acumulador, observe la *figura 3.9*, donde $(m_{\text{máx}}, m_{\text{mín}})$ y $(b_{\text{máx}}, b_{\text{mín}})$ son los rangos esperados de los valores esperados de la pendiente y la ordenada al origen. La célula de coordenadas (i,j) , con un valor de acumulador $A(i,j)$, corresponde al cuadrado asociado con las coordenadas del espacio parámetros (m_i, b_j) .

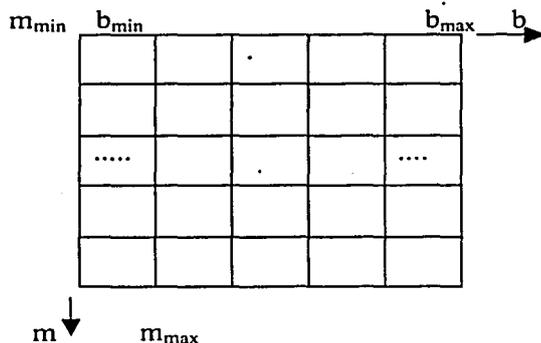


Figura 3.9 La célula acumulador usada en la transformada de Hough

Inicialmente estas células se inicializan en cero. Después para cada punto (x_k, y_k) se fija el parámetro m igual a cada uno de los valores permitidos para subdividir el eje X . Y se resuelve para la b correspondiente utilizando la ecuación $b = -mx_k + y_k$. Las b resultantes se redondean al valor más próximo permitido por el eje X . Si una elección de m_p resulta ser la solución b_q , se fija $A(p,q) = A(p,q) + 1$. Al final de este procedimiento, un valor de M en $A(i,j)$ corresponde a M puntos del plano XY situados en la línea $y = mx_i + b_j$.

Un problema que aparece al utilizar la ecuación $y = mx + b$ para representar una línea es que tanto la pendiente como la ordenada al origen tienden a infinito cuando la línea se acerca a la vertical. Una forma de evitar esta dificultad es la de usar la representación normal de una recta:

$$x \cos \theta + y \sin \theta = \rho \qquad 3.3.4.2 - 1$$

Al utilizar esta representación, en lugar de líneas rectas, ahora los lugares geométricos son curvas de tipo sinusoidales en el plano $\rho\theta$. Como antes, M puntos colineales situados sobre una línea recta $x \cos \theta_i + y \sin \theta_i = \rho_i$ produce M curvas sinusoidales que cortan a (ρ_i, θ_i) en el espacio de parámetros. Al incrementar θ y resolver para ρ se obtienen M entradas en el acumulador de $A(i,j)$ asociado con la célula determinada por (ρ_i, θ_i) .

El rango del ángulo θ es de $\pm 90^\circ$ medidos con respecto al eje x . Es por eso que para una línea horizontal el ángulo $\theta = 0^\circ$, siendo ρ igual a la X positiva, de forma similar, si la línea es vertical, $\theta = 90^\circ$ y ρ igual a la línea vertical.

3.3.5 Umbralización.

La umbralización es uno de los métodos más usados en la segmentación de imágenes, a continuación se dará una breve explicación de dicho método.

Así entonces, desde el punto de vista de la segmentación, la umbralización es un método que produce regiones uniformes dentro de una imagen basado en un criterio de umbral: T . La operación de umbralización puede pensarse como una prueba de envolvente de la función T , la cual es definida como sigue:

$$T = T\{x,y,A(x,y),f(x,y)\} \quad 3.3.5.-1$$

Donde $f(x,y)$ es el nivel de gris del píxel (x,y) y $A(x,y)$ denota una propiedad local sobre la vecindad del píxel[].

La umbralización de una imagen $g(x,y)$ se define como:

$$g(x,y) = \left\{ \begin{array}{l} 1 \text{ si } f(x,y) \text{ es } > T \\ 0 \text{ si } f(x,y) \text{ es } < T \end{array} \right\}$$

El valor de la función de T puede ser definida para lograr alguno de los siguientes tres procesamientos:

- a) Umbralización global: $T = T\{f(x,y)\}$
Donde T es dependiente del nivel de gris del píxel en x,y .
- b) Umbralización local: $T = T\{A(x,y),f(x,y)\}$
Donde T es dependiente de alguna propiedad de vecindad de un píxel como por ejemplo un valor de nivel de gris.
- c) Umbralización dinámica: $T = T\{x,y, A(x,y),f(x,y)\}$
Donde T es dependiente de las coordenadas del píxel, en adición de los otros dos criterios.

El más sencillos de los procedimientos es el umbralización global: Para realizarlo, el histograma se particiona utilizando un umbral único, T . La segmentación se completa entonces recorriendo a la imagen píxel por píxel y etiquetando cada píxel como perteneciente al fondo o no, dependiendo del valor de nivel de gris de ese píxel con respecto al valor T del umbral.

Como ya se observa, el éxito de estos métodos radica en cómo se puede dividir el histograma.

3.3.5.1 Umbralización: Método de White y Rohrer.

Uno de los aspectos más importantes para la selección de un umbral es la capacidad de identificar los picos en un histograma, ya que casi siempre se tiene una imagen en donde hay un amplio rango de distribuciones de intensidad[27].

Un método para mejorar la forma de los histogramas es considerar solamente aquellos píxeles que están situados en el límite entre los objetos y el fondo (o cerca de él). Pero esto no lleva a pensar en como diferenciar los objetos del fondo, y la respuesta es simple, la utilización de gradiente y el laplaciano nos permite obtener los lugares en donde se encuentran los bordes. Además el empleo del laplaciano puede proporcionar información sobre si un píxel está situado en la parte oscura o clara de un borde.

Así, sabiendo esto se puede generar una imagen de tres niveles de la siguiente forma:

$$S(x,y) = \begin{cases} 0 & \text{si } \nabla f < T \\ + & \text{si } \nabla f \geq y \nabla^2 \geq 0 \\ - & \text{si } \nabla f \geq y \nabla^2 < 0 \end{cases} \quad \text{Procedimiento de White y Rohrer.}$$

donde los símbolos 0,+,- representan tres niveles de gris cualquiera distintos, T es un umbral, el gradiente y el laplaciano se calculan para todos los puntos(x,y). Para un objeto oscuro sobre un fondo claro produce una imagen s(x,y) en la que todos los píxeles que no están en un borde (para ∇f menor que T) están marcados con 0, todos los del lado oscuro de un borde están marcados con +, y todos los del lado claro están marcados con -. Los símbolos se invierten en el caso de segmentar un objeto claro sobre un fondo oscuro.

Estos fueron algunos métodos de segmentación usados para obtener un mapa de contornos de las imágenes y detección de límites a las cuales se les aplica dichas técnicas, en el siguiente apartado se estudiarán otras técnicas de segmentación, las cuales no generan mapas de contornos, o se basan en las técnicas de umbralización ya que estas técnicas se aplican en regiones de la imagen, y es por eso que reciben el nombre de técnicas de segmentación por regiones.

3.4 Segmentación de Imágenes (procesos basados en regiones).

3.4.1 Métodos basados en crecimiento de regiones.

Como su nombre lo indica, el método de crecimiento de regiones es un procedimiento que agrupa píxeles o subregiones dentro de regiones más grandes. La sencillez de este método radica en la agregación de píxeles, que comienza con un conjunto de puntos generadores a partir de los que van creciendo los renglones o columnas al agregar a cada uno de estos puntos los píxeles próximos que tiene alguna propiedad similar (por ejemplo: nivel de gris, textura o color).

A continuación se enuncia un procedimiento general para llevar a cabo este tipo de segmentación:

Empleando dos tipos de salida se obtiene una segmentación que consiste en, al menos dos regiones : R1 que esta asociado al generador (a,b), y R2 asociada al generador (c,d). La propiedad P que se utiliza para que un píxel se incluya en una de las dos regiones es que la diferencia absoluta entre el nivel de gris del píxel en estudio sea menor que un umbral T. Si alguno de los píxeles cumple simultáneamente los dos umbrales, entonces se asigna a la región R1.

Pero este tipo de procedimiento genera algunos conflictos, el más inmediato es la selección de los generadores iniciales que representen las regiones de elección, y también la selección de los umbrales adecuados para decir si un píxel pertenece a una región o no.

Cuando no se dispone de este tipo de información , el procedimiento consiste en calcular para cada píxel el mismo conjunto de propiedades que se utilizarán al final para asignar píxeles a las regiones durante el proceso de crecimiento. Si el resultado de estos cálculos muestra agrupaciones de valores, los píxeles cuyas propiedades los sitúan cerca del centro de estas agrupaciones se pueden utilizar como puntos generadores.

3.4.1.1 Algoritmo de Brice y Fenema.

Brice y Fenema han desarrollado un método de crecimiento de regiones. En la primera fase del proceso, los pares de píxeles cuantizados son combinados en grupos llamados regiones atómicas, donde los píxeles están agrupados en una vecindad de orden cuatro y todos tienen la misma amplitud .

Si se tienen dos regiones R1 y R2 y ambas son adyacentes con perímetros P1 y P2 respectivamente, que han sido previamente marcados. Después de la primera etapa de crecimiento de la región, una región podría contener subregiones previamente marcadas de diferente valor de amplitud. También tenemos que C denota la longitud del límite común y la letra D representa la longitud de la porción de C que es más pequeña que un valor ξ_1 . Las regiones R1 y R2 están unidas si:

$$\frac{D}{\text{MIN}\{P1,P2\}} > \xi_2$$

Donde ξ_2 es una constante con el valor típico de 0.5. Este valor previene que haya regiones de adyacencia del mismo tamaño, pero permite pequeñas regiones que puedan ser absorbidas por regiones más grandes. La segunda regla fusiona límites comunes débiles que permanecen después de la aplicación de la primera regla. Dos regiones adyacentes son unidas si:

$$\frac{D}{C} > \xi_3$$

Donde ξ_3 tiene el valor de $\frac{1}{4}$.

Este tipo de algoritmo es razonablemente exacto para realizar una segmentación de escenas simples con pocos objetos y pequeña textura, pero no tiene buen desempeño cuando se trata de escenas más complejas.

3.4.1.2 Algoritmo DN.

Hay dos vías para decir si un píxel es alojado con sus vecinos en una región. Generalmente se prefiere usar una vecindad de orden 8 por que es más general e incluye regiones de vecindad de orden 4 como un caso especial. Pero los algoritmos basados en vecindad 4 son más rápidos, pero solamente se usan cuatro vecinos para realizar el algoritmo[23].

Se especifica un umbral tomando en cuenta la similaridad de un píxel con sus vecinos. Si hay alguna diferencia estos no son tomados como parte de la entidad y serán procesados después.

A continuación se describe un algoritmo para realizar la segmentación cumpliendo lo anterior.

Los píxeles en la imagen son procesados en un orden normal, es decir de izquierda a derecha(en los renglones) y de arriba a bajo (renglón por renglón). Si los datos son procesados al revés (de derecha a izquierda), se obtendría una segmentación (igualmente válida) diferente. Dos parámetros son especificados por el usuario, un umbral diferente DN, t , y la desviación estándar del umbral σ .

Se usan tres píxeles con una vecindad de 4 (observe la *figura 3.10*) para tomarla decisión de adicionar al píxel de estudio o no a una región, o inicializar una nueva. El algoritmo se resume a continuación:



Figura 3.10 Ejemplo de la vecindad usada para aplicar el algoritmo DN.

Tabla de decisión del algoritmo de DN.

Case	If	Then
1 Iniciar una nueva región	$ DN - DU > t$ $ DN - DN_i > t$	L = nueva marca Iniciar nueva región
2 Unión con la región de arriba	$ DN - DU \leq t$ $ DN - DN_i > t$	L = Lu
3 Unión con la región de la izquierda.	$ DN - DU > t$ $ DN - DN_i \leq t$	L = Li
4 Unión con la región de la izquierda y la de arriba	$ DN - DU \leq t$ $ DN - DN_i \leq t$ Li = Lu	L = LU
5 Marcar y unir	$ DN - DU > t$ $ DN - DN_i > t$ Li ≠ Lu	Li = Lu L = Lu

3.4.2 Métodos basados en división y fusión de regiones.

Los algoritmos anteriormente explicados incrementan las regiones a partir de un conjunto de puntos generadores. Una alternativa para este método consiste en subdividir la imagen inicialmente en un conjunto de regiones arbitrarias disjuntas y después fusionar y/o dividir las regiones intentando satisfacer las condiciones establecidas para lograr todas las condiciones de la segmentación (tales condiciones fueron expuestas en la sección 3.2).

Así, supóngase que R representa a la imagen completa y se selecciona un predicado P (predicado lógico) como ya se definió en la sección 3.2. Para una imagen cuadrada, un método para hacer la segmentación de R, consiste en dividirla de forma sucesiva en subregiones cada vez más pequeñas de forma de que para cada P_i , $P(R_i) = \text{verdadero}$. Esto es, si $P(R) = \text{falso}$, se divide la imagen en cuadrantes. Si P es falso para cualquier cuadrante, se divide el cuadrante en subcuadrantes, y así sucesivamente. Esta técnica es conocida como árbol cuádruple (*figura 3.11*). Esto se debe a que el nodo raíz del árbol es la imagen total, y cada nodo corresponde a una subdivisión de la misma.

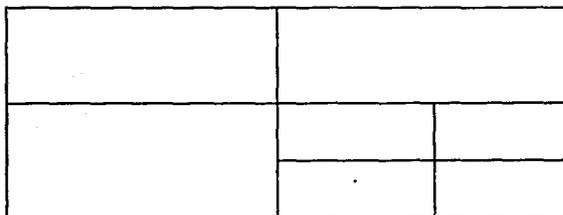


Figura 3.11 Aquí se presenta un árbol cuádruple de orden 2.

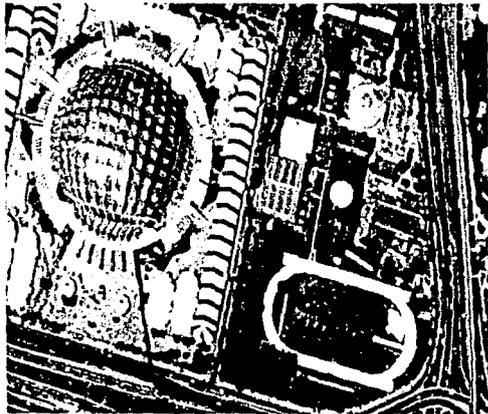
El único problema que presenta este método es que si sólo se utiliza la subdivisión, la partición final probablemente contenga regiones adyacentes con propiedades idénticas. Sin embargo este inconveniente se puede remediar, permitiendo tanto divisiones como fusiones de regiones.

A continuación se brinda un resumen del algoritmo:

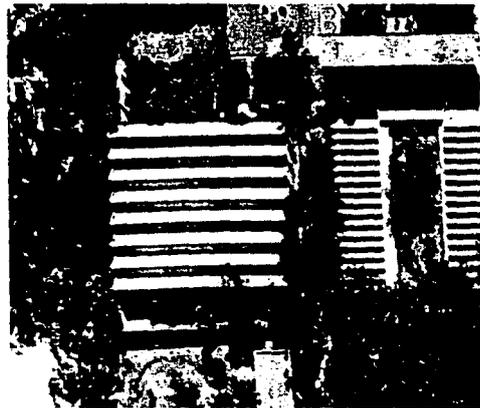
1. Dividir en cuatro cuadrantes disjuntos cualquier región R_i donde $P(R_i) = \text{Falso}$.
2. Fusionar las regiones adyacentes R_i y R_k para las que $P(R_i \cup R_k) = \text{verdadero}$.
3. Parar el algoritmo cuando no sea posible realizar más fusiones ni divisiones.

Ejemplos.

Para ejemplificar la utilización conjunta de las herramientas para la segmentación se usaran las siguientes imágenes:



a) deportes.bmp



b) usar.bmp



c) berlin.bmp

A las cuales se les aplicaran los siguientes procedimientos para su segmentación. Primero se les aplicara el operador de Frei y Chen para obtener su mapa de contornos, y después se segmentarán en 4 clases ocupando el algoritmo de DN. Para obtener los resultados de la imagen segmentada con contornos, sólo se hizo la resta de la imagen segmentada menos la imagen de contornos.

E1)

Aquí se presentan los resultados de la imagen deportes.bmp.

Imagen Original

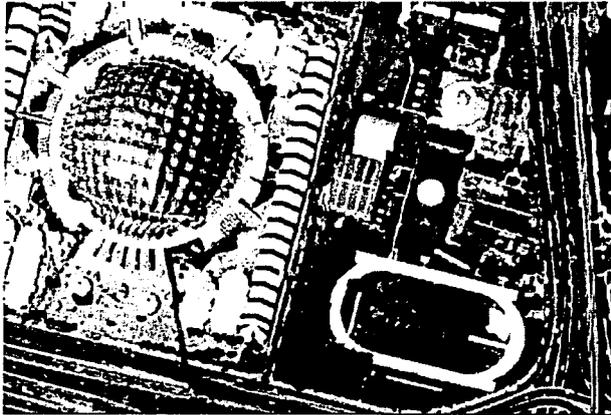


Imagen de contornos

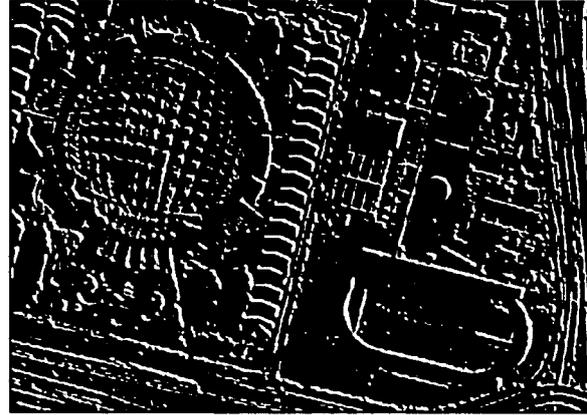
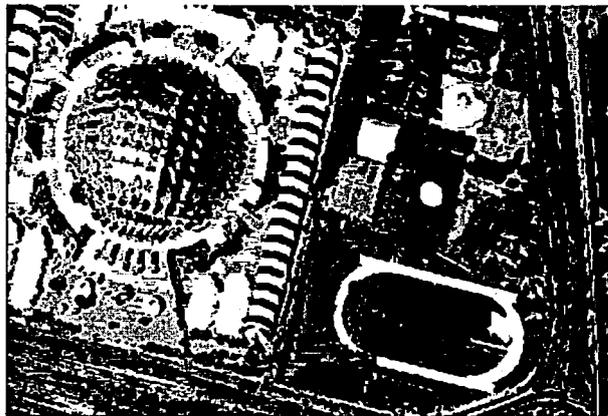


Imagen segmentada con contornos superpuestos.



E2)

Resultados de la imagen usar.bmp

Imagen original

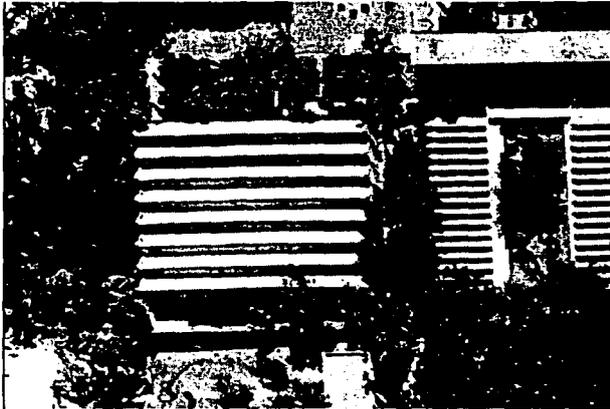


Imagen de contornos

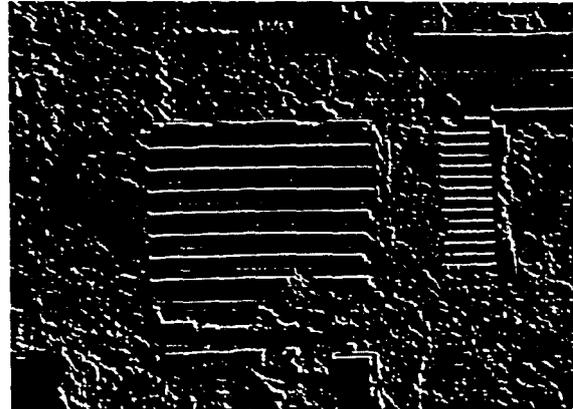
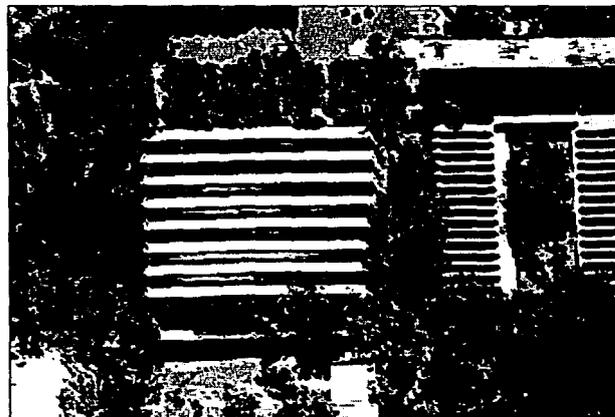


Imagen de segmentada con contornos superpuestos.



E3)

Resultados de la imagen berlin.bmp

Imagen original



Imagen de contornos



Imagen segmentada con contornos superpuestos.



Conclusiones.

Como se observa, los resultados en lo que respecta a encontrar un mapa de contornos de las imágenes, son los esperados, pues el operador de Frei y Chen nos muestran un mapa en donde los contornos pueden ser detectados. Pero, en la parte de segmentación de la imagen, los resultados son pobres, algo de antemano esperado, pues el algoritmo que se utilizó (el DN), aunque está clasificado dentro de los procedimientos de crecimientos de regiones, no hace una segmentación muy exhaustiva, ya que deja algunas regiones sin segmentar, al no cumplir estas con algunas de las características de dicho algoritmo.

Es por eso que este capítulo se cataloga como introductorio al estudio de otros algoritmos de segmentación de imágenes, los cuáles se basan en estudios estadísticos, variables aleatorias, y campos aleatorios de Markov, los cuales serán revisados en la siguiente sección.

4

Teoría de los Campos Aleatorios de Markov (MRF)

4.1 Probabilidad y variables aleatorias.

Frecuentemente en los experimentos el interés radica en una función del experimento y no en el resultado propiamente dicho. Por ejemplo en el contexto de comunicación a través de mensajes, el interés del analista podría centrarse en el número de transmisiones exitosas de mensajes en un tiempo dado, más que en el estado de la transmisión de cada mensaje.

4.1.1 Definición.

Una variable aleatoria(VA) es una función de los elementos de un espacio de muestra S . Convencionalmente se representa por una letra mayúscula. Cualquier valor de particular de una VA se representa por una letra minúscula[19].

En otras palabras, si se tiene un experimento con un espacio muestral S , y X es una función que asigna un número (entero, real, etc.) $X(S)$. A cada resultado $s \in S$, entorno a $X(S)$ se le denomina variable aleatoria.

4.1.2 Función de distribución acumulada.

Por otro lado una VA se dice continua si toma valores de un conjunto continuo de posibles valores.

Sea F , una función definida para cualquier valor real b de una VA X , con b en el intervalo $(-\infty, \infty)$ de la siguiente manera: $F_X(b) = P(X \leq b)$, es conocida como función de distribución acumulada de X y denota la probabilidad de que X puede tomar un valor menor o igual a b .

Propiedades de $F_X(\cdot)$

1. $F(b)$ es una función no decreciente en b .
2. $\lim F_X(b)=1$ cuando $b \rightarrow \infty$.
3. $\lim F_X(b)=0$ cuando $b \rightarrow -\infty$.
4. $P\{X_1 < X < X_2\} = F_X(X_2) - F_X(X_1)$
5. $F_X(X_1) \leq F_X(X_2)$

Para variables aleatorias discretas se define la función de densidad de X con $P(a) = P\{X=a\}$.

Si $X = \{x_1, x_2, \dots\}$ en Z , entonces $p(x_i) > 0$ con $i=1, 2, \dots$ también se debe de cumplir que:

Si $X = \{x_1, x_2, \dots\}$ en Z , entonces $p(x_i) > 0$ con $i=1, 2, \dots$ y $p(x) = 0$ para todo x que no esta en X . Además se debe cumplir que:

$$\sum_{i=1}^{\infty} p(X_i) = 1 \quad 4.1.2-1$$

$$F(a) = \sum_{x \leq a} p(x_i) \quad 4.1.2-2$$

4.1.2.1 Momentos de variables aleatorias.

Si x es una variable aleatoria discreta, el valor esperado de X , denotado por $E(x)$, es definida por:

$$E(x) = \sum_{i=1}^n x_i P(x_i) = \mu_x \quad 4.1.2.1-1$$

donde x_1, \dots, x_n son todas las posibles salidas de x y $P(x)$ es la probabilidad de que $x = x_i$. El valor esperado $E(x)$ es además llamado la media de X y también puede ser denotada como μ y μ_k .

Si x es una variable aleatoria continua, el valor esperado de x es:

$$E(x) = \int_{-\infty}^{\infty} xp(x)dx = \mu_x \quad 4.1.2.1-2$$

donde $p(x)$ es la densidad de x . El valor esperado $E(x)$ es además denotado como μ_x .

El segundo momento de una variable aleatoria x es definida por $E(x^2)$, pero el segundo momento central o variancia de la variable aleatoria x es definido por:

$$\sigma_x^2 = E[(x - E(x))^2] = E[(x - \mu_x)^2] \quad 4.1.2.1-3$$

Esto es la desviación cuadrada de x sobre la media y describe lo esparcido de los valores de la variable aleatoria sobre la media. Pero el valor más usado es la desviación estándar σ_x que es definida como la raíz cuadrada de la variancia σ_x^2 .

4.1.3 Funciones de distribución de VA discretas.

4.1.3.1 Distribución Binomial.

En una distribución binomial, se asume que la probabilidad de suceso es igual para cada evento, y la probabilidad se denota por θ . El número total de sucesos k obtenidos en n intentos es llamada variable aleatoria binomial. Esta función de distribución es dada por:

$$P(k) = \binom{n}{k} \theta^k (1-\theta)^{n-k}, \quad k = 0, 1, 2, \dots, n \quad 4.1.3.1-1$$

donde:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad 4.1.3.1-2$$

El símbolo $\binom{n}{k}$ es llamado el coeficiente binomial.

4.1.3.2 La distribución de Poisson.

Este tipo de distribución es usada cuando se tiene una entrada discreta y una salida de valores infinita. Un conjunto de salida se dice infinito si este conjunto puede ser listado por una secuencia infinita. Por ejemplo, el conjunto de enteros no negativos se contabiliza como infinito por que puede ser listado en un infinito listado $\{0, 1, 2, \dots, n\}$. La distribución de Poisson es la siguiente:

$$P(n) = \frac{e^{-\lambda} \lambda^n}{n!} \quad 4.1.3.2-1$$

donde $n = 0, 1, 2, \dots, n$, y se observa que es una distribución del tipo discreto. El valor $P(n)$ es interpretada como la probabilidad de que exactamente n eventos podrían ocurrir en un intervalo de tiempo y λ es el número de eventos que ocurren en la longitud de tiempo, así λ es la media de la variable aleatoria n. Como ocurre con la distribución binomial, la distribución de Poisson asume que sólo valores enteros son posibles.

4.1.4 Funciones de distribución de VA continuas.

4.1.4.1 Función de densidad exponencial.

Esta función de densidad es importante para áreas de estudio tales como: el decaimiento de las partículas radioactivas. Y su expresión es:

$$p(t) = \beta \cdot e^{-\beta \cdot t}, t \geq 0. \quad 4.1.4.1-1$$

donde t es el tiempo en que un evento ocurre, y t puede tomar un valor de 0 a ∞ . El parámetro β es una proporción constante, que controla el promedio de decaimiento.

4.1.4.2 Función de densidad normal.

Esta distribución también es denotada como función de densidad Gaussiana. Y su ecuación es la siguiente:

$$p(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-0.5 \left(\frac{x-\mu}{\sigma} \right)^2} \quad 4.1.4.2-1$$

El parámetro μ denota la media de x, y σ denota la desviación estándar de x. Para una distribución normal, μ es localizada en el centro de simetría de la densidad normal.

4.1.5 Covarianza.

Si (x,y) es una variable aleatoria conjunta, donde μ_x y σ_x son la media y la desviación estándar de x, y μ_y y σ_y son la media y la desviación estándar de y. La covarianza de x e y es definida como:

$$\sigma_{x,y} = E[(x - \mu_x)(y - \mu_y)] \quad 4.1.5-1$$

La covarianza indica como x e y tienden juntas a variar. El valor de la covarianza depende de cómo cada variable tiende a desviarse de su media., que es representada por las desviaciones estándar tanto de x como de y, y además depende del grado de asociación entre x e y.

4.2 Teoría Bayesiana.

El concepto básico de la teoría Bayesiana es que todas las probabilidades son condicionales. Es por eso que en su teoría se utiliza el símbolo de $P(A|.)$ donde "." es el evento en el que se basa A [22].

A continuación se dan algunas definiciones de interés propios de esta teoría.

Probabilidad Conjunta: Si las variables aleatorias A y B son discretas, la función de distribución conjunta de las variables aleatorias (A,B) es la probabilidad $P(A,B)$ de que tanto A como B ocurran. Esto es, la distribución conjunta da la probabilidad de cada posible combinación de salidas de las variables aleatorias.

Probabilidad Condicional: Dado un evento B con probabilidad de ocurrencia distinta a 0, la probabilidad condicional de A dado B es definida como:

$$P(A|B) = \frac{P(A \cap B) / NtB}{P(B) / NtB} = \frac{P(A \cap B)}{P(B)} \quad 4.2-1$$

donde NtB = número total de posibles resultados de B .

Independencia estadística de dos eventos: Sea un espacio muestral S donde se definen dos eventos A y B ambos $\neq 0$. A y B son independientes sí y solo sí:

$$P(A|B) = P(A). \quad 4.2-2$$

Esto nos dice que, el conocimiento de que ocurra B no influye en la probabilidad de A .

Si:

$$P(A|B) = P(A) = \frac{P(A \cap B)}{P(B)} \quad 4.2-3$$

Al igual que:

$$P(A \cap B) = P(A)P(B) = P(B \cap A) \quad 4.2-4$$

Con estas definiciones se puede enunciar a continuación el famoso Teorema de Bayes, también conocido como la regla de la probabilidad inversa.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ de 4.2-4 tenemos que } P(A \cap B) = P(B \cap A) \text{ y aplicando 4.2-3}$$

Se obtiene lo siguiente:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad 4.2-5$$

Que lleva el nombre de Teorema de Bayes

Las probabilidades de A y B son llamadas probabilidades *a priori* por que, estas dan las probabilidades de las clases en la población antes de saber los valores de los "rasgos" o características deseadas. Las probabilidades condicionales de los miembros de las clases $P(B|A)$ son llamadas probabilidades *a posteriori* por que estas dan las probabilidades de las clases después de observar los valores de los "rasgos" o características deseadas.

Riesgo de Bayes. El riesgo de Bayes de una regla de decisión δ con respecto a una distribución a priori P es dada por:

$$r(P, \delta) = \{R(Y, \delta)\} \quad 4.2-6$$

donde Y es una variable aleatoria sobre Θ con distribución P.

Regla de decisión de Bayes. Dada una distribución a priori P, δ_0 es una regla de decisión de Bayes con respecto P si:

$$r(P, \delta_0) = \inf_{\delta} r(P, \delta) \quad 4.2-7$$

Donde el valor de $r(P, \delta_0)$ recibe el nombre de regla de decisión de Bayes.

4.3 Antecedentes en los MRF.

La teoría de los campos aleatorios de Markov es una rama de la teoría de la probabilidad para analizar la dependencia espacial o contextual de fenómenos físicos. Es usada en la etiquetación visual para establecer distribuciones de probabilidad.

4.3.1 Sistemas de vecindad y cliques.

Antes de definir lo que es un sistema de vecindad se definirán algunos términos usados en esta teoría:

Vecinos: Dos puntos s_i y s_j son vecinos si tienen en contorno $e_{ij} \in \mathcal{E}$ que los une.

Clique: Un subconjunto $C \subseteq S$ es un clique si cada par de distintos sites en C son vecinos. C denota un conjunto de cliques.

Y por lo tanto un sistema de vecindad se define de la siguiente manera:

$$N = \{ N_i \mid \forall i \in S \}$$

donde N_i es el conjunto de sites vecinos. La vecindad esta relacionada con las siguientes propiedades:

- 1) Un site no es vecino de si mismo. $i \notin N_i$.
- 2) La relación de la vecindad es mutua. $i \in N_{i'} \Leftrightarrow i' \in N_i$.

Para una rejilla regular S , la vecindad de i es definida como un grupo de sites cercados por un radio r , es decir:

$$N_i = \{ i' \in S \mid [dist(pixel_i, pixel_{i'})]^2 \leq r, i' \neq i \}$$

donde la $dist(A,B)$ denota la distancia Euclidiana entre A y B y r toma un valor entero.

Un sistema de vecindad de primer orden, es conocida como un sistema de vecindad cuatro (vecindad - 4), cada site interior tiene cuatro vecinos, como se muestra en la *figura 4.1a*, donde x denota al site considerado y los espacios en blanco son sus vecinos. Un sistema de vecindad de segundo orden, es también llamada vecindad de orden ocho (vecindad - 8) *figura 4.1 b*, donde el site tiene ocho vecinos[15].



Figura 4.1 Se muestra los sistemas de vecindad de a) primer, y b) segundo orden.

Para una rejilla irregular S , el conjunto vecino N_i de i es definido de la siguiente manera:

$$N_i = \{ i' \in S \mid [dist(rasgo_i, rasgo_{i'})]^2 \leq r, i' \neq i \}$$

En general, el tamaño del conjunto de vecinos de i puede ser de varios tamaños.

4.3.2 Campos aleatorios de Markov.

Si se tiene una familia de variables aleatorias $F = \{F_1, \dots, F_m\}$ definidas en un espacio S , en que cada variable aleatoria F_i toma un valor f_i en C . Donde la familia F es llamado un campo aleatorio. Se usa la notación $F_i = f_i$ que significa que el evento F_i toma el valor f_i y la notación $\{ F_1 = f_1, \dots, F_m = f_m \}$ que denota el evento unión. Por simplicidad el evento unión es abreviado como $F = f$, donde $f = \{f_1, \dots, f_m\}$ es una configuración de F . Para un conjunto discreto de etiquetas C , la probabilidad de que la variable aleatoria F tome el valor de f se expresa de la siguiente manera $P(F_i = f_i)$, abreviado $P(f_i)$, y la probabilidad de la unión es denotada por $P(F = f) = P(F_1 = f_1, \dots, F_m = f_m)$ y abreviado $P(f)$. Para un C continuo, se tienen las funciones de densidad, $p(F_i = f_i)$ y $p(F = f)$ [15].

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

$$P(f) > 0, \forall f \in F. \quad \text{Propiedad de positividad.}$$

$$P(f_i | f_{S-\{i\}}) = P(f_i | f_{N_i}) \text{ ó}$$

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

Donde $S - \{i\}$ es el conjunto diferencia, $f_{S-\{i\}}$ denota el conjunto de etiquetas en los sites en $S - \{i\}$, y $f_{N_i} = \{f_{i'} | i' \in N_i\}$, η_{ij} es la vecindad del píxel (i,j).

En otras palabras, un campo de Markov X tiene la propiedad de que la distribución condicional de x en un píxel en particular, dados los valores de todos lo demás píxeles en la imagen completa, es igual a la distribución condicional dados sólo los valores de sus vecinos.

Un MRF puede tener otras propiedades tales como: isotropía y homogeneidad. Así entonces, se dice que un MRF es homogéneo si $P(f_i | f_{N_i})$ es indiferente a la posición relativa que tiene el site i en S. La propiedad de isotropía tiene que ver con los potenciales de los cliques.

El potencial para un clique de dos píxeles se define como:

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

donde el parámetro β es una constante positiva. De acuerdo con la definición de la distribución de Gibbs, x es el índice del tipo de región a determinar para un píxel considerado y esta puede tener cualquier valor.

4.3.3 Campos aleatorios de Gibbs.

Un conjunto de variables aleatorias F es un Campo Aleatorio de Gibbs (Gibbs random field GRF) en S con respecto a N si y solo sí esta configuración obedece a una distribución de Gibbs, donde una distribución de Gibbs toma la siguiente forma:

$$P(f) = Z^{-1} \exp(-T^{-1}U(f)) \quad 4.3.3-1$$

Donde:

$$Z = \sum_{f \in F} e^{-\frac{1}{T} U(f)} \quad 4.3.3-2$$

es una constante normalizada llamada función partición, T es una constante llamada la *temperatura* que se asume como 1 y U(f) es la función de energía. La energía:

$$U(f) = \sum_{c \in C} V_c(f) \quad 4.3.3-3$$

es la suma de los potenciales de los cliques $V_c(f)$ sobre todos los cliques posibles. El valor de $V_c(f)$ depende de la configuración local del clique c .

Se dice que un GRF es homogéneo si $V_c(f)$ es independiente de la posición relativa del clique c en S . Y cumple con la propiedad de isotropía si $V_c(f)$ es independiente de la orientación de c .

4.3.4 Equivalencia entre MRF y GRF.

Un MRF se caracteriza por una propiedad local (Propiedad de Markov) mientras que un GRF es caracterizado por su propiedad global (distribución de Gibbs). El teorema de Hammersley y Clifford establece la equivalencia entre estos dos campos. El teorema expone que F es un MRF en S con respecto a N si y solo si F es un GRF en S con respecto a N .

De forma más formal, el teorema dice:

X es un MRF con respecto al sistema de vecindad v si y solo si $\Pi(w) = P(w)$ es una distribución de Gibbs con un potencial de Gibbs V de su vecino más cercano, que es[22]:

$$\Pi(w) = \frac{1}{Z} \exp\left(-\sum_{c \in C} V_c(w)\right) \quad 4.3.4-1$$

El principal beneficio que esta equivalencia da, es que provee de una simple vía para especificar un MRF, especificando los potenciales en lugar de las características locales.

4.4 Estimación Bayesiana.

Para realizar este proceso, lo primero que se debe de generar es un estimador para el campo de etiquetas. Se definen tanto la probabilidad condicional como la de intersección en términos de las distribuciones a priori y a posteriori:

$$P_{X,F}(w,f) = P_{F|X}(f|w)P_X(w) \quad 4.4-1$$

$$P_{X|F}(w|f) = \frac{P_{X,F}(w,f)}{P_F(f)} = \frac{P_{F|X}(f|w)P_X(w)}{P_F(f)} \quad 4.4-2$$

Sabiendo que la probabilidad $P(f)$ es constante, se puede escribir lo siguiente:

$$P_{X|F}(w|f) \propto P_{F|X}(f|w)P_X(w) \quad 4.4-3$$

De esto se tiene que el riesgo de Bayes es dado por:

$$r(P_X, \delta) = E(R(w, \delta(f))) \quad 4.4-4$$

donde $R(w, \delta(f))$ es una función de costo. Y este estimador se podría tener un mínimo, obteniéndolo de la siguiente manera:

$$w = \arg \min_{w \in \Omega} \int_{w' \in \Omega} R(w, w') P_{X|F}(w' | f) dw' \quad 4.4-5$$

4.4.1 Estimador Máximo a posteriori.

El estimador MAP (maximun a posteriori) es el estimador que con más frecuencia se usa en el procesamiento de imágenes. La función de costo es definida por:

$$R(w, w') = 1 - \Delta_w(w') \quad 4.4.1-1$$

donde $\Delta_w(w')$ es la masa de Dirac en w' . Claramente, esta función tiene el mismo costo para cualquier configuración diferente a w' . El estimador MAP toma la siguiente forma:

$$w^{MAP} = \operatorname{argmax}_{w \in \Omega} P_{X|F}(w|f) \quad 4.4.1-2$$

Este tipo de ecuación combinatorial requiere (para su solución) de algoritmos especiales como son el Recocido Simulado.

4.4.2 Modos Marginales a posteriori.

Se define la función de costo del estimador MPM (Marginal a Posteriori Modes):

$$R(w, w') = \sum_{s \in S} (1 - \Delta_{w_s}(w'_s)) \quad 4.4.2-1$$

Obsérvese que esta ecuación es relacionada al número de sites $s \in S$ tal que $w_s \neq w'_s$. Y la solución de la ecuación es:

$$\forall s \in S : w_s^{MPM} = \underset{w_s \in \Delta}{\operatorname{argmax}} P_{X_s|F}(w_s|f) \quad 4.4.2-2$$

qué da la configuración que aumenta al máximo a cada site la probabilidad a posteriori marginal $P_{X_s|F}(\cdot|f)$

4.4.3 Mean Field (MF).

Se tiene la siguiente función de costo:

$$R(w, w') = \sum_{s \in S} (w_s - w'_s)^2 \quad 4.4.3-1$$

Y el estimador es el siguiente:

$$\forall s \in S : w_s^{MF} = \int_{w \in \Omega} w_s P_{X|F}(w|f) dw \quad 4.4.3-2$$

que no es más que el valor esperado condicional de X dado $F = f$ que es la media del campo de X .

4.4.4 Función discriminante de Bayes.

Por el Teorema de Bayes, nosotros sabemos que:

$$P(w_i|x) = \frac{P(x|w_i)p(w_i)}{p(x)} \quad 4.4.4-1$$

donde $p(x) = \sum_i p(x|w_i)p(w_i)$, $I = 1, 2, \dots, M$, es la probabilidad de que x ocurra sin tener en cuenta a la clase a que pertenece, $p(w_i)$ es la probabilidad a priori de la clase w_i y $p(x|w_i)$ es la probabilidad de que ocurra w_i con respecto a x [15].

La función de pérdida L_{ij} es definida como la pérdida, costo o multa por decidir que $x \in w_j$ cuando $x \in w_i$. De igual forma, se define a la medida condicional del riesgo $r_k(x)$ como sigue:

$$r_k(x) = \sum_{i=1}^M L_{ik} p(w_i | x) \quad 4.4.4-2$$

Sustituyendo la ecuación 4.4.4-1 en $r_k(x)$, se tiene:

$$r_k(x) = p(x)^{-1} \sum_{i=1}^M L_{ik} p(x|w_i)p(w_i) \quad 4.4.4-3$$

Dado que $p(x)$ es común para todo $r_j(x)$, $j = 0, 1, \dots, M$ se puede eliminar de la ecuación de riesgo, y obtener el siguiente mínimo:

$$\min r_k(x) = \min \sum_{i=1}^M L_{ik} p(x|w_i)p(w_i) \quad 4.4.4-4$$

conocido como función discriminante de Bayes.

Si se tiene un problema de dos clases, se plantea lo siguiente:

$$r_1(k) = L_{11}p(x|w_1)p(w_1) + L_{21}p(x|w_2)p(w_2) \text{ y}$$

$$r_2(k) = L_{12}p(x|w_2) p(w_1)+ L_{22}p(x|w_2)p(w_2)$$

Se dice que $x \in w_1$ sí :

$$L_{11}p(x|w_1)p(w_1) + L_{21}p(x|w_2)p(w_2) < L_{12}p(x|w_2) p(w_1)+ L_{22}p(x|w_2)p(w_2)$$

ó

$$\frac{p(x|w_1)}{p(x|w_2)} > \frac{(L_{21} - L_{22})p(w_2)}{(L_{12} - L_{11})p(w_1)} \quad 4.4.4-5$$

este es la llamada regla de máxima verosimilitud.

Si se generaliza a varias clases:

$$L_{ik} = \frac{p(x|w_k) > p(w_i)}{p(x|w_i) p(w_k)} \quad 4.4.4-6$$

de donde el criterio de asignación es el siguiente:

$$x \in w_k \text{ si } L_{ki} > 1$$

ó $p(x|w_k)p(w_k) > p(x|w_i)p(w_i)$ entonces, la función discriminante se puede expresar como sigue:

$$d_k(x) = p(x|w_k)p(w_k) \quad 4.4.4-7$$

una forma alternativa sería la siguiente:

$$d'_k(x) = \log p(x|w_k) + \log p(w_k) \quad 4.4.4-8$$

Extendiendo esto a una forma general, para que $x \in w_k$ sí:

$$r_k(x) = \sum_{i=0}^M L_{ik} p(x|w_i) p(w_i) \quad 4.4.4-9$$

la regla de máxima verosimilitud se da cuando:

$$x \in w_i \quad \text{si } r_i(x) < r_j(x)$$

4.4.4.1 Función discriminante óptima para modelos distribuidos normalmente.

Una función de densidad normal multi - variada para un patrón de M clases se puede representar de la siguiente manera:

$$P(X|W_k) = \frac{1}{(2\pi)^{n/2} |C_k|^{1/2}} \exp[-0.5(X - m_k)^T C_k^{-1} (X - m_k)] \quad 4.4.4.1-1$$

$\approx N(m_k, C_k) \quad k = 1, 2, \dots M; n = \text{dimensión del modelo.}$

donde N es la función de densidad, m_k es el vector media, y C_k es la matriz de covarianza de la clase K , definidos por los valores esperados del modelo de la clase k . Esto es

$$m_k = E_k[X] \quad 4.4.4.1-2$$

y

$$C_k = E_k[(x - m_k)(x - m_k)^T] \quad 4.4.4.1-3$$

De esto se observa que el centro de los agrupamientos se obtiene del vector de media obtenido de las muestras y , la forma de los agrupamientos lo determina la matriz de covarianzas.

A continuación se muestran algunos ejemplos de diferentes formas agrupaciones.

4.4.4.1.1 Función discriminante óptima.

De la ecuación $d'_k(x) = \log p(x|w_k) + \log p(w_k)$ cuando le aplicamos este discriminante a la ecuación 4.4.4.1-1 obtenemos lo siguiente:

$$d''_k = \frac{n \log(2\pi)}{2} - \frac{1}{2} \log|C_k| - \frac{1}{2} (x - m_k)^T C_k^{-1} (x - m_k) + \log p(w_k) \quad 4.4.4.1.1-1$$

es claro que el primer término del lado derecho de la ecuación es el mismo para todo k , y puede ser eliminado. Por lo tanto la función discriminante se puede reducir a:

$$d'''_k = \log p(w_k) - \frac{1}{2} \log|C_k| - \frac{1}{2} (x - m_k)^T C_k^{-1} (x - m_k) \quad 4.4.4.1.1-2$$

La ecuación 4.4.4.1.1-2 se aplica cuando las matrices de covarianza son diferentes para cada clase (o agrupación). Pero a continuación se presentan dos simplificaciones.:

Caso 1. Cuando las matrices de covarianza son iguales para las diferentes clases. El significado físico de esto es que las clases (agrupaciones) son del mismo tamaño y de forma similar, pero las clases están centradas en diferentes medias. Si se expande la ecuación 4.4.4.1.1-2 tenemos:

$$d_k(x) = -\frac{1}{2}x^T C^{-1} x - \frac{1}{2} m_k^T C^{-1} m_k + x^T C^{-1} m_k + \log p(w_k) - \frac{1}{2} \log |C|$$

el primer y último término del lado derecho de la ecuación anterior es la misma para todas las clases, por lo tanto la función discriminante es la siguiente:

$$d_k(x) = -\frac{1}{2} m_k^T C^{-1} m_k + x^T C^{-1} m_k + \log p(w_k) \quad 4.4.4.1.1-4$$

Caso 2. Cuando las diagonales de las matrices de covarianza son de la siguiente forma: $\sigma_k^2 = |C_k|$. El significado físico de este caso es que las agrupaciones tienen el mismo número de componentes a lo largo de los ejes principales, y su forma del agrupamiento es una esfera. Si se sustituye $|C_k|$ por σ_k^2 en 4.4.4.1.1-4 tenemos lo siguiente:

$$d_k(x) = \frac{\log p(w_k)}{2} - \frac{1}{2} \log \sigma_k^2 - \frac{1}{2} \frac{(x - m_k)^T (x - m_k)}{\sigma^2} \quad 4.4.4.1.1-5$$

Cuando las variables son estadísticamente independientes, y tienen la misma varianza σ^2 , entonces $\sigma_k = \sigma_j = \sigma$ para todo j, k esto es:

$$C_k = C_j = \sigma^2 I \quad y$$

$$d_k(x) = -\frac{1}{2} \frac{x^T x}{\sigma^2} - \frac{2x^T m_k + m_k^T m_k}{\sigma^2} + \log p(w_k) - \frac{1}{2} \log(\sigma^2) \quad 4.4.4.1.1-6$$

Obsérvese que tanto $x^T x$ como $\frac{1}{2} \log(\sigma^2)$ son los mismos para cualquier valor de k . Por lo tanto, estos valores pueden ser eliminados, tomando la ecuación 4.4.4.1.1-6 la siguiente forma:

$$d_k(x) = -\frac{x^T m_k}{\sigma^2} + \frac{m_k^T m_k}{2\sigma^2} + \log p(w_k) \quad 4.4.4.1.1-7$$

obteniéndose otra función discriminante lineal.

4.5 Ejemplos de modelos Markovianos.

Para los propósitos de segmentación de imágenes, las características de la imagen deben modelarse de tal manera que esta puede dividirse en K regiones diferentes, cada una siguiendo una distribución normal caracterizada por una media μ_k y una varianza σ_k^2 . Si se toma en cuenta la distribución de Gibbs y se hace a $T=1$, se tiene una mediada de probabilidad de que un píxel (i,j) pertenezca a l región q_k :

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

Este modelo se compone de dos partes. La primera depende sólo de las estadísticas de los datos y de lo cerca que se encuentre de los mismos. La segunda depende de las propiedades de los píxeles vecinos y el central.

4.5.1 Modelo de Ising.

El modelo de Ising es un modelo simple para ferromagnetismo que predice las transiciones de fase. Los giros son discretos y solamente tiene dos posibles estados. Este modelo fue introducido por Lenz en 1920 y resuelto por Ising en 1925[15].

Este es el modelo de más bajo orden, y como es de esperarse sólo usa dos etiquetas para resolverse. Es muy empleado gracias a su bajo costo y simplicidad. Se basa en la utilización de la energía de Gibbs como potenciales de cliques de parejas de sites. Con los potenciales de cliques de dos sites, la energía toma la siguiente forma:

$$U(f) = \sum_{i \in S} V_1(f_i) + \sum_{i \in S} \sum_{i' \in N_i} V_2(f_i, f_{i'}) \quad 4.5.1-1$$

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

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

$$U(f) = \sum_{\{i\} \in C_1} f_i G_i(f_i) + \sum_{\{i,i'\} \in C_2} \beta_{i,i'} f_i f_{i'} \quad 4.5.1-2$$

La ecuación anterior es llamada "auto-modelos". Los auto-modelos pueden ser clasificados de acuerdo a las suposiciones hechas sobre las funciones individuales f_i .

Un auto-modelo se dice que es un modelo auto-logístico si, las f_i toman valores en el conjunto de etiquetas discretas $L = \{0,1\}$, o $L = \{-1, 1\}$. La energía correspondiente toma la siguiente forma:

$$U(f) = \sum_{\{i\} \in C_1} \alpha_i f_i + \sum_{\{(i,j)\} \in C_2} \beta_{i,j} f_i f_j \quad 4.5.1-3$$

donde $\beta_{i,j}$ pueden ser vistos como los coeficientes de interacción. Cuando N es el sistema de vecindad más cercano en una red (en dos dimensiones los cuatro vecinos más cercanos, y en una los dos vecinos más cercanos), el modelo auto-logístico se reduce al modelo de Ising.

4.5.2 Modelo de Potts.

Recordando que el modelo de Ising fue propuesto para la solución de un problema en el cuál solo se presentaban dos caso: giro arriba, giro abajo ($q=2$). Más generalmente, si se considera un sistema de giros, cada giro esta apuntando a las igualmente espaciadas direcciones q . Estos vectores son la combinación de q vectores unitarios apuntando a q direcciones simétricas diferentes de un hiper - tetraedro en $q-1$ dimensiones.

La energía para el modelo de Potts se puede escribir de la siguiente manera:

$$U(w) = \sum_{i,j} J(\Theta_{i,j})$$
 donde $J(\Theta)$ es periódica en 2π y $\Theta_{i,j}$ es el ángulo entre dos giros(o estados) vecinos en i y j . Para el caso de $q=2$, el modelo de Potts es equivalente al modelo de Ising.

4.6 Métodos locales y globales de minimización.

4.6.1 Métodos locales.

Un problema de minimización puede ser categorizado como continuo o combinatorial según sea la etiqueta del conjunto C (discreta o continua); y también puede ser categorizado como reprimido o no reprimido de acuerdo a si f esta dentro o no del espacio entero.

La búsqueda de un mínimo local es la base para muchos algoritmos de minimización. La idea es muy simple y natural: en un punto $f \in F$, se busca en para donde movernos en la vecindad $N(f)$, si f' tiene una mejora, $E(f') < E(f)$ se reemplaza f por f' , este proceso continua hasta que ninguna mejora pueda ser hecha. Cuando $E(f)$ es el más bajo, el algoritmo ha convergido a un mínimo local.

Hay diferentes formas de generar un candidato f' . En el caso continuo, f' puede escogerse basándose en el gradiente: $f' = f - \mu \nabla E(f)$ donde $\mu > 0$ es el tamaño del paso. Esto lleva a un punto de gradiente cero, que es necesario para la minimización. En el caso discreto, hay varios métodos para generar $f' \in N(f)$ para cada $E(f') < E(f)$.

4.6.1.1 Minimización con etiquetas discretas.

4.6.1.1.1 Modos condicionales iterados(ICM).

Dada la dificultad para maximizar la probabilidad conjunta de un MRF, Besag propuso en algoritmo determinístico llamado Modos Condicionales Iterados(ICM) que maximiza la probabilidad condicional local secuencialmente. El algoritmo ICM usa la estrategia "ávida" en la maximización local. Dado el dato d y las otras etiquetas $f_{s-(i)}$, las actualizaciones del algoritmo son secuencialmente cada $f_i^{(k)}$ en $f_i^{(k+1)}$ maximizar la probabilidad condicional $P(f_i|d, f_{s-(i)})$, con respecto a f_i .

Dos suposiciones son hechas para el cálculo de $P(f_i|d, f_{s-(i)})$: Primero, los componentes d_1, d_2, \dots, d_m son condicionalmente independientes.

La segunda suposición es que f depende de las etiquetas de la vecindad local(propiedad de Markov). De las dos suposiciones y del teorema de Bayes, se tiene lo siguiente:

$$P(f_i|d, f_{s-(i)}) \propto p(d_i, f_i)P(f_i|f_{N_i})$$

Obviamente, $P(f_i|d_i, f_{N_i}^{(k)})$ es más fácil de maximizar que $P(f|d)$, que es lo que busca ICM. Maximizar a q es equivalente a minimizar el potencial usando la regla siguiente:

$$f_i^{(k+1)} \leftarrow \underset{f_i}{\operatorname{argmin}} V(f_i|d_i, f_{N_i}^{(k)})$$

donde

$$V(f_i|d_i, f_{N_i}^{(k)}) = \sum_{i \in N_i} V(f_i | f_i^{(k)} + V(d_i | f_i))$$

El resultado obtenido por el algoritmo ICM depende mucho del operador inicial $f^{(0)}$. Pero el valor exacto del operador lógicamente no se conoce. Una opción para darle un valor inicial a $f^{(0)}$ es el estimador de máxima probabilidad:

$$f^{(0)} = \underset{f}{\operatorname{argmax}} p(d|f)$$

4.6.2 Métodos de minimización globales.

La minimización global requiere de:

- (1) Encontrar todos los mínimos locales y
- (2) Demostrar que no hay más mínimos locales.

Dos métodos que comúnmente son usados para obtener los mínimos locales: búsqueda aleatoria y recocido. Algunos ejemplos de métodos de búsqueda aleatoria son el algoritmo Metrópolis y el Muestreador de Gibbs, en estos algoritmos, la energía casi siempre descende y ocasionalmente aumenta.

El recocido es a menudo incorporado dentro de los métodos de búsqueda local para superar los problemas del mínimo local. Hay dos tipos de recocido: el determinístico y el estocástico.

4.6.2.1 Muestreador de Gibbs.

Se trata de un algoritmo iterativo basado en la construcción a partir de un estado inicial cualesquiera en la cuál se espera encontrar buenas propiedades de convergencia. Para este algoritmo se usa una serie de sites $S(n)[1]$, donde:

El procedimiento es el siguiente:

1. Barrido de la imagen.
2. En cada uno de los píxeles se calculan las probabilidades locales en función de la vecindad y de la estructura Markoviana del campo.
3. En función de las probabilidades condicionales locales a hacer un sorteo aleatorio de estado.
4. El procedimiento es repetido en todos los píxeles de la imagen, esto es, para todos los sites del campo de Markov.
5. El procedimiento tiene que ser repetido un gran número de veces para obtener buenos resultados.

Algoritmo:

1. Inicialización de la simulación: $x^{(0)}$.
2. En cada etapa (n):

- a) Escoger el site $r=s^{(n)}$. Su valor es $x_r^{(n)}$ y su configuración de vecindad es $V_r^{(n)}$.
- b) Sortear aleatoriamente su nuevo valor $x_r^{(n+1)}$, según una probabilidad de transición de Gibbs:

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

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

$$\lim P(X^{(n)}=w | X^{(0)}=\eta) = \Pi(w)$$

La expresión anterior se denomina Teorema de relajación o muestreador de Gibbs, esto es, cuando n tiende al infinito la configuración w sigue siendo una cadena de Markov.

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

4.6.2.2 Algoritmo de Metrópolis (a temperatura constante).

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

Algoritmo:

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

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

en este último paso existen dos eventualidades:

- a) $\Delta U \leq 0$: puesto que la energía disminuye, el cambio es aceptado incondicionalmente, entonces $X_r^{(n+1)} = \xi$.

- b) $\Delta U > 0$: el cambio es aceptado o rechazado efectuando un sorteo aleatorio con las probabilidades respectivas

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

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

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

ΔU corresponde a la variación tanto de la energía local como de la energía global asociada a la transición $X_r^{(n+1)} = \xi$. Tal caso resulta de la estructura Markoviana de la distribución $P(Y=y | X=x)$.

4.6.2.3 Algoritmo de Recocido Simulado.

Los algoritmos de recocido simulado ("simulated annealing") fueron introducidos por Cerny y Kirkpatrick, para la optimización de problemas combinatorios con mínimos locales.

Estos algoritmos derivan de la analogía termodinámica con el proceso metalúrgico del recocido: cuando se enfría un metal fundido suficientemente despacio, tiende a solidificarse en una estructura de mínima energía (equilibrio térmico); a medida que disminuye la temperatura, las moléculas tienen menos probabilidad de moverse de su nivel energético; la probabilidad de movimiento se ajusta a la función de Boltzmann $e^{-U/kT}$, dependiente de la temperatura, del salto de energía y de una constante.

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

$\Omega_0 = \{w \in \Omega\}$ donde w es la energía mínima.

Algoritmo

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

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

- c) Para toda configuración inicial η de la imagen y para toda configuración w , se tiene que:

$$\lim_{n \rightarrow \infty} P(x(n)=w|x(0)=\eta) = \Pi(w)$$

Algoritmo:

1. Inicializar aleatoria de la solución: $x^{(0)}$.

Escoger una temperatura inicial $T^{(0)}$ suficientemente elevada

2. En cada etapa (n):

- Se escoge el site $r=s^{(n)}$, su valor es $x_r^{(n)}$, y su configuración de vecindad es $V_r^{(n)}$.
- Se efectúa un sorteo aleatorio uniforme de una variable $\xi \in \Omega$.

3. Criterio Metrópoli: Calcular la variación de la energía local asociada:

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

Y se pueden presentar dos casos:

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

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

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

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

Cuando T es mayor a un calor determinado entonces volver al paso 2.

Comentario:

La convergencia del algoritmo sobre un mínimo global de la función de energía, se da cuando:

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

Capítulo No.4 Teoría de los Campos Aleatorios de Markov (MRF)

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

En la práctica se simula la probabilidad $p = \exp -\frac{\Delta U}{T_n}$ efectuando un sorteo aleatorio uniforme de una variable $\eta \in [0,1]$.

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

Los descensos de temperatura pueden ser de la forma

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

e incluso de $T^{(n+1)} = \alpha T^{(n)}$.

La temperatura inicial $T^{(0)}$ debe ser tan alta que virtualmente todas las transiciones propuestas son aceptadas.

Ejemplos.

Para la presentación de los resultados, se usara un clasificador bayesiano, un algoritmo de método global (el recocido simulado), y el algoritmo de k medias para realizar una segmentación en k diferentes clases. El algoritmo de k – medias es fácil de implementar, a continuación se dará una pequeña descripción del mismo:

Éste es el algoritmo de agrupamiento más popular. También es llamado el *algoritmo de las medias móviles* porque en cada iteración se recalculan los centros de los agrupamientos. Por esta razón se incorpora el índice t a la notación que se está empleando de manera que con $S_i(t)$ se indica el conjunto de patrones asociados al agrupamiento S_i en la iteración t y mediante $Z_i(t)$ se señala el valor de su centro en esa iteración.

Este algoritmo requiere un único parámetro, K , que es el número de agrupamientos que debe encontrar. Puede plantearse en tres pasos:

Algoritmo de k-medias.

Inicialización.

Consiste en inicializar arbitrariamente los centros de los K grupos.

Asignación y actualización de los centros.

En este paso se asigna cada patrón al grupo más cercano y se recalculan los centros en base a esta asignación.

¿Convergencia?

En el paso anterior algunos patrones pueden cambiar de agrupamiento y en consecuencia, los centros de éstos. Si esto ocurre, se trata de repetir el paso 2 hasta que no se modifiquen los centros. Cuando no haya modificaciones se considera que se ha encontrado una buena partición y se termina el agrupamiento.

Se usarán 4 diferentes imágenes para hacer las comparaciones, las imágenes son de 512 X 512 píxeles y en escala de 256 valores de grises.

Las imágenes de prueba son las siguientes:

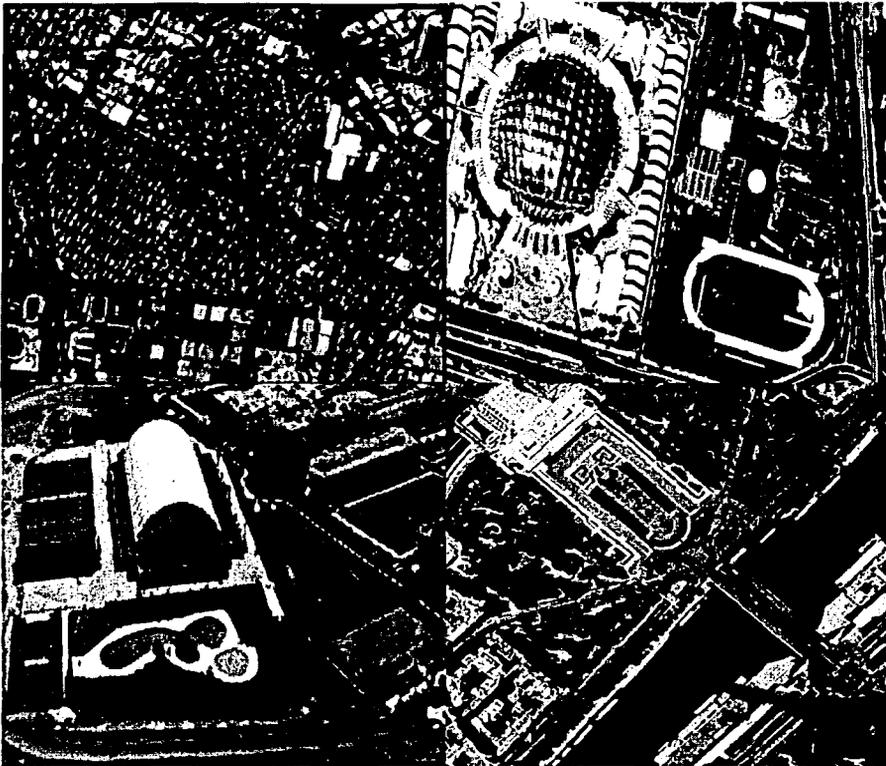


Fig 1. Imágenes utilizadas: a) cacho.bmp, b) deportes.bmp, c) palacio.bmp y d) paris3.bmp

El procedimiento es el siguiente: A cada imagen se le aplicarán cuatro diferentes formas de segmentación, el primer procedimiento es el algoritmo DN(explicado en la sección 3.4.12), el segundo es el clasificador óptimo bayesiano(definido en la sección 4.4.4.1.1), en tercer lugar se segmentará por medio del algoritmo k – medias y el último procedimiento es un algoritmo contextual por campos de Markov (empleando el modelo de Potts) este último se presenta en el Anexo 1 .

La presentación de la figura de resultados es la siguiente: en la esquina superior izquierda se encuentra la imagen segmentada por medio del algoritmo DN (a), a su derecha el resultado de la segmentación por medio de un clasificador bayesiano (b), en la esquina inferior izquierda, se encuentra el resultado al aplicar el algoritmo de k – medias (c), a su derecha la segmentación obtenida por medio de un algoritmo por campos de Markov (d).

E1)

Figura cacho.bmp, esta figura se segmentó en 5 clases.

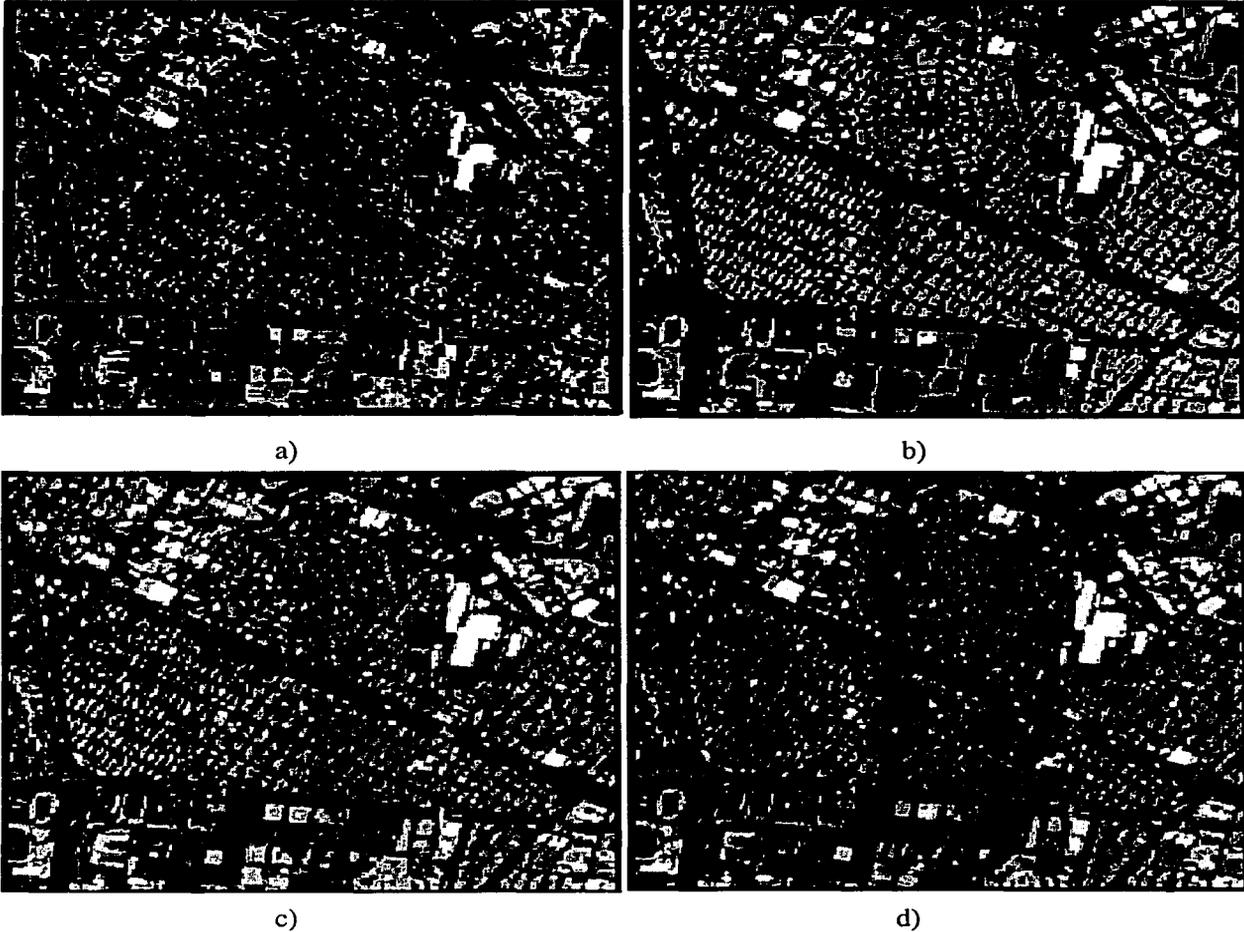


Figura No.2 Resultados de los distintos procedimientos para realizar la segmentación a) Algoritmo DN, b) Algoritmo k – medias, c) Clasificador bayesiano óptimo, d) Campos de Markov .

Obsérvese que una de las diferencias más marcadas entre sí es la intensidad de los valores de gris que se tiene en los resultados, algo que se esperaba, pues en k – medias y el algoritmo DN no se pueden escoger las medias para hacer la segmentación, y otra diferencia marcada, es que el algoritmo de k- medias no diferencia las vecindades, característica que el clasificador bayesiano y el algoritmo por campos de Markov si se observa. Pero la gran desventaja de este último es la gran cantidad de tiempo que necesita para procesarse.

E2)

Figura deportes.bmp, esta figura se segmentó en 5 clases.

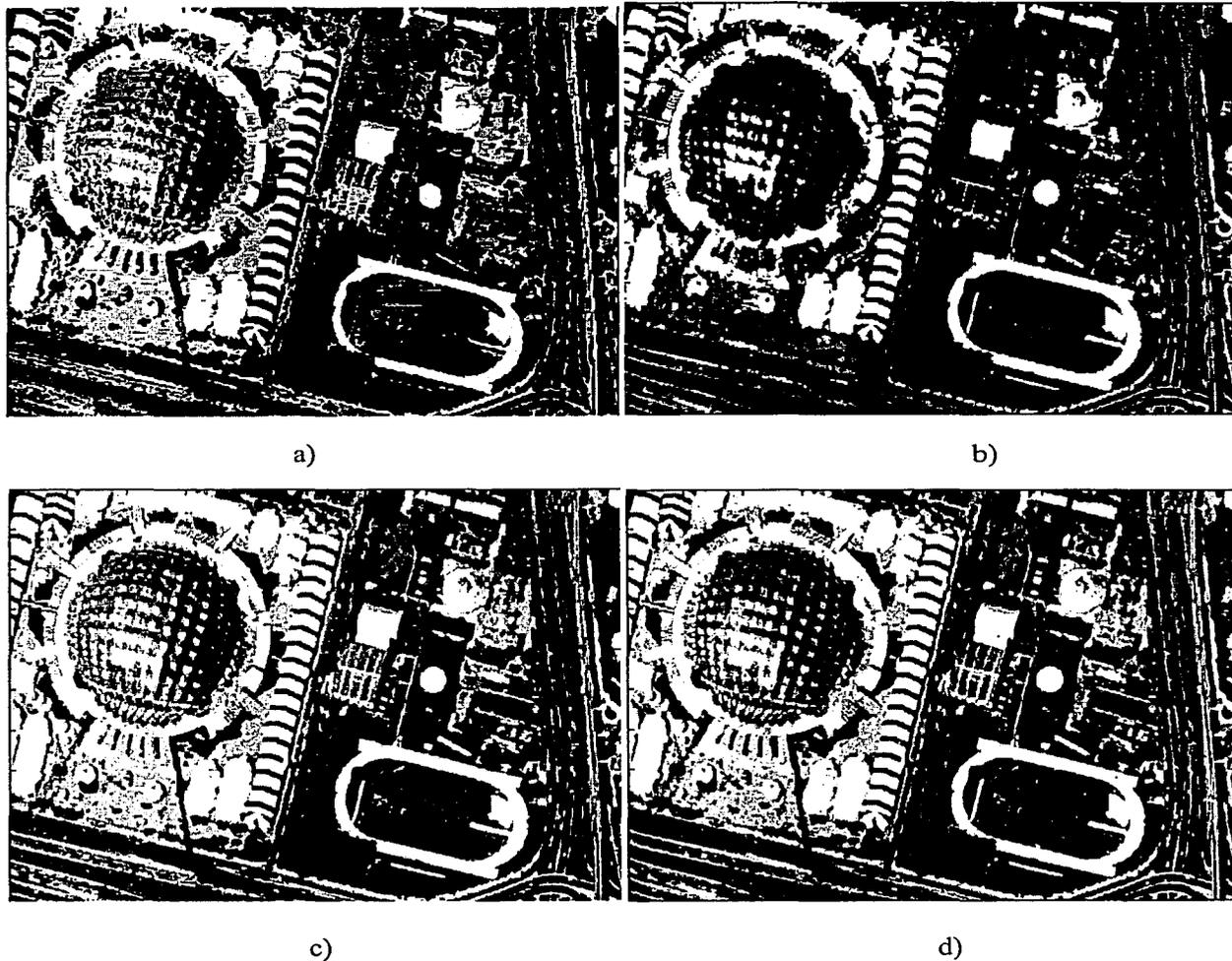


Figura No.3 Resultados de las diferentes segmentaciones. a) Algoritmo DN, b) Algoritmo k – medias, c) Clasificador bayesiano óptimo, d) Campos de Markov .

E3)

Figura palacio.bmp, esta figura se segmento en 5 clases.

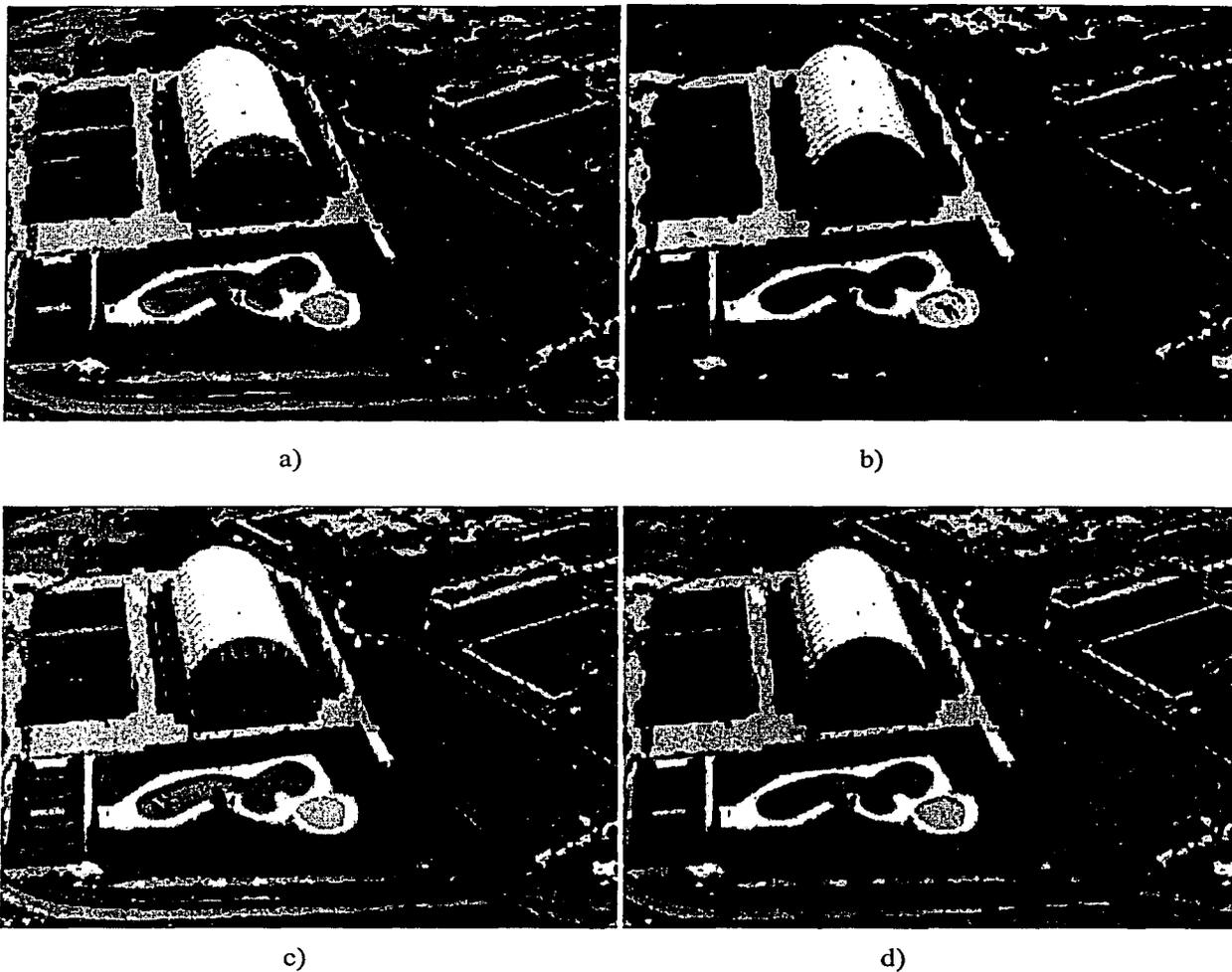


Figura No.4 Resultados de las distintas formas de segmentación. a) Algoritmo DN, b) Algoritmo k-medias, c) Clasificador bayesiano óptimo, d) Campos de Markov .

E4)

Figura paris3.bmp, esta figura se segmentó en 6 clases.

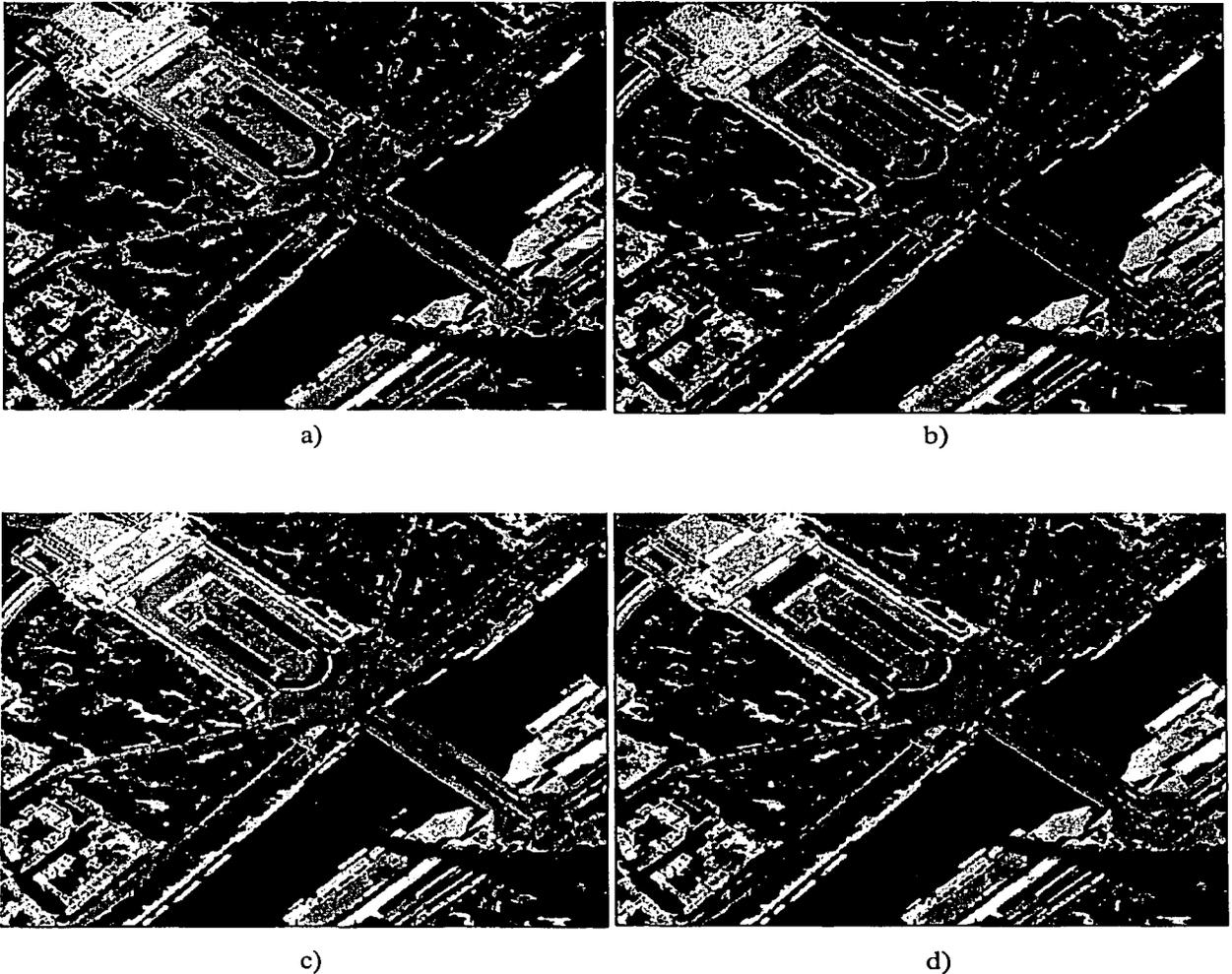


Figura No.5 Resultados de la segmentación. a) Algoritmo DN, b) Algoritmo k – medias, c) Clasificador bayesiano óptimo, d) Campos de Markov .

Conclusiones

Para concluir, diré, que todos los resultados anteriores fueron obtenidos por métodos de segmentación supervisada, puesto que en todos los algoritmos se debe de especificar el número de clases (k - medias y DN), y en otros, además se deben de dar los puntos de donde se deben de obtener las medias(segmentación bayesiana y campos de Markov), para después hacer la segmentación.

Los resultados pueden ser clasificados como buenos o malos según la percepción de cada usuario, pero lo que si se trata de dejar en claro, es que estos resultados dependen de los datos de la imagen original (por ejemplo: ventanas de entrenamiento) que se tomen en cuenta para obtener las medias y del número de clases en las que se decida hacer la segmentación, es decir el resultado depende del usuario, de la forma en que introduzca la información y de cómo analice los resultados.

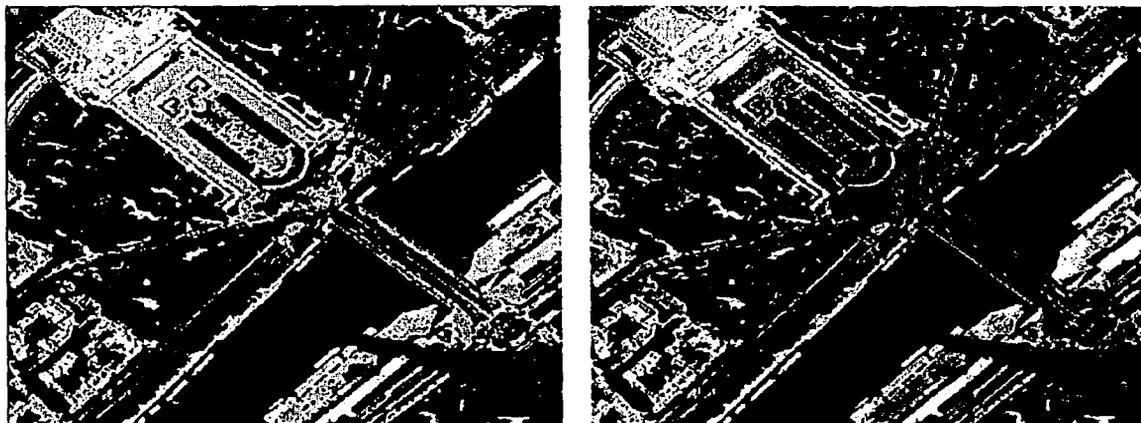


Figura No.6 Diferentes resultados de una segmentación.

Como puede observarse en la *figura No.6*, los resultados son diferentes, aunque se aplicó el mismo algoritmo de segmentación: clasificador bayesiano, y se segmentó con el mismo número de clases (6), pero se tomaron diferentes ventanas de entrenamiento para hacer la segmentación, la diferencia más marcada, es que en algunas partes, los árboles no se diferencian del pasto, aunque si se observan las dos imágenes, ambas son "buenas" segmentaciones.

Sin embargo, un resultado no debe de ser ambiguo, y al contrario, debe de ser lo más exacto posible.

Es por eso, que en el siguiente capítulo se discutirán dos formas de obtener, en forma automática y no supervisada el número de medias y su valor para hacer una segmentación, a partir de dos algoritmos: El algoritmo EM y el SEM.

5

Estimación de Parámetros: Algoritmos EM y SEM.

5.1 Introducción.

Como se ha expuesto anteriormente, para realizar la segmentación de una imagen, se necesita, aparte de conocer el número de clases en la que se segmentará la imagen, la media y en algunos casos, la variancia de dichas clases.

En los capítulos anteriores, estos valores el usuario, programador o investigador los proponía al utilizar algoritmos de segmentación tales como: k – medias, segmentación bayesiana o recocido simulado. A continuación se presentarán dos de las formas más usuales para encontrar dichos valores (μ, σ) de una manera no supervisada, es decir sin que halla la necesidad de ser introducidos para la ejecución del algoritmo.

5.2 Modelando a la imagen.

Se debe de mencionar que cada píxel de una imagen se puede describir en términos de una imagen píxel o contexto. Una imagen píxel es sólo referida a la intensidad (niveles de gris), y una imagen contexto se refiere a las clases etiquetadas (número de miembros) de la misma.

También se define a una región de una imagen como un grupo de píxeles que tiene la misma media y varianza.

5.2.1 Modelado por píxel.

Dada una imagen digital constituida por $N = m \times n$ píxeles, asumimos que esta imagen contiene k regiones y que cada píxel es descompuesto dentro de una imagen píxel x y una imagen contexto l . Si se desconoce la información con respecto a la clasificación espacial de los píxeles, la imagen contexto (por ejemplo: el etiquetado de píxeles) se pueden tratar como variables aleatorias y se pueden ser descritas usando una distribución multinomial con parámetros desconocidos Π_k . Dado que estos parámetros reflejan la distribución de el número total de píxeles en cada región, Π_k puede ser interpretada como la probabilidad a priori de la etiquetación de los píxeles determinada por la información de contexto global.

La medida de la probabilidad marginal para cualquier píxel de la imagen, puede ser obtenida escribiendo la densidad de probabilidad conjunta de x y de l y sumando la probabilidad conjunta de todas las posibles salidas de l , por ejemplo para calcular[5].

$$p(x_i) = \sum_l p(x_i, l) \tag{5.2.1-1}$$

resultando en una suma de la siguiente forma general:

$$p(x_i) = \sum_{k=1}^K \Pi_k p_k(x_i) \quad i=1, \dots, N \tag{5.2.1-2}$$

donde x_i es el nivel de gris del píxel i , $p_k(x_i)$ son las densidades de probabilidad condicional de regiones tomando en cuenta al factor Π_k , satisfaciendo $\Pi_k > 0$, y $\sum_{k=1}^K \Pi_k = 1$.

La función de densidad de probabilidad Gaussiana dada la región k es definida por:

$$p_k(x_i) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}\right) \quad i=1, \dots, N \tag{5.2.1-3}$$

donde μ_k y σ_k^2 son la media y la variancia del k -esimo kernel Gaussiano y K es el número de componentes Gaussianos.

La imagen entera puede aproximarse por un campo aleatorio X independiente e idénticamente distribuido. La correspondiente función de densidad de probabilidad (fdp) conjunta es:

$$P(x) = \prod_{i=0}^N \sum_{k=1}^K \Pi_k p_k(x_i) \tag{5.2.1-4}$$

donde $x = [x_1, x_2, \dots, x_N]$, y $x \in X$. Basada en la medida de la probabilidad conjunta de píxel imagen , la función de probabilidad puede ser expresada como:

$$\ell(r) = \prod_{i=1}^N p_r(x_i) \quad 5.2.1-5$$

donde $r: \{K, \alpha, \Pi_k, \mu_k, \sigma_k, k=1, \dots, K\}$ denota el modelo de parámetros.

5.3 Estimación de parámetros.

Cuando se tiene una función de probabilidad, el objetivo en dicha función es estimar los parámetros del modelo para maximizar la función de probabilidad.

El método más popular para realizar esta estimación es el algoritmo EM (expectation maximization). El algoritmo EM primero calcula la probabilidad a posteriori Bayesiana de los datos, basándose en la observación, y obtiene la estimación de los parámetros actuales (paso E). Después, las actualizaciones de los parámetros estimados utilizando teoremas de medias (paso M). El procedimiento se mueve de un lado a otro entre estos dos pasos. Las iteraciones sucesivas incrementan la probabilidad de que los parámetros estimados se acerquen a los valores reales[29].

5.3.1 Algoritmo EM.

Forma general.

Y denota el espacio muestra de las observaciones, y sea $y \in \mathbb{R}^m$ denota una observación de Y . X denota el espacio fundamental y sea $x \in \mathbb{R}^n$ una salida de X , con $m < n$. El dato x es llamado *dato completo*. El dato completo x no es observado directamente, sólo se puede observar por medio de y , donde $y = y(x)$, e $y(x)$ es uno de tantos mapeos. Una observación y determina un subconjunto de X , que es denotada como $X(y)$. La siguiente figura explica lo anterior.

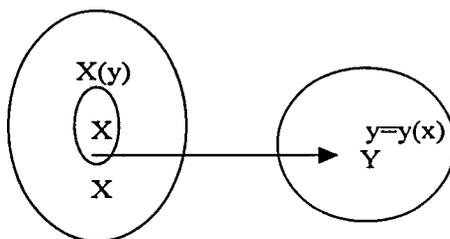


Figura 1. El punto y es la imagen de x , y el conjunto $X(y)$ es el mapeo de y .

La función de densidad de probabilidad (fdp) de el dato completo es $f_x(x|\theta) = f(x|\theta)$, donde $\theta \in \Theta \subset R^r$. Aquí la fdp se asume como una función continua de θ y diferenciable. La probabilidad máxima (ML) de θ es usada para decidir si el dato pertenece a la región Θ . La fdp del dato incompleto es:

$$g(y|\theta) = \int_{x(y)} f(x|\theta) dx \quad 5.3.1-1$$

Si $l_y = g(y|\theta)$ denota la función de probabilidad y

$$L_y(\theta) = \log g(y|\theta) \quad 5.3.1-2$$

denota el logaritmo de la función de probabilidad.

La idea básica detrás del algoritmo EM es que se desea encontrar θ al maximizar $\log f(x|\theta)$ pero no se tiene el dato x para calcular el logaritmo. Sabiendo esto, se maximiza la expectativa del $\log(x|\theta)$ dado el dato y nuestro actual estimador θ . Esto puede ser expresado en dos pasos:

Sea $\theta^{[k]}$ nuestro estimador de los parámetros en la k - ésima iteración.

Para el paso E se calcula lo siguiente:

$$Q(\theta|\theta^{[k]}) = E[\log f(x|\theta)|y, \theta^{[k]}] \quad 5.3.1-3$$

Para el paso M, sea $\theta^{[k+1]}$ el valor de θ que maximiza a $Q(\theta|\theta^{[k]})$:

$$\theta^{[k+1]} = \arg \max_{\theta} Q(\theta|\theta^{[k]}) \quad 5.3.1-4$$

Es importante hacer notar que la maximización es con respecto al primer argumento de la función Q .

El algoritmo EM consiste en escoger un inicial $\theta^{[k]}$, y después aplicar los pasos E y M hasta la convergencia. Se finaliza el algoritmo cuando los parámetros dejan de variar, por ejemplo cuando se cumple que $\|\theta^{[k]} - \theta^{[k-1]}\| < \xi$ para un ξ de cierto valor.

Algoritmo EM

Se inicializa el modelo suponiendo valores para m_k u σ_k . Siendo k el número de clases. La imagen se trata como un arreglo de $p(i,j) = 1,2,\dots,n$ elementos, manejándose como VA de tipo Gaussiano, con media y variancia definidas como[20]:

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \quad 5.3.1-5$$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \quad 5.3.1-6$$

Se piensa que el dato observado es la mezcla de x estados, siendo $p(k)$ la probabilidad del k - ésimo estado. Se asume que la mezcla (o probabilidad condicional) puede ser descrita por una variable Gaussiana de la siguiente forma:

$$p(x_i|k) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}} \quad 5.3.1-7$$

Utilizando la regla de Bayes tenemos:

$$p(x_i) = \sum_k p(x_i | k) p(k) \quad 5.3.1-8$$

Si se desea saber si un dato viene de un estado en particular, se aplica la regla de Bayes y obtenemos lo siguiente:

$$p(k|x_i) = \frac{p(x_i | k) p(k)}{p(x_i)} \quad 5.3.1-9$$

Una vez expuesto la anterior, hasta que los valores converjan se debe de hacer lo siguiente:

Paso E

Obtener la probabilidad de la clase k_i de la siguiente forma:

$$p(k) = \frac{1}{n} \sum_i P(k | x_i) \quad 5.3.1-10$$

donde k es el número de la clase, x_i es el píxel en estudio y n es el número total de elementos del arreglo(en nuestro caso de la imagen). Y $p(k|x_i)$ se obtiene de la siguiente forma:

$$p(k|x_i) = \frac{p(x_i | k) P(k)}{\sum_j p(x_i | j) P(j)} \quad 5.3.1-11$$

Paso M

Una vez encontrados los valores anteriores se obtiene los valores de m_{k+1} y σ_{k+1} de la siguiente forma:

Ahora la probabilidad de ver todos los datos, dado que las muestras son generadas independientemente, es la siguiente:

$$p(x_1, \dots, x_n) = \prod_i \sum_k p(x_i | k) p(k) \quad 5.3.1-12$$

para elegir los parámetros de cada iteración se usa un estimador de máxima verosimilitud en el logaritmo de la ecuación de probabilidad dada por:

$$\log L = \log(p(x_1, \dots, x_n)) = \sum_i \log \sum_k p(x_i | k) p(k) \quad 5.3.1-13$$

Si aplicamos derivadas parciales con respecto a los parámetros μ_k y σ_k , igualamos a 0 y obtenemos su máximo:

$$\frac{\partial \log(L)}{\partial \mu_k} = \sum_i p(k | x_i) \frac{(x_i - \mu_k)}{\sigma_k^2} = 0 \quad 5.3.1-14$$

$$\frac{\partial \log L}{\partial \sigma_k^2} = \sum_i p(k | x_i) \frac{(x_i - \mu_k)^2 - \sigma_k^2}{2\sigma_k^3} = 0 \quad 5.3.1-15$$

De estas ecuaciones despejamos a μ_k y a σ_k y obtenemos las ecuaciones para hacer la estimación[29].

$$\mu_{k+1} = \frac{\sum_i P(k | x_i) x_i}{\sum_i P(k | x_i)} \quad 5.3.1-16$$

$$\sigma_{k+1} = \frac{\sum_i P(k | x_i) (x_i - m_k)^2}{\sum_i P(k | x_i)} \quad 5.3.1-17$$

Para ver la convergencia, se propone un valor de ξ no mayor a 0.5. Es decir el algoritmo se detiene cuando:

$$\|m_{k+1} - m_k\| < \xi \qquad 5.3.1-18$$

5.3.2 Algoritmo SEM.

Se basa en el algoritmo EM, pero tiene la diferencia de se inicializa con un número N de clusters o clases y en cada intervalo se hace el paso E y después el paso M, para saber si el píxel pertenece o no a dicho cluster[18].

Tiene tres pasos: inicialización paso S, paso E y paso M.

Inicialización:

Se define a M como el número de clases.

Se asume, que en cada iteración (u observación) Y_i , una probabilidad $P_m^o(Y_i)$. En ausencia de alguna información, se toma para $[P_m^o(Y_i)]$, $1 < m < M$ la distribución uniforme. Lo anterior denota el numero de iteraciones, para cada $n \geq 0$.

Paso S:

Para cada Y_i se saca del grupo de clases $(e_1, e_2, \dots, e_{Mn})$ un elemento de acuerdo a la distribución $P_1^n(Y_i), \dots, P_k^n(Y_i)$. Esta extracción define una partición $Q_1^n, Q_2^n, \dots, Q_k^n$ de la muestra y_1, y_2, \dots, y_N .

Paso M

La idea del algoritmo SEM es la de suponer que cada Y_i pertenece a Q_m^n , para cada $1 \leq m \leq M$, es realizado de acuerdo a la distribución definida por f_m :

$$\text{donde } f_m = \sum_{p=1}^k \Pi_p f_p$$

densidad correspondiente a la clase e_m . Que es denotada como:

$$C_m^n \text{card}(Q_m^n), \text{ con } Q_m^n = (y_{1,m}^n, y_{2,m}^n, \dots, y_{C_m^n, m}^n),$$

después se hace la estimación de la media μ_m y la matriz de covariancia Γ_m^{n+1} . y probabilidad a priori de las frecuencias.

Paso E

Para cada Y_i , se define la siguiente distribución $[P_1^{n+1}(y_i), \dots, P_k^{n+1}(y_i)]$ de el grupo de clases para la distribución a posteriori basada en los parámetros actuales

$$(\hat{\Pi}_m^{n+1}, \mu_m^{n+1}, \Gamma_m^{n+1}), 1 \leq m \leq K$$

$$P_m^{n+1}(y_i) = \frac{\prod_m^{n+1} f_m^{n+1}(y_i)}{\sum_{q=1}^k \prod_q^{n+1} f_q^{n+1}(y_i)} \quad 5.3.2-1$$

donde f_m^{n+1} designa a la distribución Gaussiana correspondiente a $(\mu_m^{n+1}, \Gamma_m^{n+1})$.

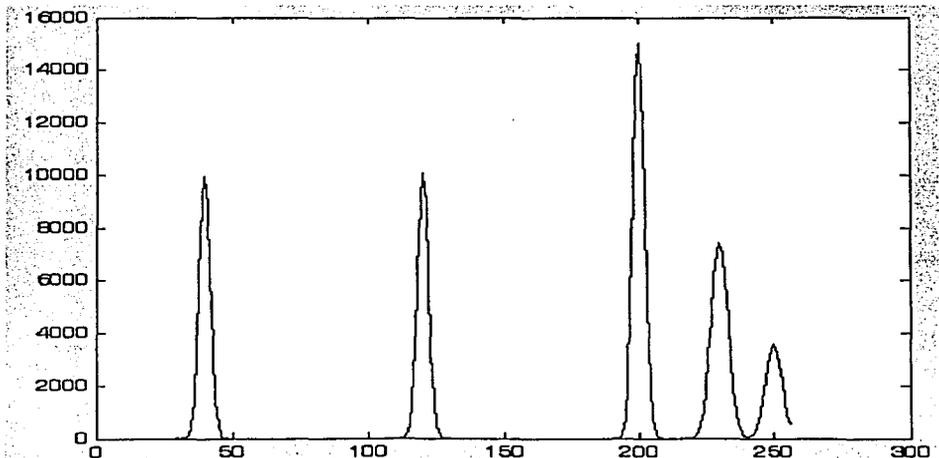
Después regresar al paso S.

5.4 Cálculo de los parámetros de una imagen por medio del algoritmo EM.

Para observar como se desenvuelve el algoritmo EM, se realizarán varios experimentos, con diferentes condiciones, tales como datos con modos dispersos y mezclado, para efectos de evaluar la confiabilidad del mismo.

Los datos que serán analizados son de tipo Gaussiano con medias y desviaciones estándar propuestas en diferentes combinaciones, para observar como se comporta el algoritmo EM. Los datos que se analizarán en cada caso son arreglos matriciales de 512×512 elementos (lo cuál nos da como resultado 262,144 datos), obtenidos por la aplicación de un generador de números aleatorios.

1) El histograma de la matriz de datos es el siguiente:



Capítulo No.5 Estimación de parámetros: Algoritmos EM y SEM

Al terminar de procesar el algoritmo, se obtuvieron los siguientes resultados:

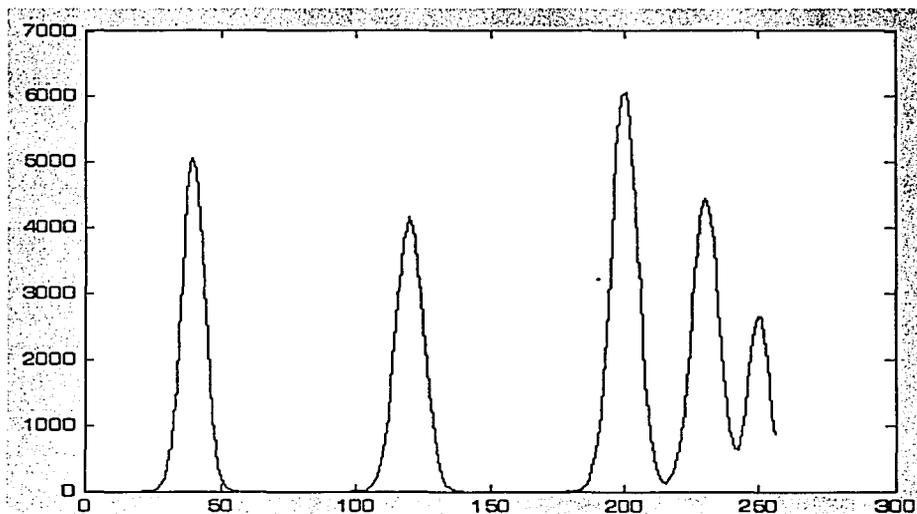
Tabla 1. Resultados.

Media real	Media obtenida	Error	Desviación estándar real	Desviación estándar obtenida	Error
40	40	0	2	2	0
120	120	0	2	2	0
200	200	0	2	2	0
230	230	0	3	3	0
250	250	0	3	3	0

Podemos observar que no hay ningún error al obtener las medias aplicando el Algoritmo EM.

En el siguiente experimento, se incrementará la variancia en los datos y observaremos que efectos tiene en los resultados.

2) El histograma de los datos es el siguiente:



Los resultados arrojados por el Algoritmo EM son los siguientes:

Tabla 2. Resultados

Media real	Media obtenida	Error	Desviación estándar real	Desviación estándar obtenida	Error
40	41	2.5	4	2.39	40.25
120	121	0.8333333333	5	3.19	36.2
200	200	0	5	3.13	37.4
230	234	1.739130435	5	7.61	52.2
250	249	0.4	4	2.92	27

Lo primero que observamos es que este cambio arroja un incremento en el porcentaje de error de las desviaciones estándar calculadas.

Los datos de abajo tienen las mismas característica que los datos del inciso 2, pero se cambio el valor de una media y el de todas las desviaciones estándar.

3) El histograma de los datos es el siguiente:

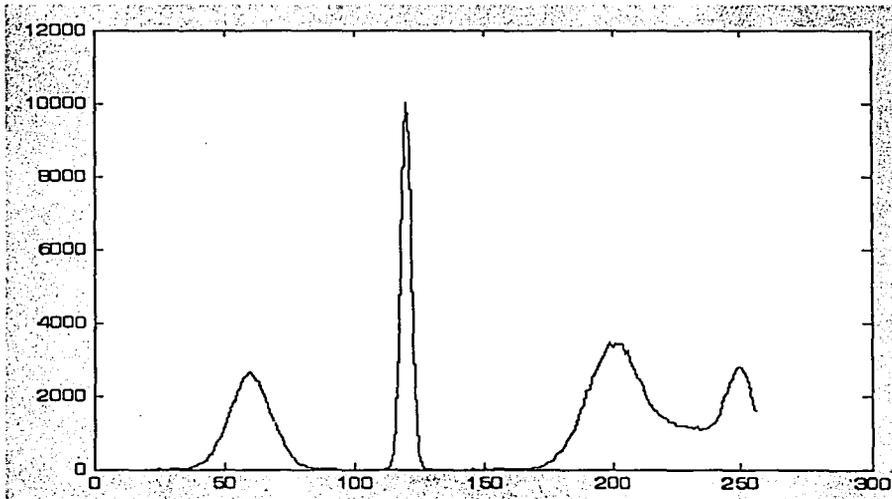


Tabla 3. Resultados.

Media real	Media obtenida	Error	Desviación estándar real	Desviación estándar obtenida	Error
60	61	1.666666667	8	5	37.5
120	120	0	2	2	0
200	200	0	10	6	40
230	233	1.304347826	20	19	5
	204			7	
250	250	0	5	4	20

En la prueba siguiente, los datos tienen por característica principal que las dos últimas medias están más juntas y las desviaciones estándar son ahora pequeñas.

4) Histograma de los datos.

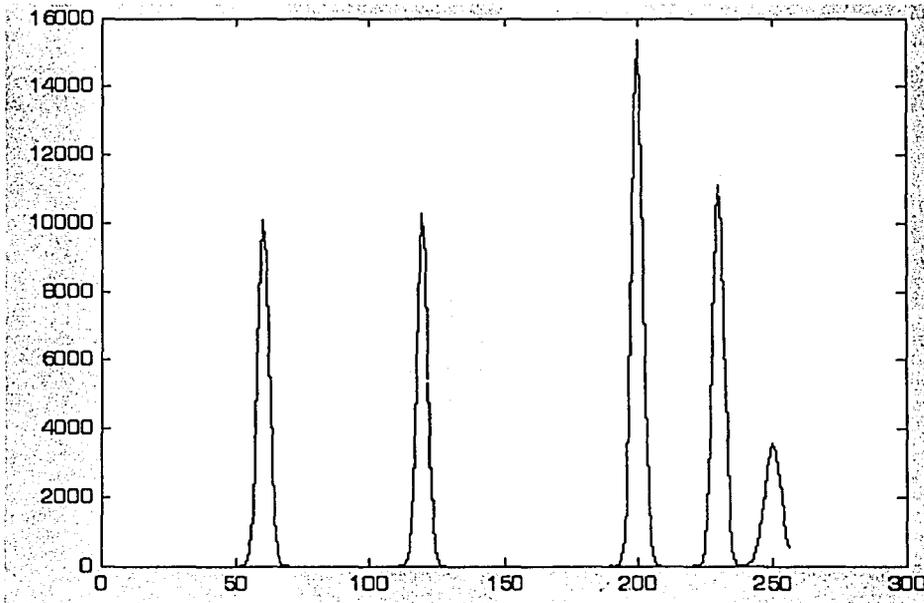


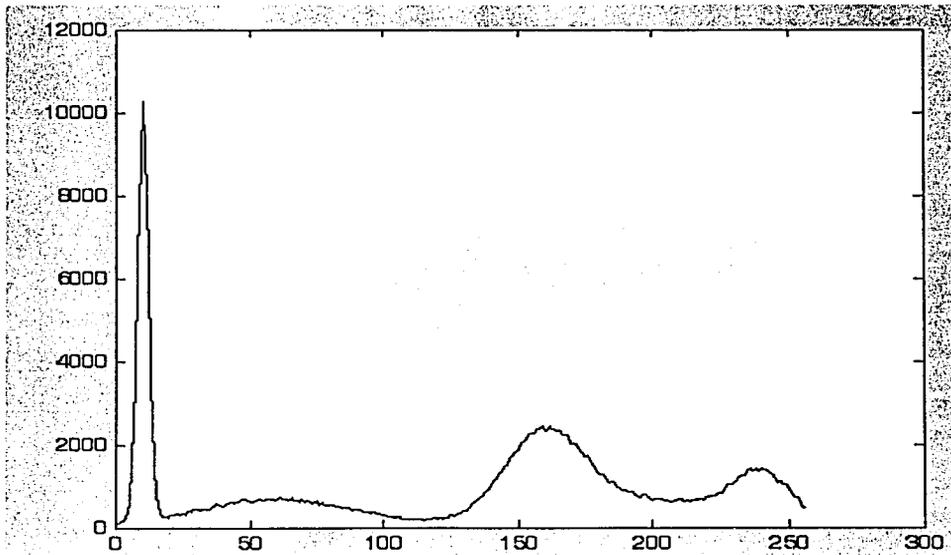
Tabla 4. Resultados

Media real	Media obtenida	Error	Desviación estándar real	Desviación estándar obtenida	Error
60	60	0	2	1.99	0.5
120	120	0	2	2	0
200	200	0	2	1.99	0.5
230	230	0	2	2.01	0.5
250	250	0	3	4.48	49.3333333

Se observa que en las medias no hay errores al calcularlas, pero si los hay cuando las campanas gaussianas están muy cercanas y las variancias son significativas.

El último caso es un histograma con los datos muy dispersos.

5) Histograma de los datos.



Capítulo No.5 Estimación de parámetros: Algoritmos EM y SEM

Lo primero que se observa es que los errores que se presentan al calcular las medias no son muy significativos, cosa contraria en el cálculo de las desviaciones estándar.

Tabla 5. Resultados.

Media real	Media obtenida	Error	Des real	Des obtenida	Error
10	10	0	2	2	0
60	55	8.333333333	30	10	66.6666667
	70		10		
160	160	0	15	8	46.6666667
	163		9		
200	221	10.5	35	26	25.7142857
240	241	0.416666667	10	6	40

Conclusiones.

Como se observa en los anteriores experimentos, el algoritmo EM converge hacia los valores reales de los parámetros si los datos no están muy dispersos, pero cuando los datos lo están, el algoritmo converge sobre los picos que presenta el mismo histograma, presentando "nuevas" medias y desviaciones estándar, las cuales se detectan en los datos obtenidos por la aplicación del generador de números aleatorios.

El error mayor para el cálculo de las desviaciones estándar se explica, por que al momento de que los datos están más dispersos, y como el algoritmo calcula los parámetros (las medias y las desviaciones estándar) basándose en los picos del histograma, era de esperarse que la dispersión de los datos no pueda registrarse de forma muy confiable, pues esta no se puede observar en los picos del histograma.

La generación de medias y desviaciones "nuevas", no necesariamente representan un error, si no lo que nos indica es que, los datos sintéticos, posiblemente fueron generados de tal forma que en un momento se sumaron cierto numero de ellos produciendo en el histograma un pico nuevo, o un valle algo plano, que al ser procesados los datos por el algoritmo fueron encontrados. De esta forma, cuando se tenía un histograma muy denso en sus valores, se encontraban valores nuevos de medias, las cuales significaban picos o valles extensos en el mismo.

Una desventaja del algoritmo EM es que el tiempo de procesamiento del algoritmo es muy grande, porque se puede presentar el caso de que aparezcan dos medias similares en valor, por ejemplo 101 y 103, las cuáles en valor son diferentes y deben de ser tomadas en cuenta como parte de un resultado, pero para ser tomadas como parte de una segmentación, la diferencia de 2 valores de intensidad de gris no es muy significativa, por esto en el siguiente capítulo se discutirá una forma de segmentación de imágenes por medio del algoritmo EM modificado de tal forma que el tiempo de procesamiento disminuya, la idea es formar un número reducido de clases, a partir de un número mayor.

Segmentación contextual no supervisada: Esquema EM.

6.1 Introducción.

Ahora en esta capítulo, se propondrá una forma de cómo utilizar los resultados del algoritmo EM para realizar una segmentación contextual no supervisada. Se intentará probar que esta forma de segmentar cumple con las característica de calidad visual, para comprobar esto, los resultados se compararan con los obtenidos por medio de un algoritmo clásico: el k – medias.

6.2 Modelado de contexto.

En una imagen, las regiones son piezas contiguas, esto es, un píxel etiquetado toma un valor discreto, y todos los píxeles que están igualmente etiquetados están ampliamente correlacionados. Usando la equivalencia entre una distribución de Gibbs y un campo aleatorio de Markov, para una probabilidad conjunta de valor 1[29]:

$$P(1) = \frac{\exp(-U(1))}{Z_1}$$

Donde $U(1)$ es la función de energía, y la constante normalizada Z_1 es la función de partición. Un sistema de vecindad debe de ser establecida para especificar la función de clique $V_c^{(i)}(1)$ donde $U(1) = \sum_{i=1}^N V_c^{(i)}(1)$. La vecindad del píxel se denota con ∂_i . Entonces, la función de energía es la siguiente:

$$U(1) = \sum_{i=1}^N \sum_{j \in \partial_i} [\theta_i I(l_i, l_j)]$$

donde $\{\theta_i\}$ es el parámetro de Markov y $I(\dots)$ es la función indicador.

Si se aplica la ley de Bayes, se puede construir un conjunto para unir al modelo de imagen píxel con el modelo imagen contexto. Esto es, una probabilidad condicional a posteriori para la imagen contexto l , dado la observada imagen píxel x , que tiene la forma de una distribución de Gibbs dada por:

$$p(l|x) = \frac{\exp(-U(l|x))}{Z_{l|x}}$$

donde $Z_{l|x}$ es una constante normaliza. Y la función de energía correspondiente es la siguiente:

$$U(l|x) = \sum_{i=1}^N \left\{ \prod_{k=1}^K [0.5 \ln(\sigma_k^2 + (x_i - \mu_k)^2 / 2\sigma_k^2)] I(l, k) \right\} + \sum_{j \in \partial_i} [\theta_i I(l_i, l_j)]$$

Tomando en cuenta la vecindad del píxel, tenemos lo siguiente:

$$p(l_i | l_{\partial_i, x}) = \frac{\exp(-U(l_i | l_{\partial_i, x}))}{Z_i}$$

Donde Z_i es una constante normalizada, l_{∂_i} es la imagen contexto local, y la energía se define como[5]:

$$U(l_i | l_{\partial_i, x}) = \prod_{k=1}^K \left[\frac{1 \ln(\sigma_k^2)}{2} + \frac{(x_i - \mu_k)^2}{2\sigma_k^2} \right]^{r^{(l_i, k)}} + \sum_{j \in \partial_i} [\theta_i I(l_i, l_j)]$$

6.3 Algoritmo EM y Segmentación.

Lo que se propone para realizar la segmentación no supervisada es lo siguiente:

1. Se aplica el algoritmo EM a la imagen que se quiere segmentar de la siguiente manera:

Se inicializa el modelo suponiendo valores para μ_k y σ_k . Siendo k el número de clases. La imagen se trata como un arreglo de $p(i,j) = 1,2,...,n$ elementos, manejándose como VA de tipo Gaussiano, con media y variancia definidas como:

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \quad 6.3-1$$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \quad 6.3-2$$

Se piensa que el dato observado es la mezcla de x estados, siendo $p(k)$ la probabilidad del k - ésimo estado. Se asume que la mezcla(o probabilidad condicional) puede ser descrita por una variable Gaussiana de la siguiente forma:

$$p(x_i|k) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(x_i - \mu_k)^2}{2\sigma_k^2}} \quad 6.3-3$$

Si se desea saber si un dato viene de un estado en particular, se aplica la regla de Bayes y obtenemos lo siguiente:

$$p(k|x_i) = \frac{p(x_i | k)p(k)}{p(x_i)} \quad 6.3-4$$

Una vez expuesto la anterior, hasta que los valores converjan se debe de hacer lo siguiente:

Paso E

Obtener la probabilidad de la clase k_i de la siguiente forma:

$$p(k) = \frac{1}{n} \sum_i P(k | x_i) \quad 6.3-5$$

donde k es el número de la clase, x_i es el píxel en estudio y n es el número total de elementos del arreglo (en nuestro caso de la imagen). Y $p(k|x_i)$ se obtiene de la siguiente forma:

$$p(k|x_i) = \frac{p(x_i | k)P(k)}{\sum_j p(x_i | j)P(j)} \quad 6.3-6$$

Paso M

Una vez encontrados los valores anteriores se obtiene los valores de m_{k+1} y σ_{k+1} de la siguiente forma:

Ahora la probabilidad de ver todos los datos, dado que las muestras son generadas independientemente, es la siguiente:

$$p(x_1, \dots, x_n) = \prod_i \sum_k p(x_i | k) p(k) \quad 6.3-7$$

para elegir los parámetros de cada iteración se usa un estimador de máxima verosimilitud en el logaritmo de la ecuación de probabilidad dada por:

$$\log L = \log(p(x_1, \dots, x_n)) = \sum_i \log \sum_k p(x_i | k) p(k) \quad 6.3-8$$

Si aplicamos derivadas parciales con respecto a los parámetros μ_k y σ_k y obtenemos su máximo:

$$\frac{\partial \log(L)}{\partial \mu_k} = \sum_i p(k | x_i) \frac{(x_i - \mu_k)}{\sigma_k^2} = 0 \quad 6.3-9$$

$$\frac{\partial \log L}{\partial \sigma_k^2} = \sum_i p(k | x_i) \frac{(x_i - \mu_k)^2 - \sigma_k^2}{2\sigma_k^3} = 0 \quad 6.3-10$$

De estas ecuaciones despejamos a μ_k y a σ_k y obtenemos las ecuaciones para hacer la estimación.

$$\mu_{k+1} = \frac{\sum_i P(k | x_i) x_i}{\sum_i P(k | x_i)} \quad 6.3-11$$

$$\sigma_{k+1} = \frac{\sum_i P(k | x_i) (x_i - m_k)^2}{\sum_i P(k | x_i)} \quad 6.3-12$$

Antes de regresar al paso E, se verifica que el valor de las medias “vecinas” obtenidas en esta iteración difieran en valor no más de 20 niveles de gris, si se da el caso, se saca un promedio de ambas y se eliminan dejando en su lugar el valor del promedio de ambas, disminuyendo así el número de clases en cada iteración. Después de verificar lo anterior, regresar al paso E.

Para verificar que el algoritmo ya convergió sobre un valor, lo que se hace es verificar que los valores de las medias de la iteración anterior con los valores de la iteración actual no difieran a un valor ξ (en nuestro caso el valor fue de 1), esto es, el algoritmo terminará cuando:

$$\| m_{k+1} - m_k \| < \xi \quad 6.3-13$$

6.3.1 Procedimiento para la segmentación.

Para realizar la segmentación, ahora que se tienen los valores de las medias y el número de las clases, se pueden tomar en cuenta varios de los algoritmos anteriores.

Pero para llevar una coherencia en los procedimientos, recordemos que el algoritmo EM propone que la imagen es un arreglo Gaussiano y además se basa en la teoría bayesiana para obtener varios de sus parámetros, es por eso que el algoritmo de segmentación que se usará es el del clasificador bayesiano. Y para tener un resultado para comparar, se utilizara también un algoritmo de umbral.

Es decir, mi propuesta es la siguiente:

- 1) Primero se aplicará el algoritmo EM modificado para obtener el número de clases en que estadísticamente se puede segmentar la imagen, y de estas su media y desviación estándar.
- 2) Una vez obtenidos los datos anteriores sólo se necesita un algoritmo de segmentación para presentar un resultado para nuestros ojos de tal forma que seamos capaces de analizarlos y dar una conclusión. Para realizar esto se proponen dos procedimientos:
 - a. Un algoritmo de umbral: Donde se propone que las medias obtenidas por el algoritmo EM sean consideradas como umbrales para realizar la segmentación.
 - b. Un clasificador bayesiano: Para presentar el resultado de la imagen segmentada se aplica un clasificador óptimo bayesiano, ya que el algoritmo EM, tiene como premisa que los datos son de tipo Gaussiano y toma muy en cuenta a la teoría Bayesiana para obtener los parámetros correspondientes de la imagen analizada.

6.3.2 Resultados

Las imágenes de prueba son de 512 X 512 píxeles, a excepción de iko.bmp que es de 329 X 303 píxeles y en 256 niveles de grises, las cuáles son la siguientes:

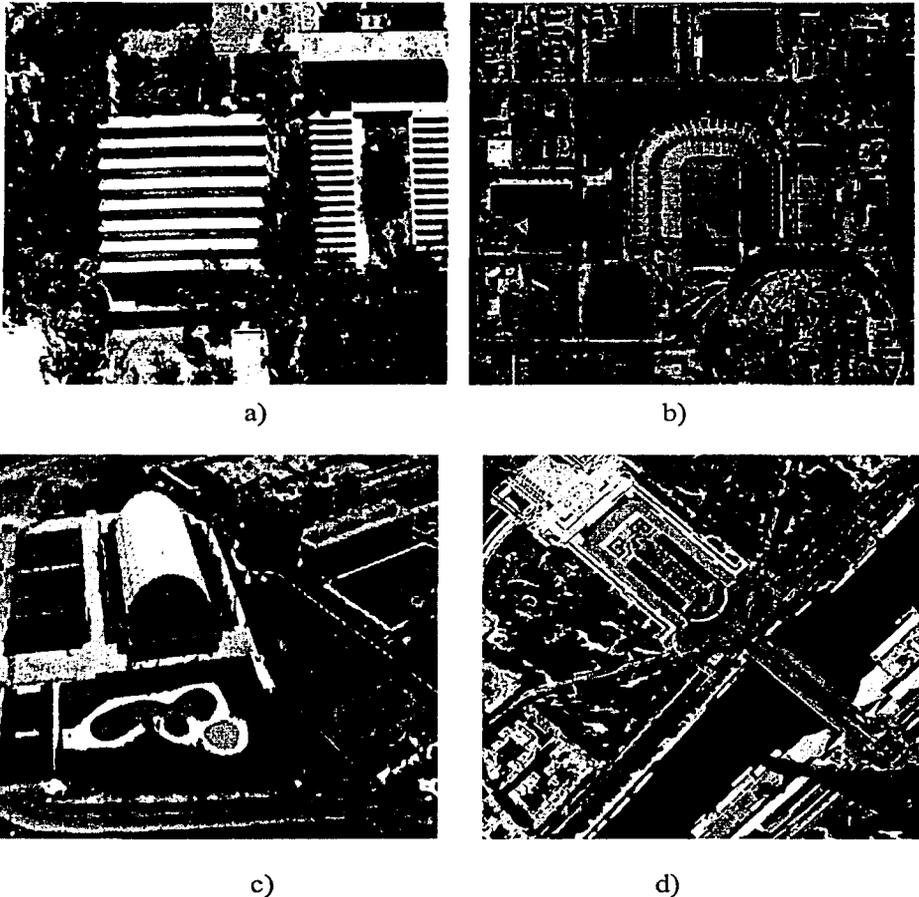


Fig 1. Imágenes utilizadas: a) usar.bmp, b) iko.bmp, c) palacio.bmp y d) paris3.bmp

El procedimiento es el siguiente: Se toma la imagen, se le aplica el algoritmo EM y después para presentar un resultado visual, se utiliza el algoritmo de umbral descrito en el Anexo1. Otra alternativa teóricamente más funcional, lo representa el esquema óptimo bayesiano descrito en la sección 4.4.4.1.1. Para su aplicación, se emplean también los resultados del algoritmo de estimación EM.

Resultado No. 1

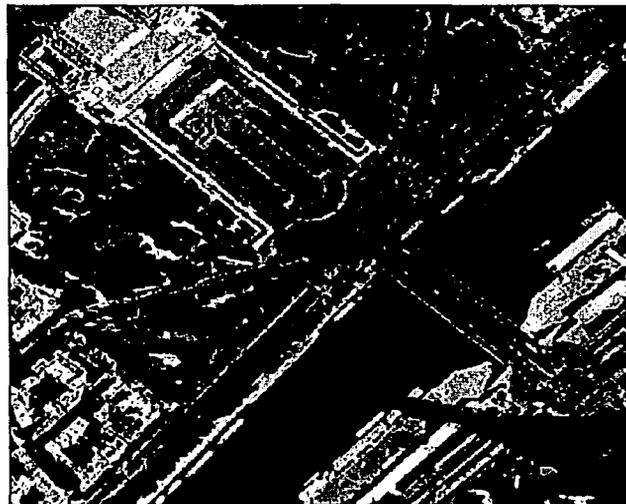
La primera imagen de prueba corresponde a:
1.a) Paris3.bmp:



a)



b)



c)

Fig. No.2 a) Imagen original y b) Resultado del algoritmo EM - umbral y c) Resultado del algoritmo EM - bayesiano.

Capítulo No.6 Segmentación contextual no supervisada.

1.b) De los resultados del algoritmo de umbral, la tabla no.1 presenta el valor en niveles de gris de la clase, el número de píxeles que pertenece a tal clase y el porcentaje que representan para la imagen original.

Tabla No.1.

Valor	No. de elementos de la imagen	Porcentaje
28	46959	17.9134369
66	95025	36.2491608
120	46097	17.58461
159	31888	12.1643066
192	18354	7.00149536
233	23811	9.08317566

De los resultados del algoritmo del clasificador bayesiano óptimo, la tabla no. 2 presenta el valor en niveles de gris de la clase, el número de píxeles que pertenece a tal clase y el porcentaje que representan para la imagen original.

Tabla de resultados No.2.

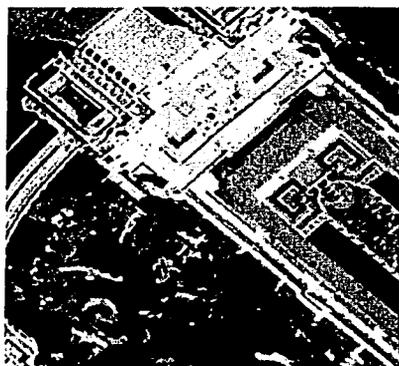
Valor	No. de elementos de la imagen	Porcentaje
28	69567	26.5377045
66	83720	31.9366455
120	43270	16.5061951
159	28784	10.9802246
192	21968	8.38012695
233	14835	5.65910339

1.c) Se realizará una comparación numérica de los datos obtenidos al realizar los dos tipos de segmentación. En la tabla no. 3, se observa la diferencia en porcentaje de los dos métodos con respecto al número total de píxeles de la imagen. Los resultados son los siguientes:

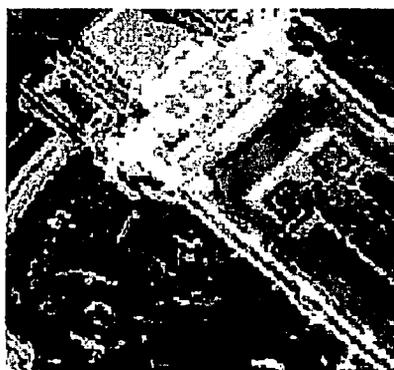
Tabla No. 3.

No. de clase	Porcentaje de desviación
1	8.62426758
2	4.31251526
3	1.07841492
4	1.18408203
5	1.37863159
6	3.42407227

1.d) Ahora se muestran ampliaciones de los resultados obtenidos.



a)



b)

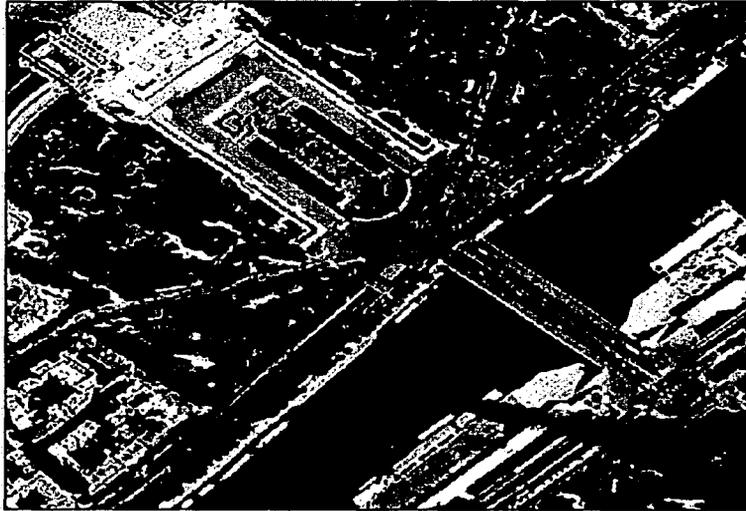


c)

Figura No.3 Comparación de los resultados de la segmentación empleando ventanas ampliadas de 120 X 120 píxeles. a) Imagen original, b) Algoritmo de umbral y c) Algoritmo del clasificador óptimo bayesiano.

Se observa que ambas segmentaciones tienen como característica que se pueden identificar fácilmente regiones tales como árboles, sombras y construcciones. Las diferencias que se presentan en el número de píxeles que pertenecen a cada clase, se explica de la siguiente forma: el algoritmo de umbrales no toma en cuenta a la variancia y el algoritmo del clasificador óptimo bayesiano si, es por eso que el número de píxeles asociados a una clase en cada algoritmo es diferente.

1.e) Por ultimo se aplica el algoritmo de k – medias y se compara su resultado contra los obtenidos anteriormente.



a)



b)



c)



d)

Figura No.4 Resultados: a) Imagen original, b) Segmentada por k – medias, c) Segmentada por EM - umbral y d) Segmentada por EM - bayesiano.

A continuación se presentan ventanas ampliadas de 120 X 120 píxeles.

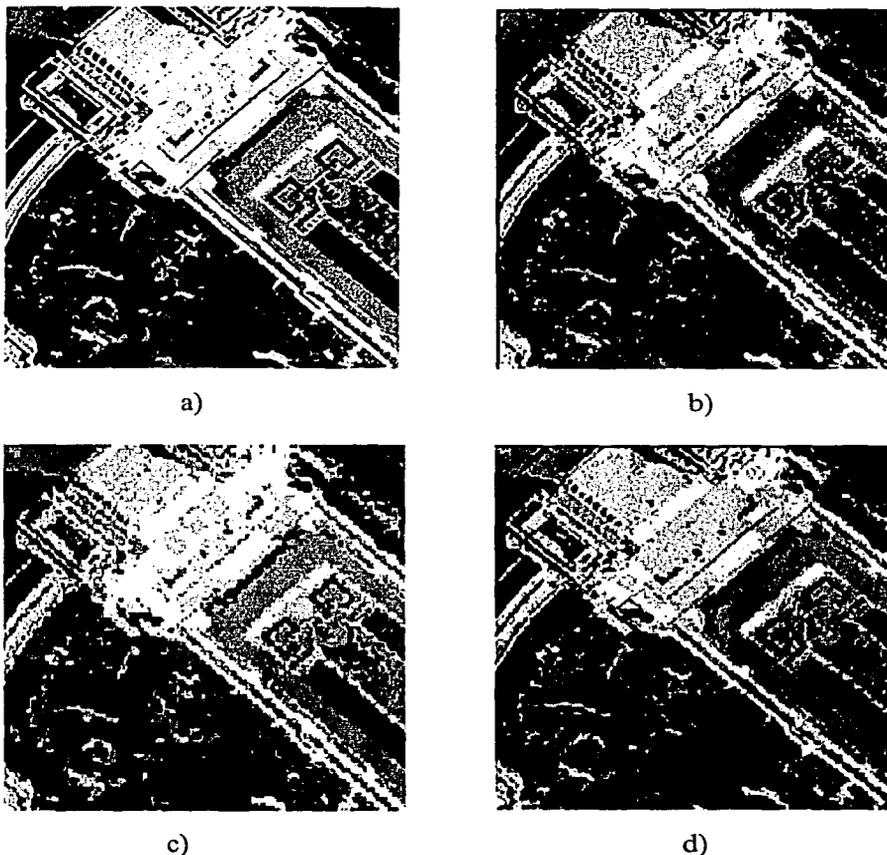


Figura No. 5 Acercamiento de los resultados, a) Imagen original, b) Segmentada por k – medias, c) Segmentada por EM - umbral y d) Segmentada por EM - bayesiano.

En la tabla no. 4 se observan los valores de las medias calculadas a partir de k –media y del algoritmo EM, así como su diferencia porcentual.

Tabla no.4

Clase No.	Valor obtenido por k – medias	Valor obtenido por EM	Diferencia porcentual
1	33	28	15.15151515
2	68	66	2.941176471
3	111	120	8.108108108
4	152	159	4.605263158
5	192	192	0
6	230	233	1.304347826

Capítulo No.6 Segmentación contextual no supervisada.

Y en la figura No.6 se muestran las medias obtenidas por los algoritmos presentados

Medias obtenidas por EM —————

Medias obtenidas por k – medias —————>

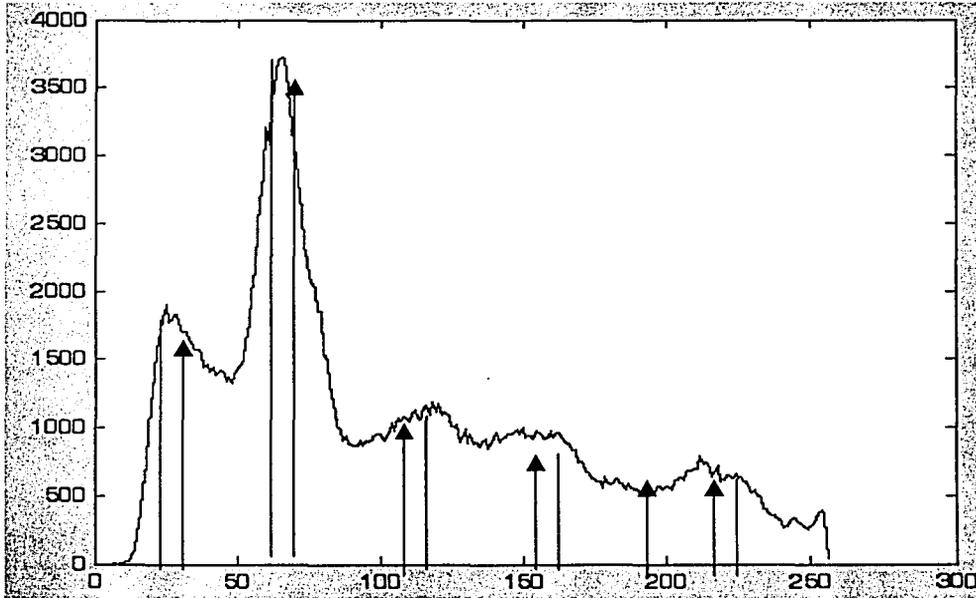


Figura No. 6 Histograma de la imagen paris3.bmp

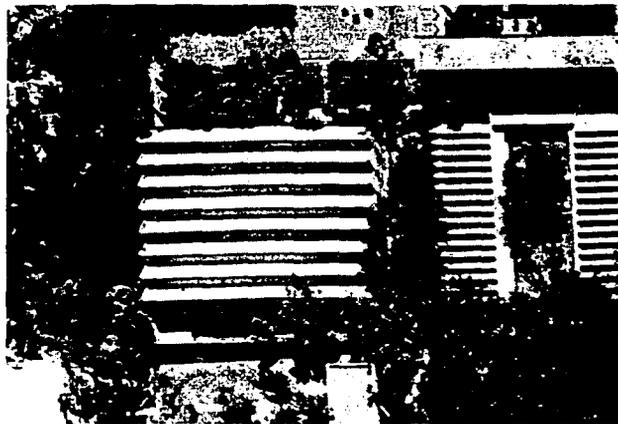
Tabla no. 5 Comparación de los píxeles segmentados por los tres métodos que se presentaron.

No. de Clase	Método de k – medias		Método de EM - umbral		Método de EM - bayesiano	
	No. de elementos	Porcentaje	No. de elementos	Porcentaje	No. de elementos	Porcentaje
1	51113	19.30949287	46959	17.7401928	69567	26.28105355
2	87348	32.99836799	95025	35.8985886	83720	31.62778046
3	42475	16.04622522	46097	17.4145461	42475	16.04622522
4	37137	14.0296331	31888	12.0466634	37137	14.0296331
5	23024	8.698017408	18364	6.93756044	23024	8.698017408
6	21047	7.95114543	23811	8.99533063	21047	7.95114543

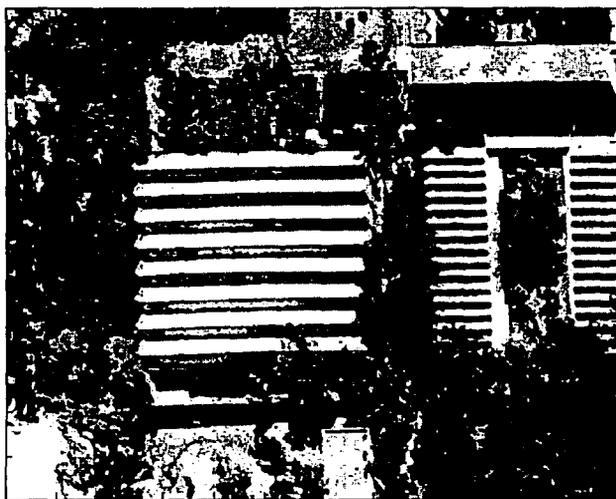
Resultados No. 2

La segunda imagen de prueba es:

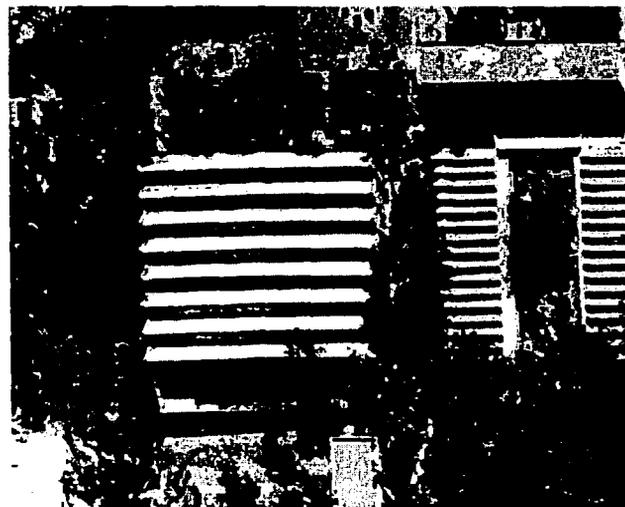
2.a) Usar.bmp:



a)



b)



c)

Fig. No.7 a) Imagen original y b) Resultado del algoritmo EM - umbral y c) Resultado del algoritmo EM - bayesiano.

Capítulo No.6 Segmentación contextual no supervisada.

2.b) De los resultados del algoritmo de umbral, la tabla no.6 presenta el valor en niveles de gris de la clase, el número de píxeles que pertenece a tal clase y el porcentaje que representan para la imagen original.

Tabla No.6.

Valor	No. de elementos de la imagen	Porcentaje
26	92149	29.1050763
72	88643	27.9977133
120	46690	14.7469426
166	22515	7.11131747
199	24064	7.600566
242	44257	13.9784844

De los resultados del clasificador optimo bayesiano, la tabla no.7 presenta el valor en niveles de gris de la clase, el número de píxeles que pertenece a tal clase y el porcentaje que representan para la imagen original.

Tabla No.7.

Valor	No. de elementos de la imagen	Porcentaje
26	120455	38.0454695
72	76225	24.0755129
120	36385	11.4921291
166	22589	7.13469022
100	28850	9.11221447
242	33816	10.6807156

2.c) Se realizará una comparación numérica de los datos obtenidos al realizar los dos tipos de segmentación. En la tabla no. 8, se observa la diferencia en porcentaje de los dos métodos con respecto al número total de píxeles de la imagen. Y los resultados son los siguientes:

Tabla No. 8

No. de clase	Porcentaje de desviación
1	8.89
2	3.90
3	3.23
4	0.02
5	1.50
6	3.28

2.d) Ahora se muestran ampliaciones de los resultados obtenidos.



a)



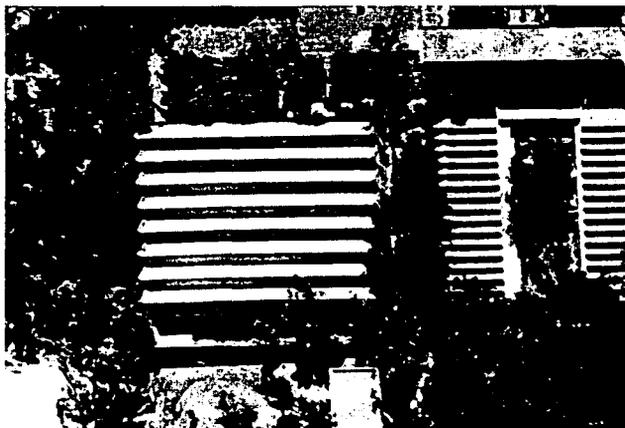
b)



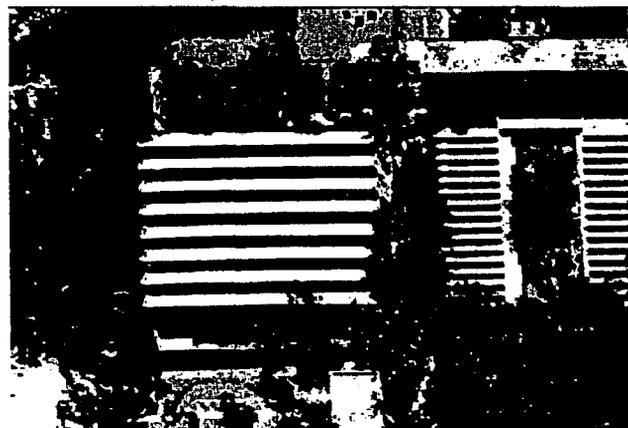
c)

Figura No.8 Comparación de los resultados de la segmentación empleando ventanas ampliadas de 120 x 90 píxeles . a) original, b) umbral y, c) Bayesiana.

2.e) Por ultimo se comparara el resultado obtenido contra uno obtenido por el algoritmo de k - medias.

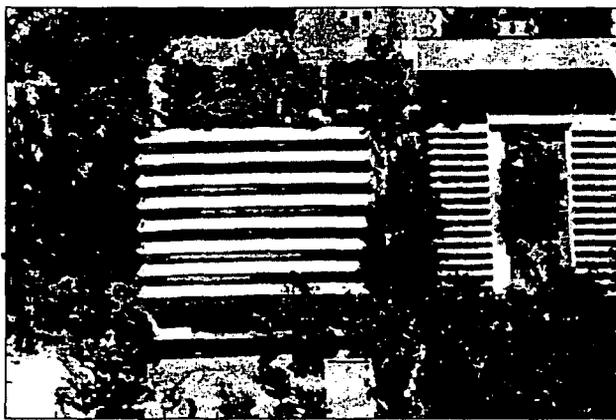


a)

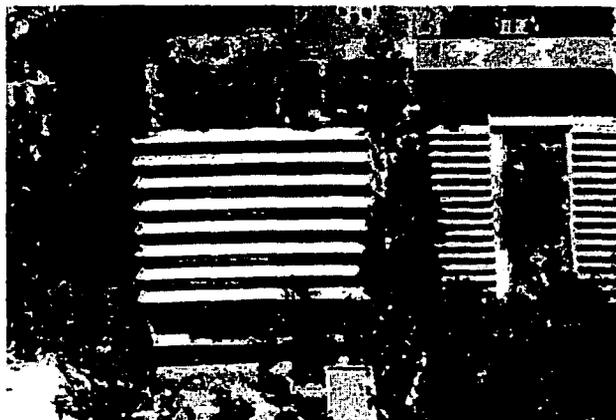


b)

TESIS CON
FALLA DE ORIGEN



c)



d)

Figura No.9. Resultados: a) Imagen original, b) Segmentada por k – medias, c) Segmentada por EM - umbral y d) Segmentada por EM - bayesiano.



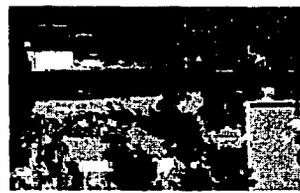
a)



b)



c)



d)

Figura No.10 Acercamiento a los resultados, a) Imagen original, b) Segmentada por k – medias, c) Segmentada por EM - umbral y d) Segmentada por EM - bayesiano.

En la tabla no. 9 se observan los valores de las medias calculadas a partir de k –media y del algoritmo EM, así como su diferencia porcentual.

Tabla no.9

Clase No.	Valor obtenido po k - medias	Valor obtenido por EM	Diferencia porcentual
1	22	26	18.18181818
2	59	72	22.03389831
3	97	120	23.71134021
4	139	166	19.42446043
5	189	199	5.291005291
6	236	242	2.542372881

Capítulo No.6 Segmentación contextual no supervisada.

Y en la figura No.11 se muestran las medias obtenidas por los algoritmos presentados

Medias obtenidas por EM —————

Medias obtenidas por k – medias —————>

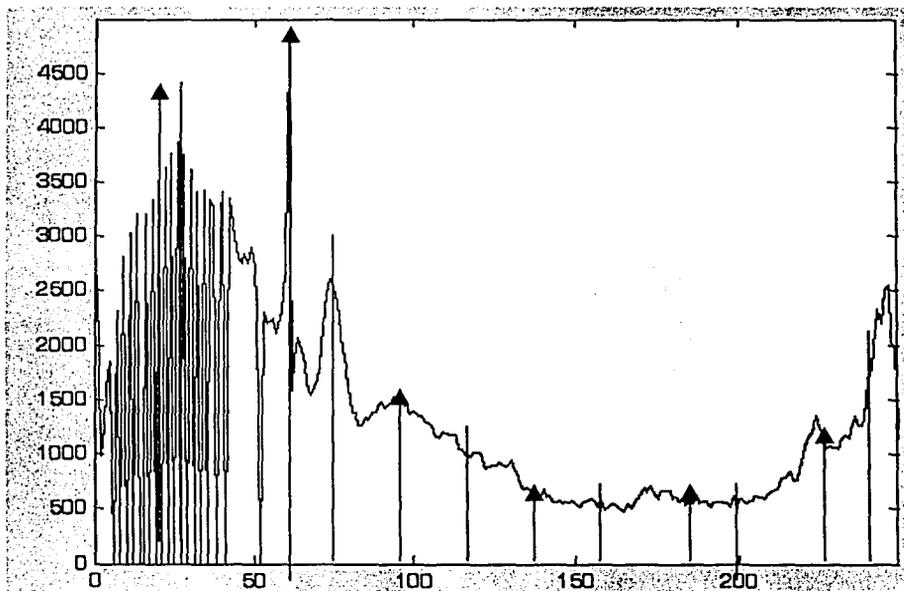


Figura No. 11 Histograma de la imagen usar.bmp

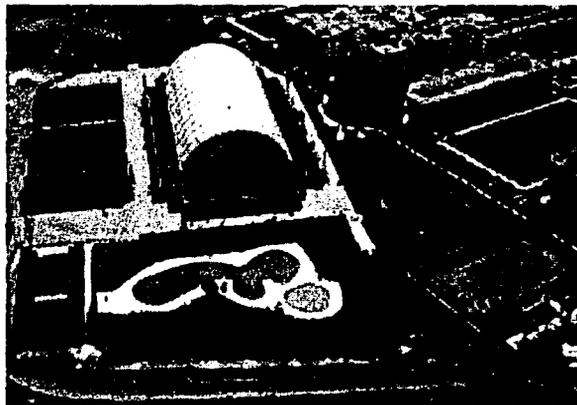
Tabla no. 10 Comparación de los píxeles segmentados por los tres métodos que se presentaron.

No. de Clase	Método de k - medias		Método de EM - umbral		Método de EM - bayesiano	
	No. de elementos	Porcentaje	No. de elementos	Porcentaje	No. de elementos	Porcentaje
1	68483	21.5140	92149	28.9487	120455	33.8410
2	86138	27.0603	88643	27.8473	76225	23.9461
3	53043	16.6635	46690	14.6677	36385	11.4303
4	31316	9.83796	22515	7.07311	22589	7.09636
5	28516	8.95833	24064	7.55973	28850	9.06326
6	50822	15.9657	44257	13.9033	33816	10.6233

Resultados No. 3

La tercera imagen de prueba es:

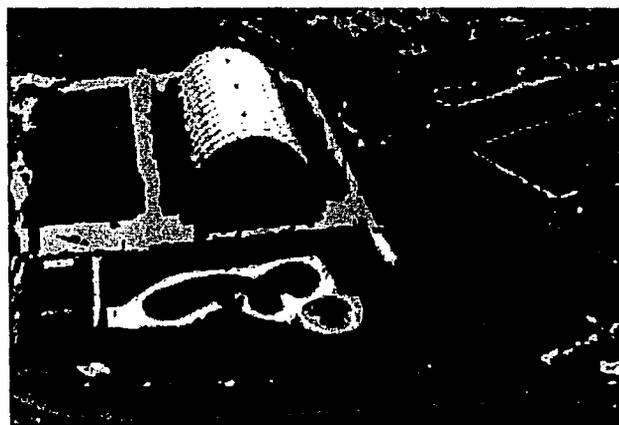
3.a) Palacio.bmp:



a)



b)



c)

Fig. No.12 a) Imagen original y b) Resultado del algoritmo EM - umbral y c) Resultado del algoritmo EM - bayesiano.

Capítulo No.6 Segmentación contextual no supervisada.

3.b) De los resultados del algoritmo de umbral, la tabla no.11 presenta el valor en niveles de gris de la clase, el número de píxeles que pertenece a tal clase y el porcentaje que representan para la imagen original.

Tabla de resultados No.11.

Valor	No. de elementos de la imagen	Porcentaje
75	118339	45.142746
101	52365	19.9756622
127	49978	19.065094
194	28962	11.0481262
242	12500	4.76837158

De los resultados del algoritmo del clasificador óptimo bayesiano, la tabla no. 12 presenta el valor en niveles de gris de la clase, el número de píxeles que pertenece a tal clase y el porcentaje que representan para la imagen original.

Tabla de resultados No. 12.

Valor	No. de elementos de la imagen	Porcentaje
75	133279	50.8419037
101	55082	21.0121155
127	40492	15.4464722
194	22869	8.72383118
242	10422	3.97567749

3.c) Se realizará una comparación numérica de los datos obtenidos en porcentaje al realizar los dos tipos de segmentación. Y los resultados son los siguientes:

Tabla No. 13.

No. de clase	Porcentaje de desviación
1	5.69915771
2	1.03645325
3	3.61862183
4	2.32429504
5	0.79269409

3.d) Ahora se presentan ampliaciones de los resultados obtenidos.

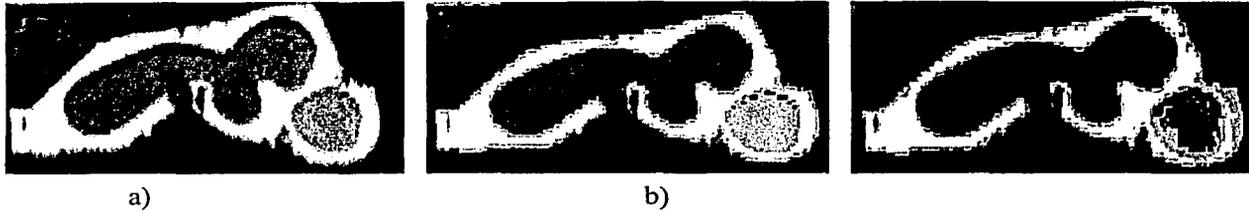
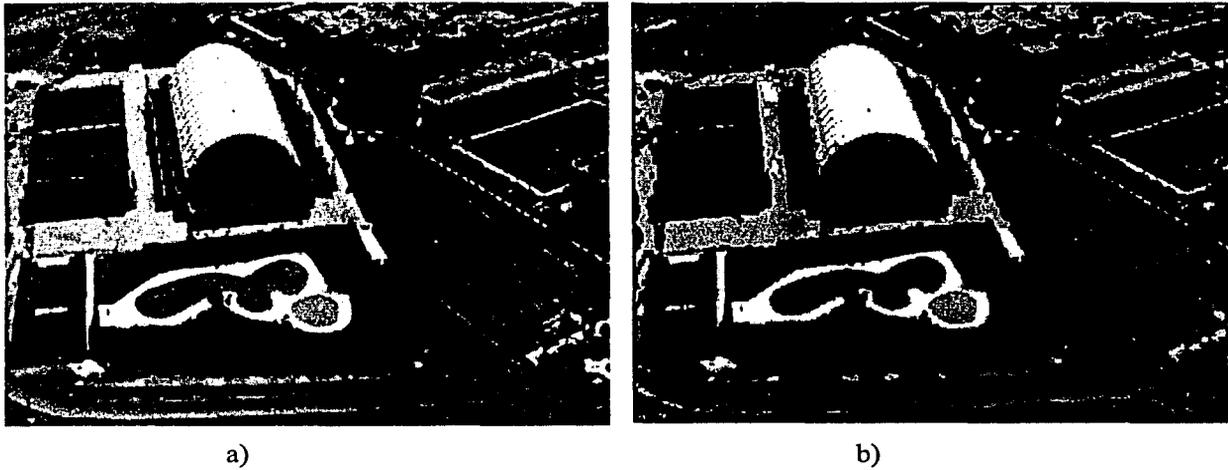


Figura No.13 Comparación de los resultados de la segmentación empleando ventanas ampliadas de 120 x 80 píxeles . a) original b) umbral y c) bayesiana

3.e) Por ultimo se comparara el resultado obtenido contra uno obtenido por el algoritmo de k - medias.





c)

d)

Figura No.14 Resultados: a) Imagen original, b) Segmentada por k – medias, c) Segmentada por EM - umbral y d) Segmentada por EM - bayesiano.

Acercamiento de los resultados



a)



b)



c)



d)

Figura No.15 Acercamiento a los resultados, a) Imagen original, b) Segmentada por k – medias, c) Segmentada por EM - umbral y d) Segmentada por EM - bayesiano.

Capítulo No.6 Segmentación contextual no supervisada.

En la tabla no. 13 se observan los valores de las medias calculadas a partir de k –media y del algoritmo EM, así como su diferencia porcentual.

Tabla no.13

Clase No.	Valor obtenido po k – medias	Valor obtenido por EM	Diferencia porcentual
1	63	75	19.04761905
2	99	101	2.02020202
3	135	127	5.925925926
4	185	194	4.864864865
5	235	242	2.978723404

Y en la figura No.16 se muestran las medias obtenidas por los algoritmos presentados

Medias obtenidas por EM —————

Medias obtenidas por k – medias —————>

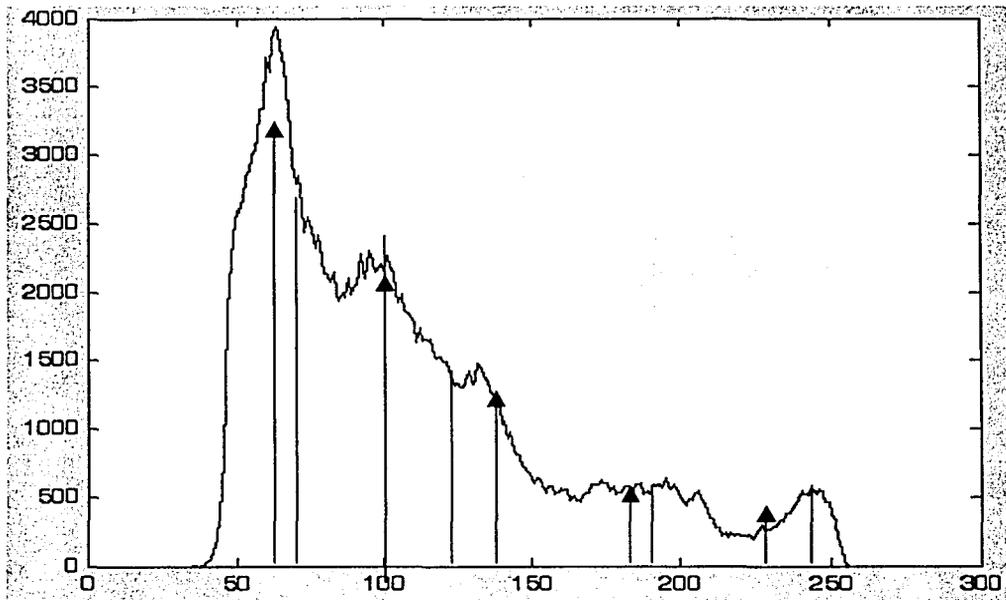


Figura No. 16 Histograma de la imagen palacio.bmp

Capítulo No.6 Segmentación contextual no supervisada.

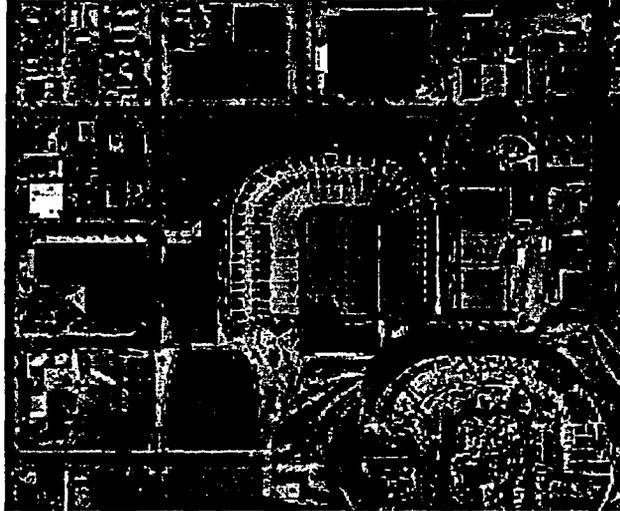
Tabla no. 14 Comparación de los píxeles segmentados por los tres métodos que se presentaron.

No. de Clase	Método de k – medias		Método de EM - umbral		Método de EM - bayesiano	
	No. de elementos	Porcentaje	No. de elementos	Porcentaje	No. de elementos	Porcentaje
1	104127	39.3371464	118339	44.7061624	133279	50.35020249
2	71416	26.97956963	52365	19.782474	55082	20.80890353
3	45139	17.05263237	49978	18.880712	40492	15.29708656
4	26953	10.18231685	28962	10.9412778	22861	8.636439192
5	14509	5.481216755	12500	4.7222558	10422	3.937227998

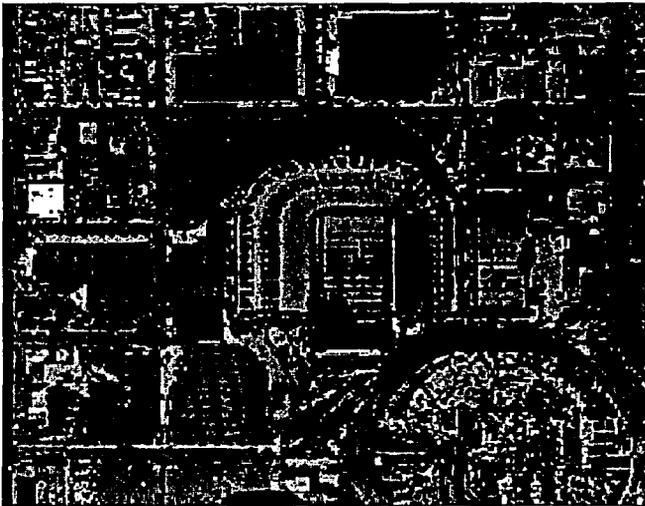
Resultados No. 4

La cuarta imagen a segmentar es:

4.a) lko.bmp



a)



b)



c)

Fig. No.17 a) Imagen original y b) Resultado del algoritmo EM - umbral y c) Resultado del algoritmo EM - bayesiano.

Capítulo No.6 Segmentación contextual no supervisada.

4.b) De los resultados del algoritmo de umbral, la tabla no.15 presenta el valor en niveles de gris de la clase, el número de píxeles que pertenece a tal clase y el porcentaje que representan para la imagen original.

Tabla de resultados No.15.

Valor	No. de elementos de la imagen	Porcentaje
20	38452	14.6682739
63	81977	31.2717438
105	95807	36.5474701
151	43478	16.5855408
238	2430	0.92697144

De los resultados del algoritmo del clasificador óptimo bayesiano, la tabla no.16 presenta el valor en niveles de gris de la clase, el número de píxeles que pertenece a tal clase y el porcentaje que representan para la imagen original.

Tabla de resultados No. 16.

Valor	No. de elementos de la imagen	Porcentaje
20	54866	20.929718
63	93864	35.8062744
105	91169	34.7782135
151	20619	7.86552429
238	1626	0.62026978

4.c) Se realizará una comparación numérica de los datos obtenidos al realizar los dos tipos de segmentación. En la tabla no. 17 se observa la diferencia en porcentaje de los dos métodos con respecto al número total de píxeles de la imagen. Y los resultados son los siguientes:

Tabla No. 17.

No. de clase	Porcentaje de desviación
1	6.26144409
2	4.53453064
3	1.76925659
4	8.72001648
5	0.30670166

4.d) Ahora se muestran ampliaciones de los resultados obtenidos.



a)



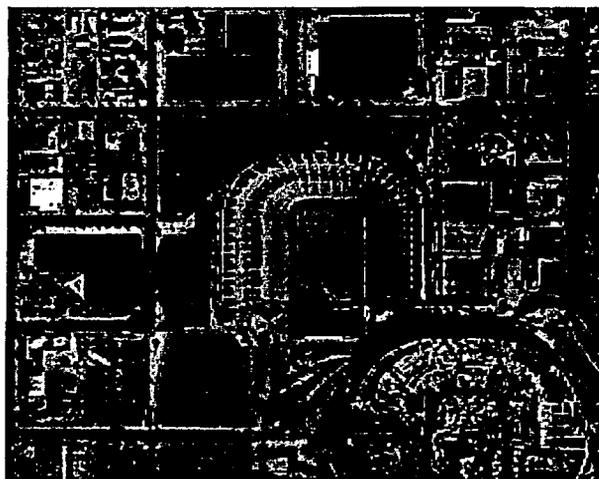
b)



c)

Figura No. 18. Comparación de los resultados de la segmentación empleando ventanas ampliadas de 80 X 120 píxeles. a) original b) umbral y c) bayesiana

4.e) Por ultimo se comparara el resultado obtenido contra uno obtenido por el algoritmo de k - medias.



a)



b)



c)



d)

Figura No.19. Resultados: a) Imagen original, b) Segmentada por k – medias, c) Segmentada por EM - umbral y d) Segmentada por EM - bayesiano.

Acercamiento de los resultados



a)



b)



c)



d)

Figura No. 20. Acercamiento de los resultados, a) Imagen original, b) Segmentada por k – medias, c) Segmentada por EM - umbral y d) Segmentada por EM - bayesiano.

Capítulo No.6 Segmentación contextual no supervisada.

En la tabla no. 18 se observan los valores de las medias calculadas a partir de k –media y del algoritmo EM, así como su diferencia porcentual.

Tabla no.18

Clase No.	Valor obtenido po k - medias	Valor obtenido por EM	Diferencia porcentual
1	34	20	41.17647059
2	76	63	17.10526316
3	108	105	2.777777778
4	142	151	6.338028169
5	194	238	22.68041237

Y en la figura No.21 se muestran las medias obtenidas por los algoritmos presentados

Medias obtenidas por EM —————

Medias obtenidas por k – medias —————>

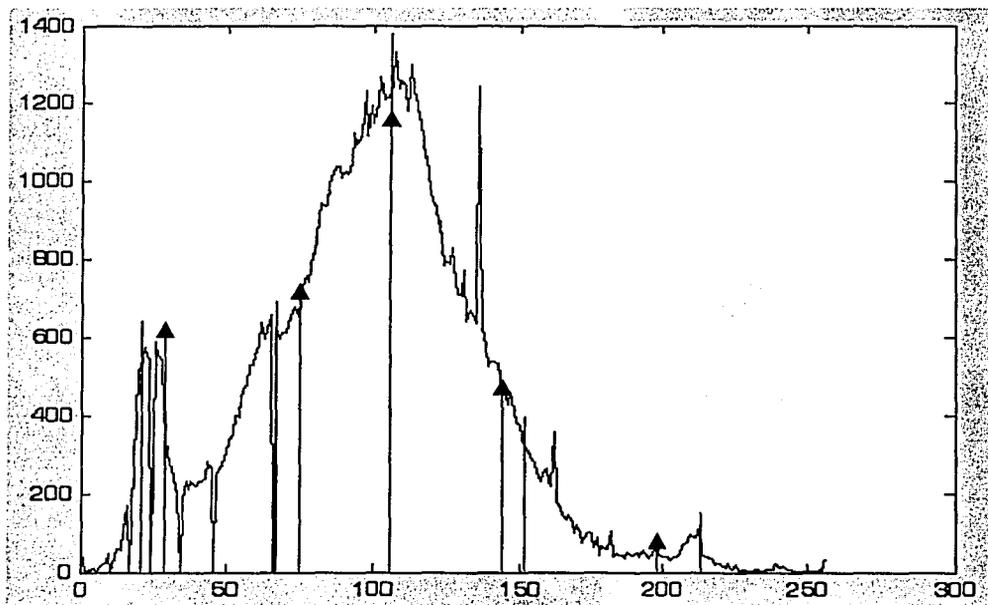


Figura No. 22 Histograma de la imagen iko.bmp

Capítulo No.6 Segmentación contextual no supervisada.

Tabla no. 19 Comparación de los píxeles segmentados por los tres métodos que se presentaron.

No. de Clase	Método de k - medias		Método de EM - umbral		Método de EM - bayesiano	
	No. de elementos	Porcentaje	No. de elementos	Porcentaje	No. de elementos	Porcentaje
1	60511	22.85987367	38452	14.5264144	54866	20.72730295
2	80697	30.48575012	81977	30.9693091	93864	35.45998549
3	69818	26.37587645	95807	36.1940129	91169	34.44186714
4	44467	16.7987639	43478	16.425139	20619	7.789455392
5	6651	2.512617868	2430	0.91800653	1626	0.614271035

Conclusiones

Como se puede observar, el resultado de la segmentación (en calidad) depende de que tipo de algoritmo de segmentación se utilice, pero lo importante de este capítulo es ver los resultados de una clasificación no supervisada, pues estos son comparados con los resultados del capítulo anterior, estos serán muy parecidos, pero la gran ventaja de estos, es que el resultado de la segmentación no depende de los puntos o número de clases que se tomen, si no que el algoritmo EM decide que medias serán utilizadas para realizar una segmentación, esto tiene una gran ventaja, pues el observador o investigador no tiene que proporcionarle datos como: número de clases y medias al algoritmo cada que se desee segmentar una imagen. Lo anterior conduce a un procesamiento automático y de gran ventaja en el caso de procesamiento de secuencias de imágenes o de imágenes adquiridas con alta cadencia.

Conclusiones generales

La percepción remota es un dominio de gran actividad de investigación. Esta área nos permite un gran campo de desarrollo de procesos y algoritmos para realizar la segmentación de las imágenes.

Un apartado dentro de la segmentación de imágenes, es el desarrollo de algoritmos, técnicas y procesos para realizar una segmentación no supervisada. Y es que este tipo de segmentación es una herramienta muy poderosa en la clasificación de imágenes, no sólo de satélites, si no también en la de imágenes biomédicas y en la visión de robots. Ya que la importancia de este tipo de segmentación, es que con este tipo de procedimiento, se pueden ir sensando los cambios, pero no sólo de una imagen, si no de un conjunto cronológico de ellas, y eso es una buena ventaja, pues permite encontrar errores o cambios en un conjunto de imágenes temporales.

Por estas razones se realizó el presente trabajo, de donde las conclusiones más importantes son las siguientes:

Al presentar los resultados de la segmentación supervisada, podemos observar que en los algoritmos con esta característica se debe de especificar el número de clases (por ejemplo k - medias y DN), y en otros, además se deben de dar los puntos de donde se deben de obtener las medias (segmentación Bayesiana y recocido simulado), mediante ventanas de entrenamiento para después hacer la segmentación.

En los algoritmos supervisados, los resultados dependen fuertemente de los datos de entrenamiento y del número de clases en las que se decida hacer la segmentación, es decir el resultado depende del usuario, de la forma en que introduzca la información y de cómo analice los resultados.

Tal dependencia ocasiona que estos algoritmos no puedan ser aplicados a estaciones de monitoreo remotas, pues a estos se les tiene que indicar ciertas característica para su funcionamiento. Y estas características son canalizadas por el operador o investigador según su experiencia y apreciación visual de los fenómenos reflejados en las imágenes de percepción remota.

Conclusiones Generales

Claramente, el resultado no debe de ser ambiguo, y al contrario, debe de ser lo más exacto posible.

Es por esto, que se realizaron estudios para obtener información de cuales eran las principales formas de obtener el número de medias y su valor para hacer una segmentación, a partir de un algoritmo llamado algoritmo EM (expectation- maximization).

Una de las características del algoritmo EM es que converge hacia los valores reales de los parámetros si los datos no están muy dispersos, pero cuando los datos lo están, el algoritmo converge sobre los picos que presenta el mismo histograma, presentando "nuevas" medias y desviaciones estándar, las cuales se detectan en los datos generados por medio de la distribución gaussiana.

Aunque da resultados muy exactos al calcular las medias preponderantes en la imagen, tiene un error mayor en el cálculo de las desviaciones estándar, pero esto se puede explicar de la siguiente manera: al momento de que los datos están más dispersos, y como el algoritmo calcula los parámetros (las medias y las desviaciones estándar) basándose en los picos del histograma, era de esperarse que la dispersión de los datos no pueda registrarse de forma muy confiable, pues esta no se puede observar en los picos del histograma.

También el algoritmo EM tiene otra característica : la generación de medias y desviaciones "nuevas", no necesariamente representan un error, si no lo que nos indica es que, los datos sintéticos, posiblemente fueron generados de tal forma que en un momento se sumaron cierto numero de ellos produciendo en el histograma un pico nuevo, que al ser procesado por el algoritmo fue encontrado. De esta forma, cuando se tenía un histograma muy denso en sus valores, se encontraban valores nuevos de medias, las cuales significaban picos en el mismo y por ende fueron encontrados.

En simulación con datos sintéticos, al observar el histograma, uno se da cuenta que en el se observan picos nuevos que no deberían (en teoría de aparecer) y el algoritmo EM los localiza dejando muy en claro que si aparece un pico nuevo, el algoritmo lo detectará y dará su valor.

Al hacer referencia en la presentación de los resultados, se tiene en claro que esta presentación visual tiene que ver con que algoritmo de segmentación se escoja para presentarlo, es por eso que los mismo resultados se dejan a la consideración del lector, pero logramos con este trabajo disminuir la ambigüedad de la presentación de los mismo, pues ahora en la presentación de resultados no se habla del algoritmo X con el número de clases Y y las medias Z, sino que ahora sólo se hace referencia al algoritmo EM y a su segmentación por el algoritmo X. En otra palabras ahora se pide como quieres el adorno del reloj, y no como quieres que se construya y se presente.

Conclusiones Generales

Además se puede asegurar que, el resultado de la segmentación (en calidad) depende de que tipo de algoritmo de segmentación se utilice, pero lo importante de este trabajo es ver los resultados de una clasificación no supervisada, pues estos son comparados con los resultados de la segmentación supervisada, y son muy parecidos, pero la gran ventaja de estos, es que el resultado de la segmentación no depende de las zonas de entrenamiento o del número de clases que se tomen, si no que el algoritmo EM decide que medias serán utilizadas para realizar una segmentación, esto tiene una gran ventaja, pues el observador o investigador no tiene que proporcionarle datos como: número de clases y medias al algoritmo cada que se desee segmentar una imagen, y más si esta es una secuencia de imágenes.

Es por eso que pienso que la herramienta que se está dando en esta tesis, es muy aplicable a cualquier trabajo posterior que tome como tema la segmentación de imágenes, pues ahora ya se tiene un procedimiento que da respuesta a ciertas preguntas, siendo estas las primeras que se abordan cuando se va a realizar una segmentación:

- ¿ En cuantas clases se va a segmentar la imagen?
- ¿Cuál es el valor de las media?
- ¿ Qué ventanas se toman para obtener el valor de las medias?
- ¿ Utilizar o no el valor de las desviaciones estándar de las ventanas?

Y que una vez respondidas, las investigaciones se pueden enfocar al desarrollo de algoritmos para generar un resultado visual mucho mejor. Es por eso que espero que este trabajo sirva de base para estudios posteriores. Ya que este campo tiene un infinito número de caminos en donde uno puede centrarse para hacer una investigación.

GRACIAS

Glosario

Algoritmo EM: Algoritmo Expectation - Maximization. Este procedimiento obtiene los valores de medias y desviaciones estándar de un cúmulo de datos, gracias a un proceso iterativo.

Algoritmo SEM: Algoritmo Stochastic Expectation – Maximization. Este procedimiento no sólo obtiene el valor de las medias y de las desviaciones estándar de un cúmulo de datos, si no que también obtiene el número de clases en las que se pueden etiquetar los datos.

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

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

Conectividad: Relación geométrica que un píxel establece con sus vecinos inmediatos.

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

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

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

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

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

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

Igualación del histograma: Consiste en obtener una distribución uniforme de la función de densidad de probabilidad de los niveles de gris de una imagen.

Imagen: Una imagen digital es un arreglo bidimensional de enteros cuyos índices de filas y columnas denotan la posición de un punto, y el valor del punto corresponde a un nivel de gris.

Glosario

Laplaciano: Operador diferencial que aplicado a una imagen, genera una buena definición de bordes en cualquier dirección. Las zonas de cambios abruptos de intensidad se resaltan mientras que las planas desaparecen.

Mapa de contornos: Imagen en donde sólo se observan contornos.

MRF: Abreviación de Markov Random Fields, cuya traducción al español es: Campos Aleatorios de Markov.

Píxel: Elemento pictórico de la imagen.

Región: Conjunto de píxeles con propiedades similares, conectados entre sí.

Segmentación: Separación de una imagen en regiones disjuntas, cuyos píxeles mantienen un valor dentro de un rango.

Umbral: Valor de un píxel de gris utilizado como referencia para separar dos regiones.

Bibliografía

1. Barilla M.E, Escalante Boris, Franco J.A, Moctezuma M, y Parmiggiani F. **SAR Images Filtering and Segmentation: A Multiresolution and Contextual Approach.**
2. Bendel James C y Keller James, Krisnapuram Raghu and R. Pal Nikhil. **Fuzzy models and algorithms for pattern recognition and image processing.** Kluwer Academic Publishers.
3. C. Gonzáles Rafael y E. Woods Richard. **Tratamiento digital de imágenes.** Addison Wesley/Díaz de Santos. 1992.
4. Childers D.G. **Probability and Random Processes Using MATLAB: with applications to continuous and discrete time systems.** Irwin – McGraw – Hill, 1997.
5. Derr Alan, Gish Herbert, Ostendorf Mari, Roucos Salim. **Stochastic Segment Modelling Using The Estimate – Maximiza Algorithm.**
6. Franco J.A, Moctezuma M, y Parmiggiani F. **Aproximación Contextual a la Segmentación de Fotografías Aéreas.** SIARP 1999. La Habana, Cuba. 1999.
7. Franco J.A, Moctezuma M, y Parmiggiani F. **Remote Sensing Images Segmentation: A Mixed Approach Adapted for Fine Structure Detection.** SIARP 2000. Lisboa, Portugal, 2000.
8. G.W. Awcock y R. Thomas. **Applied Image Processing.** McGraw – Hill. 1995.
9. Gdalyahu Yoram, Weinshall Daphna and Werman Michael. **Stochastic Image Segmentation By Typical Cuts.**
10. Gose Earl y Johnsonbaugh Richard and Jost Steve. **Pattern Recognition and Image Analysis.** Prentice Hall PTR. 1996.
11. Hecht Eugene y Zajac Alfred. **Optica.** Addison Wesley Longman. 1998.
12. **Introductory remote sensing. Principies and Concepts.**
13. Kato, Zoltan. **Modélisations markoviennes multirésolutions en vision par ordinateur. Application á la segmentation d'images SPOT.** Thèse pour obtenir le titre de Docteur es Science. L'Université de Nice Sophia Antipolis, France. 1994.

Bibliografía y Referencias

14. Lagendijk Reginald L y Blemond John. **Iterative identification and restoration of images.** Kluwer Academic Publishers.
15. Li. S.Z. **Markov Random Field Modeling in Computer Vision.** Springer. Tokio 1995.
16. Marhic Nicole and Pieczynski Wojciech. **Estimation of Mixture And Unsupervised Segmentation Of Images.** IEEE. 1991.
17. Marhic Nicole, Masson Pascale and Pieczynski Wojciech. **Malange De Lois Et Segmentation Non Supervisee Des Donnees Spot. Statistique et Analyse des Données.** 1991.
18. Masson Pascale and Pieczynski Wojciech. **SEM Algorithm And Unsupervised Statistical Segmentation Of Satellitary Images.**
19. Moctezuma Flores Miguel. **Apuntes de señales aleatorias.** México 1998.
20. Moon Todd K. **The Expectation Maximization Algorithm.** IEEE Signal Processing Magazine. 1996.
21. Paul J. Gibson. **Routledge.** 2000.
22. Peebles P.Z. **Probability, Random Variables and Random Signal Principies.** McGraw – Hill.
23. Reinhard Klette y Piero Zamperoni. **Handbook of Image Processing Operators.** John Wiley and Sons 1996.
24. Robert A. Schowengerdt. **Remote sensing, Models and Methods for image processing.** Academic Press. 1997.
25. Russ, C. John. **The Image Processing Handbook.** CRC Press. 1995.
26. Samadani Ramin and F. Vesecky John. **A Finite Mixtures Algorithm For Finding Clases in Images.** IEEE, 1991.
27. Thomassi Lillesand and Ralph Kieffer. **Remote sensing and image interpretation** Edit John Wiley and Sons Inc. 1994.
28. Wang Yue and Tülay Adah. **Stochastic model based image analysis.** Signal processing for magnetic resonance imaging and spectroscopy 2000.

Referencias

29. Algoritmo EM

[http:// www.cs.brown.edu/research/ai/dynamics/tutorial/documents/Em.html](http://www.cs.brown.edu/research/ai/dynamics/tutorial/documents/Em.html).

31. Algoritmo del recocido simulado.

http://www.depi.itchihuahua.edu.mx/mirror/itch/academic/industrial/sabiticorita/_privat

32. Agencia Espacial Europea.

<http://www.esrin.esa.it/esport/esaCP/index.html>

33. Dispersión de Rayleigh.

<http://www.us.es/fisica/dispersionluz.htm>

34. Electromagnetismo.

<http://orbita.starmedia.com/~fortiz/ConceptosBasicos/tema03Espectro.htm>

35. Introducción al Algoritmo EM

<http://www.student.comp.brad.ac.uk/~ksichan/Intro.html>

36. Procesamiento de Imágenes de Radar.

http://sputnik.dpi.inpe.br/spring/usuario_spa/radar_in.htm

37. Satélites Spot.

http://ceos.cnes.fr:8100/cdrom98/ceos1/satellit/spotsys/english/ss_paylo.htm

38. Sensores.

http://info.uned.es/ca-bergara/spring/usuario_spa/sensorr.htm#sistemas

39. Spot

<http://www.spot.com>

40. Teledet

<http://www.teledet.com.uy/>

Referencias de las figuras de prueba.

Berlin45.bmp imagen de 256 niveles de gris y de un tamaño de 660 x 630 píxeles.

Imagen cedida por el Laboratorio de Procesamiento Digital de Imágenes, Facultad de Ingeniería de la UNAM.

Cacho.bmp imagen de 256 niveles de gris y de un tamaño de 512 x 512 píxeles.

Base de datos de fotos aéreas de alta resolución del Valle de la ciudad de México de la Facultad de Ingeniería.

Deportes.bmp imagen de 256 niveles de gris y de un tamaño de 512 x 512 píxeles.

Base de datos de fotos aéreas de alta resolución del Valle de la ciudad de México de la Facultad de Ingeniería.

Iko.bmp imagen de 256 niveles de gris y de un tamaño de 303 x 329 píxeles.

<http://snig.cnig.pt/ROT/rot2/ikonos.html>

Palacio.bmp imagen de 256 niveles de gris y de un tamaño de 512 x 512 píxeles.

<http://www.pgb.es/marbella/html/iinstalaciones.html>

Paris3.bmp imagen de 256 niveles de gris y de un tamaño de 512 x 512 píxeles.

Base de datos de fotos aéreas de alta resolución del Valle de la ciudad de México de la Facultad de Ingeniería.

Usar.bmp imagen de 256 niveles de gris y de un tamaño de 512 x 512 píxeles.

Base de datos de fotos aéreas de alta resolución del Valle de la ciudad de México de la Facultad de Ingeniería.

Anexo 1

1.- EL RECOCIDO SIMULADO (*SIMULATED ANNEALING*)

Los campos de Markov, además de ser un conjunto de metodologías para el análisis y la representación de imágenes, están también ligados a un método específico de optimización: el *recocido simulado*. Este algoritmo permite resolver el problema de estimación en el sentido del MAP. Muy importante además, el recocido simulado nos asegura (respetando sus fundamentos) la convergencia hacia un **óptimo global** del problema.

En la práctica se emplea un algoritmo de Metrópolis, con un esquema de temperatura variable, a fin de minimizar la función de energía *a posteriori* $U(Y|X)$. El objetivo del problema de optimización consiste en encontrar un elemento de Ω_0 donde:

$$\Omega_0 = (\omega \in \Omega, \text{ donde } \omega \text{ es de energía mínima})$$

1.1 PLANTEAMIENTO:

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

$$T_n \rightarrow 0 \text{ cuando } n \rightarrow \infty$$

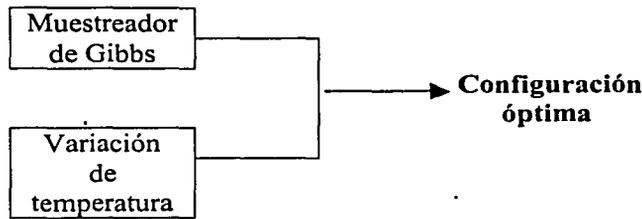
$$T_n \geq \frac{N \Delta U}{\log(n)} \text{ para } n \geq 2$$

donde N es el número de *sites* en la imagen

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

$$\lim_{n \rightarrow \infty} P(Y^{(n)} = \omega | Y^{(0)} = \eta) = \pi(\omega)$$

donde $\pi(\omega)$ corresponde a la medida de probabilidad de Gibbs definida por los campos de Markov. Así, existe una convergencia hacia una realización dada, con una probabilidad que sigue la distribución de Gibbs global.



1.2 ALGORITMO DE RECOCIDO SIMULADO

1.- Inicialización

- determinación aleatoria de la solución $Y^{(0)}$
- escoger una temperatura inicial $T^{(0)}$ suficientemente elevada

2.- Barrer secuencialmente la imagen

En cada etapa (n),

- escoger un site: $r = s^{(n)}$, su valor es $Y_r^{(n)}$ y su configuración de vecindad es $V_r^{(n)}$.
- efectuar el sorteo de una variable aleatoria con distribución uniforme: $\xi \in \Omega$.

3.- Criterio de Metrópolis

- calcular la variación de la energía local asociada al site r:

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

- dos eventualidades pueden surgir:

- si $\Delta U \leq 0$, $Y_r^{(n+1)} = \xi$

- si $\Delta U > 0$, $Y_r^{(n+1)} = \xi$ con probabilidad $\Pi(Y_r^{(n+1)} \rightarrow \xi) = \exp\left(-\frac{\Delta U}{T_n}\right)$

4.- Continuar con la etapa 2

5. Serie de temperaturas decrecientes hacia 0, por tanto: $\lim_{n \rightarrow \infty} T_n = 0$.

Comentarios:

- como precedentemente, $\Delta U = U(\xi | V_r^{(n)}, X) - U(Y_r^{(n)} | V_r^{(n)}, X)$ representa tanto la variación local de energía, como la variación global de la energía asociada a la transición $Y_r^{(n+1)} = \xi$.
- la convergencia del algoritmo hacia el mínimo global de la función de energía, independientemente de la configuración inicial ha sido probada en [3] cuando: $T^{(n)} > \frac{c}{\log(n+1)}$, para n suficientemente grande y donde c es una constante que depende de la expresión de la función de energía a minimizar, $U(Y | X)$.

Anexo 1

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

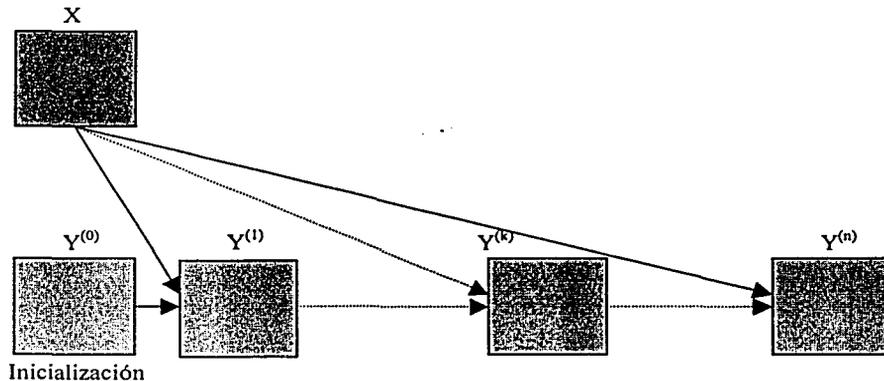
Entonces:

$$\begin{cases} \text{si } 0 \leq \eta \leq p, & Y_r^{(n+1)} = \xi \\ \text{en caso contrario, } & Y_r^{(n+1)} = Y_r^{(n)} \end{cases}$$

- en la práctica, los esquemas de descenso de temperatura son del tipo $T^{(n)} = \frac{c}{\log(n+1)}$, o incluso $T^{(n+1)} = \alpha \cdot T^{(n)}$, donde α es una constante, tal que $0 < \alpha < 1$.

2. SIMULACIÓN

Sea X la imagen original (en niveles de gris) e Y la imagen segmentada (por ejemplo, en cuatro clases), la simulación toma como datos iniciales a X y a la imagen de clases originada en la inicialización, $Y^{(0)}$. La figura siguiente muestra la esquematización del algoritmo de simulación, para el caso de n iteraciones:

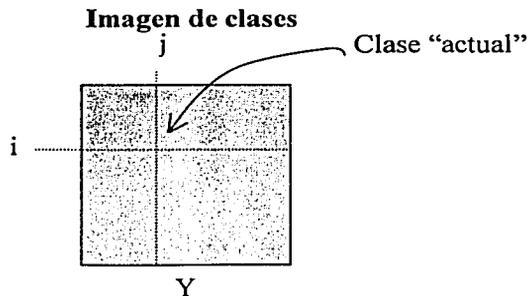


Para el caso de la energía $U(Y)$ se considera al modelo de Potts, el cual se aplica en vecindades en 8-conexidad.

2.1 ALGORITMO

do $N_{iter}=1, N_{iter}=n$

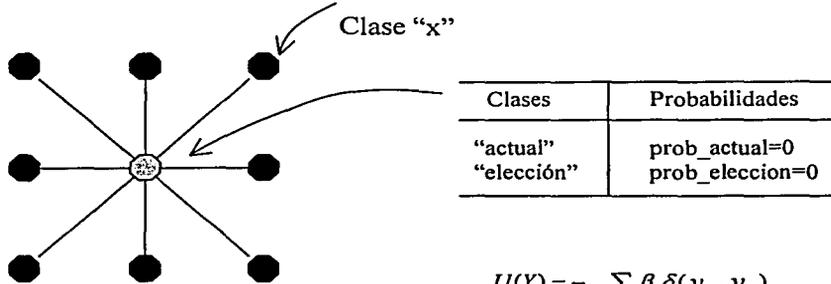
Barrido de la imagen:
(por líneas y por columnas)



Modelo de Potts:

Anexo 1

- Sorteo aleatorio de una clase \rightarrow "elección" en el rango $[0, 3]$



$$U(Y) = - \sum_{c1=\{g,h\}} \beta \delta(y_g, y_h)$$

Análisis en 8-conexidad:

- si "actual" = "x": $\text{prob_actual} = \text{prob_actual} - \beta[\text{actual}]$
- si "elección" = "x": $\text{prob_elección} = \text{prob_elección} - \beta[\text{elección}]$
- $\text{prob_actual} = \text{prob_actual} + \text{potencial}[\text{actual}][i][j]$
- $\text{prob_elección} = \text{prob_elección} + \text{potencial}[\text{elección}][i][j]$

donde el potencial $[\][\][\]$ representa la energía condicional $U(X|Y)$ e $[i][j]$ son las coordenadas del píxel analizado.

- si $\text{prob_elección} \leq \text{prob_actual}$: $Y[i][j] = \text{"elección"}$
en caso contrario
 - $p = \exp - [(\text{prob_elección} - \text{prob_actual}) / T^{(k)}]$
 - sorteo de una variable aleatoria ξ_2 , uniformemente distribuida en el rango $[0, 1]$
 - si $\xi_2 < P$: $Y[i][j] = \text{"elección"}$

Fin del barrido

Incrementar la temperatura:

$$T(n) = T(n-1) * 0.95$$

Fin do N_{iter}

3. Algoritmo de Umbrales.

Este algoritmo toma como un umbral a el valor de las media de cada clase, para hacer la segmentación se realiza lo siguiente:

- Se toma el píxel $p(i,j)$ a segmentar y se obtienen la distancia en valor absoluto con cada una de las medias que se tengan.
- El píxel toma el valor de la clase cuya distancia absoluta sea menor a las demas.