



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE INGENIERIA

“ESQUEMA ANISOTRÓPICO DE FILTRADO Y DE ANÁLISIS ESTOCÁSTICO”

TESIS

PARA OBTENER EL TÍTULO DE
“INGENIERO EN TELECOMUNICACIONES”

PRESENTA

JIMENEZ JIMENEZ JAMES ESTEBAN



DIRECTOR DE TESIS
DR. MIGUEL MOCTEZUMA FLORES

CIUDAD UNIVERSITARIA

2014



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.

Agradecimientos

En primer lugar quisiera agradecer a mis padres no solo por sus sabios consejos, sino también por todo el apoyo que me proporcionaron a lo largo de toda la carrera. Por estar a mi lado cuando más lo necesite y ayudarme a salir de momentos difíciles. Es por ellos que hoy estoy aquí. Muchísimas gracias.

A mis hermanas Jessica y Jezabel por estar ahí siempre que las necesite, por ser tan comprensivas y por creer en mí. Ahora espero sinceramente que sigan mi ejemplo y puedan terminar sus respectivas carreras y al igual que ustedes ahí estaré para apoyarlas.

A mis abuelos quienes siempre han estado a mi lado mientras atravesaba este largo y difícil camino, y que sin sus palabras de aliento no estaría escribiendo esto ahora mismo. Principalmente quisiera agradecer a mi abuela materna, quien desde que era pequeño siempre tuvo grandes expectativas en mí y nunca dudo que sería ingeniero.

Y finalmente al Dr. Miguel Moctezuma Flores por darme la oportunidad de realizar esta tesis, por todos sus conocimientos, su paciencia y su ayuda.

Índice

Introducción.....	3
1. Capítulo 1 Principios de probabilidad	
1.1 Axiomas de probabilidad.....	7
1.2 Variables aleatorias.....	8
1.3 Funciones de probabilidad de densidad y distribución.....	9
1.4 Distribución gaussiana.....	12
1.5 Regla de Bayes.....	15
2. Capítulo 2 Campos aleatorios de Markov	
2.1 Conceptos.....	17
2.2 Topologías.....	23
2.2.1 Vecindades.....	23
2.2.2 Cliques.....	24
2.2.3 Propiedades de Markov.....	24
2.3 Modelos de Markov.....	29
2.3.1 Modelo de Potts.....	32
3. Capítulo 3 Filtrado anisotrópico	
3.1 Antecedentes.....	34
3.2 Esquemas de filtrado lineal y no lineal.....	36
3.3 Ecuación diferencial de difusión de calor.....	45
3.4 Esquemas de filtrado anisotrópico.....	47
3.5 Discretización de la ecuación diferencial.....	50
3.6 Representaciones escala-espacio y sus parámetros asociados...52	
4. Capítulo 4 Optimización por Campos de Markov	
4.1 Fundamentos.....	56
4.2 Muestreador de Gibbs.....	57
4.3 Esquema de Metrópolis.....	58
4.4 Recocido simulado.....	59
4.5 Parámetros: vecindades, cliques y temperatura.....	62
5. Capítulo 5 Esquema estocástico	
5.1 Planteamiento y esquema general.....	64
5.2 Funciones de energía.....	70
5.3 Minimización.....	74
6. Resultados	
6.1 Imágenes de prueba.....	77
6.2 Comparación.....	85
7. Conclusiones.....	87
Bibliografía.....	89

Introducción

Objetivos

- Con fundamentos probabilísticos, definir un esquema para el procesamiento de datos de radar de apertura sintética.
- Diseñar y aplicar un esquema estocástico de segmentación para la detección de derrames petroleros.

Fundamentación

El fundamento teórico reside en métodos no lineales expresados mediante ecuaciones diferenciales, la teoría de Bayes y la maximización de probabilidades *a posteriori*. Como aplicación, se pretende la detección de zonas del derrame petrolero provocado por la plataforma petrolera DeepWater Horizon en el Golfo de México, ocurrido en la primavera de 2010. Se cuenta con un banco de imágenes del satélite ENVISAT/ASAR de alta resolución. Se emplearán esquemas típicos de segmentación puntual a fin de realizar comparaciones en la funcionalidad de los resultados del esquema propuesto.

Definición del problema

En el presente tema de tesis, se pretende el estudio de técnicas estocásticas para la segmentación de imágenes adquiridas por el satélite ENVISAT. Las interacciones de la atmosfera con la superficie del océano y la degradación producida por señales altamente coherentes dificultan enormemente el empleo y la aplicación de las señales de radar. El material biológico también puede provocar las falsas detecciones llamadas look-alikes es por eso que hay la necesidad de implementar varios algoritmos para así poder disminuir estas falsas detecciones y el ruido que pudiera contener la imagen del satélite. En este caso se segmenta la imagen además de usar principalmente dos algoritmos la difusión anisotrópica y el recocido simulado, ambos algoritmos implementados para obtener una mejor calidad en la detección de zonas importante.

A continuación una breve descripción de la segmentación de imágenes, la difusión anisotrópica y el recocido simulado.

La segmentación de imágenes es el proceso más importante en un gran número de aplicaciones. Se trata de particionar la imagen en diferentes regiones significativas con características homogéneas usando las discontinuidades o similitudes de los componentes de la imagen, analizando los píxeles vecinos y de esa manera hallar cierta similitud entre píxeles y de esa manera poder agruparlos en “clases”.

La segmentación de imágenes es un proceso muy efectivo para la detección de las similitudes en las imágenes, sin embargo existen diversos factores que afectan la calidad y la fidelidad de la imagen, introduciendo ruido a la imagen haciendo más difícil el procesamiento y la partición de la imagen, es por este motivo que hay la necesidad de aplicar más de un algoritmo para asegurar la correcta detección de las regiones significativas de la imagen. En la presente tesis se usa dos algoritmos que ayudan a reducir el ruido presente en la imagen. Estos algoritmos son la difusión anisotrópica y el recocido simulado, además de la segmentación de la imagen en clases.

Difusión anisotrópica

En el procesamiento de imágenes, la difusión anisotrópica, también llamada difusión Perona-Malik, es una técnica destinada a reducir el ruido de la imagen sin eliminar partes importantes del contenido de la imagen, por lo general los bordes, líneas u otros detalles que son importantes para la interpretación de la imagen. La difusión anisotrópica se asemeja al proceso que crea un espacio de escala, donde una imagen genera una serie de imágenes parametrizadas de forma sucesiva cada vez más borrosas con base un proceso de difusión. Cada una de las imágenes resultantes en esta familia se dan como una convolución entre la imagen y un filtro gaussiano isotrópico. Este proceso de difusión es una transformación lineal y el espacio invariante de la imagen original. La difusión anisotrópica es una generalización de este proceso de difusión, se produce una serie de imágenes con parámetros, pero cada imagen resultante es una combinación de la imagen original y un filtro que depende del contenido local de la imagen original. Como consecuencia de ello, la difusión anisotrópica es una transformación no lineal y el espacio variante de la imagen original.

Aunque la serie de imágenes resultantes pueden ser descritas como una combinación entre la imagen original y los filtros espacio variante, el filtro adaptado localmente y su combinación con la imagen no tiene que ser realizado en la práctica.

La difusión anisotrópica se implementa normalmente por medio de una aproximación de la ecuación de difusión generalizada, cada imagen nueva se calcula mediante la aplicación de esta ecuación a la imagen anterior.

Por lo consiguiente, la difusión anisotrópica es un proceso iterativo en el que se utiliza un conjunto de cálculo relativamente simple para calcular cada imagen sucesiva y este proceso continua hasta que un grado de suavizado suficiente se obtiene.

Recocido Simulado

El método de Simulated Annealing (SA) o recocido simulado es un algoritmo probabilístico que permite resolver problemas de optimización. Se basa en una analogía de un proceso termodinámico, conocido como *annealing*, consistente en el lento enfriamiento de una sustancia desde su estado líquido hasta su estado sólido. Durante este proceso sus átomos forman un compuesto sólido puro que corresponde al estado de mínima energía de la sustancia. Luego, la sustancia es calentada hasta alcanzar su estado líquido y gradualmente su temperatura es disminuida hasta alcanzar su estado sólido. La analogía entonces se realiza interpretando una solución del problema de optimización como un estado sólido-liquido en equilibrio, entendiendo que la energía que está siendo minimizada en el proceso termodinámico, corresponde con la función objetivo del problema a resolver.

El procedimiento esquematizado busca una optimización reduciendo gradualmente el valor de la función objetivo. A partir de una solución actual se busca una solución vecina a la que siempre se acepta en el caso de ser mejor y se acepta eventualmente en el caso de ser peor que la actual respetando una función de probabilidad definidas en este caso la función de probabilidad de Boltzmann. El algoritmo cuenta con dos ciclos, en el primer ciclo se va disminuyendo la temperatura, mientras que en el segundo, se realiza una búsqueda local aleatoria a una temperatura dada.

Después de haber dado una introducción de este trabajo, a continuación se presenta una breve explicación de los temas cubiertos en cada uno de los capítulos.

En el primer capítulo se exponen las herramientas matemáticas, en donde se incluyen definiciones de términos y expresiones matemáticas necesarias que se usarán a lo largo del desarrollo de la tesis.

En el segundo capítulo se presentan la explicación teórica de los campos de Markov.

En el tercer capítulo se explica los antecedentes necesarios para poder entender el funcionamiento y la aplicación del algoritmo de recocido simulado.

En el cuarto capítulo se presenta formalmente el algoritmo recocido simulado, sus parámetros y su funcionamiento.

En el quinto capítulo se expone de forma clara el problema a resolver.

Y finalmente, en el sexto capítulo se presenta la implementación y los resultados arrojados por los algoritmos.

1. PRINCIPIOS DE PROBABILIDAD

1.1 AXIOMAS DE PROBABILIDAD

Existen dos maneras diferentes de definir probabilidad. La primera se basa en la teoría de conjuntos y los axiomas fundamentales; este método es el más conocido de los dos y quizá es más complejo de entender desde el punto de vista práctico. El segundo se fundamenta principalmente en el método de la *frecuencia relativa*, que se basa más en el sentido común y las observaciones científicas y de ingeniería y por lo tanto es más fácil de entender. Sin embargo este capítulo se enfoca en el primer modo y de ese modo poder entender los axiomas de probabilidad.

Si definimos que a cada suceso sobre un espacio muestral le asignamos un número no negativo denominado *probabilidad*. Por lo tanto, la probabilidad es función de una serie de sucesos definidos. Podemos adoptar la notación $P(A)$ para definir la “probabilidad del suceso A”. Esta probabilidad del suceso A debe cumplir con ciertas condiciones para asegurar que este valor de probabilidad sea válido.

Sea A cualquier suceso definido sobre un espacio muestral S. Entonces los dos primeros axiomas son:

$$\text{Axioma 1:} \quad P(A) \geq 0 \quad (1.1-1a)$$

$$\text{Axioma 2:} \quad P(S) = 1 \quad (1.1-2b)$$

El primero nos asegura que la probabilidad del suceso A va a ser mayor a cero y por lo tanto un número positivo. El segundo axioma nos dice que el propio espacio muestral es un suceso y, dado que es el suceso que comprende a todos, debe tener la probabilidad más alta posible, que se selecciona como la unidad. Por esta razón S se conoce como el *suceso seguro*. Alternativamente, el conjunto nulo \emptyset es un suceso con ningún elemento; se conoce como el *suceso imposible* y su probabilidad es 0.

El tercer axioma se aplica a N sucesos A_n , $n = 1, 2, \dots, N$, donde N posiblemente sea infinito, definido sobre un espacio muestral S y teniendo la propiedad $A_m \cap A_n = \emptyset$ para todo $m \neq n$. Es decir,

$$\text{Axioma 3:} \quad P\left(\bigcup_{n=1}^N A_n\right) = \sum_{n=1}^N P(A_n) \quad (1.1-3c)$$

El tercer axioma establece que la probabilidad del suceso definido como la unión de cualquier número de sucesos mutuamente excluyentes es igual a la suma de las probabilidades de cada suceso.

1.2 VARIABLES ALEATORIAS

Se define una *variable aleatoria* como una *función* real de los elementos de un espacio muestral \mathcal{S} . Una variable aleatoria se representa mediante una letra mayúscula y cualquier valor concreto de dicha variable aleatoria mediante una letra minúscula. Una variable aleatoria X puede considerarse una función que aplica todos los elementos del espacio muestral a puntos de la recta real o a algunas partes de ella.

Por lo tanto, una variable aleatoria puede ser prácticamente cualquier función que se desee. Sin embargo, se requiere que no sea multivaluada. Es decir, todo punto de \mathcal{S} debe corresponderse sólo con un valor de la variable aleatoria.

Ahora bien se deben cumplir con dos condiciones importantes para que una función X sea una variable aleatoria. En primer lugar, el conjunto $\{X \leq x\}$ será un suceso para cualquier número real x . La probabilidad de este suceso, expresada como $P\{X \leq x\}$, es igual a la suma de las probabilidades de todos los sucesos elementales correspondientes a $\{X \leq x\}$.

La segunda condición es que las probabilidades de los sucesos $\{X = \infty\}$ y $\{X = -\infty\}$ sean igual a 0, es decir, $P\{X = -\infty\} = 0$ y $P\{X = \infty\} = 0$.

Sin embargo esta condición no impide que X sea $-\infty$ o ∞ para algunos valores de s ; sólo requiere que la probabilidad del conjunto de dichos valores s sea cero.

Variable aleatoria discreta

Una *variable aleatoria discreta* es aquella que sólo toma valores discretos. El espacio muestral para una variable aleatoria discreta puede ser discreto, continuo, o incluso una mezcla de puntos discretos y continuos.

Variable aleatoria continua

Una *variable aleatoria continua* es aquella que tiene un rango continuo de valores. No se puede obtener a partir de un espacio muestral discreto debido a que se

requiere que todas las variables aleatorias sean funciones univaluadas para todos los puntos del espacio muestral.

Variable aleatoria mixta

Una *variable aleatoria mixta* es aquella para la que algunos de sus valores son discretos y algunos son continuos. Habitualmente, el caso mixto es el tipo menos importante de variable aleatoria, aunque aparece en algunos problemas de gran importancia práctica.

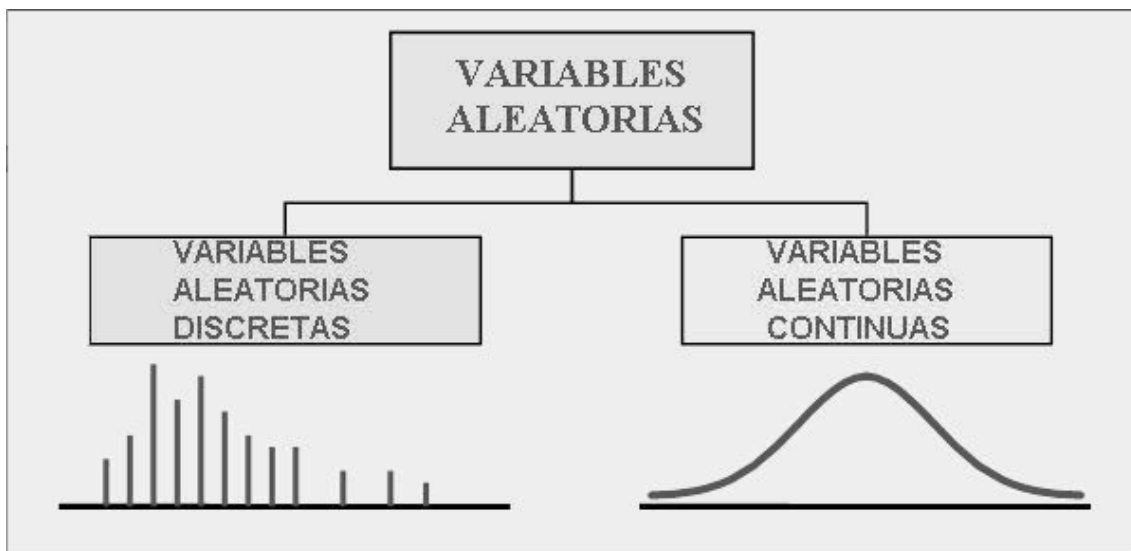


Figura 1.2-1 Esquema de división de las variables aleatorias.

1.3 FUNCIONES DE PROBABILIDAD DE DENSIDAD Y DE DISTRIBUCIÓN

Función de distribución

La probabilidad $P\{X \leq x\}$ es la probabilidad del suceso $\{X \leq x\}$. Es un número que depende de x ; es decir, es una función de x . A esta función se le denomina $F_X(x)$, *función de distribución de probabilidad acumulativa* de la variable aleatoria X .

Por lo tanto,

$$F_X(x) = P\{X \leq x\}$$

$$(1.3-1)$$

Es muy común llamar a $F_X(x)$ simplemente como función de distribución de X . En donde el argumento x es cualquier número real perteneciente al rango de $-\infty$ a ∞ .

La función de distribución tiene algunas propiedades específicas derivadas del hecho de que $F_X(x)$ es una probabilidad. Son las siguientes:

$$(1) F_X(-\infty) = 0 \quad (1.3-2a)$$

$$(2) F_X(\infty) = 1 \quad (1.3-2b)$$

$$(3) 0 \leq F_X(x) \leq 1 \text{ para todo } x \in \mathbb{R} \quad (1.3-2c)$$

$$(4) F_X(x_1) \leq F_X(x_2) \text{ para } x_1 < x_2 \quad (1.3-2d)$$

$$(5) P\{x_1 < X \leq x_2\} = F_X(x_2) - F_X(x_1) \quad (1.3-2e)$$

$$(6) F_X(x^+) = F_X(x) \quad (1.3-2f)$$

Las primeras tres propiedades son fáciles de justificar. La primera nos dice que si tomamos el límite de x cuando x tiende a menos infinito se obtendrá como resultado 0, es decir:

$$\lim_{x \rightarrow -\infty} F_X(x) = \lim_{x \rightarrow -\infty} F_X(-\infty) = 0 \quad (1.3-3)$$

La segunda esta muy relacionada con la primera excepto que ahora se toma el límite cuando x tiende a infinito, es decir:

$$\lim_{x \rightarrow \infty} F_X(x) = \lim_{x \rightarrow \infty} F_X(\infty) = 1 \quad (1.3-4)$$

La tercera propiedad establece que la función de distribución acumulada únicamente puede tomar valores entre cero y uno.

La cuarta propiedad establece que $F_X(x)$ es una función no decreciente de x .

La quinta propiedad establece que la probabilidad de que X tenga valores mayores que determinado número de x_1 pero menores o iguales a otro número x_2 es igual a la diferencia de $F_X(x)$ calculada para dichos dos puntos. Esto se justifica a partir del hecho de que los sucesos $\{X \leq x_1\}$ y $\{x_1 < X \leq x_2\}$ son mutuamente excluyentes, por lo que la probabilidad del

suceso $\{X \leq x_2\} = \{X \leq x_1\} \cup \{x_1 < X \leq x_2\}$ es la suma de las probabilidades $P\{X \leq x_1\}$ y $P\{x_1 < X \leq x_2\}$.

La sexta propiedad establece que $F_x(x)$ es una función continua por la derecha, es decir:

$$\lim_{h \rightarrow 0^+} F_x(x+h) = F_x(x) \quad \text{o} \quad \lim_{x \rightarrow x_0^+} F_x(x) = F_x(x_0)$$

(1.3-5)

Función de densidad

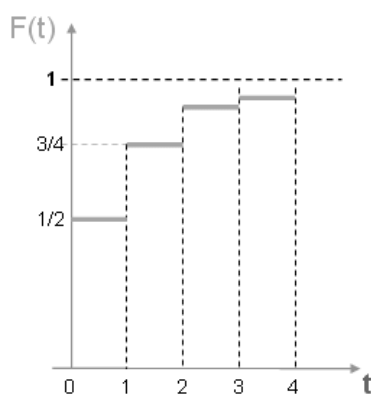
La *función de densidad de probabilidad*, $f_x(x)$, se define como la derivada de la función de distribución:

$$f_x(x) = \frac{dF_x(x)}{dx}$$

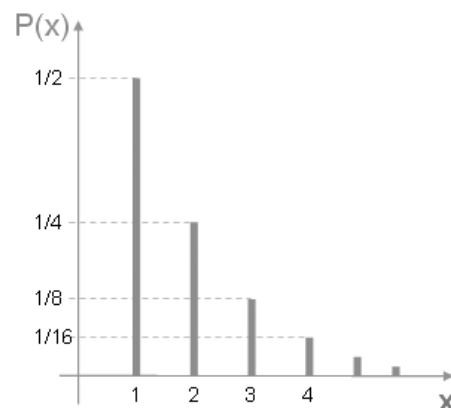
(1.3-6)

A menudo se denomina a $f_x(x)$ simplemente función de densidad de la variable x .

Si la derivada de $F_x(x)$ existe, entonces $f_x(x)$ existe y está dada por la expresión de arriba. Sin embargo, pueden existir puntos en los que $dF_x(x)/dx$ no esté definida. Por ejemplo, una variable aleatoria continua tiene una distribución continua $F_x(x)$, pero $F_x(x)$ presenta esquinas (puntos de cambio abrupto en su pendiente). En dichos casos, se dibuja $f_x(x)$ como una función con discontinuidades de tipo escalón.



Función de distribución



Distribución de probabilidades

Propiedades de la función de densidad

Las distintas propiedades que $f_x(x)$ satisface son las siguientes:

$$(1) \quad 0 \leq f_x(x) \quad \text{para} \quad \text{todo} \quad x$$

(1.3-7a)

$$(2) \quad \int_{-\infty}^{\infty} f_x(x) dx = 1$$

(1.3-7b)

$$(3) \quad F_x(x) = \int_{-\infty}^x f_x(\xi) d\xi$$

(1.3-7c)

$$(4) \quad P\{x_1 < X \leq x_2\} = \int_{x_1}^{x_2} f_x(x) dx$$

(1.3-7d)

La primera propiedad nos indica que la función de densidad no sea negativa, tomando el hecho de que una probabilidad no puede ser un valor negativo.

La segunda propiedad nos dice que el área bajo la curva de la función de densidad es uno, tomando el hecho de que la probabilidad máxima que podemos obtener es uno.

La tercera propiedad sirve para relacionar $F_x(x)$ y $f_x(x)$, simplemente es otro modo de escribir la expresión (1.3-6).

Finalmente la cuarta propiedad relaciona la probabilidad de que X tome valores desde x_1 a x_2 (inclusive) con la función de densidad.

1.4 DISTRIBUCIÓN GAUSSLANA

Se dice que una variable aleatoria X es *gaussiana* si su función de densidad tiene la forma

$$f_x(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-(x-\mu)^2/2\sigma_x^2}$$

(1.4-1)

Donde $\sigma_x > 0$ y $-\infty < \mu < \infty$ son constantes reales. Esta función se ha dibujado en la figura de abajo. Su valor máximo $(2\pi\sigma_x^2)^{-1/2}$ se produce en $x = \mu$.

Su “amplitud” alrededor del punto $x = \mu$ esta relacionada con σ_x .

La función decrece a 0.607 veces su máximo en $x = \mu + \sigma_x$ y $x = \mu - \sigma_x$

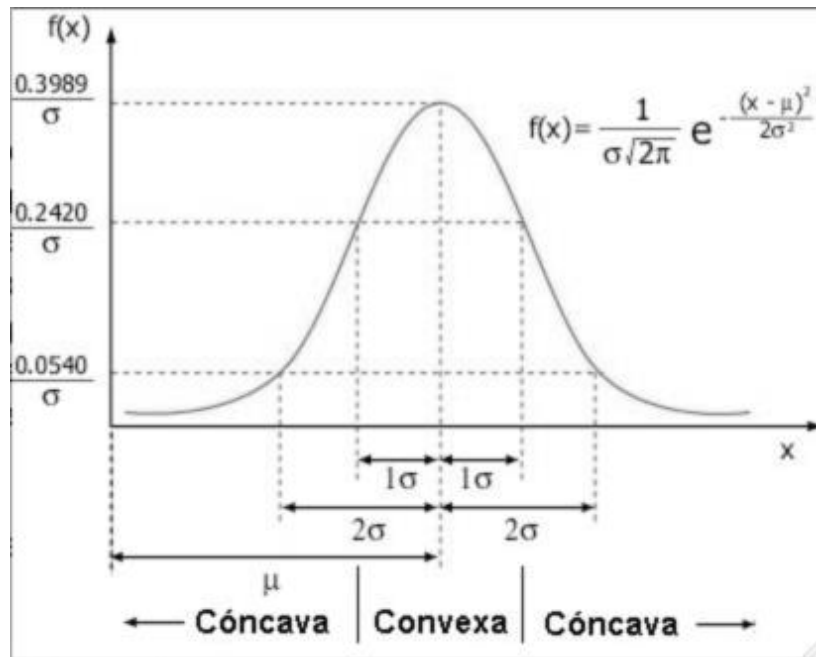


Figura 1.4-1 Función de densidad de una variable aleatoria gaussiana.

La densidad gaussiana es la más importante de todas las densidades y se emplea prácticamente en todas las áreas de las ciencias y la ingeniería. Esta importancia radica en que permite describir de forma precisa muchas magnitudes importantes del mundo real, especialmente cuando dichas magnitudes son el resultado de pequeños efectos aleatorios, que se combinan para crear la magnitud de interés. Por ejemplo, la tensión en una resistencia a la salida de un amplificador puede ser aleatoria debido a una corriente aleatoria que es el resultado de la contribución de otras corrientes aleatorias que se producen en el amplificador, etc.

La función de distribución se obtiene a partir de (1.3-7c) usando (1.4-1). La integral es:

$$F_x(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \int_{-\infty}^x e^{-(\xi-\mu)^2/2\sigma_x^2} d\xi$$

(1.4-2)

La integral no tiene una solución cerrada conocida y debe evaluarse por métodos numéricos o aproximaciones. Para hacer que los resultados estén disponibles de forma general, se puede desarrollar un conjunto de tablas de $F_x(x)$ para distintos valores de x con μ y σ_x como parámetros. Sin embargo, este método tiene valor limitado porque existe un número infinito de posibles combinaciones de μ y σ_x , lo que requiere un número infinito de tablas.

Es por este motivo que es necesario emplear otro método. La función $F_x(x)$ también puede evaluarse mediante aproximaciones. En primer lugar, escribimos la función como:

$$F_x(x) = 1 - Q(x)$$

(1.4-3)

En donde:

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\xi^2/2} d\xi$$

(1.4-4)

A esta función se le conoce como función Q se obtiene tomando en cuenta que $\mu = 0$ y $\sigma_x = 1$ para hacer más fácil la ecuación (1.4-2). Y al igual que $F_x(x)$, $Q(x)$ no tiene solución cerrada conocida.

Pero existe una excelente aproximación dada por:

$$Q(x) \approx \left[\frac{1}{(1-a)x + a\sqrt{x^2+b}} \right] \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

(1.4-5)

donde a y b son constantes y $x \geq 0$.

Se ha demostrado que esta aproximación proporciona un error relativo absoluto mínimo para cualquier $x \geq 0$, cuando $a=0.339$ y $b=5.510$. Con estos valores de a y b , la aproximación es igual al valor verdadero de $Q(x)$ con un error absoluto máximo del 0.27% de $Q(x)$.

1.5 REGLA DE BAYES

Para poder describir la regla de Bayes es necesario primero definir algunos conceptos

Eventos Mutuamente Excluyentes

Dos eventos A y B son mutuamente excluyentes si no pueden ocurrir al mismo tiempo.

$$P(A \cup B) = P(A) + P(B) \quad \text{si } P(A \cap B) = 0$$

(1.5-1)

Definición de probabilidad condicional

Se llama probabilidad condicional a la probabilidad de que un suceso se cumpla habiéndose cumplido ya otro. Se denota “la probabilidad de A sabiendo que B se ha cumplido”. Dicha probabilidad se calculará de la siguiente forma:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

(1.5-2)

Entonces a ecuación (1.5-2) puede escribirse para una serie de sucesos como:

$$P(B_n|A) = \frac{P(B_n \cap A)}{P(A)}$$

(1.5-3)

$$P(A|B_n) = \frac{P(A \cap B_n)}{P(B)}$$

$$(1.5-4)$$

Definición de probabilidad total

Sean dos sucesos que cumplen que:

- Ambos sucesos son mutuamente excluyentes.

$$B_m \cap B_n = \emptyset \quad (1.5-5)$$

- a unión de todos ellos es un suceso seguro. L

$$\bigcup_{n=1}^N B_n = S$$

$$(1.5-6)$$

Sea B_1, B_2, \dots, B_n un sistema completo de sucesos tales que la probabilidad de cada uno de ellos es distinta de cero, y sea A un suceso cualquier del que se conocen las probabilidades condicionales $P(A|B_n)$, entonces la probabilidad de suceso A viene dada por la siguiente expresión:

$$P(A) = \sum_{n=1}^N P(A|B_n) P(B_n)$$

$$(1.5-7)$$

El teorema de Bayes se obtiene igualando las ecuaciones 1.5-3 y la 1.5-4 y despejando $P(B_n|A)$:

$$P(B_n|A) = \frac{P(A|B_n)P(B_n)}{P(A)}$$

$$(1.5-8)$$

La otra forma es obtiene sustituyendo $P(A)$ con la expresión dada por la ecuación 2.5-7

$$P(B_n|A) = \frac{P(A|B_n)P(B_n)}{P(A|B_1)P(B_1) + \dots + P(A|B_N)P(B_N)}$$

(1.5-9)

2. CAMPOS ALEATORIOS DE MARKOV

2.1 CONCEPTOS

La segmentación es un proceso importante en una visión artificial, ya que divide una escena en un conjunto de regiones disjuntas, (de manera que se puedan discriminar unas regiones de otras), basándose para ello en características de similitud. Es el primer nivel en la tarea de interpretación de la imagen y afecta significativamente procesos posteriores.

Los algoritmos de segmentación generalmente están basados en la consideración de dos criterios: la homogeneidad de la región y la discontinuidad entre regiones disjuntas adyacentes. Los campos aleatorios de Markov (CAM) tienen una gran capacidad para integrar grandes cantidades de información visual; han sido aplicados a problemas tales como restauración y segmentación de imágenes.

El procesamiento de datos en el sistema puede enfocarse desde dos puntos de vista:

- Alteración píxel a píxel de los datos en una escala global (operaciones individuales).
- O operaciones basadas en múltiples puntos (transformaciones de vecindad)

Operaciones individuales

El proceso consiste en obtener el valor del píxel de una locación dada en la imagen, modificarlo con una operación lineal o no lineal, y colocar el valor del nuevo píxel en la correspondiente localización de la nueva imagen. Conocido como operador umbral, crea una imagen de salida binaria a partir de una escala de grises, donde el nivel de transición está dado por el parámetro de entrada P_1 . La función de transformación es la siguiente:

$$q = \begin{cases} 0 & \text{si } p \leq p_1 \\ 255 & \text{si } p > p_1 \end{cases}$$

(2.1-1)

Transformaciones de vecindad

Son las transformaciones para modificar una imagen de entrada, bien en otra imagen distinta, o bien en una matriz de valores transformados, que no constituyen realmente una imagen, pero contienen la información necesaria que se espera de la transformación. Se dice que todo píxel p de coordenadas (x, y) , de vecindad 4 tiene cuatro píxeles que establecen con él la siguiente relación: horizontal: $(x - 1, y)$ y $(x + 1, y)$, y verticalmente: $(x, y - 1)$ y $(x, y + 1)$; naturalmente, existen las excepciones dadas cuando el píxel (x, y) es un punto de la frontera o esquina de la imagen, en cuyo caso algunos de los vecinos definidos anteriormente no existen.

Como se vio anteriormente las operaciones individuales son más complejas que las transformaciones no solo por la cantidad de operaciones que se debe realizar para transformar la imagen, sino también por las operaciones que deben de aplicársele a la imagen ya sea una operación lineal o no lineal. Es por este motivo que nos enfocaremos únicamente a las operaciones de vecindad.

Operaciones de Vecindad

Consisten básicamente, en transformar el valor de un píxel p en la posición (x, y) , teniendo en cuenta los valores de los píxeles vecinos. Por ejemplo, si se considera una vecindad $E_8(p)$, se realiza una suma ponderada (un promedio) con los valores de los ocho vecinos, y el resultado de dicha suma es el valor del nuevo píxel q de la imagen de salida en la misma posición (x, y) . Por último, se establecen los valores de ponderación, definiendo una *máscara* con valores constantes. Dicha máscara es realmente un filtro, por lo que, dependiendo de su naturaleza, así será el resultado final. Por ejemplo, si se define la siguiente máscara:

$$\begin{bmatrix} 1.0 & 2.0 & 1.0 \\ 0.0 & 1.2 & 0.0 \\ -1.0 & -2.0 & -1.0 \end{bmatrix}$$

El valor del píxel $q(x, y)$ vendría dado por la siguiente suma ponderada, con los factores de ponderación definidos por la máscara:

$$q(x, y) = 1 * p(x - 1, y - 1) + 2 * p(x, y - 1) + 1 * p(x + 1, y - 1) + 0 * p(x - 1, y) + 1.2 * (x, y) + 0 * p(x + 1, y) - 1 * p(x - 1, y + 1) - 2 * p(x, y + 1) - 1 * p(x + 1, y + 1)$$

(2.1-2)

Realizando esta transformación sobre todos los píxeles de la imagen original con la misma máscara, se obtendrá una nueva imagen de salida cuya dimensión es inferior a la original.

La figura (2.1-2) muestra el resultado de aplicar la operación dada en la máscara sobre la imagen original de la figura (2.1-1), con la máscara dada por la expresión (2.1-2). Como se puede apreciar, el resultado es una especie de repujado en el relieve donde se marcan los bordes respecto del resto de la imagen.



Figura 2.1-1 Imagen original

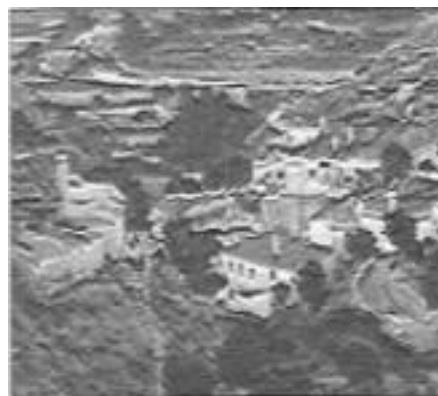


Figura 2.1-2 Aplicación de la máscara a la imagen original

Definición de Campos Aleatorios de Markov

Formalmente los Campos Aleatorios de Markov se pueden definir de la siguiente manera:

⇒ Sea S un conjunto finito con N estados y sea $G = \{G_s, s \in S\}$ el conjunto de vecinos S , es decir:

1. $s \in G, \forall s \in S$
2. $s \in G$, si y solo si $r \in G, \forall r, s \in S$

⇒ Sea $F = \{F_s, s \in S\}$ una familia de variables aleatorias indexadas por $s \in S$, que toman valores en algún conjunto finito $\{Q_s\}$.

⇒ Se llamará configuración a toda realización posible $f: (f_{s_1}, \dots, f_{s_N}), f_{s_i} \in Q_{s_i}$.

⇒ Sea Ω el conjunto de todas las configuraciones posibles y sea P una medida de probabilidad en Ω .

⇒ F es un MRF con respecto a G si:

1. $P(F = f) > 0, \forall f \in \Omega$ ($F = f$) denota el evento $(F_s = f_s \forall s \in S)$.
2. $P(F_s = f_s | F_r = f_r, r \neq s) = P(F_s = f_s | F_r = f_r, r \in G_s)$ para algún $s \in S$.

Campos aleatorios de Markov (CAM) para segmentación de imágenes

Los CAM se fundamentan en procesos de decisión estocásticos con aplicaciones en un gran número de disciplinas en especial medicina y robótica;

proporcionan una manera conveniente y constante de modelar entidades dependientes del contexto, tales como píxeles de la imagen y otras características especiales correlacionadas; son muy eficientes por las siguientes razones:

- Hacen más fácil el cálculo de medidas de ejecución cuantitativas para caracterizar cómo trabajan los algoritmos de análisis de imagen
 - Los modelos de CAM pueden ser usados para incorporar información contextual a priori, o bien restricciones de manera cuantitativa
-
- Los algoritmos basados en CAM tienden a ser locales y conducen a una implementación de hardware paralelo de manera natural.

Definición de proceso estocástico: es un concepto matemático que sirve para caracterizar y estudiar todo tipo de fenómenos aleatorios (estocásticos) que evolucionan, generalmente, con el tiempo. Sirven por lo general para realizar grandes series de muestreo, reducen el tiempo de procesamiento en las computadoras y son muy utilizados en investigaciones científicas.

Problema del etiquetado

El problema de etiquetado consiste en asignar una etiqueta del conjunto de etiquetas L a cada estado de S , y su solución consiste en la asignación de un conjunto de etiquetas a los píxeles (estados) de la imagen. Sea S un conjunto discreto de valores con m píxeles:

$$S = \{1, 2, \dots, m\} \quad (2.1-3)$$

Un *estado* o “*site*” representa un punto o una región del espacio euclidiano, tal como un píxel en una imagen o una característica como una línea o una superficie.

El conjunto de estados puede ser clasificado en términos de su homogeneidad o regularidad. Se puede denotar al conjunto de localizaciones de los píxeles de una imagen 2D de tamaño $n \times n$ como:

$$S = \{(i, j) \mid 1 \leq i, j \leq n\} \quad (2.1-4)$$

El conjunto de etiquetas, L puede ser un conjunto discreto o continuo. En el caso continuo, el conjunto de etiquetas puede corresponder a \mathbb{R} o a un intervalo compacto de \mathbb{R} .

$$L_c = [x_1, x_n] \subset \mathbb{R} \quad (2.1-5)$$

Una etiqueta es un evento que tal vez ocurre a un estado. Para el caso discreto, las etiquetas se toman dentro de un conjunto discreto de M valores.

$$L_d = [I_1, \dots, I_M] = \{1 \dots M\} \quad (2.1-6)$$

El conjunto de etiquetas puede ser caracterizado en términos de su continuidad. Además de la continuidad, otra característica esencial de un conjunto es su orden. En un conjunto discreto como, por ejemplo, la intensidad del tono de gris (0,1,...,255), podemos, ordenar los elementos según los valores de intensidad.

Ordenar las etiquetas no solamente es importante para clasificar problemas de etiquetado, sino que afecta también a la elección de algoritmos de etiquetado y por lo tanto a la complejidad computacional.

Se llamará etiquetado a:

$$f = \{f_1, \dots, f_m\} \quad (2.1-7)$$

Donde a cada píxel de S se le asigna una única etiqueta de L con la función mapping o de mapeo:

$$f : S \rightarrow L \quad (2.1-8)$$

En términos de CAM, se llama configuración al etiquetado de esta forma, una configuración de etiquetas puede corresponder a una imagen, a un mapa de borde de la imagen, etc.

Cuando todos los estados tienen el mismo conjunto de etiquetas Λ , el conjunto de todas las configuraciones posibles, es el producto cartesiano:

$$F = \Lambda * \Lambda * \dots * \Lambda = \Lambda^m \quad (2.1-9)$$

En donde m es el tamaño de S .

En ciertas circunstancias, el conjunto de etiquetas puede no ser común a todos los estados. En un caso extremo, cada estado i puede tener su propio conjunto de etiquetas posibles, A_i y esto nos lleva al siguiente espacio de configuraciones:

$$F = A_1 * A_2 * \dots * A_m \quad (2.1-10)$$

2.2 TOPOLOGIAS

Las topologías de los CAM describen las relaciones contextuales entre estados, para posteriormente servir como base y establecer distribuciones de probabilidad.

2.2.1 Vecindades

Los estados de S están relacionados mediante lo que se conoce como un sistema de vecinos $N = \{N_i | \forall i \in S\}$, donde N_i es el conjunto de los estados vecinos de $i \in S$.

Para una rejilla regular S , el conjunto de vecinos de i está definido como el conjunto de estados próximos dentro de un radio r , es decir,

$$N_{i,j} = \{(i,j) \in S | [dist(pixel_i, pixel_j)]^2 \leq r, i \neq j\} \quad (2.2-1)$$

En donde $dist(pixel_i, pixel_j)$ es la distancia euclidiana entre el $pixel_i$ y el $pixel_j$ y r es un número entero.

Para cada estado interior, representado por (i,j) , de vecindad 4 está definido como:

$$N_{i,j} = \{(i-1,j), (i+1,j), (i,j-1), (i,j+1)\} \quad (2.2-2)$$

Para una vecindad 8 se tendría lo siguiente:

$$N_{i,j} = \left\{ \begin{array}{l} (i-1, j), (i+1, j), (i, j-1), (i, j+1), (i+1, j+1), \\ (i-1, j-1), (i+1, j-1), (i-1, j+1) \end{array} \right\} \quad (2.2-3)$$

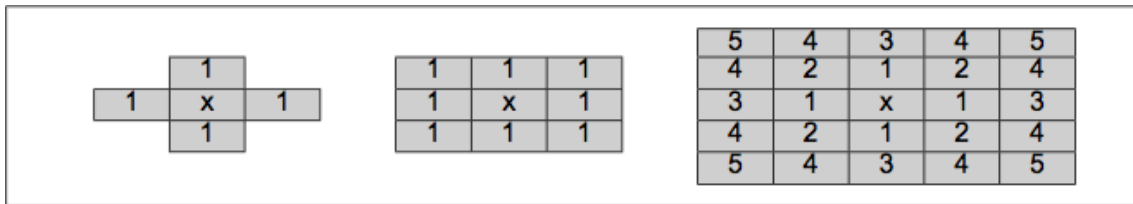


Figura 2.2.1-1 Esquema de vecindad 2,4 y 8

2.2.2 Cliques

Un *clique* C para $\{S, N\}$ se define como un subconjunto de S tal que C consiste en un único estado $C = \{i\}$ o un par de estados $C = \{i, j\}$, o tres $C = \{i, j, k\}$ y así sucesivamente.

El tipo de clique para $\{S, N\}$ de una rejilla regular viene determinado por su tamaño, dimensión y orientación. La figura 2.2.2-1 muestra distintos tipos de cliques para sistemas de vecinos de primer y segundo orden. A medida que aumenta el orden del sistema de vecinos, el número de cliques crece rápidamente y también el procesamiento requerido.

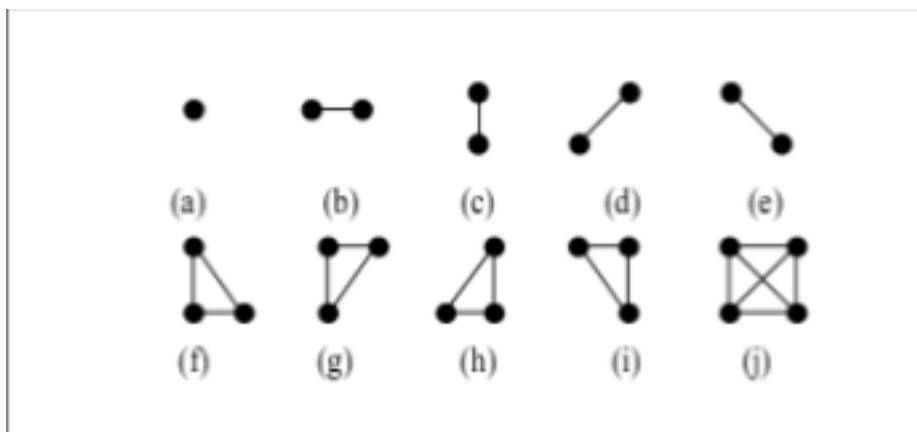


Figura 2.2.2-1 Cliques para vecindad 4 y 8.

2.2.3 Propiedades de los Campos aleatorios de Markov

Campos Aleatorios de Markov

Sea $F = \{F_1, \dots, F_m\}$ una familia de variables aleatorias definidas en S , donde cada variable F_i toma un valor de f_i en L . Se llamará campo aleatorio a la familia de variables aleatorias F . Se usará la notación $F_i = f_i$, para denotar que el suceso F_i toma el valor de f_i .

Para un conjunto discreto de etiquetas L , se denota por $P(F_i = f_i) = P(f_i)$ a la probabilidad de que la variable F_i tome el valor de f_i y se denota la probabilidad conjunta por $P(F = f) = P(F_1 = f_1, \dots, F_m = f_m) = P(f)$.

Una familia de variables aleatorias F se dice que es un CAM sobre S , con respecto al sistema de vecindad N , si y solo si se cumple que:

$$1. \quad P(F = f) > 0 \quad (2.2-3)$$

$$2. \quad P(f_i | f_{S-\{i\}}) = P(f_i | f_N) \quad (2.2-4)$$

(Propiedad Markoviana)

En donde $f_{S-\{i\}}$ denota al conjunto de etiquetas en todos los estados exceptuando a $\{i\}$ y $f_{N_i} = \{f_{i'} | i' \in N_i\}$ es el conjunto de etiquetas que pertenecen a los estados de la vecindad de i .

La condición 1 es para exigir que F sea un campo aleatorio, y es llamada positividad. La condición 2 es llama característica local. Se dice que la probabilidad de un suceso local a L condicionando a todos los sucesos restantes es equivalente a aquella condicionada a los sucesos vecinos de L .

Se define un Campo Aleatorio de Markov discreto sobre una trama finita como una colección de variables aleatorias, que corresponden a píxeles de una rejilla, cuya distribución de probabilidad es tal que la probabilidad condicional de que una variable determinada tenga un valor particular, dado los valores del resto de las variables, es idéntica a la probabilidad condicional dados los valores del campo en un conjunto pequeño de píxeles, que son los vecinos de ese píxel determinado.

Un Campo Aleatorio de Markov puede tener otras características tales como la homogeneidad e isotropía. Se dice que es homogéneo si $P(f_i|f_{N_i})$ se calcula si importar la posición relativa del estado i en S .

Hay dos aproximaciones para especificar un Campo Aleatorio de Markov, en términos de probabilidades condicionales $P(f_i|f_{N_i})$ y en términos de probabilidad conjunta $P(f)$. Esta aproximación mediante la probabilidad conjunta se discute en vista de las desventajas de la aproximación mediante la probabilidad condicional. Afortunadamente, un resultado teórico sobre la equivalencia entre los Campos Aleatorios de Markov y la distribución de probabilidad de Gibbs proporciona medios de especificar la probabilidad conjunta de un Campo Aleatorio de Markov matemáticamente manejables.

Campos Aleatorios de Gibbs

Se dice que un conjunto de variables aleatorias F es un campo aleatorio de Gibbs en S , con respecto a N (sistema de vecinos), si y solo si sus configuraciones siguen una distribución de Gibbs. Una distribución de Gibbs tiene la siguiente función de densidad:

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

(2.2-5)

Donde,

- U(f)= función de energía
- Z=constante de normalización
- T=temperatura

La constante de normalización se define como:

$$Z = \sum_f e^{-\frac{1}{T}U(f)}$$

(2.2-6)

Mientras que la función de energía se define como:

$$U(f) = \sum_c V_c(f)$$

(2.2-7)

C es el clique y $V_c(f)$ es la función potencial del clique

$U(f)$ es la suma de todas las funciones cliques potenciales, $V_c(f)$, para todos los cliques posibles C , donde el valor de $V_c(f)$ depende de la configuración local del clique c .

Se dice que un campo aleatorio de Gibbs es homogéneo si $V_c(f)$ es independiente de la posición relativa del clique c en S . Es isotrópico si la posición de V_c es independiente de la orientación de c .

$P(f)$ mide la probabilidad de la ocurrencia de una configuración determinada. Las configuraciones más probables son aquellas con energías más bajas. La temperatura T controla la forma de la distribución. Cuando la temperatura es alta, todas las configuraciones tienden a estar igualmente distribuidas. Con la temperatura cercana a cero, la distribución se concentra alrededor del mínimo global de energía.

La energía de una distribución de Gibbs se puede expresar como la suma de varios términos correspondientes a cliques de diferentes tamaños, es decir:

$$U(f) = \sum_{c \in \mathcal{C}} V_c(f) = \sum_{\{i\} \in \mathcal{C}} V_1(f_i) + \sum_{\{i,j\} \in \mathcal{C}} V_2(f_i, f_j) + \sum_{\{i,j,k\} \in \mathcal{C}} V_3(f_i, f_j, f_k) + \dots$$

(2.2-8)

La ecuación anterior implica una distribución homogénea de Gibbs porque V_1, V_2, \dots , son independientes de las localizaciones de i, j . En las distribuciones no homogéneas de Gibbs, las funciones clique se deben considerar de la forma: $V_1(i, f_i), V_2(i, j, f_i, f_j)$, etc.

Un caso importante es cuando se consideran los cliques de tamaño dos como máximo, entonces

$$U(f) = \sum_{i \in S} V_1(f_i) + \sum_{i \in S} \sum_{j \in N_i} V_2(f_i, f_j)$$

(2.2-9)

La función de energía sería expresada como la suma de dos funciones de potencial de clique.

La equivalencia Markov-Gibbs

Un Campo Aleatorio de Markov se caracteriza por sus propiedades locales (propiedades Markovianas) mientras que un Campo Aleatorio de Gibbs se caracteriza por sus propiedades globales (la distribución de probabilidad de Gibbs). El teorema de Hammersley-Clifford establece la equivalencia entre estos dos tipos de características.

Teorema de Hammersley-Clifford: F es un Campo Aleatorio de Markov en S con respecto a N si y sólo si F es un Campo de Gibbs en S con respecto a N.

Sea $P(f)$ una distribución de Gibbs en S con respecto al sistema de vecinos N. Consideremos la probabilidad condicional

$$P(f_i | f_{N_i}) = \frac{P(f_i, f_{N_i})}{P(f_{N_i})} = \frac{P(f)}{\sum_{f'_i \in L} P(f')}$$

(2.2-10)

Donde $f' = f_1, \dots, f_{i-1}, f'_i, \dots, f_m$ es cualquier configuración que contenga todos los estados de f menos, posiblemente, el i.

Sustituyendo $P(f) = \frac{1}{Z} e^{-\sum_{c \in C} V_c(f)}$ lo cual también proporciona una fórmula para calcular la probabilidad condicional de las funciones potenciales, tenemos que:

$$P(f_i | f_{N_i}) = \frac{e^{-\sum_{c \in C} V_c(f)}}{\sum_{f'_i \in L} e^{-\sum_{c \in C} V_c(f)}}$$

(2.2-11)

Dividimos L en dos conjuntos A y B de manera que A es el conjunto de cliques que contienen a i y B el conjunto de cliques que no contienen a i . Entonces podemos escribir:

$$P(f_i | f_{N_i}) = \frac{[e^{-\sum_{c \in A} V_c(f)}][e^{-\sum_{c \in B} V_c(f)}]}{\sum_{f'_i \in L} [e^{-\sum_{c \in A} V_c(f')}][e^{-\sum_{c \in B} V_c(f')}]}$$

(2.2-12)

Como $V_c(f) = V_c(f')$ para cualquier clique c que no contenga a i , $e^{-\sum_{c \in B} V_c(f)}$ se simplifica en el numerador y en el denominador. Por lo tanto, esta probabilidad depende solamente de los potenciales de los cliques que contienen i ,

$$P(f_i | f_{N_i}) = \frac{[e^{-\sum_{c \in A} V_c(f)}]}{\sum_{f'_i \in L} [e^{-\sum_{c \in A} V_c(f')}]}$$

(2.2-13)

Es decir, depende de la etiqueta en los vecinos de i . Esto prueba que un campo aleatorio de Gibbs es un campo aleatorio de Markov.

El valor práctico del teorema es que proporciona una manera simple de especificar la probabilidad conjunta. Se puede definir la probabilidad. Se puede definir la probabilidad conjunta $P(F=f)$ especificando las funciones clique potenciales $V_c(f)$ y eligiendo la funciones potenciales apropiadas para el conocimiento deseado del sistema. De esta manera, se codifica el conocimiento a priori sobre las interacciones entre las etiquetas.

Para calcular la probabilidad conjunta de un Campo Aleatorio de Markov, que tiene una distribución de Gibbs, es necesario evaluar Z . Dado que la suma es un número combinatorio de configuraciones, el costo computacional es muy elevado. La evaluación explícita se puede evitar mediante la probabilidad máxima en modelos de visión en Campos Aleatorios de Markov cuando $U(f)$ no tiene parámetros desconocidos.

Sin embargo, esto no es verdad cuando la estimación de los parámetros es también una parte del problema. En el último caso, la función de energía $U(f) = U(f|\theta)$ es también función de los parámetros θ , así también $Z = Z(\theta)$. Para evitar la gran dificultad planteada, en la práctica se calcula una aproximación de la probabilidad conjunta.

2.3 MODELOS DE MARKOV

A continuación se definen algunos modelos útiles de CAM para modelar características de la imagen tales como regiones y texturas. El interés principal se centra en sus distribuciones de probabilidad condicional y conjunta, y en sus correspondientes funciones de energía. Dependiendo de cómo esta de definida la función de energía se pueden agrupar en:

- ✓ Modelos Auto-
- ✓ Logístico Multinivel Modelo

Auto modelos

Las restricciones de contexto entre dos etiquetas son un primer nivel a tener en cuenta en la información del contexto. Son muy utilizadas debido a su simplicidad y bajo costo computacional. Se codifican en la función de energía de Gibbs como funciones clique potenciales para cada par de estados.

Entonces la función clique potencial queda definida por:

$$U(f) = \sum_{i \in S} V_1(f_i) + \sum_{i \in S} \sum_{j \in N_i} V_2(f_i, f_j)$$

(2.3-1)

A la cual se le puede llamar energía de segundo orden porque se calcula para cada clique formada por dos estados. Se puede especificar un Campo Aleatorio de Markov o un Campo Aleatorio de Gibbs mediante la selección apropiada de V_1 y V_2 .

Cuando $V_1(f_i) = f_i G_i(f_i)$ y $V_2(f_i, f_j) = \beta_{i,j} f_i f_j$, donde $G_i(*)$ son funciones arbitrarias y $\beta_{i,j}$ son constantes que reflejan la interacción entre cada par de estados i y j , entonces la función de energía es:

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

(2.3-2)

El modelo anterior recibe el nombre de auto-modelo. Los auto-modelos se pueden clasificar más a fondo según las superposiciones hechas sobre las f_j .

Modelo Logístico Multinivel

El modelo auto-logístico se puede generalizar al modelo logístico multinivel (MLL), también llamado modelo de Strauss. Hay $M (> 2)$ etiquetas discretas en el conjunto de etiquetas $L = \{1, \dots, M\}$. Un clique potencial depende del tipo c (tamaño, dimensión y orientación posible) del clique y de la configuración local $f_c = \{f_i \mid i \in c\}$. Para cliques que contiene más de un estado ($c > 1$), se definen las funciones cliques potenciales MLL como;

$$V_c(f) = \begin{cases} \zeta_c & \text{si todos los estados de } c \text{ tiene la misma etiqueta} \\ -\zeta_c & \text{en otro caso} \end{cases}$$

En donde ζ_c es el potencial para los cliques del tipo c ; para cliques de un solo estado, depende de la etiqueta asignada al estado.

$$V_c(f) = V_c(f_i) = \alpha_i \text{ si } f_i = l \in L_d$$

(2.3-3)

Donde α_i es el potencial para la etiqueta l .

Si se asume que un modelo MLL es de segundo orden, si y sólo si α (para cliques de un estado) y β (para cliques de dos estados) son parámetros distintos de cero. La función clique potencial para cliques de dos estados es:

$$V_2(f_i, f_j) = \begin{cases} \beta_c & \text{si los estados del clique } \{i, j\} = c \in C_2 \text{ tienen la misma etiqueta} \\ -\beta_c & \text{en otro caso} \end{cases}$$

Donde β_c es el parámetro β para los cliques del tipo c y C_2 es el conjunto de cliques de dos estados. Para un sistema de cuatro vecinos hay cuatro tipos de clique de dos estados y cuatro diferentes β_c . Cuando el modelo es isotrópico los cuatro toman el mismo valor. Cuando el modelo MLL es isotrópico, la probabilidad condicional puede ser expresada como:

$$P\left(f_i = \frac{l}{f_{N_i}}\right) = \frac{e^{-\alpha_i - \beta_{ni}(l)}}{\sum_{l=1}^M e^{-\alpha_i - \beta_{ni}(l)}} \quad (2.3-4)$$

Donde $n_i(l)$ es el número de estados en N_i que están etiquetados por l . Cuando solamente hay dos etiquetas, 0 y 1, se reduce a:

$$P(f_i | f_{N_i}) = \frac{e^{\alpha_i f_i + \sum_{j \in N_i} \beta_{i,j} f_i f_j}}{\sum_{f_i \in \{0,1\}} e^{\alpha_i f_i + \sum_{j \in N_i} \beta_{i,j} f_i f_j}} = \frac{e^{\alpha_i f_i + \sum_{j \in N_i} \beta_{i,j} f_i f_j}}{1 + e^{\alpha_i + \sum_{j \in N_i} \beta_{i,j} f_j}} \quad (2.3-5)$$

2.3.1 Modelo de Potts

En caso de que se quisiera segmentar una imagen es escala de grises, es posible modelar las interacciones entre los píxeles a efecto de segmentar en varias clases. La finalidad del modelo de Potts es asociar a cada píxel una etiqueta de un conjunto finito. Si se denota X a la imagen de datos originales y por Y la imagen segmentada, el proceso de segmentación consiste en maximizar la probabilidad condicional $P(X|Y)$, que de acuerdo a la regla de Bayes se tiene $P(Y|X) = P(X|Y)P(Y)$.

$P(Y)$ está referido como el modelo a priori mientras que $P(X|Y)$ está referido como el término de manejo de datos. En estudios prácticos, el modelado probabilístico se basa en el análisis de energía de los cliques. Esto produce funciones de energía relacionadas a la función potencial. La función de energía asociada con este modelo es la siguiente:

$$U(Y) = - \sum_{\{i,j\} \in C} \beta \delta_{y_i y_j} \quad (2.3-1)$$

Los cliques tienen dos elementos, definiendo interacciones estrechamente pares. El coeficiente β define la probabilidad de homogeneidad de la solución, esto es, el mayor de esos términos, el más probable de los píxeles adyacentes que pueden tener la misma etiqueta.

Los términos de manejo de datos están definidos por las funciones de costo dependiendo de la etiqueta l y denotada por f^l . Los parámetros inducidos son directamente extraídos de los datos.

El potencial asociado a los cliques de primer orden se define como:

$$U(X|Y) = \sum_{c=\{i\}} \sum_l f^{-1}(x^i) \partial_{y^i} = 1$$

(2.3-2)

Donde x_i e y_i son los datos y los valores de etiquetas respectivamente en el sitio i .

El problema de segmentación consiste en minimizar la energía global:

$$U(X|Y) + U(Y)$$

(2.3-3)

La estructura del MAP, el proceso de minimización está desarrollado por una técnica estocástica como el Recocido Simulado (convergiendo próximamente con un mínimo global). Remarcando la minimización puede incluso ser desarrollada por técnicas determinísticas como los Modos de Iteración Condicional (rápido, pero converge solo aproximándose al mínimo local).

3. FILTRADO ANISOTRÓPICO

3.1 ANTECEDENTES

Las operaciones espaciales o de vecindad se definen en un entorno “EN” (vecindad) del punto a transformar (m_0, n_0)

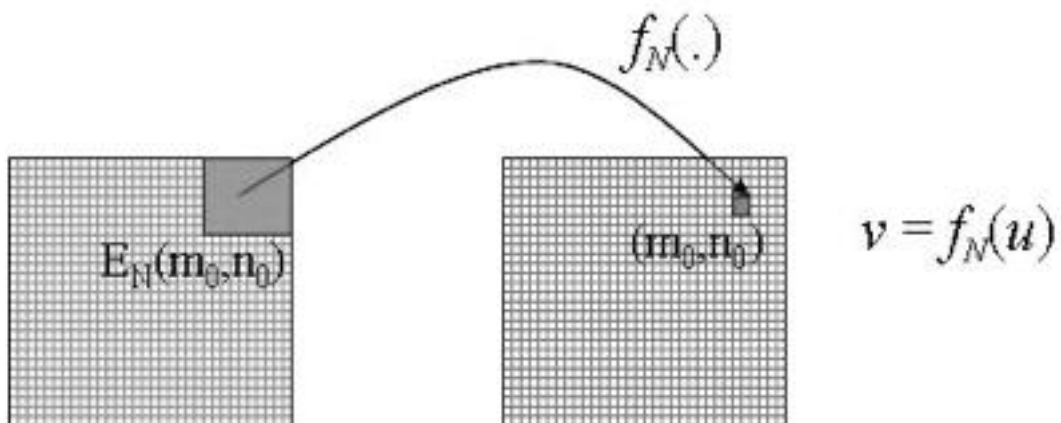


Figura 3.1-1 Esquema de transformación de puntos.

La herramienta habitual son las operaciones basadas en máscaras espaciales (plantillas, ventanas, kernel o filtros FIR): array pequeño en relación a la imagen (3x3, 5x5, 7x7,...). Los valores de los coeficientes determinan el proceso de transformación.

- Filtrado o basada en máscaras:

- ✓ Máscara: matriz de coeficientes de la combinación lineal.

⇒ El entorno del punto (x, y) que se considera en la imagen I para obtener $Q(x, y)$ está determinado por el tamaño y la forma de la máscara.

⇒ El tipo de filtrado está determinado por el contenido de la máscara.

- ✓ Matemáticamente, la operación de la máscara se puede escribir como:

$$O(x, y) = \sum_{k \in \mathbb{E}_N} \sum_{l \in \mathbb{E}_N} h(k, l) I(x - k, y - l) \quad (3.1-1)$$

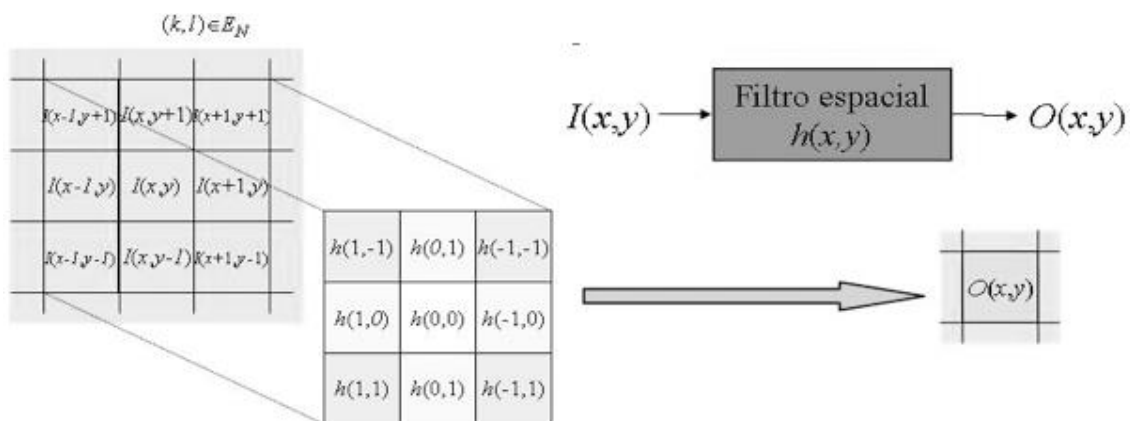


Figura 3.1-2 Ejemplificación de la transformación de puntos.

- T
tratamiento de límites de la imagen
- ✓ P
uede aplicarse la máscara extendiendo la imagen con un marco de ceros de la anchura adecuada.
- ✓ E
sto puede tener efectos no deseados (por ejemplo de difuminación en los límites de la imagen) pero, en general, poco significativos si la máscara es pequeña en comparación con el tamaño de la imagen.

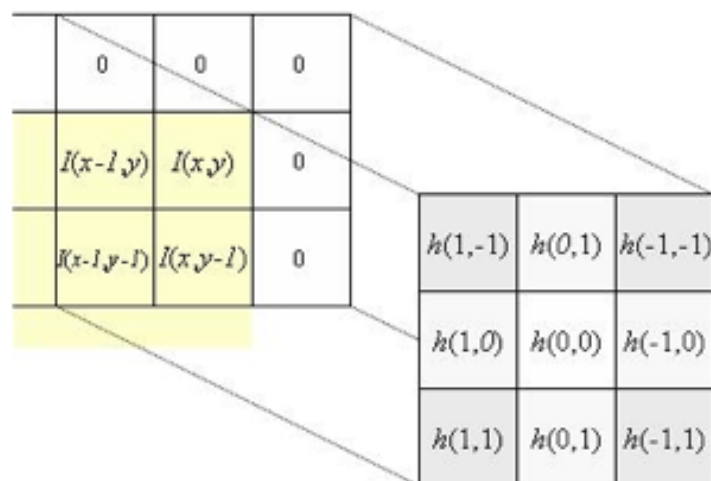


Figura 3.1-3 Ejemplificación de los límites de la máscara.

3.2 ESQUEMAS DE FILTRADO LINEAL Y NO LINEAL.

Filtros lineales

Filtrado paso-bajo

Se utiliza básicamente para desenfocar, suavizar y eliminar el ruido. Y todos los coeficientes de la máscara deben ser positivos y la suma entre sus elementos debe ser uno. Ejemplo.

$$\frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1/8 & 0 \\ 1/8 & 1/2 & 1/8 \\ 0 & 1/8 & 0 \end{bmatrix}$$

Como se mencionó en el párrafo anterior se utiliza para reducir ruido, aunque también producen un difuminado, tanto mayor cuanto mayor sea el tamaño de la máscara.

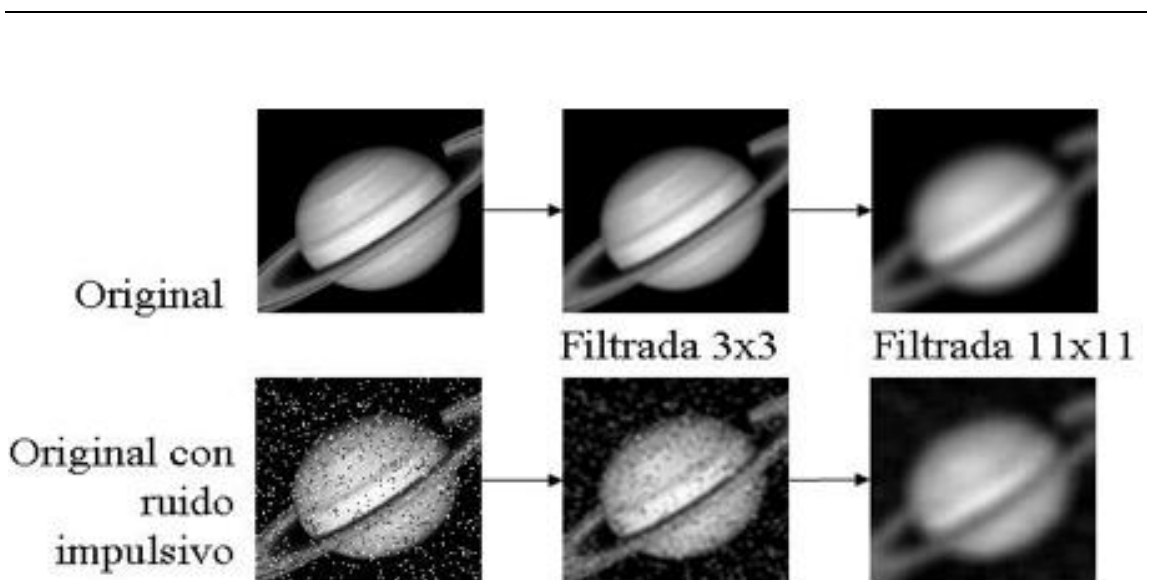


Figura 3.2-1 Filtrado paso-bajo de una imagen

Filtrado paso-alto

Se utiliza principalmente para resaltar bordes, enfocar y para la detección de piezas, objetos, etc. Los coeficientes deben sumar 0. En general con este tipo de filtrado se reduce bastante el contraste, y como los coeficientes deben sumar 0 hay valores negativos que producen escalas o recortes.

Existen algunos tipos de máscaras para este tipo de filtrado, los más sencillos son las máscaras de derivación direccional:

✓

Robert

s (gradiente cruzado)

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

✓

Prewitt

$$\begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$$

✓

Sobel

$$\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

Los operadores Prewitt y Sobel son separables: pueden construirse como una combinación de filtros en direcciones ortogonales (uno de ellos paso alto y el otro paso bajo).

Aplicando un cuantificador, pueden utilizarse para detección de bordes direccionales.



Figura 3.2-2 Detección de bordes direccionales.

También pueden diseñarse filtros no direccionales, es decir, con coeficientes positivos en el centro y negativos en la periferia. Ejemplo.

$$\frac{1}{9} \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \quad \begin{bmatrix} -.5 & -.5 & -.5 \\ -.5 & 4 & -.5 \\ -.5 & -.5 & -.5 \end{bmatrix}$$

Filtrado paso banda

Al igual que el filtrado paso-alto se usa principalmente para realzar bordes, etc. Los coeficientes deben sumar 0. Y como en el filtrado anterior se reduce mucho el contraste y aparecen valores negativos lo que provoca escalas o recortes

Se pueden implementar a partir de dos filtros paso-bajos de diferente frecuencia de corte.

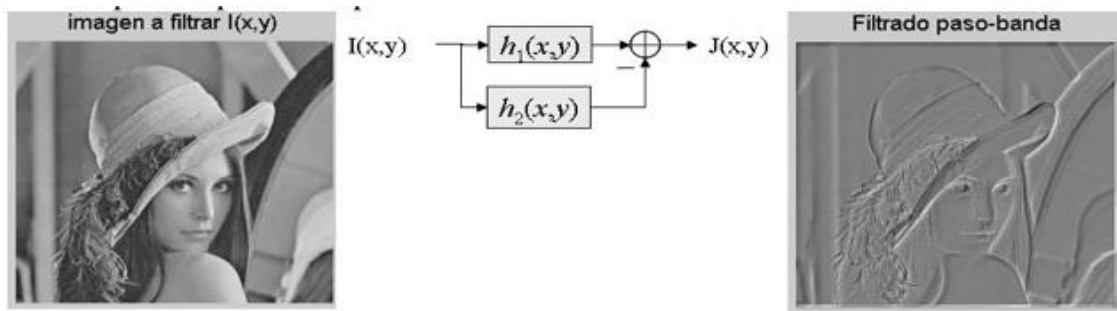


Figura 3.2-3 Filtrado paso-banda de una imagen.

Suavizado Gaussiano

Sea una imagen de escala de grises f ser representado por una de valor real función $f \in L^1(\mathbb{R}^2)$. Una forma muy utilizada para suavizar f es mediante el cálculo de la convolución.

$$(K_\sigma * f)(x) = \int K_\sigma(x - y)f(y)dy \quad (3.2-1)$$

Donde K_σ denota la gaussiana de dos dimensiones con una anchura (desviación estándar) de $\sigma > 0$:

$$K_\sigma(x) = \frac{1}{2\pi\sigma^2} \cdot \exp\left(\frac{-|x|^2}{2\sigma^2}\right) \quad (3.2-2)$$

Hay varias propiedades en este método:

En primer lugar se observa que desde $K_\sigma(x) \in C^\infty(\mathbb{R}^2)$ obtenemos $K_\sigma * f \in C^\infty(\mathbb{R}^2)$, incluso si f es sólo absolutamente integrable.

A continuación, vamos a investigar el comportamiento en el dominio de la frecuencia. En la definición de la transformación de Fourier F por:

$$(Ff)(w) = \int f(x) \exp(-i(w, x)) dx$$

(3.2-3)

Por el teorema de convolución se obtiene que:

$$(F(K_\sigma * f))(w) = (FK_\sigma)(w) \cdot (Ff)(w)$$

(3.2-4)

Dado que la transformada de Fourier de una gaussiana es de nuevo una gaussiana en forma de:

$$(FK_\sigma)(w) = \exp\left(-\frac{|w|^2}{2/\sigma^2}\right)$$

(3.2-5)

Equivalencia a un filtrado de difusión lineal

Es un resultado clásico que para cualquier acotada $f \in C(\mathbb{R}^2)$ el proceso de difusión lineal es el siguiente:

$$\delta_t u = \Delta u \tag{3.2-6}$$

$$u(x, 0) = f(x) \tag{3.2-7}$$

Y posee la siguiente solución:

$$|u(x, t)| = \begin{cases} f(x) & (t = 0) \\ (K_{\sqrt{2t}} * f)(x) & (t > 0) \end{cases}$$

(3.2-8)

Esta solución es única, siempre nos restringimos a las funciones que cumplen con la siguiente expresión:

$$|u(x, t)| \leq M \cdot \exp(\alpha|x|^2) \quad (M, \alpha) > 0$$

(3.2-9)

Depende continuamente en la imagen inicial F con respecto a $\|\cdot\|_{L^\infty(\mathbb{R}^2)}$, y debe cumplir con el principio del máximo-mínimo.

$$\inf_{\mathbb{R}^2} f \leq u(x, t) \leq \sup_{\mathbb{R}^2} f \quad \text{en } \mathbb{R}^2 \times [0, \infty)$$

(3.2-10)

A partir de la ecuación (3.2-8) se observa que el tiempo t está relacionada con la anchura espacial $\sigma = \sqrt{2t}$ de la gaussiana. Por lo tanto, suavizar las estructuras de orden σ requiere para detener el proceso de difusión en el momento

$$T = \frac{1}{2} \sigma^2$$

(3.2-11)

Filtros no lineales

Filtros derivativos

Se obtienen a partir de las máscaras de derivación.

$$\nabla f = \begin{cases} \frac{\delta f}{\delta x} \\ \frac{\delta f}{\delta y} \end{cases}$$

(3.2-12)

$$\text{mag}(\nabla f) = \left[\left(\frac{\delta f}{\delta x} \right)^2 + \left(\frac{\delta f}{\delta y} \right)^2 \right]^{-1/2}$$

(3.2-13)

$$\begin{bmatrix} z_1 & z_2 & z_3 \\ z_4 & z_5 & z_6 \\ z_7 & z_8 & z_9 \end{bmatrix} \rightarrow \nabla_{z_5} f = [(z_5 - z_6)^2 + (z_5 - z_8)^2]^{-1/2}$$

$$\approx (|z_5 - z_6|) + (|z_5 - z_8|)$$

(3.2-14)

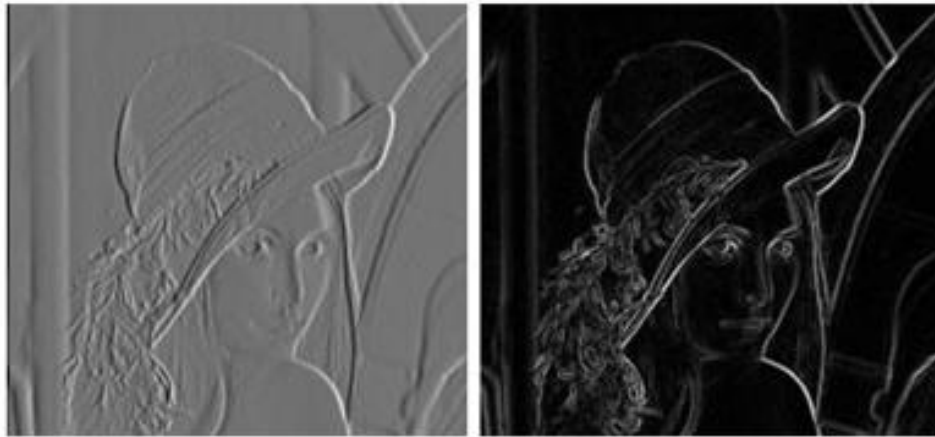


Figura 3.2-4 Ejemplos de imágenes con filtrado derivativo.

Filtros de estadísticos ordenados: mínimo, máximo, mediana

Los filtros de estadísticos ordenados son filtros en los que la operación a realizar no es lineal, funcionan ordenando los valores en la vecindad de cada punto de menor a mayor y obteniendo algún valor a partir de la lista ordenada. Ejemplos:

- ✓ mínimo: selecciona el valor más pequeño. M
- ✓ máximo: selecciona el valor más alto. M
- ✓ mediana: selecciona el valor en la posición intermedia. M

El filtro de mediana suele utilizarse para eliminar ruido impulsivo preservando los bordes de la imagen.

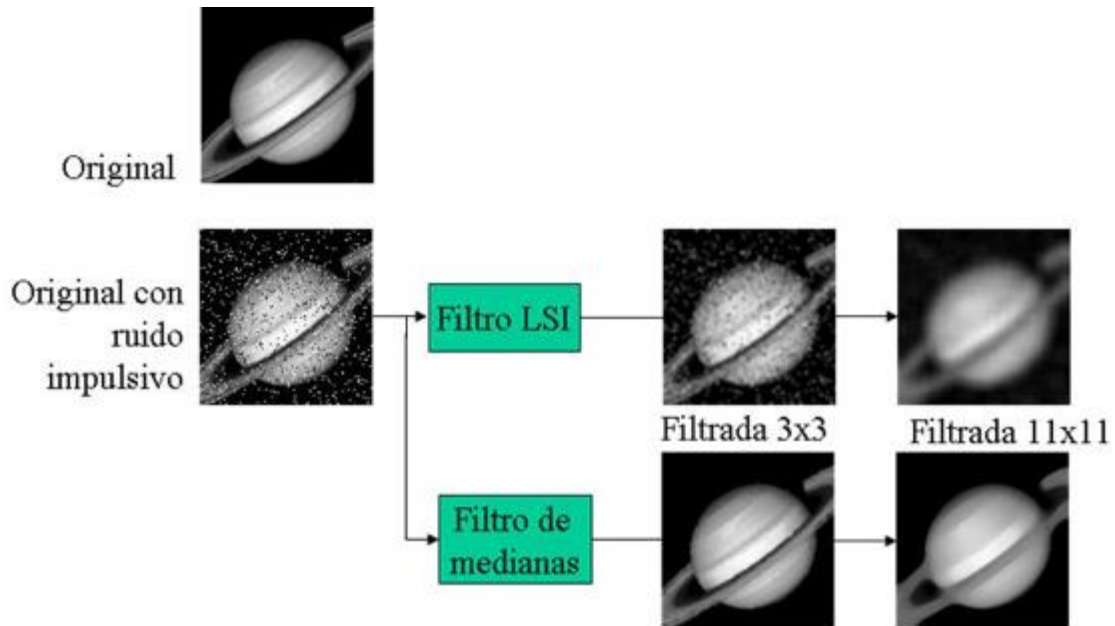


Figura 3.2-5 Ejemplo de una imagen con filtro de mediana

El modelo de Perona-Malik

Perona y Malik proponen un método de difusión no lineal para evitar la confusión y los problemas de localización de difusión lineal de filtrado. Se aplica un proceso no homogéneo que reduce la difusividad en esos lugares que tienen una mayor probabilidad de ser bordes. Esta probabilidad se mide por $|\nabla u|^2$.

El filtro Perona-Malik se basa en la ecuación:

$$\delta_t u = \text{div}(g(|\nabla u|^2) \nabla u) \quad (3.2-15)$$

Y utiliza difusividades tales como:

$$g(s^2) = \frac{1}{1 + s^2/\lambda^2} \quad (\lambda > 0) \quad (3.2-16)$$

Aunque Perona y Malik nombraron su filtro como anisotrópico, debe tenerse en cuenta que debería ser considerado como un modelo isotrópico, ya que utiliza un escalar de valor de difusividad y no un tensor de difusión.

Los experimentos de Perona y Malik eran visualmente muy impresionantes: los bordes seguían siendo estables durante un tiempo muy largo. Se demostró que la detección de bordes sobre la base de este proceso claramente supera al detector de bordes lineal Canny, incluso sin la aplicación de la no supresión de máximos y el umbral de histéresis. Esto es debido al hecho de que la difusión y la detección de bordes interactúan en un solo proceso que en lugar de ser tratados como dos procesos independientes.

Mejora en los bordes

Para estudiar el comportamiento del filtro Perona-Malik en los bordes, nos limitaremos únicamente al caso unidimensional. Esto simplifica la notación e ilustra el comportamiento principal desde cerca de un borde recto de una imagen de dos dimensiones se aproxima a una función de una variable.

Para la difusividad la ecuación (3.2-16) se deduce que la función de flujo $\Phi'(s) = sg(s^2)$ satisface que $\Phi'(s) \geq 0$ para $|s| \leq \lambda$, y $\Phi'(s) < 0$ para $|s| > \lambda$. Puesto que (3.2-15) puede escribirse como:

$$d_t u = \Phi'(u_x) u_{xx} \quad (3.2-17)$$

Se observa que a pesar de su difusividad no negativa el modelo Perona-Malik es de tipo parabólico hacia adelante para $|u_x| \leq \lambda$, y de tipo parabólico hacia atrás para $|u_x| > \lambda$.

Por lo tanto, λ desempeña el papel de un parámetro de contraste que separa hacia adelante (bajo contraste) de versiones anteriores (alto contraste) áreas de difusión.

No es difícil comprobar que el filtro Perona-Malik aumenta la pendiente en los puntos de inflexión de bordes dentro de una zona hacia atrás: Si existe una solución suficientemente suave satisface que u:

$$d_t(u_x^2) = 2u_x d_x(u_t) = 2\Phi''(u_x)u_x u_{xx}^2 + 2\Phi'(u_x)u_{xx}u_{xxx}. \quad (3.2-17)$$

Una ubicación x_0 donde u_x^2 es máxima en algún tiempo t se caracteriza por $u_x u_{xx} = 0$ y $u_x u_{xxx} \leq 0$.

Por lo tanto,

$$(\delta_t(u_x^2))(x_0, t) \geq 0 \quad \text{para } |u_x(x_0, t)| > \lambda \quad (3.2-18)$$

Con una estricta desigualdad para $u_x u_{xxx} < 0$.

En el caso de dos dimensiones, (3.2-17) se sustituye por:

$$\delta_t u = \Phi'(\nabla u) u_{\eta\eta} + g(|\nabla u|^2) u_{\xi\xi} \quad (3.2-19)$$

Donde el indicador de coordenadas ξ y η denotan las direcciones perpendicular y paralela a ∇u , respectivamente. Por lo tanto, tenemos por delante una difusión a lo largo de isofotas (es decir, las líneas de valor de gris constante) combinada con la difusión hacia delante-hacia atrás a lo largo de las líneas de flujo (líneas de variación de valor de gris máximo).

Se observa que el comportamiento de difusión de adelante-atrás no sólo se limita a la difusividad especial (4.2-16), que aparece por todas las difusividades $g(s^2)$ cuya desintegración rápida hace que las funciones de flujo no monótono del tipo $\Phi(s) = sg(s^2)$.

3.3 LA ECUACIÓN DIFERENCIAL DE DIFUSIÓN DEL CALOR

El procesamiento de imágenes digitales ha encontrado un enorme apoyo en la teoría de las ecuaciones en derivadas parciales, con aplicaciones que van desde el procesamiento de bajo nivel a los campos complejos de análisis de imágenes y visión por computadora. El procesamiento básico, como convolución para el filtrado de paso bajo, tiene una interpretación unificada por medio de ecuaciones diferenciales parciales. Con el fin de optimizar una función de energía, el procesamiento de imágenes recursiva se puede derivar como un algoritmo de descenso de gradiente, con el resultado final está determinado por el estado de equilibrio de este proceso recursivo. La imagen bajo análisis es el estado inicial de un proceso de difusión bien definida cuya evolución temporal conduce a una representación a escala-espacio.

El flujo por difusión a través de un plano dado es la respuesta a un gradiente de concentración a través del plano.

La norma local para flujo de difusión o flujo J viene dado por primera ley de difusión de Fick:

$$J = -D \cdot \nabla u \quad (3.3-1)$$

En la que el flujo de J es proporcional a la difusividad D y para el gradiente negativo de la concentración de ∇u . El signo negativo indica que el flujo de difusión se produce a partir de las regiones de mayor a regiones de menor concentración. El flujo de difusión depende de ambas variables de espacio y tiempo tales que, más en general, para una ecuación de dos variables dimensiones puede ser también expresado como la siguiente expresión $J(x, y, t) = -D(\delta u / \delta x, \delta u / \delta y)$. En una posición dada (x, y) , en el tiempo t , el flujo es proporcional a la variación espacial en la concentración del material cuyas partículas se están moviendo en la dirección apositiva señalado por los vectores de la unidad del operador de gradiente.

Un caso de interés es el transporte de masa (proceso de difusión) sin la destrucción masiva o la creación de nueva masa.

Este equilibrio se expresa por medio de la ecuación de continuidad como:

$$\delta_t u = -div J = -\left(\frac{\delta J}{\delta x} + \frac{\delta J}{\delta y}\right)$$

(3.3-2)

En una representación a escala-espacio, en un instante dado, un cambio en la concentración de u es inversamente proporcional a la variación del vector de flujo. Un cambio en la concentración nula se observa cuando el flujo se mantiene constante en la región analizada. Se obtiene la ecuación de difusión cuando la ecuación vectorial de la ley Ficks se sustituye en la ecuación de continuidad:

$$\delta_t u = div (D \nabla u) = D \left(\frac{\delta^2 u}{\delta x^2} + \frac{\delta^2 u}{\delta y^2} \right)$$

(3.3-3)

La ecuación (3.3-3) es una ecuación diferencial parcial en relación a la evolución de las concentraciones con la difusión espacial. Visto como una variable continua, la concentración depende de las variables de tiempo y espacio.

Los valores pequeños de la gradiente de concentración son producidos por materiales altamente homogéneos, en este caso, con una definición adecuada de tensor de D , una gran difusión controla la evolución en el tiempo de la concentración de u . Una constante de difusión D produce tensor de difusión homogénea o lineal, en este caso, la masa se propaga linealmente en todas las direcciones a lo largo de un medio dado. Una difusión no lineal se obtiene cuando el tensor de difusión depende de la estructura del medio de propagación.

La información radiométrica de una imagen puede considerarse como una especie de concentración y la imagen de la red es el medio de propagación. Una descripción recursiva escala-espacio se obtiene por muestreo en la ecuación tiempo (3.4-3), de tal manera que la difusión anisotrópica se realiza mediante la estimación adaptativa del parámetro D .

La difusión, como un proceso físico para el equilibrio de gradientes de concentración, proporciona las bases teóricas para la obtención de un filtro que es más adecuado para hacer frente a la heterogeneidad local de una escena del hielo marino SAR. Un punto importante es la caracterización de la información primitiva para ser procesados por el método de filtrado.

Una caracterización básica a retener es que una imagen está integrada por un conjunto de regiones uniformes de niveles de gris disjuntos y por conjuntos de bordes de las regiones vecinas.

3.4 ESQUEMAS DE FILTRADO ANISOTRÓPICO

El concepto original de difusión anisotrópica para el filtrado de la imagen fue:

- i) para detectar bordes
- ii) para suavizar regiones homogéneas a lo largo de los bordes, pero no a través de los contornos detectados.

Para ser desarrollado el filtro se basa en la ecuación de continuidad, cuando el tensor de difusión D está definido por una función de difusión adecuada.

El modelo de difusión isotrópica, el valor del tensor escalar g se denomina *difusividad* o bien *conductividad*, y en el proceso de imágenes se suele hacer que sea una función dependiente del módulo del gradiente de la imagen en cada punto.

La ecuación (3.3-3) da como resultado un proceso de:

$$\begin{cases} \frac{\delta I(x, y, t)}{\delta t} = \text{div} [g(|\nabla I(x, y, t)|) \nabla I] \\ I(t = 0) = I_0 \end{cases}$$

(3.4-4)

En donde ∇ es el operador gradiente, div es el operador divergencia, $g(\bullet)$ es la función de difusión con $|\bullet|$ siendo su magnitud, (x, y) son los índices espaciales de la imagen, I_0 es la imagen original, y $\delta I(x, y, z)/\delta t$ es el resultado obtenido por la evolución de los parámetros de tiempo t . El gradiente es una ventaja, no estimador de vanguardia y este operador de bajo nivel es el núcleo central de la filtración anisotrópica. La función $g(|\nabla I(x, y, t)|)$ controla la intensidad de desenfoque en cada píxel de la imagen de la red muestreada. El coeficiente de difusión dado por la siguiente expresión $C(x, y, z) = g(|\nabla I(x, y, t)|)$ es "grande" cuando (x, y) no es parte de un borde ($C \approx 1$ cuando $|\nabla I(x, y, t)| \approx 0$), mientras que $C(x, y, t)$ es "pequeña" cuando (x, y) es parte de un borde ($C \approx 0$ cuando $|\nabla I(x, y, t)| \approx 1$). Esto da como resultado la difusión detuvieron principalmente por los bordes. Cuando el coeficiente de difusión tiene un valor constante, la ecuación (3.4-4) realiza la difusión isotrópica, es decir, una operación de suavizado lineal equivalente a una convolución con un kernel de Gauss.

El coeficiente de difusión debe ser elegido de tal manera que la evolución en la resolución de baja escala no introduce artefactos falsos y también para producir el espacio borrosa invariantes. Para conseguir esto, dos funciones no negativas y monótona decreciente se utilizan. El resultado se da por la ecuación (3.4-4):

$$\begin{cases} g(|\nabla I|) = (1 + (|\nabla I|/K)^2)^{-1} \\ g(|\nabla I|) = \exp(-(|\nabla I|/K)^2) \end{cases}$$

(3.4-5)

Que son funciones decrecientes entre uno y cero y en donde K es un umbral de pendiente constante. La ecuación (3.4-5) complementa la interpretación paramétrica de la ecuación (3.4-5). El proceso no homogéneo reduce la difusión en píxeles, que son muy probable que sean bordes. Si K es pequeño y $|\nabla I| \gg K$, el proceso de difusión tendrá un efecto muy débil. Si $|\nabla I| \ll K$, el proceso de difusión producirá regiones borrosas y, en este caso, la iteración en el tiempo t relativa es una operación de filtro de paso bajo.

Intuitivamente, la información de una imagen dada puede ser descrita por un criterio de escala. La textura y los datos basados en píxeles estructurados se pueden analizar a diferentes escalas de resolución espacial. Micro y macro texturas son conceptos bien conocidos en la visión por computador. Con la llamada teoría del espacio de escala es posible sucesivamente para recuperar y suprimir conjuntos específicos de imagen.

Derivado de los datos que están siendo analizados, un conocimiento más específico puede ser introducido en la ecuación (3.4-4).

Para obtener el resultado en múltiples escalas, la función de difusión es:

$$g(x, y, t) = (1 + (|\nabla I|_t \cdot \sigma_t / K)^2)^{-1} \quad (3.4-6)$$

En donde σ_t es la desviación estándar calculada por las ventanas de 3x3 píxeles deslizante en la escala t .

El siguiente paso es la implementación numérica del esquema propuesto. Las ecuaciones continuas (3.4-4) y (3.4-6) pueden ser resueltos por el enfoque Eulers o por métodos numéricos recursivos.

3.5 DISCRETIZACIÓN DE LA ECUACIÓN DIFERENCIAL

En la ecuación general de difusión (3.3-3), el tensor D es una matriz simétrica definida positiva. Para el caso particular de imágenes, la ecuación de difusión se puede expresar en forma desarrollada como:

$$\delta_t u = (\delta_x \ \delta_y \ \delta_z) \begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix} (\delta_x u \ \delta_y u \ \delta_z u)^T$$

(3.5-1)

En función de los coeficientes del tensor y las derivadas parciales en coordenadas cartesianas. En total existen nueve contribuciones en el filtro:

$$\delta_t u = \delta_x(a\delta_x u) + \delta_x(b\delta_y u) + \delta_x(c\delta_z u) + \delta_y(b\delta_x u) + \delta_y(a\delta_y u) \\ + \delta_y(e\delta_z u) + \delta_z(c\delta_x u) + \delta_z(e\delta_y u) + \delta_z(f\delta_z u)$$

(3.5-2)

La resolución numérica de esta expresión se aproxima para retículos discretos mediante diferencias finitas. La manera más directa de resolver la evolución en el tiempo consiste en utilizar diferencias hacia adelante, mediante un esquema explícito que permite calcular el resultado en cada incremento de tiempo a partir de resultados anteriores. La discretización temporal queda como:

$$u^{i+1} = (I + \tau A^\tau) * u^i$$

(3.5-3)

Donde $A^\tau * u^i$ es la realización discreta del segundo término de la ecuación de difusión, $\text{div}(D\nabla u)$. El valor de τ debe hacerse lo suficientemente pequeño para asegurar la estabilidad y convergencia de la solución.

En cuanto a la aproximación discreta de los términos espaciales de la expresión (3.5-2), se han usado tanto diferencias finitas centrales como las diferencias hacia adelante y atrás, definidas para la dimensión x como:

$$\begin{aligned}\delta_x^- u &= u_{x,y,z} - u_{x-1,y,z} \\ \delta_x^+ u &= u_{x+1,y,z} - u_{x,y,z} \\ \delta_x^c u &= u_{x+1,y,z} - u_{x-1,y,z}\end{aligned}$$

(3.5-4)

Los términos de (3.5-2) que dependen de una sola variable espacial el esquema utilizado es:

$$\delta_x(a\delta_x u) = \frac{1}{2}(\delta_x^+(a\delta_x^- u) + \delta_x^-(a\delta_x^+ u))$$

(3.5-5)

Y los términos en dos variables se aproximan mediante diferencias finitas centrales:

$$\delta_x(b\delta_y u) = \delta_x^c(b\delta_y^c u)$$

(3.5-6)

De igual manera se opera con los términos equivalentes en y , z . Esta discretización es válida para cualquier tensor D , y el comportamiento del filtro dependerá del valor que tomen sus coeficientes.

4.6 REPRESENTACIONES ESCALA-ESPACIO Y PARÁMETROS ASOCIADOS

Es un hecho bien conocido que las imágenes por lo general contienen estructuras en una gran variedad de escalas. En aquellos casos en que no está claro de antemano cuál es la escala de la derecha para obtener la información descrita es deseable tener una representación de la imagen en múltiples escalas. Por otra parte, mediante la comparación de las estructuras a diferentes escalas, se obtiene una jerarquía de las estructuras de la imagen que facilita una interpretación de la imagen posterior.

Una escala-espacio es una representación de la imagen en un continuo de escalas, la incorporación de la imagen de f en una familia $\{T_t f | t \geq 0\}$ de versiones simplificadas gradualmente de la misma, que siempre que cumpla con ciertos requerimientos. La mayoría de estas propiedades se pueden clasificar como de arquitectura, suavizado (reductor información) o los requisitos de invariancia.

Un supuesto arquitectónico importante es recursividad, es decir para $t = 0$, la representación escala-espacio da a la imagen original f , y el filtrado se puede dividir en una secuencia de bancos de filtros:

$$T_0 f = f \quad (3.6-1)$$

$$T_{t+s} f = T_t(T_s f), \quad \forall s, t \geq 0 \quad (3.6-2)$$

Esta propiedad se conoce como la propiedad “semigrupo” muy a menudo. Otros principios arquitectónicos comprenden, por ejemplo, las propiedades de regularidad de T_t y el comportamiento local t tiende a 0.

Las propiedades de suavizado y la reducción de información surgen del deseo de que la transformación no debe crear artefactos cuando se pasa de bien a la representación gruesa. Por lo tanto, en una escala gruesa, no debemos tener estructuras adicionales que son causadas por el método de filtración en sí y no por las estructuras subyacentes a escalas más finas.

A menudo, estos requisitos se complementan con un supuesto adicional que es equivalente al principio de superposición, es decir, linealidad:

$$T_t(af + bg) = aT_t f + bT_t g \quad \forall t \geq 0, \quad \forall a, b \in \mathbb{R} \quad (3.6-3)$$

Representación espacio-escala Gaussiana

El históricamente primera y mejor investigada es la escala-espacio gaussiana, que se obtiene a través de convolución con gaussianas de varianza en incremento, o lo que es equivalente por filtrado de difusión.

Por lo general, un documento de 1983 por Witkin o un informe de 1980 por Stansfield se consideran como las primeras referencias a la idea de la escala del espacio lineal. Un trabajo reciente de Weickert, Ishikawa y Imiya, sin embargo, muestra que la escala del espacio es más de 20 años de edad: Una derivación axiomática de 1-D Gauss escala-espacio ya ha sido presentado por Taizo Iijima en un documento técnico a partir de 1959 seguido por una versión revista en 1962.

En el documento presentado por Iijima considera una transformación Φ que depende de un parámetro de escala σ y que transforma la imagen original $f(x)$ en una borrosa versión $\Phi[f(x'), x, \sigma]$. Esta clase de transformaciones borrosas se llama boke (desenfoque). Se asume que tiene la estructura la siguiente expresión:

$$\Phi[f(x'), x, \sigma] = \int_{-\infty}^{\infty} \phi\{f(x'), x, x', \sigma\} dx'$$

(3.6-4)

La expresión anterior debe satisfacer con cinco condicionales básicas que son las siguientes:

I. L
inealidad (con respecto a las multiplicaciones)

Si la intensidad de un patrón se convierte en A veces su intensidad original, entonces, lo mismo debería pasar con el patrón observado:

$$\Phi[Af(x'), x, \sigma] = A\Phi[f(x'), x, \sigma]$$

(3.6-5)

II. T

traducción invariante

El filtrado de una imagen traducida es lo mismo que la traducción de la imagen filtrada:

$$\Phi[f(x' - a), x, \sigma] = \Phi[f(x'), x - a, \sigma]$$

(3.6-6)

III. E

escala invariante

Si un patrón en especial se amplió por algún factor de λ , entonces existe una $\sigma' = \sigma'(\sigma, \lambda)$ de tal manera que:

$$\Phi[f(x'/\lambda), x, \sigma] = \Phi[f(x'), x/\lambda, \sigma]$$

(3.6-7)

IV. P

propiedad del semigrupo (generalizada)

Si se observa a f bajo un parámetro σ_1 y esta observación se observa después bajo un parámetro σ_2 , entonces esto es equivalente a la observación de f en virtud de un parámetro adecuado $\sigma_3 = \sigma_3(\sigma_1, \sigma_2)$:

$$\Phi[\Phi[f(x''), x', \sigma_1], x, \sigma_2] = \Phi[f(x''), x, \sigma_3]$$

(3.6-8)

V. P

reservación de la positividad

Si la imagen original es positiva, entonces la imagen observada es positiva, así:

$$\Phi[f(x'), x, \sigma] > 0 \quad \forall f(x') > 0, \quad \forall \sigma > 0$$

(3.6-9)

Bajo estos requisitos se deriva de una manera muy sistemática que:

$$\Phi[f(x'), x, \sigma] = \frac{1}{2\sqrt{\pi}\sigma} \int_{-\infty}^{\infty} f(x') \exp\left(\frac{-(x - x')^2}{4\sigma^2}\right) dx'$$

(3.6-10)

Por lo tanto, $\Phi[f(x'), x, \sigma]$ es sólo la convolución entre f y una gaussiana con desviación estándar $\sigma\sqrt{2}$.

4. OPTIMIZACIÓN POR CAMPOS ALEATORIOS DE MARKOV

4.1 FUNDAMENTOS

La idea de los algoritmos Monte Carlo vía cadenas de Markov es crear una cadena de Markov que tenga distribución estacionaria aproximada a una función de densidad de interés. Para ser más precisos, supongamos que deseamos generar una muestra a partir de la función de densidad $\pi(\theta | y)$ para $\theta \in \Theta$ pero que esto no se puede hacer directamente. Supongamos que podemos construir una cadena de Markov con espacio de estados Θ , la cual sea fácil de simular y tenga como distribución estacionaria a $\pi(\theta | y)$. Si simulamos una trayectoria de la cadena lo suficientemente grande, los valores simulados pueden usarse para resumir algunas propiedades de $\pi(\theta | y)$.

Lo que se quiere es utilizar algoritmos Monte Carlo vía cadenas de Markov que produzcan cadenas de Markov ergódicas con la distribución estacionaria de interés.

Es necesario hacer las siguientes observaciones:

- Estamos
s interesados en el caso cuando la distribución estacionaria de la cadena construida corresponde a la densidad a posteriori $\pi(\theta | y)$.
- La
propiedad ergódica significa que la distribución de θ_t converge a la distribución límite $\pi(\theta | y)$ para cualquier valor inicial θ_0 . Por lo tanto para t grande, θ_t proviene de la distribución $\pi(\theta | y)$ sin importar el punto inicial. Un problema es decidir que tan grande tomar t , pero esto puede ser establecido haciendo pruebas de hipótesis sobre la estacionariedad de la cadena.
- Las
cadenas de Markov obtenidas por los métodos Monte Carlo vía cadenas de Markov son casi siempre homogéneas en el tiempo, o sea, la

distribución de $\theta_{t_0+1}, \theta_{t_0+2}, \dots, \theta_{t_0+k} | \theta_{t_0}$ es igual que la distribución de $(\theta_1, \theta_2, \dots, \theta_k | \theta_0)$ para cualquier $t_0 \geq 0$ y $k > 0$.

- Otra importante propiedad de estabilidad es la llamada reversibilidad, que significa que la dirección del tiempo de la cadena no importa. La cadena es reversible con respecto a $\pi(\theta | y)$ si el kernel de transición satisface la condición de balance:

$$\underline{k(\theta|\theta')\pi(\theta|y) = K(\theta|\theta')\pi(\theta'|y)} \quad (4.1-1)$$

La condición de balance es una condición suficiente (no necesaria) para que $\pi(\theta|y)$ sea una densidad estacionaria asociada al kernel de transición.

De acuerdo a lo anterior, los valores generados $\theta_1, \theta_2, \dots$ no son independientes. Sin embargo, la independencia requerida para calcular algunas características de $\pi(\theta|y)$, es obtenida al tomar valores generados en iteraciones suficientemente distantes.

4.2 MUESTREADOR DE GIBBS

El muestreador de Gibbs es una versión del método de Metrópolis-Hastings donde se toma como distribución de propuesta la distribución condicionada de la distribución original de donde se quería simular el parámetro θ_i , es decir,

$$g_i(\cdot | \theta_i, \theta_{-i}) = f(\cdot | \theta_{-i}, x) \quad (4.2-1)$$

Dada esta distribución, se puede demostrar que la probabilidad de aceptar ϕ es siempre 1 y así, los valores propuestos son siempre aceptados. Aunque, en este caso, las distribuciones condicionadas deben ser todas fáciles de simular.

El algoritmo para el muestreador de Gibbs es el siguiente:

1. Comenzar con valores iniciales arbitrarios de $\theta^{(0)}$
2. Generar $r \theta_1^{(t+1)} \sim f(\theta_1 | x, \theta_2^{(t+1)}, \dots, \theta_k^{(t)})$
3. Generar $r \theta_2^{(t+1)} \sim f(\theta_2 | x, \theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_k^{(t)})$,
Generar $\theta_k^{(t+1)} \sim f(\theta_k | x, \theta_1^{(t+1)}, \dots, \theta_{k-1}^{(t+1)})$,

4.
r al paso 2

Regresa

4.3 ESQUEMA DE METRÓPOLIS

Este es el algoritmo más general. Se trata de simular un parámetro del que conocemos la densidad a posteriori $f(\theta|x)$. Se busca una densidad condicional $g(\cdot|\theta)$ o una distribución de transición (arbitraria) que sea fácil de muestrear. Después se generan observaciones de esta distribución de propuesta para decidir si pertenecen a la distribución $\theta|x$ mediante un sorteo.

El algoritmo Metrópolis-Hastings se puede resumir de la siguiente forma:

1. Definir un valor inicial de $\theta^{(0)}$ Defi
2. $t = 0$ $t = 0$
3. Generar $\phi \sim g(\cdot|\theta^{(t)})$ Gen
4. Definir
$$\alpha(\theta^{(t)}, \phi) = \min \left[1, \frac{f(\phi|x) \cdot g(\theta^{(t)}|\phi)}{f(\theta^{(t)}|x) \cdot g(\phi|\theta^{(t)})} \right]$$
 Defi
5. Tomar Tom

$$\theta^{(t+1)} = \begin{cases} \emptyset & \text{con probabilidad } \alpha \\ \theta^{(t)} & \text{en caso contrario} \end{cases}$$

6. . Ir al paso 3

$$t = t + 1$$

Este procedimiento genera una cadena de Markov con distribución de transición:

$$P(\theta^{(t+1)}|\theta^t) = \alpha(\theta^{(t+1)}, \theta^{(t)})g(\theta^{(t+1)}|\theta^{(t)})$$

(4.3-1)

Se puede demostrar que mediante este algoritmo se simula de una cadena de Markov cuya distribución estacionaria es $f(\theta|x)$.

En teoría, se puede utilizar cualquier distribución $g(\emptyset|\theta)$. Lo más importante es que sea fácil de muestrear, aunque la convergencia del algoritmo será más rápida si $g \approx f$.

4.4 RECOCIDO SIMULADO

La técnica de Recocido Simulado, es una técnica de búsqueda aleatoria introducida en 1983 por Kirkpatrick. El término de “Simulated Anneling”, se puede traducir al castellano como “Recocido Simulado”. El término Recocido según el Diccionario de la Lengua Española de la Real Academia, hace referencia al “calentamiento de metales para que adquieran nuevamente la ductilidad o temple que suelen perder al trabajarlos”

La técnica de Recocido Simulado, que de aquí en adelante denominaremos SA (por sus siglas en inglés) difiere de las técnicas tradicionales de mejora iterativa, en su capacidad para escapar de los mínimos locales gracias al empleo del criterio Metrópolis para la aceptación de los desplazamientos.

Este algoritmo pertenece a la clase de algoritmos que pretenden resolver un problema de optimización combinatoria.

Todo algoritmo SA (Figura 4.4-1(a)) se puede caracterizar por una serie de parámetros, unos específicos del problema en cuestión y otros genéricos independientes de la naturaleza del problema.

Como parámetros específicos del problema tenemos:

- Un espacio de soluciones Ω , conjunto finito de todas las soluciones del problema. Un
- Una función de costo U . Una
- Un mecanismo de generación, que permite el desplazamiento desde una solución a otra de su entorno, entendiendo como entorno de una solución aquellas soluciones a las que podemos llegar desde una solución dada aplicando el mecanismo de generación. Un

Como parámetros genéricos, llamados a veces “esquema de enfriamiento” tenemos:

- Temperatura inicial T_0 . Tempe
- Función de decremento de la temperatura o ley de evolución. Funció
- Criterio de equilibrio. Criterio
- Criterio de congelación. Criterio

Este conjunto de parámetros es independiente del tipo de problema, y controla el funcionamiento del algoritmo. En la figura 4.4-1(b) se presenta un diagrama de flujo detallado de lo que sería un algoritmo SA estándar, incluyendo un contador R de los escalones de temperatura realizados.

En nuestro caso agregamos el parámetro β que controla la homogeneidad de las regiones.

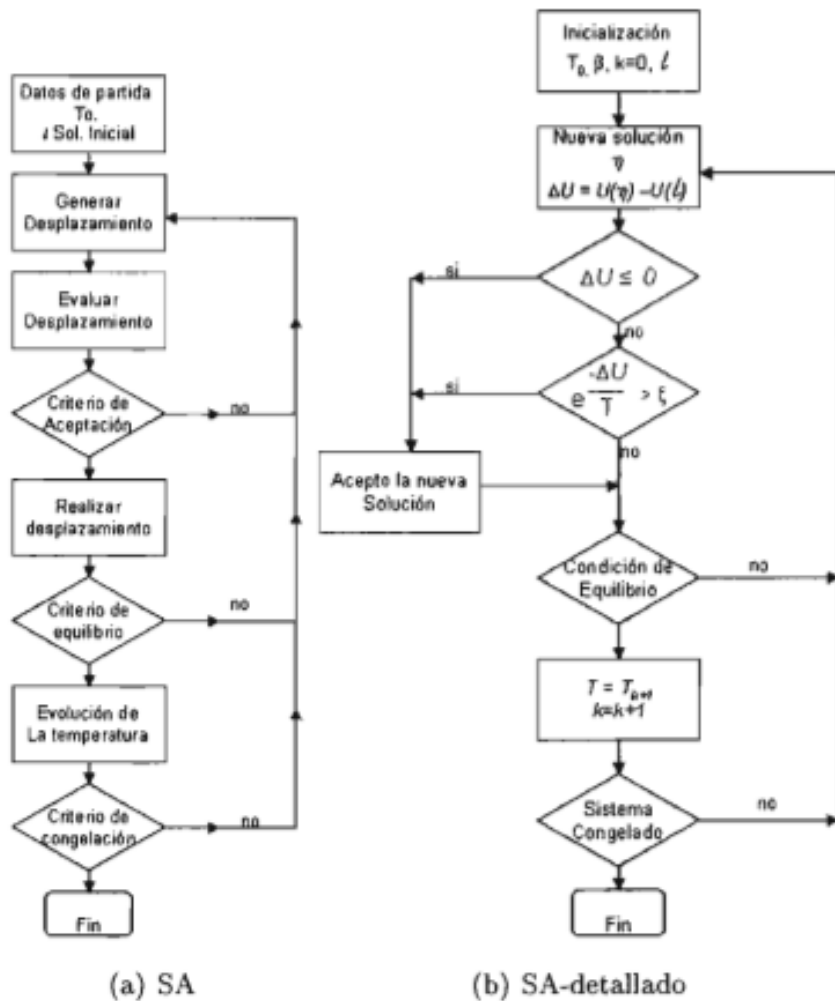


Figura 4.4-1 Algoritmo Recocido Simulado.

Algoritmo SA para segmentación de imágenes

1. Asignar
 $k=0$ e inicializar l de manera aleatoria. Escoger una temperatura inicial lo suficientemente alta $T = T_0$.
2. Construir
 uir una perturbación η de l , tal que η difiera en sólo un elemento de l .
3. (Criterio de Metrópolis)
 Calcular $\Delta U = U(\eta) - U(l)$ y aceptar η si $\Delta U < 0$, si no, aceptar η con probabilidad $e^{\frac{-\Delta U}{T}}$.
4. Ir al
 paso 2 hasta que el criterio de equilibrio sea aceptado.

5. Decrementar la temperatura $T = T_{k+1}$ e ir al paso 2 con $k=k+1$ hasta que el sistema esté congelado.

Criterio Metrópolis: Se calcula la variación de energía provocada por el movimiento. Si $\Delta E < 0$ entonces, dicho movimiento es permitido, ya que el movimiento conduce al sistema a un estado con menor nivel energético. Si $\Delta E > 0$ entonces el desplazamiento es permitido con una probabilidad $e^{-\Delta E/kT}$; para ello se genera un número aleatorio ξ entre 0 y 1, de tal forma que $\xi < e^{-\Delta E/kT}$ se desplaza la partícula a su nueva posición, en caso contrario la partícula se regresa a su posición inicial.

4.5 PARAMETROS: VECINDADES, CLIQUES Y TEMPERATURA

Temperatura inicial.

La temperatura inicial T_0 debe ser lo suficientemente grande para que casi todo (o todo) movimiento sea permitido, es decir, que la probabilidad de pasar del estado i al j (en W_{ij}) sea muy alta. Kirpatrick propone una regla empírica, escoger un valor grande de T_0 y realizar cierto número de transiciones. Si se define el coeficiente de aceptación x como el cociente entre el número de transiciones intentadas, entonces el valor obtenido de x es menor de una cierta cantidad.

Por ejemplo 0.8 entonces se dobla el valor de T_0 , procediendo así sucesivamente hasta verificar la condición impuesta.

Temperatura final.

El valor final de T es también importante, teóricamente T debería reducirse hasta 0, pero en la práctica la búsqueda converge por lo general a su óptimo local final bastante antes de ese valor nulo de la temperatura. Por lo tanto, si la temperatura de parada se fija muy baja, invertiremos mucho tiempo de búsqueda en las fases finales, que seguramente sería mejor aprovechado en temperaturas superiores. Por el contrario, si la temperatura final se fija en un valor muy alto, es posible que la búsqueda no consiga alcanzar ningún óptimo local.

La condición de parada del algoritmo se puede establecer de diferentes maneras. Una de ellas es fijar un número determinado de valores de T para los cuales se ejecuta el algoritmo. Otro criterio puede ser, establecer un coeficiente de aceptación α mínimo, de tal forma que en caso de no alcanzarse se detiene el algoritmo.

Condición de equilibrio (Longitud de la cadena de Markov).

La elección más sencilla para L_k longitud de la k -ésima cadena de Markov, es elegir un valor dependiente (de forma polinómica) del tamaño del problema, así L_k es independiente de k . Existen propuestas más elaboradas, en las que se establece que para cada valor de T_k se debería realizar un número mínimo de transiciones. Entonces en ese caso se determina L_k de tal forma que el número mínimo de transiciones γ_{\min} (siendo γ_{\min} un número fijo). Sin embargo al aproximarse T a cero, las transiciones son aceptadas cada vez con una probabilidad menor de tal forma que $L_k \rightarrow \infty$ si $T_k \rightarrow 0$.

Por ello L_k debe estar acotado superiormente por una cierta constante L , para evitar cadenas de Markov excesivamente largas para bajos valores de T .

Ley de evolución de la temperatura.

El factor más importante en el comportamiento del algoritmo es la regla de enfriamiento o decremento de la temperatura. Existen diferentes leyes de evolución de temperatura que se clasifican en función de su complejidad.

Dentro de las primeras propuestas y que aún son ampliamente usadas tenemos:

$T_k = \alpha T_{k-1}$ (Factor de enfriamiento geométrico, con $\alpha < 1$, y muy cercano, varios autores utilizan $\alpha = 0.95$)

$$T_k = \frac{T_{k-1}}{1 + \beta T_{k-1}} \text{ (donde } \beta \approx 0)$$

5 ESQUEMA ESTOCÁSTICO

5.1 PLANTEAMIENTO Y ESQUEMA GENERAL

Después de toda la explicación teórica es posible expresar de forma más clara el problema a resolver en esta tesis.

Como se menciona en la introducción el problema que se debe resolver es la correcta detección de las zonas del derrame petrolero en Golfo de México, para esto se cuenta con una serie de imágenes de alta resolución del satélite

ENVISAT/ASAR, para este caso se seleccionó únicamente una imagen para hacer la detección.

En la imagen existe cierto ruido. La parte superior parece tener un tono más claro que la parte inferior, como se aprecia en la figura (6.1-1). Este ruido hace más difícil la correcta detección de las zonas que son mar y de las zonas en donde existe petróleo.

Para poder minimizar el ruido presente en la imagen se hará primero una difusión anisotrópica para de esa manera poder disminuir el ruido de la imagen y que provoca que se puedan distinguir diferentes tonos en la misma imagen. Esto se hace con el propósito de hacer que la imagen se vea lo más uniforme posible para así poder lograr una correcta detección en las zonas de petróleo.

El objetivo de aplicar una difusión anisotrópica a la imagen original es la de suavizar dentro las regiones delimitadas por los bordes y no lo haga a través de ellos.

Para poder aplicar una difusión anisotrópica es necesario proporcionar dos parámetros importantes para definir las características de la imagen. Pietro Perona y Jitendra Malik propusieron dos parámetros importantes para este algoritmo.

✓ El primer valor es **kappa** es parámetro controla la conducción como una función del gradiente. Si el valor de kappa es bajo pequeños gradientes de intensidad son capaces de bloquear la conducción y por lo tanto la difusión a través de los bordes. Un valor grande reduce la influencia de los gradientes de intensidad en la conducción.

✓ El segundo valor es **delta** es una constante de integración y que debe estar entre el siguiente rango $0 \leq \text{delta} \leq 1/7$. Por lo general, debido a la estabilidad numérica de este parámetro se establece en su valor máximo.

En el capítulo 5 se realizan pruebas para observar con qué valor de delta se obtiene una mejor calidad en la imagen. Además de realizar las pruebas con diferentes valores de estos parámetros, también se hace una comparación con

el software *ImageJ*, que es un software que nos permite realizar una difusión anisotrópica a una imagen después de proporcionarle los valores de kappa y delta.

Después de haber aplicado la difusión anisotrópica para suavizar la imagen y hacerla lo más uniforme posible, se procede a realizar la segmentación de la imagen en *clases* para después poder aplicar el algoritmo de recocido simulado.

Además también se realizara una comparación de los resultados entre el algoritmo de recocido simulado y el algoritmo de k-means, que es un algoritmo que realiza una tarea parecida a lo que se obtiene con el recocido simulado.

Ahora se dará una breve explicación acerca del funcionamiento del algoritmo k-means, para después pasar a el concepto de *Clasificador Bayesiano Simple*.

Algoritmo k-means

El algoritmo k-means es un método de agrupamiento, que tiene como objetivo la partición de un conjunto n en k grupos en el que cada objeto pertenece al grupo más cercano a la media.

El proceso de agrupamiento del algoritmo k-means es simple. Inicialmente se determina el número de grupos k y se asume el centroide o centro de esos grupos. Para determinar los centroides hay dos alternativas prácticas: la primera es tomar de forma aleatoria K objetos como centroides iniciales y la segunda es tomar los primeros K objetos en secuencia.

Luego el algoritmo ejecuta los siguientes tres pasos hasta que alcance el criterio de convergencia, es decir que los objetos no se muevan del grupo. Los pasos son los siguientes:

1. Se determina el o los centroides iniciales de acuerdo al número de cluster esperado.

2. Se determina la distancia de cada objeto con relación a los centroides.
3. Se agrupan los objetos con base en la distancia mínima.

La figura siguiente muestra una diagrama de flujo de cómo es que funciona el algoritmo k-means.

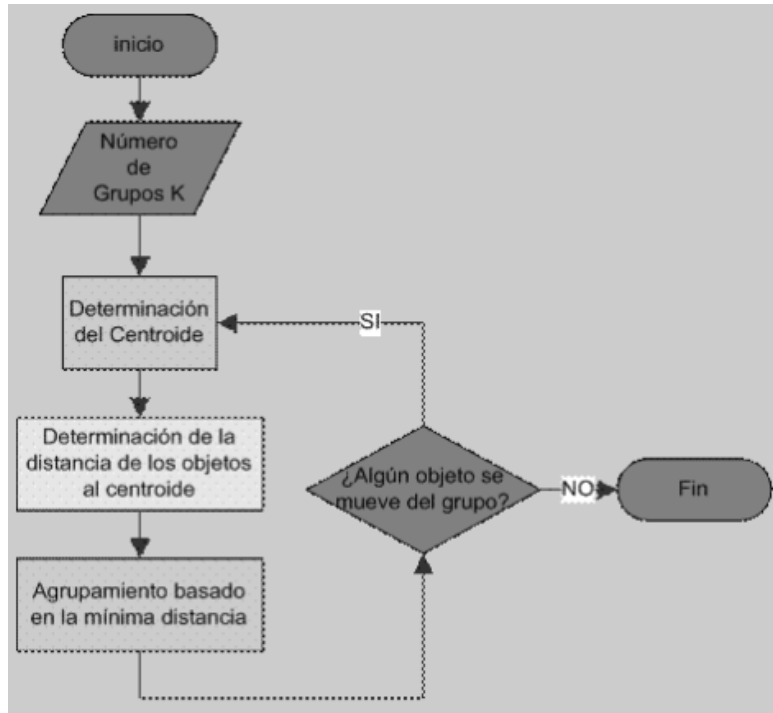


Figura 5.1-1 Diagrama de flujo del algoritmo k-means

Clasificador Bayesiano Simple

A continuación se expondrá las expresiones matemáticas necesarias para poder entender el concepto de *Clasificador Bayesiano Simple*, que es un concepto necesario para poder plantear el planteamiento general y además que nos

permite simplificar la regla de Bayes y por lo tanto también simplifica el algoritmo de recocido simulado.

A partir de la regla de Bayes y la ecuación (1.5-8), es posible escribir la probabilidad total de un evento dado un pixel en particular como:

$$P(w_i|x) = \frac{P(x|w_i)P(w_i)}{P(x)}$$

(5.1-1)

En donde:

$P(w_i|x)$ representa la probabilidad a posteriori

$P(w_i)$ es la probabilidad a priori de la clase w_i

$P(x|w_i)$ es la probabilidad conjunta de X dado w_i

$P(x)$ es la probabilidad total de X

Este último elemento puede despreciarse, ya que representa el mismo valor para un conjunto de n clases, W, W, \dots, W , compitiendo por el vector X a clasificar. Se trata de un factor de escala que se elimina por no incorporar información discriminante.

Dándonos como resultado la aproximación de la regla de Bayes.

$$P(w_i|x) \doteq P(x|w_i)P(w_i)$$

(5.1-2)

Estimar la probabilidad $P(x|w_i)$ de la ecuación (5.1-2) es complejo, pero se simplifica si se considera que los atributos son independientes dada la hipótesis:

$$P(w|x_1, x_2, \dots, x_n) = \frac{P(w)P(x_1|w)P(x_2|w) \dots P(x_n|w)}{P(x)}$$

(5.1-3)

A la ecuación (5.1-3) se le conoce como el *Clasificador Bayesiano Simple*. Como se mencionó anteriormente, no es necesario calcular el denominador, por lo que se puede aproximar como:

$$P(w|x_1, x_2, \dots, x_n) \sim P(w)P(x_1|w)P(x_2|w) \dots P(x_n|w)$$

(5.1-4)

$P(w)$ se conoce como la probabilidad a priori.

$P(x_n|w)$ es la probabilidad de los atributos dada la hipótesis.

$P(w|x_1, x_2, \dots, x_n)$ es la probabilidad a posteriori.

A través de aplicar los campos aleatorios de Markov a la aproximación de la regla de Bayes se llega a la siguiente expresión:

$$U(Y|X) \doteq U(X|Y) + U(Y)$$

(5.1-5)

El objetivo de la expresión (5.1-5) es minimizar el recocido simulado.

Ahora que ya se cuenta con las herramientas matemáticas necesarias, es posible exponer el planteamiento y el esquema general del recocido simulado.

Considerando la ecuación (5.1-1) de la regla de Bayes $P(w_i|x) = \frac{P(x|w_i)P(w_i)}{P(x)}$, siendo X la imagen original y W_i las clases con $i=1,2,\dots,5$.

En donde:

$P(w_i|x)$ representa la probabilidad a posteriori

$P(w_i)$ es la probabilidad a priori de la clase w_i

$P(x|w_i)$ es la probabilidad conjunta de X dado w_i

$P(x)$ es la probabilidad total de X

Sobre la imagen de prueba provista por el satélite ENVISAT

- Se definen 2 ventanas que identifiquen las zonas de petróleo y las zonas que son mar.
 - Sobre cada ventana se calcula las medias μ_i .
2. Funciones de energía

Se aproximan las probabilidades conjuntas del numerador de la regla de Bayes mediante las funciones de energía mostradas en la figura (5.2-2). Ellas indican la *probabilidad condicional* de un píxel dado pertenezca a la clase W_i . Se emplea el modelo de Potts para el cálculo de la energía *a priori* de clase. Considerando los parámetros $T_i = 2.5$ y $\beta = 0.35$.

3. Clasificar la imagen X aplicando 60 iteraciones

- Empleando la aproximación de la regla de Bayes $P(W_i|X) \approx P(X|W_i)P(W_i)$ para calcular la energía a posteriori:

$$U(W_i|X) \approx U(X|W_i) + U(W_i)$$

- Se analiza píxel por píxel la imagen X , clasificando al aplicar el algoritmo de recocido simulado.

4.	Mostra
r los resultados de las iteraciones 1.	
5.	Mostra
r los resultados de las iteraciones 10.	
6.	Mostra
r los resultados de las iteraciones 20.	
7.	Mostra
r los resultados de las iteraciones 40	
8.	Mostra
r los resultados de las iteraciones 60.	

5.2 FUNCIONES DE ENERGÍA

Construcción de la función de energía

Tal y como se definió en la ecuación (2.2-7) la función de energía está dada por la siguiente expresión:

$$U(f) = \sum_c V_c(f)$$

Se considera entonces que $V_c(f)$ es el potencial del clique, y que $U(f)$ es la suma de todas las funciones potenciales de los cliques, $V_c(f)$ para todos los cliques posibles, donde $V_c(f)$ depende de la configuración local del clique c , considerándose la distribución Gaussiana como un caso especial de la familia de distribuciones de Gibbs.

La función de energía permite la clasificación de cada píxel de la imagen original, esto es, decide la pertenencia de un píxel a una clase específica definida dentro de la imagen de la segmentación.

La función de energía utilizada para la segmentación Markoviana se especifica gráficamente en la figura (5.2-1). A continuación se describe el procedimiento de la construcción de las funciones de energía $U(x|w_i)$.

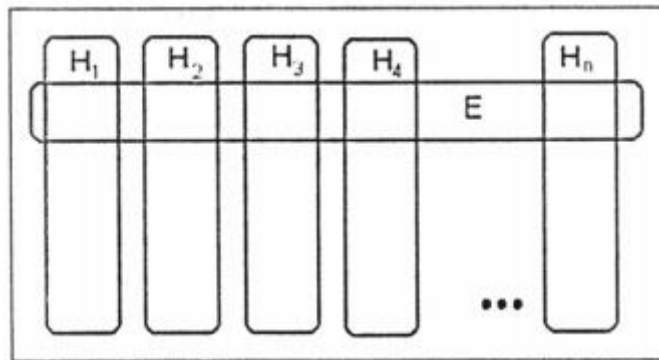


Figura 5.2-1 Representación de los conjuntos $H_i (i = 1, \dots, n)$ y E . H_i , son mutuamente excluyentes y E no es vacío.

El nivel de gris de cada píxel se proyecta sobre la función a la que pertenezca el píxel analizado, es decir, la clase w_i . El número de funciones está determinado por el número de clases definidas previamente.

Para la construcción de las funciones de energía se definen las siguientes variables:

- y:** Es la energía condicional.
- x:** Es el píxel analizado.
- ma:** Es la media de la clase analizada

S: Es la diferencia de las medias de clase. Dependiendo del caso, $S=m(\text{sig})-m(\text{act})$ ó $S=m(\text{act})-m(\text{ant})$.

m(ant): Es la media de clase anterior a la actual.

m(act): Es la media de la clase actual.

m(sig): Es la media de la clase siguiente a la actual

Pendiente positiva (clase actual)	Pendiente negativa (clase siguiente)
Función (A)	Función (B)
$y = \frac{2}{S}x - \frac{2}{S}m_a - \frac{1}{2}$	$y = -\frac{2}{S}x + \frac{2}{S}m_a + \frac{3}{2}$

- Si el nivel de gris del píxel analizado (x) está en el rango: $0 < x < m_0 + (S/4)$, la energía condicional para la clase W_0 (clase inicial) es de 0. Para las demás clases la energía será de 1.
- Si el nivel de gris del píxel analizado (x) está en el rango: $x > m_n$, la energía condicional para la clase W_n (última clase) es de 0. Para las demás clases la energía será de 1.
- Si el nivel de gris del píxel analizado (x) no está en alguno de los casos anteriores se realiza la diferencia entre las medias de las clases de la siguiente manera $S = m_x - m_a$, donde m_x es la media de la clase siguiente y m_a es la media de la clase actual. El resultado de esta diferencia se divide entre 4, es decir, $S/4$.
- Si el nivel de gris del píxel analizado (x) está en el rango: $m_a < x < m_a + (S/4)$, la energía condicional para la clase W_a (clase actual) es de 0. Para las demás clases la energía será de 1.
- Si el nivel de gris del píxel analizado (x) está en el rango: $m_a + (S/4) < x < m_a + (3/4)S$ y pertenece a la clase inicial W_0 , la energía condicional para la clase W_0 (clase inicial) se determina a través de la función (B), para ello se utilizará la diferencia de medias de la clase actual menos la clase anterior, es decir: $S=m(\text{act})-m(\text{ant})$.

Para cualquier otra clase, si el píxel analizado (x) está en el rango $m(\text{ant})+(S/4)<x<m(\text{ant})+(3/4)S$ se utiliza la función (B) y la diferencia de medias $S=m(\text{act})-m(\text{ant})$. Si el píxel analizado (x) está en el rango $m(\text{act})+(S/4)<x<m(\text{act})+(3/4)S$ se utiliza la función (A) y la diferencia de medias $S=m(\text{sig})-m(\text{act})$.

Las gráficas se muestran en la figura 5.2-2.

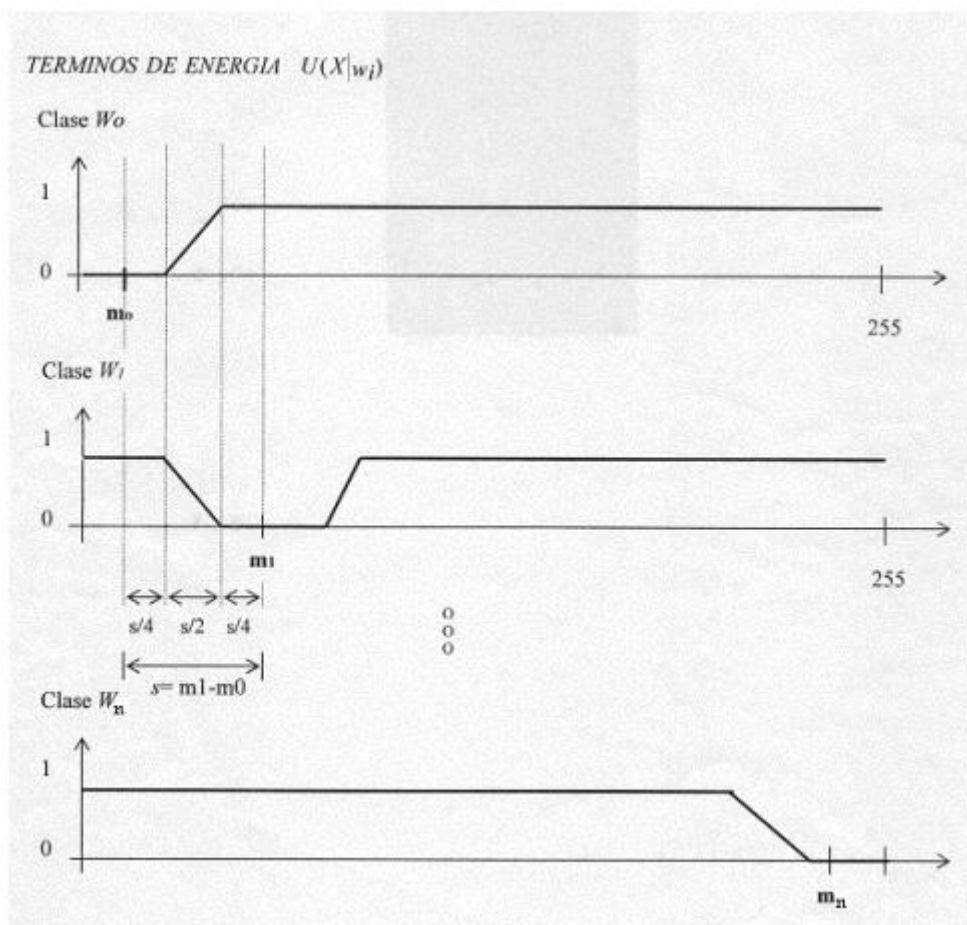


Figura 5.2-2 Funciones de energía para cada media de clase.

La energía condicional obtenida anteriormente se suma con la probabilidad de la clase en cuestión esto se aplica en el modelo de Potts. Esto permite calcular la variación de la energía local asociada al *site* r (píxel analizado).

Los métodos bayesianos asociados a la modelización Markoviana proporcionan una función de energía no convexa que debe ser minimizada. La aplicación del recocido simulado en la optimización de la función de energía a posteriori asegura la convergencia hacia el mínimo global y evita la convergencia a los mínimos locales.

5.3

MIZACION

MINI

Algoritmo de Recocido Simulado

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 aleatoria originada en el inicialización $Y(0)$. La figura (5.3-1) muestra la esquematización del algoritmo de simulación, para el caso de n iteraciones:

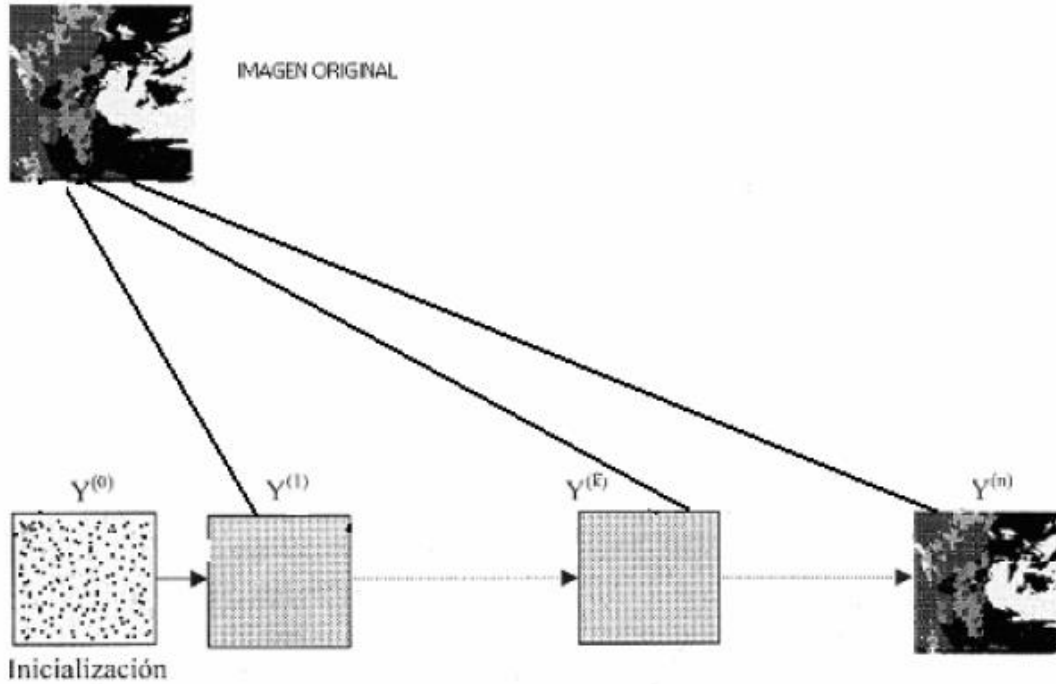


Figura 5.3-1 Obtención de la imagen fusionada a partir de una inicialización aleatoria.

1. Inicialización

- Determinación aleatoria de la solución $Y(0)$ (la clase de cada píxel debe determinarse mediante un generador de números aleatorios, con distribución uniforme en el rango del número de clases $[Z]$).
- Escoger una temperatura inicial $T(0)$ suficientemente elevada (por ejemplo, $T(0)=2.5$).

2. Hacer un contador con un número específico de iteraciones $N = 1 \dots n$

2.1 Barrer secuencialmente la imagen

En cada etapa (k)

- Escoge un píxel g de coordenadas $[i][j]$, su valor es $Y_g^{(k)}$ y su configuración de vecindad es $V_g^{(k)}$.
- Efectúa el sorteo de una variable aleatoria (clase “elección”) con una distribución uniforme en el rango del número de clases $[Z]$ (solo se varia la elección y se compara contra la actual).

2.2 Criterio de Metrópolis y modelo de Potts

-Se utiliza el algoritmo de Metrópolis de temperatura variable con el fin de minimizar la energía $U(X|Y)$ ó $U(X|W)$.

-Para el caso de la energía $U(Y)$ ó $U(w_n)$ se considera el modelo de Potts, el cual se aplica en vecindades en 8-conectividad.

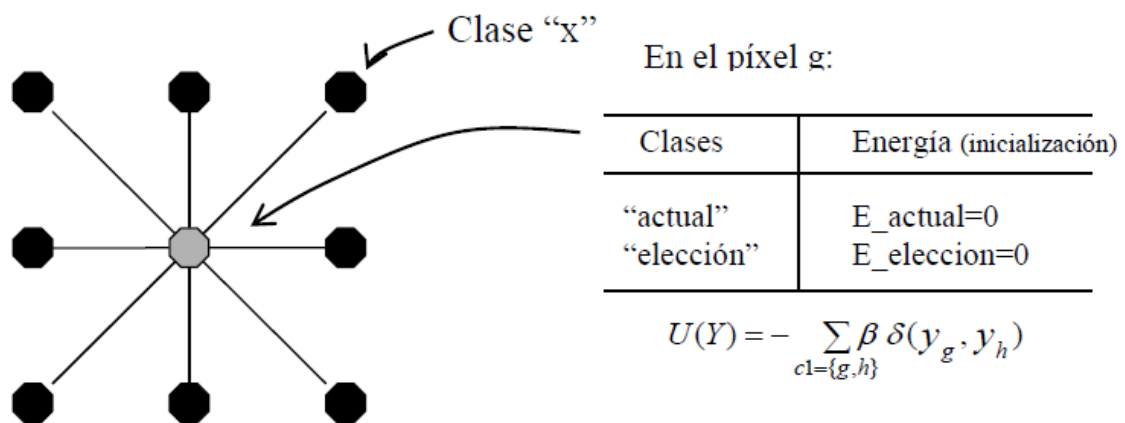


Figura 5.3-2 Vecindad $V_g^{(k)}$ asociada al píxel g (al centro de la venta en 8-conectividad).

Análisis en 8-conectividad (modelo de Potts) para el píxel $g [i][j]$ en la imagen $Y^{(k)}$

- “actual”=”x”: $E_{actual}=E_{actual}-\beta[actual]$ Si
- “elección”=”x”: $E_{elección}=E_{eleccion}-\beta[elección]$ Si

2.2 Calcular la variación de la energía local asociada al píxel g

En la imagen X, en las coordenadas [i][j], tomar el nivel de gris del píxel analizado g y proyectarlo en las funciones de la figura (5.2-2), en las clases “actual” y “elección”. Sumar la contribución a la de Potts:

- $l=E_{actual}+potencial[actual][i][j]$ E_{actual}
- $ón=E_{elección}+potencial[elección][i][j]$ E_{elecci}

Donde $potencial[actual][i][j]$ representa a la energía condicional $U(X|Y)$ e $[i][j]$ son las coordenadas del píxel analizado.

- $E_{elección} \leq E_{actual}$: $Y[i][j]=\text{“elección”}$ Si

En caso contrario:

- $p = \exp - [(E_{elección} - E_{actual}) / T]$
- de una variable aleatoria ζ_2 , uniformemente distribuida en el rango [0,1] Sorteo
- $\zeta_2 < p$, entonces $Y[i][j]=\text{“elección”}$ Si
- $Y[i][j]=\text{“actual”}$ Si no

3. Regresar al punto 2.1 hasta terminar un barrido

4. Decrementar la temperatura:

$$T(n) = T(n - 1) + 0.95$$

5. Regresa
r al punto 2 hasta terminar el número de iteraciones fijadas.
-

6 RESULTADOS

6.1 IMÁGENES DE PRUEBA

A continuación se muestra la imagen original de trabajo proporcionada por el satélite ENVISAT, es una imagen de 700x650 píxeles, en la que se aprecia una zona del mar cubierta por petróleo, debido al derrame ocurrido en 2010. En la imagen además de las zonas de petróleo y mar también hay ruido que pareciera dividir la imagen en tonos más claros para minimizar este ruido lo más posible es necesario aplicarle una difusión anisotrópica para hacer la imagen lo más uniforme posible y la detección entre las zonas de petróleo y las zonas de mar sean precisas.



Figura 6.1-1 Imagen original proporcionada por el satélite ENVIASAT/ASAR.

Tal y como se mencionó en el capítulo anterior se realizaron pruebas en la difusión anisotrópica para observar cual valor de delta es el que proporciona una imagen con una buena calidad, para después hacer una comparación con el software *ImageJ*.

Difusión Anisotrópica

Para la primera prueba se le pone a delta un valor de 0.01, un valor relativamente bajo, considerando que el valor recomendado es de 1/7. La figura (6.1-2) muestra la imagen con un valor de delta de 0.01.



Figura 6.1-2 Imagen con difusión anisotrópica y $\delta=0.01$

En la imagen se aprecia que la imagen es casi idéntica a la imagen original mostrada en la figura (6.1-1), como el objetivo de la difusión es hacer la

imagen más uniforme, es necesario incrementar el valor de delta para obtener un mejor resultado.

Para la segunda prueba se pone un valor $\delta=0.1$, un valor cercano al valor óptimo. En la figura (6.1-3) se muestra la imagen.

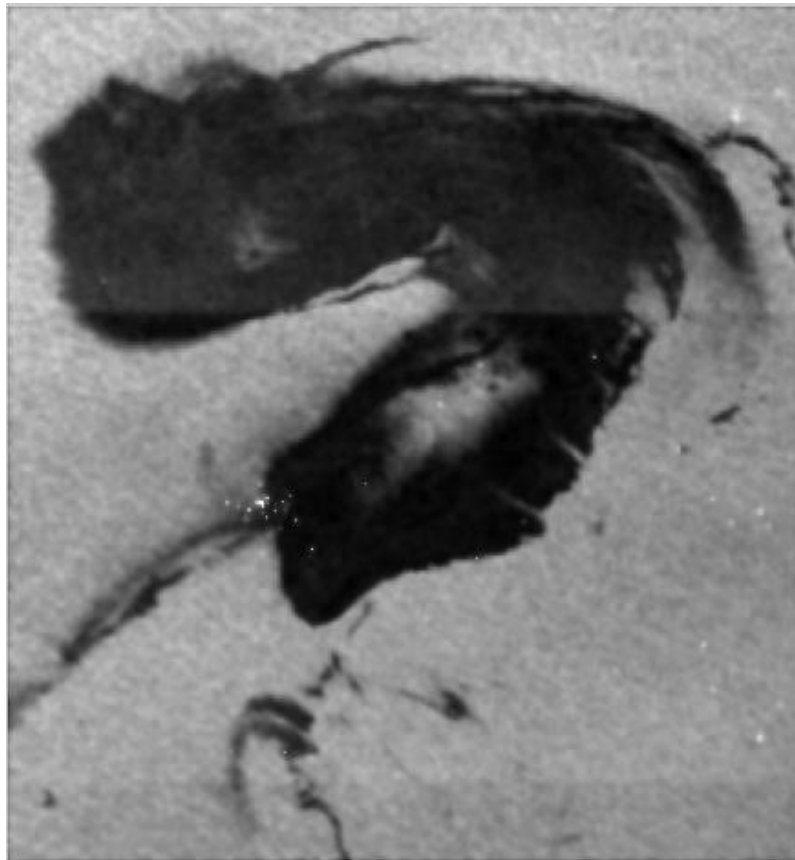


Figura 6.1-3 Imagen con difusión anisotrópica y $\delta=0.1$

En esta prueba claramente se observa que la imagen se ha suavizado bastante comparándola con la figura (6.1-2), la imagen es más uniforme sin embargo la franja que divide la imagen persiste, por lo que hay incrementar delta un poco más.

La figura (6.1-4) muestra la imagen la tercera prueba con $\delta=1/7$.

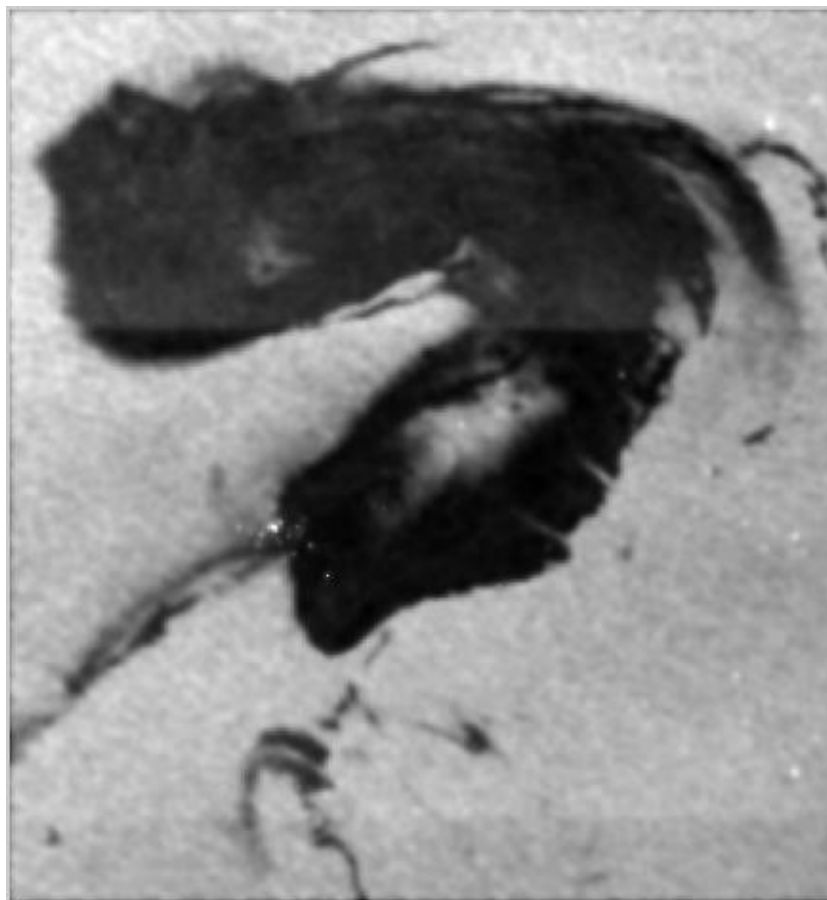


Figura 6.1-4 Imagen con difusión anisotrópica con $\delta=1/7$

Ahora la imagen está mucho más suavizada que la figura anterior, este es valor que se dejó para los algoritmos siguientes, debido a que este es el más recomendado y porque si el valor de δ se incrementa más la imagen comienza a perder claridad cómo se observara en la imagen siguiente.

Y finalmente la figura (6.1-5) muestra la imagen para la última prueba con $\delta=0.5$



Figura 6.1-5 Imagen con difusión anisotrópica y $\delta=0.5$

En esta última prueba se observa que la imagen se ha suavizado demasiado, es más difícil diferenciar las zonas que son petróleo de las zonas de mar. Es por eso que se considera a $\delta=1/7$ para las futuras pruebas.

Ahora que se ha seleccionado un valor de delta que hace la imagen más uniforme, es posible ahora aplicarle a esta imagen el algoritmo de recocido simulado.

En el capítulo anterior se mencionó que se mostrarían las imágenes en las iteraciones 1, 10, 20, 40 y 60. Por lo tanto a continuación se mostrarán las cinco imágenes correspondientes a las iteraciones y una imagen más donde se muestra los pixeles en sus respectivas clases.

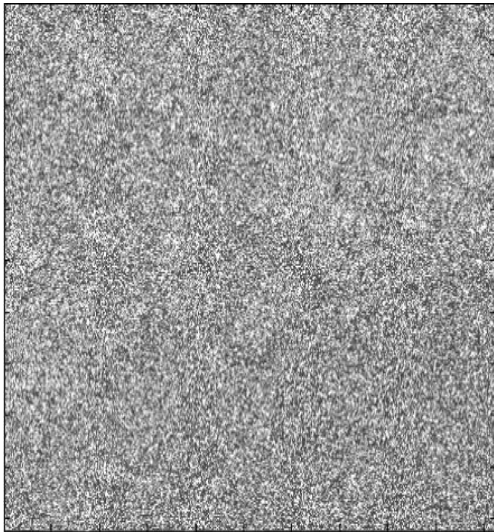


Figura 6.1-6 Imagen con 1 iteración



Figura 6.1-7 Imagen con 10 iteraciones

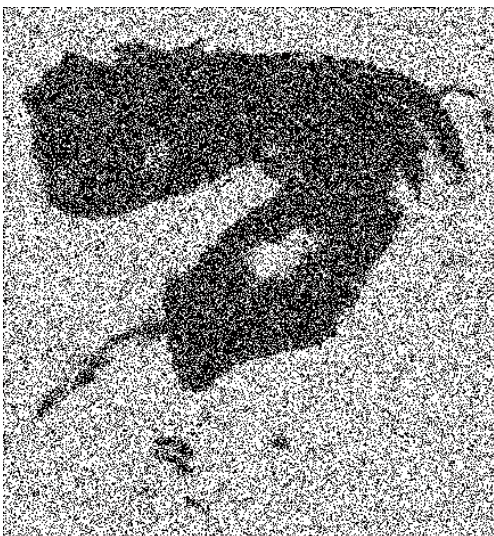


Figura 6.1-8 Imagen con 20 iteraciones



Figura 6.1-9 Imagen con 40 iteraciones

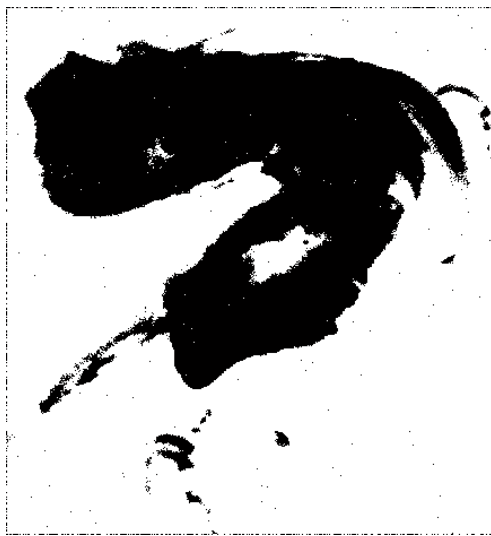


Figura 6.1-10 Imagen con 60 iteraciones

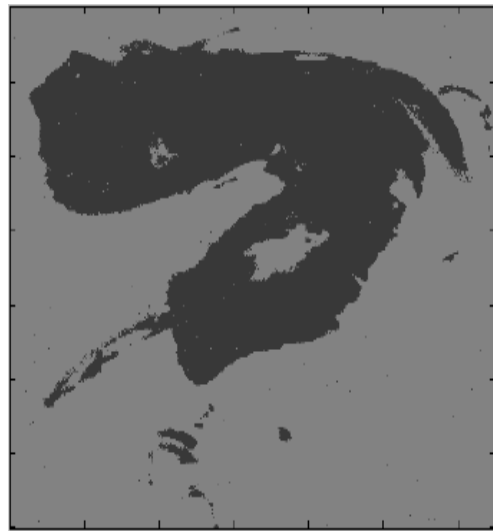


Figura 6.1-11 Imagen con tonos de gris

La figura (6.1-11) representa lo mismo que la figura (6.1-10) solo que la imagen (6.1-11) tiene los mismos tonos de gris que las medias seleccionadas en el algoritmo. Para el algoritmo se seleccionaron dos medias, ambas medias se tomaron de la figura (6.1-4), una media representa el tono de gris que tiene la zona del petróleo, mientras que la otra media representa el tono de gris que tiene la zona de mar, ambos tonos de gris después se asignan a una clase específica (clase 1 y clase 2). Ahora bien la figura (6.1-11) únicamente toma el nivel de gris que hay tanto en la zona de petróleo como en la zona de mar y los reemplaza por los valores binarios de la figura (6.1-10) (imagen en blanco y negro).

La figura (6.1-12) lo que muestra es como esta segmentada la imagen en clases, los píxeles que se encuentran de color negro pertenecen a la clase uno, representan la zona de petróleo y tienen un tono de gris de 52 y, que es el mismo tono que tiene la figura (6.1-11), para la zona de petróleo. Los píxeles de color rojo pertenecen a la clase dos, representan la zona de mar y tiene un tono de gris de 136 e igual que el tono anterior es el mismo que el de la figura (6.1-11).

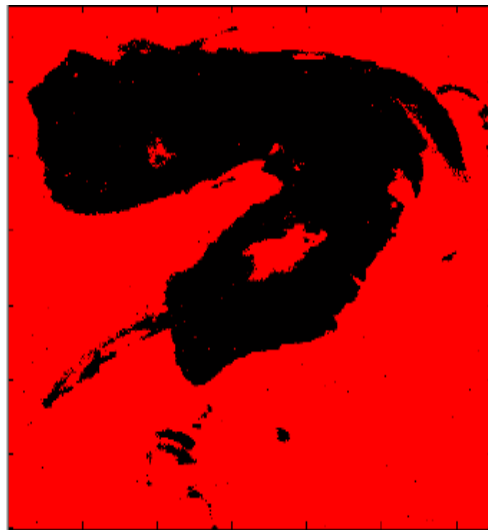


Figura 6.1-12 Imagen de clases

6.2 COMPARACION DE FUNCIONALIDAD

Primero se hará una comparación de lo obtenido con la difusión anisotrópica (con $\delta=1/7$) con el software ImageJ. Para comparar si el suavizado que realiza el software es más o menos intenso que el suavizado que se obtiene al realizar el algoritmo en Matlab.

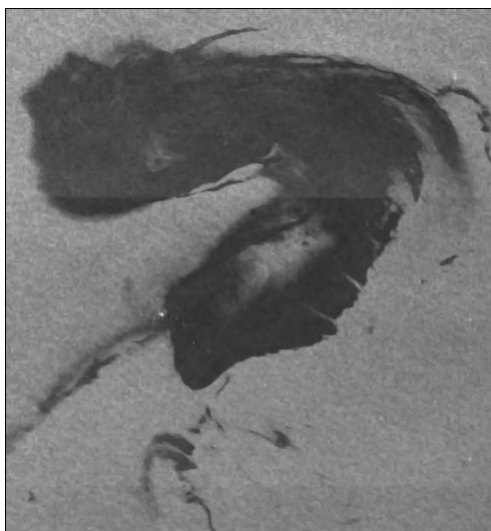


Figura 6.2-1 Imagen con difusión anisotrópica obtenida con ImageJ

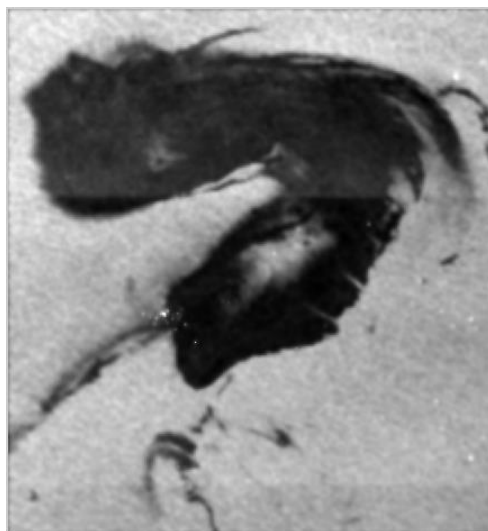


Figura 6.2-2 Imagen con difusión anisotrópica obtenida con Matlab.

Como se observa en la figura (6.2-1) la imagen tiene un suavizado apenas distinto de la imagen original (incluso con valor $\delta=0.5$ el suavizado no es tan notorio como lo es en la figura (6.2-2)), esto quizás se deba al número de iteraciones que se están usando, el valor de κ , el cual no se sabe cuál, es debido a que no aparece en el programa y tampoco es un parámetro que pueda ser modificado y sobre todo no se sabe los valores de los coeficientes que se están usando para el filtrado.

Y como el objetivo de hacer una difusión anisotrópica era suavizar la imagen para hacer menos notoria esa franja que divide a la imagen, si se usara la imagen de la figura (6.2-1), la detección sería más complicada y pudiera haber errores de las detecciones. Es por esta razón que se usó la imagen de la figura (6.2-2) y ya que aunque suavizado es más intenso la división entre los tonos de gris es menos notoria que en la figura (6.2-1)

Ahora se hace una comparación una comparación entre el resultado obtenido usando el algoritmo de recocido simulado en Matlab y lo que se obtiene usando un algoritmo que se conoce como **k-means**.



Figura 6.2-3 Imagen segmentada obtenida con el algoritmo de k-means

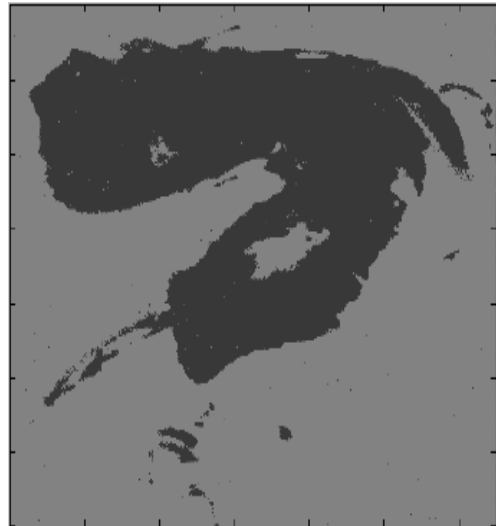


Figura 6.2-4 Imagen segmentada obtenida con el algoritmo de recocido simulado

En la figura (6.2-3) se observa que la imagen tiene mayor definición en los tonos de gris de ambas zonas, se está segmentando la imagen en dos clases al igual que en el caso de la figura (6.2-4), se logró eliminar el ruido que presentaba la imagen original y que daba la apariencia de tener una franja que dividía la imagen. Sin embargo la figura (6.2-3) tiene píxeles aislados que parecerían pertenecer a la zona de petróleo, pero esta dispersa por toda la zona de mar, lo cual hace complicado diferenciar donde termina la zona de petróleo y empieza la zona de mar.

Por otro lado la figura (6.2-4), presenta tonos de gris más tenues que la figura (6.2-3), en este caso únicamente se tomó un valor de media para cada clase y a partir de estas medias se hicieron los cálculos, a diferencia del algoritmo de k-means que recalcula las medias con cada iteración y por lo tanto se tiene una mayor exactitud en los tonos de gris. Sin embargo los pixeles dispersos no son tan abundantes como lo es en el caso de la figura (6.2-3), si existen sin embargo son menos y son casi imperceptibles, es por esta razón que se considera que el algoritmo de recocido simulado entrega un resultado más confiable que el algoritmo k-means.

Conclusiones

A lo largo de este trabajo de tesis se trataron diferentes temas, los cuales trataron varios temas como la segmentación de imágenes, el algoritmo de recocido simulado y la difusión anisotrópica principalmente. Inicialmente se las definiciones de las herramientas matemáticas y de probabilidad necesarias para entender los temas siguientes; después en el capítulo dos se trató la teoría de los Campo Aleatorios de Markov, su relación con los Campos Aleatorios de Gibbs, para después pasar a conocer las expresiones matemáticas que definen sus funciones de energía, necesarias para el algoritmo del recocido simulado.

En el capítulo tres se enfocó tratar de entender, primero la teoría detrás de los filtros lineales y no lineales, particularmente el enfoque se da en los filtros no lineales, ya que la difusión anisotrópica es una clase de filtro no lineal, después se explica la ecuación diferencial de calor, la cual es base para el algoritmo, finalmente se describe la discretización de esta ecuación y sus parámetros para así poder implementarlo y simularlo.

El capítulo cuatro trata temas con relación a la optimización, se explora con mayor profundidad el algoritmo de recocido simulado, pero ahora ya aplicado a imágenes de interés, además de que se explica también el esquema de Metrópolis, que ayuda al algoritmo de recocido simulado a obtener mejores resultados.

Finalmente en los capítulos seis y siete, se define claramente el problema y la solución propuesta. En el capítulo seis se define el problema y se plantea la solución al problema, comenzando primero con la difusión anisotrópica para suavizar la imagen y hacerla más uniforme. Una vez que la imagen ha sido suavizada es posible aplicarle el algoritmo de recocido simulado que con ayuda del modelo de Potts nos ofrece un mejor resultado.

Y en el capítulo siete se presenta la solución al problema planteado en el capítulo anterior. Se colocan prueba en ambos algoritmos para poder observar claramente con que valores se obtienen mejores resultados, pero además de las pruebas con los algoritmos, se realizan comparaciones con paquetes de software para que, de esa manera haya un punto de comparación entre los resultados que se están obteniendo mediante lo simulado con Matlab y lo que se obtiene con otro tipo de software.

Finalmente puedo concluir que existen diferentes herramientas que nos ayudan obtener el mismo resultado deseado, cada una de estas herramientas tienen sus ventajas y sus desventajas, en este trabajo de tesis únicamente se exploraron algunas, existen más para muchos otros fines, por lo que para elegir una de ellas es necesario considerar las condiciones de trabajo, el tiempo que se tenga, etc. La complejidad de estos algoritmos depende mucho de los fines para cuales vayan a ser usados, sin embargo todo esto influye de manera directa ya que al aumentar la complejidad matemática, aumenta también el tiempo de procesamiento de cómputo, pero también el resultado es mucho más preciso y exacto. Y tal como se ha visto en este trabajo de tesis no se debe realizar únicamente una sola prueba, se deben hacer varias para así obtener los resultados deseados, y también ayuda bastante tener un punto de comparación, para que de esa manera se puedan ver los mismos resultados desde otra perspectiva.

Bibliografía

- [1] Peyton Z. Peebles, Jr. “Principios de probabilidad, variables aleatorias y señales aleatorias”, Mc Graw Hill, cuarta edición, 2006.
- [2] Montgomery, Douglas, “Probabilidad y Estadística aplicadas a la Ingeniería”, Mc Graw Hill,
- [3] Wiley, J, “Simulated Annealing: Parallelization techniques”, 1992.
- [4] Aarts, E y Korst, Wiley, J, “Simulated Annealing and Boltzmann Machines”, 1989.
- [5] Sheskin, Theodore, J. “Markov chains and decisión process for engineers and managers”, CRC Press, 2011
- [6] Stroock, W. Daniel, “An introduction to Markov Processes”, Springer Verlaq, 2005
- [7] Juan E. Ortuño, Noberto Malpica, Manuel Desco, Filtros de Difusión Anisotrópica 3-D aplicados a imágenes de resonancia magnética cerebral, [en línea], Madrid, Universidad Politécnica de Madrid, 2002, Formato html, Disponible en Internet:
<http://www2.die.upm.es/im/papers/JEOCaseib02.pdf>