

Introducción a los procesos markovianos en el análisis y procesamiento de imágenes

Oscar H. Bustos

bustos@famaf.unc.edu.ar

CIEM - FaMAF - Universidad Nacional de Córdoba

Resumen

En una primera parte veremos algunos conceptos básicos referentes a la formación de imágenes de teledetección, principalmente desde sensores ubicados en plataformas espaciales. Enseguida definiremos los parámetros esenciales de los sensores en general, como distintos tipos de resolución, el significado de píxel, etc. Finalizaremos esa primera parte viendo las características principales de los sensores más usados en las múltiples e importantes aplicaciones de la Teledetección para conocer y cuidar la vida de nuestra única casa: la Tierra. Desde el punto de vista de la Estadística, en Análisis de Imágenes, se trata de encontrar la distribución subyacente a la energía de retorno captada por el sensor basándose en la distribución de la misma en los entornos de cada píxel. Desde la década del 80 del siglo pasado, muchas técnicas basadas en el uso de procesos markovianos han demostrado ser muy útiles en el análisis y procesamiento de imágenes, no sólo de las satelitales sino de toda clase de imágenes. Una buena parte de los aspectos formales del tema de nuestro interés está contenida en una rama de la Teoría de Probabilidad de creciente y rápido desarrollo que algunos llaman sistemas espaciales aleatorios con interacciones. Uno de sus objetivos importantes es la comprensión de los efectos interactuantes en sistemas aleatorios espaciales. El impulso inicial proviene de la Física Estadística. Posiblemente el antecedente más remoto de la aplicación de tales procesos es el célebre modelo de Ising que fue propuesto como una forma de explicar el fenómeno de magnetización espontánea que muestran los materiales ferromagnéticos. La cantidad de trabajos que usan procesos de Markov en el procesamiento de imágenes viene creciendo en forma exponencial en los últimos años. En este curso nos limitaremos a presentar los elementos y resultados básicos de los procesos de Markov orientados a su aplicación en el análisis y procesamiento de imágenes. Lo haremos simplificando en la mayor medida posible, la matemática involucrada en los mismos. No obstante, lo que aquí veremos será suficiente para iniciarse en el estudio del tema.

1 Introducción

En esta sección veremos algunos conceptos básicos referentes a la formación de imágenes de teledetección, principalmente desde sensores ubicados en plataformas espaciales. Definiremos los parámetros esenciales de los sensores en general, como distintos tipos de resolución, el significado de píxel, etc. Finalizaremos la sección viendo las características principales de los sensores más usados en las múltiples e importantes aplicaciones de la Teledetección para conocer y cuidar la vida de nuestra única casa: la Tierra.

1.1 Principios físicos de la teledetección

El término teledetección es una traducción del inglés (Remote Sensing), se define como el conjunto de conocimientos y técnicas utilizados para determinar características físicas y biológicas de objetos y superficies mediante mediciones a distancia, sin el contacto material con los mismos. La teledetección no engloba sólo los procesos que permiten obtener una imagen desde el aire o el espacio, sino también su posterior tratamiento en el contexto de una determinada aplicación. En otras palabras, **la teledetección se define como la técnica de adquisición y posterior tratamiento de datos de la superficie terrestre desde sensores instalados en plataformas espaciales, en virtud de la interacción electromagnética existente entre la Tierra y el sensor, siendo la fuente de radiación ya sea proveniente del sol (teledetección pasiva) o del propio sensor (teledetección activa).**

Veamos los componentes de un sistema de teledetección.

Los componentes básicos de un sistema de teledetección son: la escena, el sensor y el tratamiento de la información, con los siguientes elementos.

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

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

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

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

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

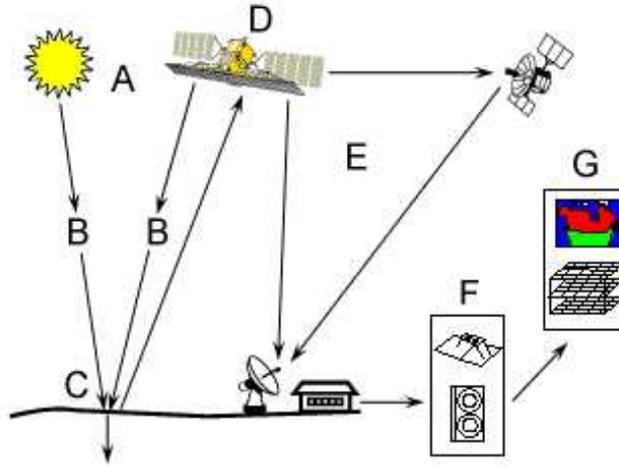


Figura 1: Componentes de un sistema de teledetección espacial: (A) Fuente de energía. (B) Atmósfera. (C) Suelo. (D) Sensor. (E) Transmisión, y procesamiento. (F) Análisis. (G) Aplicación.

Interpretación y análisis (F) - En esta etapa el intérprete, analiza la información.

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

Habitualmente se asume que la formación de imágenes en teledetección se basa en el uso de la radiación electromagnética para adquirir la información a distancia, esto es partiendo de sensores remotos.

Las propiedades de la radiación electromagnética se han explicado históricamente a través de dos teorías: 1) la que la concibe como un haz ondulatorio (Huygens, Maxwell), y 2) la que la considera como una sucesión de unidades discretas de energía, fotones o cuantos, con masa igual a cero (Planck, Einstein). Actualmente se ha demostrado que la luz puede comportarse de acuerdo a ambas teorías. Según la teoría ondulatoria, la energía electromagnética se transmite de un lugar a otro en forma de una onda continua a la velocidad de la luz y conteniendo dos campos de fuerzas ortogonales entre sí: eléctrico y magnético.

El flujo energético conforme a esta teoría puede describirse por dos elementos: longitud de onda (λ) y frecuencia (F). La primera hace referencia a la distancia entre dos picos sucesivos de una onda, mientras que la frecuencia designa el número de ciclos pasando por un punto fijo en una unidad de tiempo. Ambos

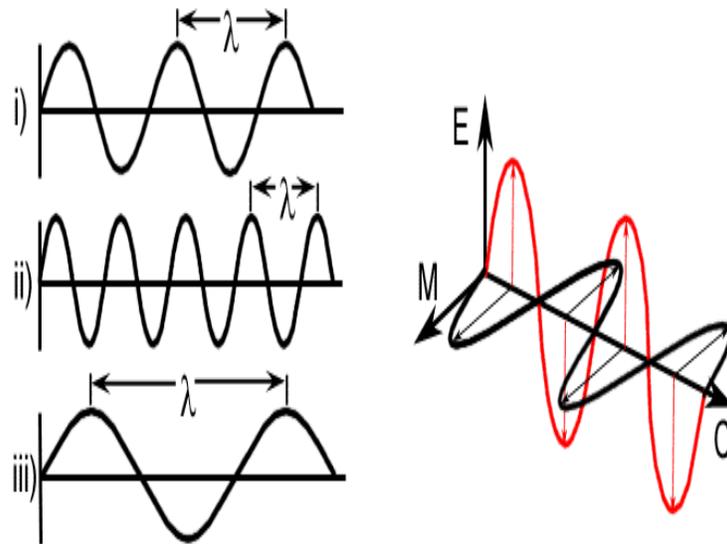


Figura 2: Las ondas eletromagnéticas.

elementos están inversamente relacionados por:

$$c = \lambda.F,$$

donde: c : velocidad de la luz, λ : longitud de onda (habitualmente medida en micrómetros, $1\mu m = 10^{-6}m$), y F : frecuencia (medida en Hertzios, ciclos por segundo).

Veamos algunos términos y unidades de medida de uso habitual en procesamiento de imágenes.

Energía radiante. Indica el total de energía radiada en todas las direcciones.

Densidad radiante. Total de energía radiada en todas las direcciones por unidad de volumen.

Flujo radiante. Total de energía radiada en todas las direcciones por unidad de tiempo.

Emitancia o excitancia radiante. Total de energía radiada (emitida) en todas las direcciones desde una unidad de área y por unidad de tiempo.

Irradiancia radiante. Total de energía radiada (incidente) sobre una unidad de área y por unidad de tiempo.

Intensidad radiante. Total de energía radiada por unidad de tiempo y por ángulo sólido. Se trata éste de un ángulo tridimensional, que refiere a la sección completa de la energía transmitida.

Radiancia. Total de energía radiada en una determinada dirección por unidad de área y por ángulo sólido de medida. Describe precisamente lo que mide el sensor.

Radiancia espectral. Indica el total de energía radiada en una determinada longitud de onda por unidad de área y por ángulo sólido de medida. También la radiancia, la emitancia e irradiancia pueden completarse con el calificativo de espectral cuando se refieren a una determinada longitud de onda.

Emisividad. Relación entre la emitancia de una superficie, y la que ofrecería un emisor perfecto, denominado cuerpo negro, a la misma temperatura.

Reflectividad. Relación entre el flujo incidente y el reflejado por una superficie.

Absortividad. Relación entre el flujo incidente y el que absorbe una superficie.

Transmisividad. Relación entre el flujo incidente y el transmitido por una superficie.

Las últimas cantidades: emisividad, reflectividad, absortividad, transmisividad, son también dependientes de la longitud de onda, en tal caso se completan con el calificativo de espectral, para referirse a su comportamiento en una banda determinada del espectro.

1.2 El espectro electromagnético

Aunque la sucesión de valores de longitud de onda es continua, suelen establecerse una serie de bandas en donde la radiación electromagnética manifiesta un comportamiento similar.

Desde el punto de vista de la teledetección, conviene destacar una serie de bandas espectrales, que son las más frecuentemente empleadas con la tecnología actual:

1) Espectro visible (0.4 a 0.7 μm). Suelen distinguirse tres bandas elementales, que se denominan azul: (0.4 a 0.5 μm), verde (0.5 a 0.6 μm), y rojo (0.6 a 0.7 μm), en razón de los colores primarios que nuestros ojos perciben a esas longitudes de onda.

2) Infrarrojo próximo (0.7 a 1.3 μm). Resulta de especial importancia por su capacidad para discriminar masas vegetales y concentraciones de humedad.

3) Infrarrojo medio (1.3 a 8 μm), en donde se entremezclan los procesos de reflexión de la luz solar y de emisión de la superficie terrestre. Resulta idóneo para estimar contenido de humedad en la vegetación y detección de focos de alta temperatura.

4) Infrarrojo lejano o térmico (8 a 14 μm), que incluye la porción emisiva del espectro terrestre, en donde se detecta el calor proveniente de la mayor parte de las cubiertas terrestres.

5) Micro - ondas (a partir de 1 μm), con gran interés por ser un tipo de energía bastante transparente a la cubierta nubosa.

1.3 Principios y leyes de la radiación electromagnética

Es importante caracterizar espectralmente las distintas cubiertas de interés. Para ello, es imprescindible ver, aunque sea superficialmente, los procesos más relevantes que permiten explicar ese procedimiento. La relación entre flujo de

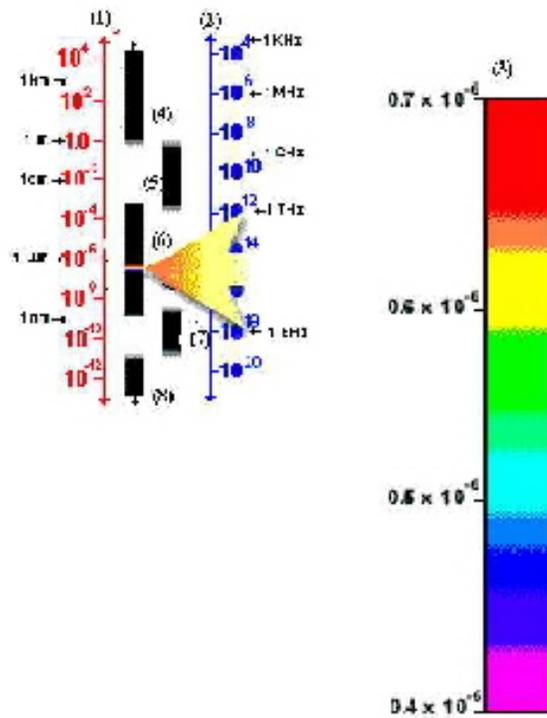


Figura 3: El espectro eletromagnético. (1) Longitud de onda en metros. (2) Frecuencia en Herz. (3) Visible. (4) Ondas de radio. (5) Micro ondas. (6) Infrarrojo. (7) Rayos X. (8) Rayos Gamma.

energía y longitud de onda puede establecerse con mayor rigor gracias a la ley de Planck que nos indica que cualquier objeto por encima del cero absoluto (-273°C) radia energía, y que ésta aumenta con la temperatura. También que, a mayor temperatura, ese cuerpo radiará con más intensidad en longitudes de onda más cortas.

A partir de la fórmula de Planck puede calcularse la longitud de onda a la que se produce la máxima emitancia de un cuerpo negro conociendo su temperatura (T). Se obtiene así la llamada ley del desplazamiento de Wien:

$$\lambda_{\max} = 2898/T \text{ } \mu\text{mK}.$$

Esta fórmula es de gran interés para seleccionar la banda más conveniente para detectar una determinada cubierta, siempre que se conozca su temperatura.

1.4 El dominio óptico del espectro

Se denomina dominio óptico del espectro a aquel grupo de longitudes de onda directamente dependientes de la energía solar.

Veamos ahora cómo la radiación interacciona con la superficie terrestre.

$$\phi_i = \phi_r + \phi_a + \phi_t,$$

donde ϕ_i : flujo incidente, ϕ_r : flujo reflejado, ϕ_a : flujo absorbido, ϕ_t : flujo transmitido.

La proporción del flujo incidente que es reflejado, absorbido y transmitido depende de las características de la superficie que se observa y de la longitud de onda a la que sea observado.

A partir de medidas de laboratorio, se han obtenido unas curvas de reflectividad espectral para las principales cubiertas terrestres:

Ahora bien, el flujo de energía recibido por el sensor no sólo depende de la reflectividad de la cubierta, sino también de otros factores externos. Los más importantes son: (i) las condiciones atmosféricas, (ii) el emplazamiento ambiental de la cubierta y (iii) la geometría de la observación.

En cuanto a las condiciones de observación, conviene tener presente que la cantidad de energía que llega al sensor depende del ángulo con que la superficie refleja la energía incidente, así como del que forman el haz incidente con la posición del sensor. La geometría de la observación está estrechamente relacionada con la rugosidad de la superficie. Podemos distinguir así diferentes tipos de coberturas (ver figura abajo): especulares: reflejan la energía con el mismo ángulo del flujo incidente; lambertianas: reflejan uniformemente en todas las direcciones.

En el caso de las superficies especulares el sensor recibe energía reflejada del suelo sólo si está situado en la dirección del ángulo de reflexión. Si la superficie es lambertiana, la radiancia es constante en cualquier ángulo de reflexión.

La mayoría de las superficies terrestres tienen un comportamiento intermedio entre las especulares y las lambertianas, dependiendo de sus características y de la longitud de onda en la que se trabaja. Si la longitud de onda es pequeña, las

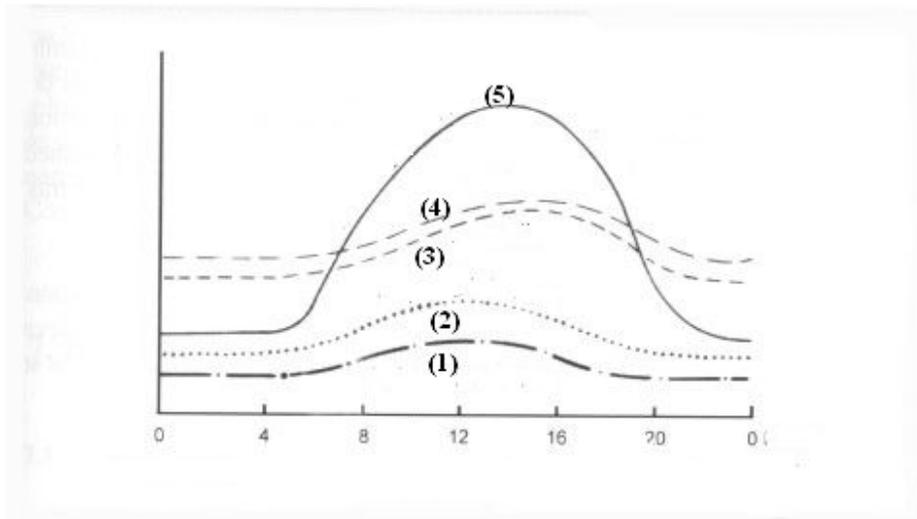


Figura 4: La reflectividad espectral según las cubiertas del suelo. (1) Suelo desnudo. (2) Agua. (3) Vegetación sana. (4) Vegetación estresada. (5) Nieve.

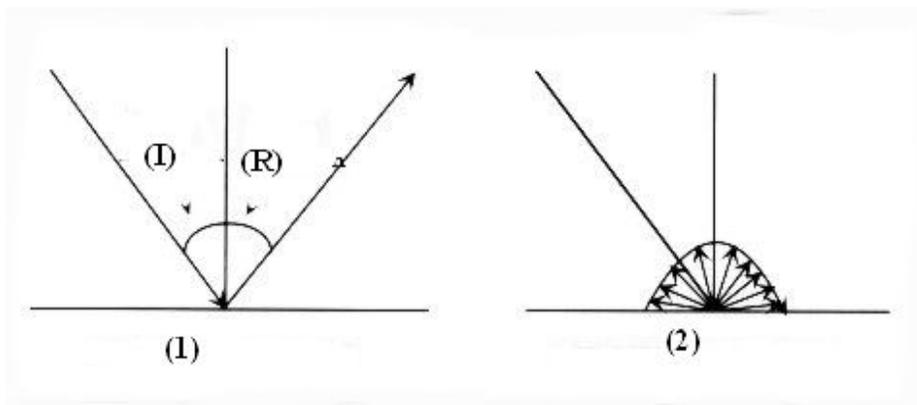


Figura 5: Principales clases de la reflectividad. (1) Reflexión especular, (I) ángulo incidente, (R) ángulo de reflexión. (2) Reflexión lambertiana.

rugosidades del terreno tenderán a difundir mucho más la energía incidente que las longitudes de onda mayores.

El comportamiento de una cubierta, especialmente en el espectro visible, está influenciado por sus propias características y también por una serie de factores externos que modifican su comportamiento espectral “teórico”. Algunos de estos factores son: ángulo de iluminación solar (depende de la fecha del año y del momento del paso del satélite), modificaciones que el relieve introduce en el ángulo de iluminación, influencia de la atmósfera (especialmente la absorción provocada por nubes), variaciones medio ambientales en la cubierta, sustrato edafológico o litológico, etc.

Vemos así que es una tarea muy compleja la caracterización de un tipo de cubierta a partir de sensores remotos. Esto hace que la llamada signatura espectral de una cierta cubierta solamente sirva como una referencia. Se llama de esa forma al modo peculiar con el que una determinada cubierta refleja o emite energía a distintas longitudes de onda. La signatura espectral es la base para discriminar dicha cubierta de otras.

Por la influencia de los factores externos las distintas cubiertas de la superficie terrestre no tienen un comportamiento espectral único y permanente, que coincida con sus curvas de reflectividad espectral y permita reconocerlas sin confusión frente a otras superficies. Con todo, es importante analizar las curvas de reflectividad patrón para interpretar las imágenes. A su vez, estas gráficas nos permiten seleccionar las bandas más conveniente para la detección de las cubiertas terrestres.

1.5 Parámetros de sensores en general

Para que un sensor pueda registrar energía reflejada o emitida desde un blanco, este sensor debe estar sobre una plataforma alejado del blanco observado.

Estas plataformas podrían estar sobre la Tierra, sobre aviones, globos, así como sobre naves espaciales o satélites fuera de la atmósfera.

Las plataformas aéreas son usualmente montadas sobre aviones. Éstas suelen utilizarse para obtener información detallada de cualquier lugar de la Tierra. En el pasado fueron la única alternativa para obtener resoluciones espaciales métricas.

Los sensores usados en tierra suelen cumplir la función de registrar información muy detallada de alguna superficie la cual, luego, es comparada con la obtenida desde aviones o satélites, es decir, suele tomarse como validación terrestre para entender mejor la imagen satelital. Éstos pueden estar sobre la mano, sobre edificios o grúas.

1.5.1 Órbitas y campos de visión

El camino seguido por un satélite es conocido como órbita.

Las órbitas de los satélites son definidas en función del objetivo de los sensores que ellos deben llevar. La órbita elegida puede variar en altitud, y su orientación y rotación relativa a la Tierra.

Los satélites que están muy altos, ven la misma parte de la Tierra siempre y tienen órbitas geoestacionarias. Tienen una altitud de unos 36.000 km, y rotan a una velocidad que se ajusta justo a la rotación terrestre, por lo que parecen estacionarios respecto de la Tierra. Este tipo de satélites tienen especialmente aplicaciones en telecomunicaciones y también meteorológicas. Por su lejanía pueden tener imágenes de un hemisferio completo de la Tierra.

La mayoría de las plataformas de sensores remotos tienen órbitas (norte-sur) que en conjunción con el movimiento de la Tierra (oeste-este), permite cubrir toda la superficie terrestre cada cierto tiempo. Se denominan órbitas casi polares ya que tienen un pequeño ángulo con el eje terrestre.

Muchos de éstos son además heliosincrónicos de tal forma de cubrir cada área del mundo a una hora local solar constante. Esto es importante para asegurar condiciones de iluminación constantes tanto para análisis multitemporales como para la realización de mosaicos.

La mayoría de las plataformas satelitales de teledetección de hoy tienen estas órbitas casi polares, lo cual significa que ellas viajan hacia el norte en la mitad del planeta y hacia el sur en la otra mitad de su órbita. Éstas se denominan pasadas ascendentes y descendentes respectivamente. Si la órbita además es heliosincrónica, la pasada ascendente suele ser del lado oscuro de la Tierra y la descendente es cuando hay iluminación solar disponible. Así los sensores de energía solar reflejada tomarán imágenes sólo en las pasadas descendentes y los activos y o térmicos también tomarán imágenes en las ascendentes.

Cuando un satélite gira alrededor de la Tierra él ve sólo una parte de la superficie terrestre. El área mirada sobre la Tierra es denominado campo de visión (swath).

El ancho de los campos de visión para los diferentes sensores varía de decenas a cientos de kilómetros.

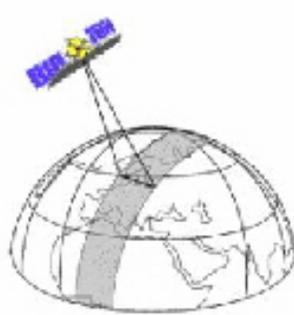
Como el satélite gira de polo a polo éste miraría siempre lo mismo si la Tierra no girara. Como éste no es el caso, el satélite parece moverse hacia el oeste (ya que la Tierra gira de oeste a este). Este movimiento permite a cada nueva campo de visión cubrir una zona diferente de la Tierra cada vez. Así, luego de varias órbitas, la Tierra es cubierta completamente.

Si comenzamos con cualquier pasada de un satélite, un ciclo completo se cumplirá cuando el satélite pase exactamente sobre el mismo lugar de la Tierra.

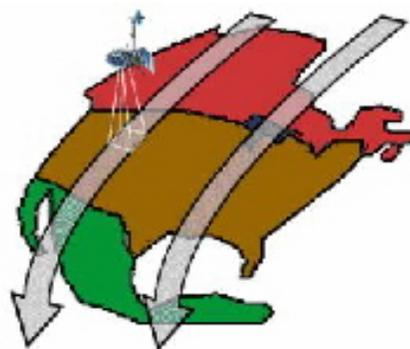
El tiempo total entre estas dos pasadas consecutivas es lo que se denomina “período de revisita”.

Si la cámara del sensor puede moverse podría mirar el mismo lugar sin tener que esperar a completar el ciclo orbital y entonces aumentar la frecuencia de muestreo de un dado lugar. El período de revisita es un punto crucial para muchas aplicaciones como por ejemplo derrames de petróleo, incendios, inundaciones etc.

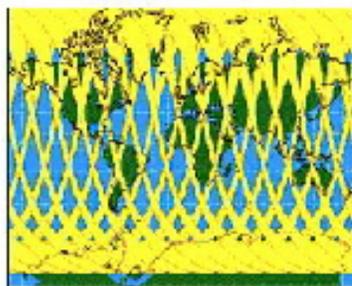
Claramente en las órbitas casi polares los campos de visión se solapan cerca de los polos permitiendo revisitas mucho más frecuentes.



(1)



(2)



(3)

Figura 6: (1) Campo de visión. (2) Campos de visión en pasadas sucesivas. (3) Campos de visión adyacentes.



Figura 7: Resolución espacial. Alta resolución (izquierda). Baja resolución (derecha).

1.5.2 Resolución espacial, tamaño del píxel y escala

La distancia entre el sensor y el blanco generalmente define en detalle la información obtenida y el área total mirada por el sensor.

Comparemos lo que vería un astronauta en el space shuttle y usted desde un avión. El astronauta podría ver todo su país pero claramente no podría diferenciar casas. Volando sobre una ciudad usted vería edificios, autos pero tendría una visión mucho más pequeña. Existe una diferencia similar entre los sensores en satélites y aviones.

Los detalles discernibles en una imagen dependen de la resolución espacial del sensor que se refiere al más pequeño objeto que es posible detectar. La resolución espacial de sensores pasivos depende principalmente de su “Campo de visión instantáneo” (Instantaneous Field of View (IFOV)).

Si bien para detectar un objeto, su tamaño deberá ser igual o mayor a la mínima celda de resolución, objetos más pequeños muchas veces son detectables si su reflectancia domina dentro de una zona del tamaño de la celda mínima.

Una imagen está formada por una matriz de elementos, o píxeles, los cuales constituyen las unidades más pequeñas de la imagen. Los píxeles de la imagen son generalmente cuadrados y representan cierta área de la imagen. Es importante distinguir entre el tamaño de píxel y la resolución espacial. Si un sensor tiene una resolución espacial de 20 m y la imagen es presentada en máxima resolución, cada píxel representa un área de 20m x 20m sobre la Tierra. Sin embargo, es posible presentar a la imagen con un tamaño de píxel mayor a la resolución espacial.

La mayor parte de las ilustraciones con imágenes tienen sus píxeles promediados para representar grandes áreas, aunque su resolución espacial original permanezca inalterada.

Las resoluciones actuales de los satélites comerciales van desde un kilómetro hasta del orden del metro.

La razón entre una distancia en una imagen o mapa a la distancia real sobre la Tierra es llamada escala. Si usted tiene un mapa con una escala de 1:100000, un objeto de 1cm de longitud sobre el mapa realmente será de 100000 cm (1km).

Mapas o imágenes con una razón “mapa-a-tierra” pequeña, se dicen de pe-

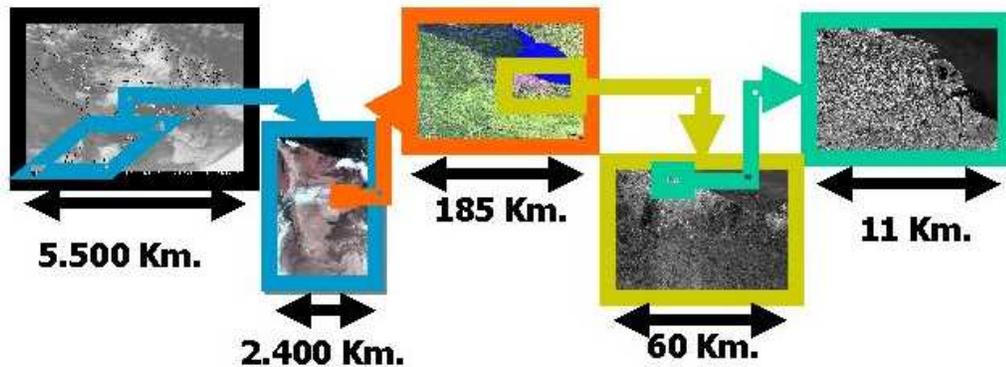


Figura 8: Campos de visión y resolución espacial de ciertos sensores. De izquierda a derecha: GOES 8 (píxel: 4,5x4,5 Km), NOAA 14 (píxel: 1,1x1,1Km), LANDSAT 5 TM (píxel: 30x30 m), SPOT 2 (píxel: 10x10 m), IKONOS (píxel: 1,1x1,1 m).

queña escala (por ejemplo 1:100000) (menos detalles), y aquéllas con cocientes más grandes (por ejemplo, 1:5000) se denominan escala grande (más detalles).

En la Figura 8 podemos ver los campos de visión y resolución espacial de distintos sensores.

1.5.3 Resolución espectral

Indica el número y ancho de las bandas espectrales que puede discriminar el sensor. Este será más eficiente cuanto mayor sea el número de bandas que ofrezca, pues facilita la caracterización espectral de las distintas cubiertas. También conviene que esas bandas sean suficientemente estrechas, con objeto de recoger la señal sobre regiones coherentes del espectro.

Entre los sensores espaciales, la menor resolución espectral corresponde al radar (generalmente un solo canal) y a los sistemas fotográficos. Los sensores óptico - electrónicos ofrecen un mayor rango de bandas; por ejemplo: 3 en el SPOT - HRV, 7 en el LANDSAT - TM.

Recientemente, podemos contar con los llamados sensores hiperespectrales. Son sensores especiales los cuales detectan y registran radiación en varios (cientos) de bandas espectrales estrechas. Tienen alta resolución espectral a causa de sus estrechos anchos de banda. Midiendo radiación sobre muchos rangos estrechos de longitudes de onda, podemos construir un espectro casi continuo de respuesta en cada píxel de la imagen. Esto permite una diferenciación fina entre blancos con respuestas diferentes que sería imposible con bandas anchas como los de las multispectrales. Ahora bien, el volumen de datos generados es muy grande, complicando tanto la manipulación como el almacenamiento de los datos.

1.5.4 Resolución radiométrica

Hace mención a la sensibilidad del sensor, esto es, a su capacidad para detectar variaciones en la radiancia espectral que recibe. En el caso de los sistemas fotográficos, la resolución radiométrica del sensor se indica por el número de niveles de gris posibles de registrar en la película. Para los óptico - electrónicos, la imagen habitualmente se presenta en formato digital, por lo que la resolución radiométrica suele identificarse con el rango de valores que codifica el sensor. Por cuanto se trata de codificación digital, habitualmente se expresa esa resolución en el número de bits que precisa cada elemento de imagen para ser almacenado (256, 128, 64, etc.)

Cuanto mayor sea la precisión radiométrica, tanto mejor podrá interpretarse la imagen.

El número de niveles más adecuado depende del método que se siga en la interpretación. Si ésta es visual, conviene considerar que el ojo humano difícilmente percibe más de 64 niveles de gris, y no más de 200.000 tonalidades.

En la interpretación digital, la sensibilidad radiométrica se puede aprovechar mejor porque la computadora es capaz de distinguir una escala numérica amplia. De hecho, en los sensores planeados para la futura estación orbital EOS, la resolución estándar será de 16 bits por píxel (65536 niveles de gris). En la Figura 9 podemos apreciar los efectos de la diferencia en la resolución radiométrica.

1.5.5 Resolución temporal

Además de las resoluciones espacial, espectral y radiométrica la resolución temporal es un parámetro crucial en las aplicaciones de sensores remotos. Se refiere al período de tiempo que lleva el sensor en completar un ciclo orbital completo. Este ciclo es generalmente de varios días.

De todos modos, debido al solapamiento de los campos de visión de órbitas adyacentes, que aumenta a medida que aumenta la latitud, algunas superficies pueden verse más frecuentemente. Campos de visión más anchos se revisitan más seguido, tienen mayor resolución temporal.

Además algunos satélites pueden mover su sensor pudiendo entonces forzar la visión de algún lugar más frecuentemente y así la resolución temporal real depende de la órbita, campo de visión, características del sensor y latitud.

La habilidad de coleccionar imágenes de la Tierra en forma frecuente es muy importante para analizar cambios en el tiempo de las características espectrales de un determinado blanco. Este tipo de análisis es conocido como análisis multitemporal.

Este concepto se refiere a la frecuencia de cobertura que proporciona el sensor. La cadencia temporal de los sistemas espaciales varía de acuerdo a los objetivos fijados para el sensor. Por ejemplo durante la primavera, la mayor parte de las especies de vegetación tienen un cambio continuo y así la habilidad de monitorear estos cambios depende de esta capacidad de revisitas continuas.

El factor temporal es importante por ejemplo en: Lugares donde usualmente está nublado y no pueda verse la superficie. Fenómenos de corta vida (inun-



Figura 9: Imágenes con el mismo escenario, resolución radiométrica de 2 bits (izquierda) y 8 bits (derecha).

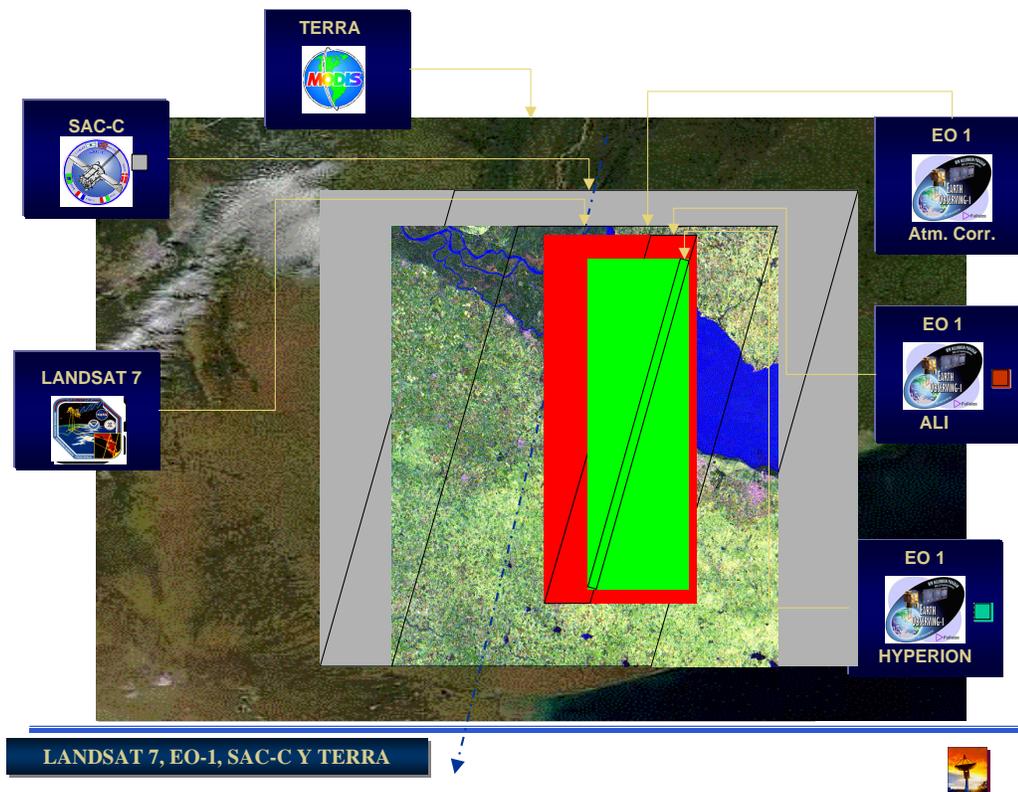


Figura 10: Resolución temporal de ciertos sensores.

daciones, derrames de petróleo). Cuando un análisis temporal es requerido (por ejemplo la expansión de deforestación de un año a otro). Cuando los cambios producidos en los aspectos nos hacen diferenciar ciertos elementos similares (trigo / maíz).

La Figura 10 muestra la resolución temporal de algunos sensores muy usados en todo el mundo.

1.5.6 Sensores en uso al día de hoy

El monitoreo del tiempo y pronóstico ha sido una de las primeras aplicaciones civiles de la teledetección. Ya desde 1960 TIROS-1 (Television and Infrared Observation Satellite - 1), se han lanzado muchos satélites con órbitas casi polares proveyendo una cobertura completa de la Tierra.

En 1966, NASA (the U.S. National Aeronautics and Space Administration) lanzó el satélite geostacionario (ATS-1) el cual pudo proveer imágenes hemisféricas cada media hora. Así por primera vez se pudo monitorear el desarrollo

y movimiento de sistemas meteorológicos.

Hoy varios países operan este tipo de satélites para monitorear las condiciones meteorológicas de la Tierra.

Generalmente hablando, tienen poca resolución espacial (comparado con los que estudian la Tierra) pero poseen una gran cobertura. Su resolución temporal es generalmente muy alta produciendo muy frecuentes observaciones de la Tierra, la humedad en la atmósfera, el cubrimiento de nubes permitiendo un análisis fino de las condiciones meteorológicas y así el pronóstico.

GOES El GOES (Geostationary Operational Environmental Satellite) siguió a la serie ATS. Ellos fueron diseñados por la NASA para la National Oceanic and Atmospheric Administration (NOAA) para proveer al United States National Weather Service imágenes frecuentes de la superficie terrestre y del cubrimiento de nubes.

La serie de satélites GOES ha sido utilizada en meteorología por más de 20 años. Estos satélites forman parte de una red global de satélites espaciados unos 70° de longitud alrededor de la Tierra proveyendo una cobertura casi completa.

Dos satélites GOES en órbitas geostacionarias a unos 36000 km arriba del ecuador, miran cada uno casi un tercio de la Tierra. Uno está a 75°W longitud y monitorea toda América y la mayor parte del Atlántico. El otro, a 135°W de longitud monitorea toda América del Norte y el Pacífico. Juntos cubren desde 20°W a 165°E de longitud.

A manera de ejemplo veamos la siguiente figura que muestra una toma completa de la Tierra realizada por GOES 7 en la que se ve al Huracán Andrews aproximándose a Louisiana el 25 de agosto de 1992.

La segunda generación iniciada por el GOES-8 (1994) con sus mejoras permiten ver la Tierra cada 15 minutos. Este incremento en la resolución temporal junto a mejoras en la resolución radiométrica y espacial han mejorado notablemente la capacidad de los pronósticos meteorológicos. GOES-8 tiene instrumentos que le dan la capacidad de tomar imágenes hemisféricas e imágenes de pequeñas áreas. Éstas tienen una resolución radiométrica de 10 bits y pueden transmitirse directamente al usuario en tierra.

NOAA AVHRR La NOAA es también responsable de otra serie de satélites esencialmente meteorológicos pero útiles para otros fines. Éstos tienen órbitas casi polares heliosincrónicas a (830-870 km de altura).

Dos satélites que proveen cobertura global cada 6 hs. Uno, cruza el ecuador a la mañana temprano de norte a sur mientras el otro lo hace a la tarde.

El sensor principal es el Advanced Very High Resolution Radiometer (AVHRR). Éste detecta la radiación en el visible, en el IR cercano, medio y térmico. Tiene un campo de visión de 3000 km.

Veamos ahora aplicaciones de este sensor:

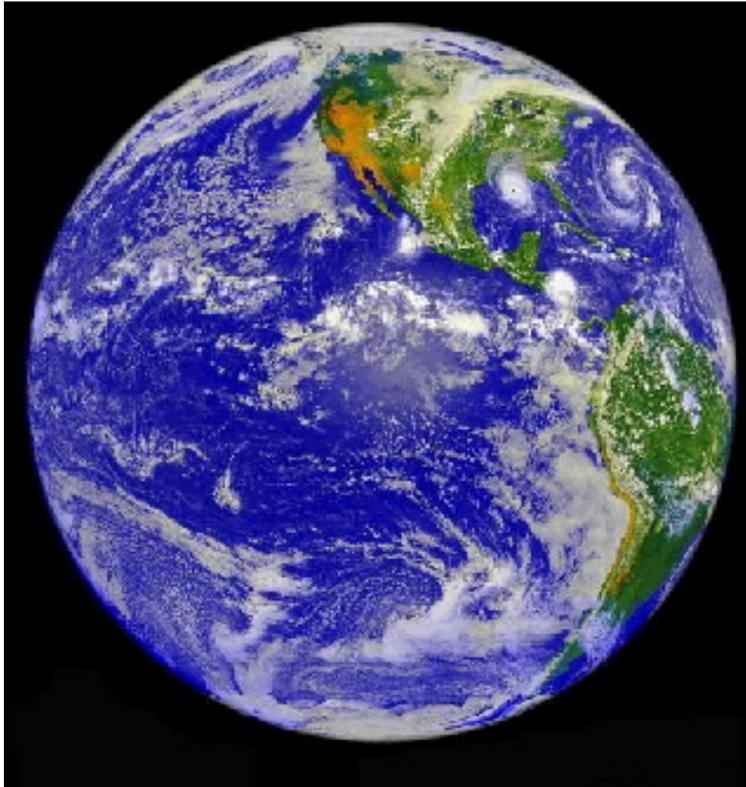


Figura 11: Visión de la Tierra tomada por el GOES 7 el 25 de agosto de 1992.

Banda	Longitud de onda	Aplicación
1	0.58 - 0.68 μm	Manejo de nubes, nieve y hielo.
2	0.725 - 1.1 μm	Inventario de agua, vegetación y uso para agricultura.
3	3.55 - 3.93 μm	Temperatura de la superficie de la Tierra, actividad volcánica e incendios forestales.
4	10.3 - 11.3 μm	Temperatura de la superficie del mar y humedad del suelo.
5	11.5 - 12.5 μm	Temperatura de la superficie del mar y humedad del suelo.

Existen otros sensores similares, uno de Japón y otro de Europa con la serie Meteosat. Ambos son geoestacionarios y están sobre el ecuador. Ambos toman imágenes cada media hora. El primero tiene dos bandas: 0.5- 0.75 μm (1.25 km de resolución), y 10.5 - 12.5 μm (5 km de resolución). Meteosat tiene 3 bandas: visible: 0.4 - 1.1 μm (2.5 km de resolución), mid-IR 5.7 - 7.1 μm , (5 km de resolución), e IR térmico: 10.5 - 12.5 μm , (5 km de resolución).

LANDSAT Aunque muchos sistemas meteorológicos satelitales también son usados para supervisar la superficie de la Tierra, no son óptimos para el trazado de un mapa detallado de la superficie de Tierra. El primer satélite diseñado expresamente para supervisar la superficie de la Tierra, LANDSAT-1, fue lanzado por la NASA en 1972. LANDSAT ha sido diseñado para probar la viabilidad de recoger datos de observación de Tierra multi-espectrales desde una plataforma satelital no tripulada. Desde entonces, este programa ha recogido una gran cantidad de datos del mundo entero de varios LANDSAT.

El éxito de LANDSAT se debe a varios factores: una combinación de sensores con bandas espectrales adaptadas a la observación de la Tierra, resolución espacial y buena cobertura regional. La larga vida útil del programa ha proporcionado un archivo voluminoso de datos de recursos de la Tierra que facilitan la supervisión de largo plazo, registros históricos e investigación. Todos los satélites de LANDSAT son colocados en órbitas casi polares, heliosincrónicas.

Los tres primeros satélites (LANDSATs 1-3) estaban en altitudes alrededor de 900 kilómetros y tenían períodos de 18 días mientras los satélites posteriores están alrededor de 700 kilómetros y tienen períodos de 16 días. Todos los satélites de LANDSAT cruzan el ecuador por la mañana para optimizar las condiciones de iluminación. El instrumento más popular en los primeros días del LANDSAT, era el MultiSpectral Scanner (MSS) luego el Thematic Mapper (TM). Ambos tenían un campo de visión de 185 km, con escenas de 185 km x 185 km.

La resolución espacial de TM es de 30 m para todas las bandas excepto la banda del infrarrojo térmico que tiene 120 m. Todos los canales son registrados sobre una gama de 256 números digitales (8 bits). Los datos del MSS y del TM son usados para una amplia variedad de aplicaciones, incluyendo la gestión de recursos, el trazado de mapas, la supervisión medioambiental, y la detección de cambio (p.ej. la supervisión de talas de bosques).

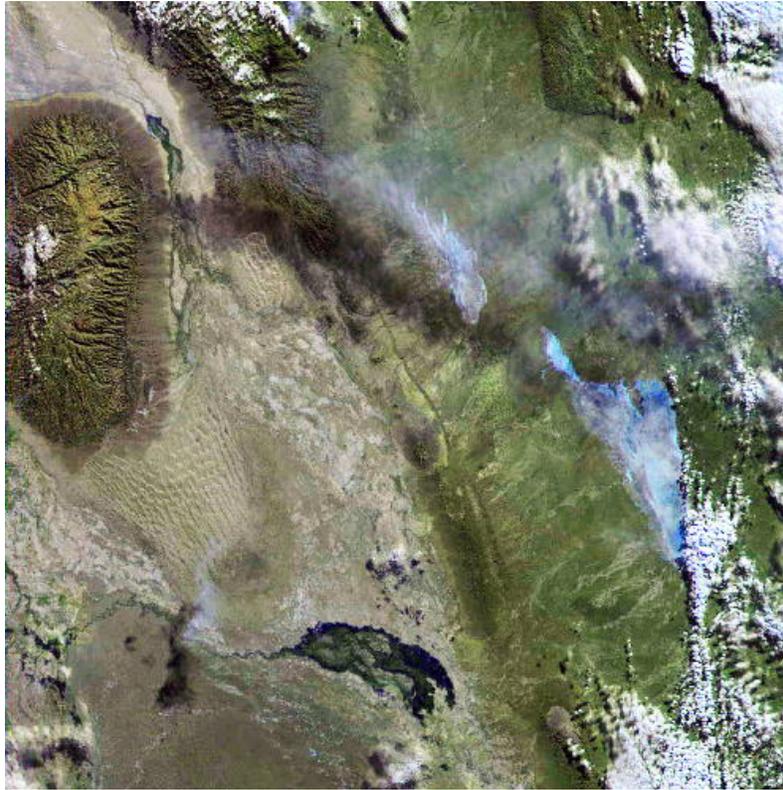


Figura 12: Imagen LANDSAT 7 ETM, Track/Frame: 231/82. Fecha: 13/05/2003. CONAE (Falda del Carmen, Córdoba).

En la Figura 12 vemos una imagen obtenida por el LANDSAT 7.

SPOT SPOT (Système Vierte l'Observation de la Terre) es una serie de satélites de observación de la Tierra diseñados y lanzados por el CNES (Centre d'Études National Spatiales) de Francia, con el apoyo de Suecia y Bélgica. El SPOT -1 ha sido lanzado en 1986, con versiones sucesivas posteriores cada tres o cuatro años. Todos los satélites están en órbitas sincronizadas con el movimiento del sol, casi polares, en altitudes alrededor de 830 kilómetros encima de la Tierra, que causa la repetición de la órbita cada 26 días. Cruzan el ecuador alrededor de las 10:30 de la mañana, hora solar local. El SPOT ha sido diseñado para ser un proveedor comercial de datos de observación de la Tierra, y fue el primer satélite en usar tecnología "along-track".

Cada uno de los satélites SPOT tiene un par de sistemas de alta resolución en las bandas del visible (HRV), que pueden ser manejados por separado y simultáneamente. Cada HRV es capaz de tomar datos en un canal solo de



Figura 13: Imagen SPOT 5 - 2,5 m de resolución de Canberra, Australia, febrero de 2003.

alta resolución espacial pancromático (PLA), o en una resolución espacial mayor en modo multiespectral (MLA) de tres canales. Cada sensor along-track HRV consiste en cuatro colecciones lineales de detectores: 6000 elementos para la grabación de modo panchromatic en una resolución espacial de 10 m, y 3000 elementos para cada una de las tres bandas multiespectrales, en 20 m de resolución espacial. El ancho del campo de visión para ambos modos es 60 kilómetros en el nadir.

CBERS El primer satélite desarrollado en Brasil, el CEBERS-1, fue lanzado exitosamente por el cohete chino Larga Marcha 4B del Centro de Lanzamiento de Taiwan el 14 de octubre de 1999. Su órbita es heliosincrónica a una altura de 778 km, efectúa aproximadamente 14 vueltas a la Tierra por día y consigue obtener una cobertura completa de la superficie terrestre cada 26 días.

Los satélites CBERS-1 y 2 están equipados con cámaras para observaciones ópticas de toda la Tierra además de un sistema de recolección de datos am-

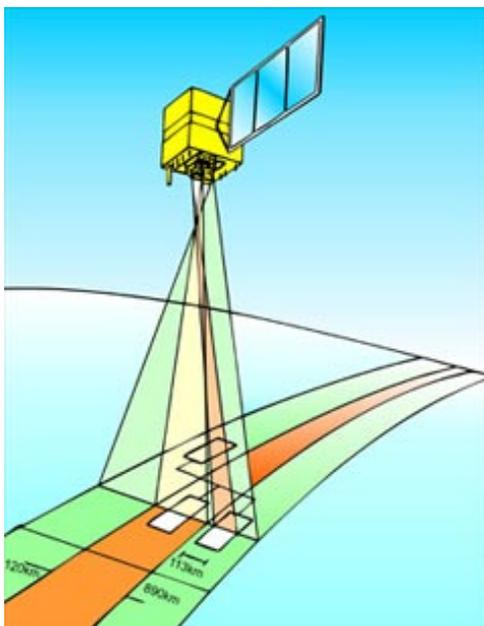


Figura 14: Los diferentes campos de visión del CBERS

bientales. Estos sistemas son únicos debido al uso de cámaras que combinan características especiales para resolver una gran variedad de escalas temporales y espaciales típicas de nuestro ecosistema.

SAC-C El SAC-C es el primer satélite argentino de observación de la Tierra, diseñado para el estudio de ecosistemas terrestres y marinos, la medición de la temperatura y contenido de vapor de agua de la atmósfera, la medición del campo magnético terrestre, estudios de la estructura y dinámica de la atmósfera e ionósfera y la determinación de componentes de onda larga del campo gravitatorio terrestre.

La carga principal del SAC-C está compuesta de un barredor multiespectral de resolución media (Multispectral Medium Resolution Scanner - MMRS), provisto por la CONAE, un conjunto de magnetómetros para mediciones escalares y vectoriales del campo magnético terrestre (Magnetic Mapping Payload - MMP) desarrollado y construido por un consorcio formado por la NASA/JPL y el Danish Space Research Institute (DSRI) y un receptor GPS de posicionamiento global (Gps OccuLtation and Passive reflection Experiment-GOLPE), provisto por el JPL/NASA. Adicionalmente, el SAC-C dispone de una cámara panorámica de una resolución de 35 metros (HRTC), una cámara de sensibilidad alta (HSTC) desarrolladas por la CONAE, dos cargas de ensayos tecnológicos provistos por Italia para experimentos de navegación y control de órbita,



Figura 15: Imagen MMRS del SAC- C - Path 228 (parcial) - 15 de octubre de 2003.

un instrumento francés para determinar el efecto de partículas de alta energía en componentes electrónicos de última generación (ICARE), y un sistema de recolección de datos ambientales (DCS) constituido por un sistema de estaciones distribuidas en el territorio nacional.

A bordo del SAC-C hay diez instrumentos que se utilizan para llevar a cabo diversos estudios relacionados con las Ciencias de la Tierra y experimentos tecnológicos con el objeto de mejorar el desarrollo de futuras misiones espaciales.

Sistemas de radar. Debido al éxito obtenido en el uso de imágenes de radar aerotransportadas, fue natural tratar de complementar la información dada por sensores ópticos instalados en satélites por medio de radares colocados en tales transportadores.

SEASAT, lanzado en 1978, fue el primer radar SAR en teledetección de uso civil. Operaba en banda L (23.5 cm) con polarización HH y una resolución espacial de 25 metros. Si bien fue diseñado con el principal objetivo de formar imágenes de océanos y témpanos también coleccionó una gran cantidad de imágenes terrestres. Sin embargo era de utilidad limitada en estas superficies a causa de la interacción del relieve con las características del radar.

Aunque estuvo operativo apenas tres meses, demostró el valor, con respecto al gran volumen de datos, de la información posible por medio del uso de radares transportados por satélites.

A partir del éxito de la misión SEASAT y el ímpetu proveniente de los éxitos de las misiones SAR aerotransportados, la Agencia Espacial Europea (ESA) lanzó el ERS-1 en julio de 1991. Este satélite llevaba a bordo: un radar altímetro, un radiómetro infrarrojo y sondeador acústico de micro ondas, y un instrumento activo en la zona de micro ondas del espectro (en banda C

(5.66 cm)). Este último se puede operar de modo tal que permite medir la reflectividad de la superficie oceánica y también la velocidad y dirección del viento sobre dicha superficie. Operado como radar de abertura sintética tiene un campo de visión de 100 km y una resolución espacial aproximada de 30 m, siendo VV su polarización.

Un segundo satélite, ERS-2, fue lanzado en abril de 1995 llevando un sensor activo de micro ondas similar al de ERS-1. Tanto este último como ERS-2 proveyeron y provén a la comunidad mundial con una enorme cantidad de datos SAR desde el espacio exterior. Se están usando tales datos en una amplia gama de aplicaciones tanto en ambiente marino como terrestre. Al igual que SEASAT ambos sensores presentan limitaciones para sus aplicaciones terrestres.

La National Space Development Agency of Japan (NASDA), lanzó el satélite JERS-1 en febrero de 1992. Lleva un SAR operando en la banda L (23.5 cm) con polarización HH. El ancho del campo de visión es de 75 km y una resolución espacial es de aproximadamente 18 metros tanto en rango como en azimut. Este SAR tiene cierto poder de penetración a través de la vegetación y otros tipos de cobertura.

Todo lo referente a imágenes SAR dio un salto gigante con el lanzamiento del satélite canadiense RADARSAT, el 4 de noviembre de 1995, a cargo de la Canadian Space Agency (CSA). Este satélite transporta un SAR que opera en la banda C (5.6 cm) con polarización HH. El campo de visión puede variar entre 35 y 500 km con una resolución espacial variable entre los 10 y los 100 metros. La tecnología de avanzada de los sensores que transporta RADARSAT permite formar imágenes de la misma región con mayor frecuencia que los anteriores SAR y también la adquisición de datos específicos para ciertas investigaciones geográficas. Finalmente, RADARSAT suministra imágenes diarias de toda la región Ártica y con la frecuencia de tres días de cualquier parte de Canadá.

2 Procesos de Markov sobre grafos

2.1 Introducción y notaciones

En Análisis de Imágenes se trata de encontrar la distribución subyacente a la energía de retorno captada por el sensor basándose en las distribuciones de la misma en los entornos de cada píxel.

En primer lugar, nos preguntamos:

¿Cómo construir un modelo matemático conveniente para dar cuenta de esas estructuras de interdependencia locales entre los valores de la energía de retorno en píxeles cercanos en el espacio?

La segunda cuestión que, naturalmente, nos planteamos es: ¿Existe un modelo matemático que simultáneamente sea adecuado para describir la distribución global de la energía de retorno y las estructuras de interdependencia locales?

Es claro que un tal modelo debe ser único para cada estructura de interdependencia locales.

Supongamos que para cada punto de coordenadas geográficas (s_1, s_2) en un cierto escenario S ; $X(s_1, s_2)$ representa el valor de la variable X en (s_1, s_2) cuya representación en un formato conveniente convenimos en llamar “imagen”. En este sentido X puede no solamente representar el valor de la energía de retorno a partir de un punto de S ; sino valores de magnitudes diversas que conjuntamente son consideradas en el momento de la formación de una imagen. Por ello es razonable considerar que X tiene rango en un conjunto E .

Por “análisis” de la imagen formada a partir de X entendemos “tener una idea lo más certera posible” de la distribución de X sobre E cuando (s_1, s_2) se mueve a lo largo de S .

Desde la década del 80 del siglo pasado, muchas técnicas basadas en el uso de procesos markovianos han demostrado ser muy útiles en el análisis y procesamiento de imágenes, no sólo de las satelitales sino de toda clase de imágenes.

Una buena parte de los aspectos formales del tema de nuestro interés está contenida en una rama de la Teoría de Probabilidad de creciente y rápido desarrollo que algunos llaman sistemas espaciales aleatorios con interacciones. Uno de sus objetivos importantes es la comprensión de los efectos interactuantes en sistemas aleatorios espaciales.

El impulso inicial proviene de la Física Estadística. Posiblemente el antecedente más remoto de la aplicación de tales procesos es el célebre modelo de Ising que fue propuesto como una forma de explicar el fenómeno de magnetización espontánea que muestran los materiales ferromagnéticos. Por alguna razón Ising no pudo concluir exitosamente su trabajo, pero su modelo recibe todavía la atención de una gran comunidad de investigadores en todo el mundo. La cantidad de trabajos que usan procesos de Markov en el procesamiento de imágenes viene creciendo en forma exponencial en los últimos años.

En este curso nos limitaremos a presentar los elementos y resultados básicos de los procesos de Markov orientados a su aplicación en el análisis y procesamiento de imágenes. Lo haremos simplificando en la mayor medida posible, la matemática involucrada en los mismos. No obstante, lo que aquí veremos

será suficiente para iniciarse en el estudio del tema.

Suponemos que el lector tiene los conocimientos de Probabilidad y Estadística que se dan habitualmente en los primeros años de las carreras que tienen tales disciplinas en sus planes de estudio.

A seguir, un poco de la notación que usaremos a lo largo de este curso.

S será un subconjunto finito de \mathbb{Z} o $\mathbb{Z} \times \mathbb{Z}$. Es el soporte de la imagen. Esto es, S representa el conjunto de píxeles.

E es un conjunto finito. Si estamos trabajando con imágenes binarias puede ser:

$$E = \{-1, +1\},$$

o bien

$$E = \{0, 1\},$$

o también

$$E = \{\text{blanco, negro}\}.$$

Si trabajamos con imágenes de radar (un solo canal espectral) podría ser:

$$E = \{0, 1, \dots, 255\},$$

con el significado habitual de que 0 es "negro", 255 es "blanco" y los números entre ambos son tonos de gris.

Si trabajamos con imágenes de color, esto es de tres canales espectrales: uno para el rojo, otro para el verde y uno para el azul, bien podría considerarse

$$E = \{(e_1, e_2, e_3) / e_i \in \{0, 1, \dots, 255\}, i = 1, 2 \text{ y } 3\}.$$

En fin, si trabajamos con imágenes multispectrales (LANDSAT, SAC-C, etc) o hiperespectrales con K canales espectrales disponibles, sería adecuado tomar

$$E = \{(e_1, \dots, e_K) / e_i \in \{0, 1, \dots, 255\}, i = 1, \dots, K\}.$$

Pondremos

$$E^S = \{x / x \text{ es una función con dominio en } S \text{ y rango en } E\}.$$

Este será el conjunto de la totalidad de imágenes que se quiere considerar. Esto es, un punto x de E^S será una "imagen" desde el punto de vista formal.

Como E y E^S son conjuntos finitos, las σ -álgebras, \mathcal{E} sobre E y \mathcal{E}^S sobre E^S será la de todos los subconjuntos correspondientes, $\mathcal{P}(E)$ y $\mathcal{P}(E^S)$ respectivamente.

Finalmente, si X es un conjunto y $A \subset X$, pondremos $\mathbf{1}_A$ para denotar la función definida sobre X por:

$$\mathbf{1}_A(x) = \begin{cases} 0 & \text{si } x \notin A \\ 1 & \text{si } x \in A \end{cases}.$$

2.2 Conceptos y resultados preliminares

Definición 1 Sea π una probabilidad nunca nula sobre E^S . Para cada $s \in S$ sea $\pi_s : E^S \rightarrow [0, 1]$ dada por

$$\pi_s(x) = \pi(\{y \in E^S / y(s) = x(s)\} | \{y \in E^S / y(t) = x(t), t \in S, t \neq s\}).$$

Llamaremos **familia de características locales asociadas a π** a la familia de funciones $\{\pi_s / s \in S\}$.

Proposición 2 Sean π y μ probabilidades nunca nulas sobre E^S . Si $\pi_s = \mu_s$ para todo $s \in S$, entonces $\pi = \mu$.

Este resultado no es válido en general si S no es finito.

Definición 3 Para cada $s \in S$ sea $V_s \subset S$ no vacío tal que:

- (i) $s \notin V_s$.
- (ii) $s \in V_t \iff t \in V_s$.
- (iii) $S = \bigcup_{s \in S} V_s$.

Diremos que $\mathcal{V} = \{V_s / s \in S\}$ es un **sistema de vecindades sobre S** .

Definición 4 Sea \mathcal{V} un sistemas de vecindades sobre S . Al par (S, \mathcal{V}) lo llamaremos **grafo sobre S con sistemas de vecindades \mathcal{V}** .

Definición 5 Sea $\mathcal{G} = (S, \mathcal{V})$ un grafo. Diremos que una **probabilidad π nunca nula sobre E^S es \mathcal{G} -markoviana** si

$$\pi_s(x) = \pi(\{y \in E^S / y(s) = x(s)\} | \{y \in E^S / y(t) = x(t), t \in V_s\}).$$

Ejercicio 6 Para cada $s \in S$ sea

$$V_s^0 = S \setminus \{s\} = \{t \in S / t \neq s\}.$$

Entonces $\mathcal{V}^0 = \{V_s^0 / s \in S\}$ es un sistemas de vecindades sobre S . Al grafo $\mathcal{G}^0 = (S, \mathcal{V}^0)$ lo llamaremos **grafo trivial**.

Definición 7 Diremos que una función $V : E^S \times \mathcal{P}(S) \rightarrow \mathbb{R}$ es un **potencial** si:

- (i) $V(x, \emptyset) = 0$, para todo $x \in E^S$.
- (ii) Si $\Lambda \subset S$, $x \in E^S$, $y \in E^S$ e $y(s) = x(s)$ para todo $s \in \Lambda$, entonces $V(x, \Lambda) = V(y, \Lambda)$.

Sea ϖ un punto de E .

Definición 8 Sea V un **potencial**. Diremos que V está **normalizado con respecto al punto ϖ** , si

$$\{\Lambda \subset S, x \in E^S, x(t) = \varpi \text{ para algún } t \in \Lambda\} \Rightarrow V(x, \Lambda) = 0.$$

Definición 9 Sea $\mathcal{G}=(S, \mathcal{V})$ un grafo. Diremos que $\Lambda \subset S$ no vacío es \mathcal{G} -**completo** si: $\#(\Lambda) = 1$ o es cierta la siguiente afirmación:
 $\{s \in \Lambda, t \in \Lambda, s \neq t\} \Rightarrow t \in V_s$.

Definición 10 Sea $\mathcal{G}=(S, \mathcal{V})$ un grafo. Diremos que $\Lambda \subset S$ no vacío es un \mathcal{G} -**clique** si: Λ es \mathcal{G} -completo maximal. Esto es:
 $\{\Delta \subset S \text{ es } \mathcal{G}\text{-completo y } \Lambda \subset \Delta\} \Rightarrow \Lambda = \Delta$.

Notación 11 Sea $\mathcal{G}=(S, \mathcal{V})$ un grafo. Pondremos $\mathcal{C}(\mathcal{G})$ para denotar la familia de todos los \mathcal{G} -completos.

Ejemplo 12 Sea $\mathcal{G}^0=(S, \mathcal{V}^0)$ el grafo trivial. Entonces

$$\mathcal{C}(\mathcal{G}^0) = \{\Lambda \subset S / \#(\Lambda) \geq 1\}.$$

Ejemplo 13 Sea $N \geq 1$ entero. $S = 0, 1, \dots, N$. Definimos: $V_0 = \{1\}$, $V_n = \{n-1, n+1\}$ para $n = 1, \dots, N-1$ y $V_N = \{N-1\}$. Entonces $\mathcal{V} = \{V_0, V_1, \dots, V_N\}$ es un sistema de vecindades. Si $\mathcal{G}=(S, \mathcal{V})$ entonces

$$\mathcal{C}(\mathcal{G}) = \{\{n\} / n \in S\} \cup \{\{n, n+1\} / n = 0, \dots, N-1\}.$$

Definición 14 Sea $V : E^S \times \mathcal{P}(\mathcal{E}^S) \rightarrow \mathbb{R}$ un potencial. Llamaremos **función de energía** asociada a V a la función $H_V : E^S \rightarrow \mathbb{R}$ definida por

$$H_V(x) = \sum_{\emptyset \neq \Lambda \subset S} V(x, \Lambda).$$

Definición 15 Sean: $\mathcal{G}=(S, \mathcal{V})$ un grafo; $V : E^S \times \mathcal{P}(S) \rightarrow \mathbb{R}$ un potencial. Diremos V es un \mathcal{G} -**potencial** si:
 Λ no es \mathcal{G} -completo $\Rightarrow V(x, \Lambda) = 0$, para todo $x \in E^S$.

Definición 16 Sean: $\mathcal{G}=(S, \mathcal{V})$ un grafo; $V : E^S \times \mathcal{P}(S) \rightarrow \mathbb{R}$ un \mathcal{G} -potencial. Diremos que una probabilidad π sobre E^S es una **distribución de Gibbs** con respecto a \mathcal{G} y V , si

$$\pi(\{x\}) = \frac{1}{Z_V} \exp(-H_V(x)),$$

para todo $x \in E^S$, donde a

$$Z_V = \sum_{y \in E^S} \exp(-H_V(y))$$

la llamaremos **función de partición** de π .

Sea ahora (Ω, \mathcal{A}, P) un espacio de probabilidades. Sea $X : \Omega \rightarrow E^S$ una función medible que llamaremos también **campo aleatorio** con valores en E^S . Sea P_X la probabilidad sobre E^S dada por:

$$P_X(\{x\}) = P(X = x).$$

Supondremos que $P_X(\{x\}) > 0$, para todo $x \in E^S$. También, si

Definición 17 Sea $\mathcal{G}=(S, \mathcal{V})$ un grafo. Diremos que X es un **campo \mathcal{G} -markoviano** (o un **\mathcal{G} -MRF**) si P_X es \mathcal{G} -markoviana.

Definición 18 Sean: $\mathcal{G}=(S, \mathcal{V})$ un grafo; $V : E^S \times \mathcal{P}(S) \rightarrow \mathbb{R}$ un \mathcal{G} -potencial. Diremos que X es un **campo de Gibbs** con respecto a \mathcal{G} y V , si P_X es una distribución de Gibbs con respecto a \mathcal{G} y V .

3 Ejemplos de campos markovianos y de Gibbs

Recordemos la definición de cadena de Markov y veamos cómo es posible considerarla como un ejemplo de campo markoviano.

Definición 19 Sean: (Ω, \mathcal{A}, P) un espacio de probabilidades; $S = \{0, 1, \dots, N\}$ con $N \geq 1$ entero. Sea $X : \Omega \rightarrow E^S$ un campo aleatorio. Diremos que X es una **cadena de Markov con respecto a P** o una **P -cadena de Markov** o simplemente una **cadena de Markov** si no hay lugar a confusión, con respecto de estados igual a E , si para todo $n = 0, 1, \dots, N$, y todo $x \in E^S$ se cumple que

$$\begin{aligned} & P(X(n+1) = x(n+1) | X(n) = x(n), \dots, X(0) = x(0)) \\ &= P(X(n+1) = x(n+1) | X(n) = x(n)). \end{aligned}$$

En tal caso, para cada $n = 1, \dots, N$, definimos $P_n : E \times E \rightarrow [0, 1]$ por

$$P_n(a, b) = P(X(n) = b | X(n-1) = a),$$

la llamaremos **probabilidad de transición de $X(n-1)$ en el estado a a $X(n)$ en el estado b** .

También llamaremos **probabilidad inicial de la cadena** a la probabilidad P_0 sobre E dada por

$$P_0(\{a\}) = P(X(0) = a),$$

para todo $a \in E$.

Proposición 20 Sean: (Ω, \mathcal{A}, P) un espacio de probabilidades; $S = \{0, 1, \dots, N\}$ con $N \geq 1$ entero; $X : \Omega \rightarrow E^S$ una cadena de Markov con respecto a P . Definimos $V_0 = \{1\}$, $V_n = \{n-1, n+1\}$ para $n = 1, \dots, N-1$ y $V_N = \{N-1\}$ y $\mathcal{V} = \{V_0, V_1, \dots, V_N\}$. Sea $\mathcal{G}=(S, \mathcal{V})$. Entonces X es un **\mathcal{G} -MRF**.

Ejemplo 21 Modelo de Ising 1-D. Sean: (Ω, \mathcal{A}, P) un espacio de probabilidades; $S = \{0, 1, \dots, N\}$ con $N \geq 1$ entero; $E = \{-1, +1\}$; $X : \Omega \rightarrow E^S$ una cadena de Markov con respecto a P con probabilidades de transición dadas por

$$P_n(a, b) = \frac{\exp(\beta ab)}{\exp(\beta) + \exp(-\beta)},$$

donde $\beta > 0$, para todo $n = 1, \dots, N$ y todo a y b en E . La probabilidad inicial es

$$P_0(\{-1\}) = P_0(\{+1\}) = \frac{1}{2}.$$

Definimos V_n para $n = 0, 1, \dots, N$, $\mathcal{V} = \{V_0, V_1, \dots, V_N\}$ y $\mathcal{G} = (S, \mathcal{V})$ como en la proposición anterior.

Afirmación 1: Sea $\mathcal{C} = \{\{0, 1\}, \{1, 2\}, \dots, \{N-1, N\}\}$. Si $\Lambda \in \mathcal{C}$, entonces Λ es \mathcal{G} -completo.

Afirmación 2: Sea $V : E^S \times \mathcal{P}(E^S) \rightarrow \mathbb{R}$ dado por

$$V(x, \Lambda) = \begin{cases} 0 & \text{si } \Lambda \notin \mathcal{C} \\ \beta x(n)x(n+1) & \text{si } \Lambda = \{n, n+1\}, \text{ con } n = 0, 1, \dots, N-1 \end{cases} .$$

Entonces V es un \mathcal{G} -potencial y su función de energía está dada por:

$$H_V(x) = \beta \sum_{n=0}^{N-1} x(n)x(n+1).$$

Afirmación 3:

$$P_X(\{x\}) = \frac{\exp(-H_V(x))}{2 \cdot (\exp(\beta) + \exp(-\beta))^N},$$

para todo $x \in E^S$.

Luego, X es un campo de Gibbs con respecto a \mathcal{G} y V .

Ejemplo 22 Modelo de Ising 2-D. Sean: $N \geq 2$ entero y

$$S = \{(s_1, s_2) / 1 \leq s_i \leq N, i = 1, 2\} \subset \mathbb{Z} \times \mathbb{Z}.$$

Afirmación 1: Para cada $s = (s_1, s_2)$ en S , sea

$$V_s = \{t = (t_1, t_2) / \|s - t\| = |s_1 - t_1| + |s_2 - t_2| = 1\}.$$

Entonces $\mathcal{V} = \{V_s / s \in S\}$ es un sistema de vecindades sobre S . Pongamos $v(s) = \#(V_s)$ para todo $s \in S$. Sea $\mathcal{G} = (S, \mathcal{V})$.

Ejercicio 2: Determinar \mathcal{C} , la familia de todos los \mathcal{G} -cliques. Verificar que $\#(\mathcal{C}) = 2n(n-1)$ y que si $\Lambda \in \mathcal{C}$ entonces $\#(\Lambda) = 2$.

Sean $E = \{-1, +1\}$, $k > 0$, $T > 0$, H, J y m números reales.

Ejercicio 3: Sea $V : E^S \times \mathcal{P}(S) \rightarrow \mathbb{R}$ dado por

$$V(x, \Lambda) = \begin{cases} 0 & \text{si } \Lambda \notin \mathcal{C} \\ \frac{J}{kT} x(s)x(t) + \frac{mH}{kT} \left(\frac{x(s)}{v(s)} + \frac{x(t)}{v(t)} \right) & \text{si } \Lambda = \{s, t\}, \text{ con } \Lambda \in \mathcal{C} \end{cases} .$$

Entonces V es un \mathcal{G} -potencial y su función de energía está dada por:

$$H_V(x) = \frac{J}{2kT} \sum_{\|s-t\|=1} x(s)x(t) + \frac{mH}{kT} \sum_{s \in S} x(s).$$

Ejemplo 23 Modelos binarios de vecinos más próximos. Sean $\mathcal{G} = (S, \mathcal{V})$ y \mathcal{C} como en el Ejemplo anterior.

Sea $E = \{0, 1\}$. Para cada $s \in S$ sea α_s un número real y para cada $(s, t) \in S \times S$ sea $\beta_{(s,t)}$ un número real. Supongamos que:

$$\begin{cases} \beta_{(s,t)} = 0 & \text{si } \|s - t\| \neq 1 \\ \beta_{(s,t)} = \beta_{(t,s)} & \text{si } (s, t) \in S \times S \end{cases} .$$

Ejercicio 1: Sea $V : E^S \times \mathcal{P}(S) \rightarrow \mathbb{R}$ dado por

$$V(x, \Lambda) = \begin{cases} 0 & \text{si } \Lambda \notin \mathcal{C} \\ \beta_{(s,t)}x(s)x(t) + \alpha_s \frac{x(s)}{v(s)} + \alpha_t \frac{x(t)}{v(t)} & \text{si } \Lambda = \{s, t\}, \text{ con } \Lambda \in \mathcal{C} \end{cases} .$$

Entonces V es un \mathcal{G} -potencial y su función de energía está dada por:

$$H_V(x) = \frac{1}{2} \sum_{\|s-t\|=1} \beta_{(s,t)}x(s)x(t) + \sum_{s \in S} \alpha_s x(s).$$

Casos particulares

1) $\alpha_s = \alpha$, y para todo $s = (s_1, s_2) \in S$

$$\begin{cases} \beta_{(s,t)} = \beta_v & \text{si } t = (s_1 \pm 1, s_2) \\ \beta_{(s,t)} = \beta_h & \text{si } t = (s_1, s_2 \pm 1) \end{cases} .$$

Este modelo es llamado **auto-logístico** por algunos autores.

2) $\alpha_s = \alpha$, y para todo $s \in S$

$$\begin{cases} \beta_{(s,t)} = 0 & \text{si } \|s - t\| \neq 1 \\ \beta_{(s,t)} = \beta & \text{si } \|s - t\| = 1 \end{cases} .$$

Este modelo es llamado **isotrópico**.

4 Relación entre distribuciones markovianas y de Gibbs

Consideremos elegido como referencia un cierto punto, digamos ϖ , de E .

Para cada $\Lambda \subset S$ y cada $x \in E^S$ definimos $x^\Lambda \in E^S$ por

$$x^\Lambda(s) = \begin{cases} \varpi & \text{si } s \notin \Lambda \\ x(s) & \text{si } s \in \Lambda \end{cases} .$$

Proposición 24 Sea π una probabilidad nunca nula sobre E^S .

Sea $VC(\pi) : E^S \times \mathcal{P}(S) \rightarrow \mathbb{R}$ dado por

$$VC(\pi)(x, \Lambda) = \begin{cases} 0 & \text{si } \Lambda = \emptyset, \text{ para todo } x \in E^S \\ \sum_{\Delta \subset \Lambda} (-1)^{\#\Lambda \setminus \Delta} \ln(\pi(\{x^\Delta\})) & \text{si } \Lambda \neq \emptyset, \text{ para todo } x \in E^S \end{cases} .$$

Sea $\mathcal{G}^0=(S, \mathcal{V}^0)$ el grafo trivial. Entonces:

a)

$$VC(\pi)(x, \Lambda) = \sum_{\Delta \subset \Lambda} (-1)^{\#(\Lambda \setminus \Delta)} \ln (\pi (\{x^\Delta\}) / \pi (\{\varpi^S\})),$$

siendo ϖ^S el elemento de E^S dado por

$$\varpi^S(s) = \varpi,$$

para todo $s \in S$.

b) $VC(\pi)$ es un \mathcal{G}^0 -potencial normalizado con respecto al punto ϖ , llamado el **potencial canónico de π** .

c) π es una distribución de Gibbs con respecto a \mathcal{G}^0 y $VC(\pi)$.

d) Si U es otro \mathcal{G}^0 -potencial normalizado con respecto al punto ϖ tal que π es una distribución de Gibbs con respecto a \mathcal{G}^0 y U entonces U y $VC(\pi)$ coinciden.

Proposición 25 Sean: π una probabilidad nunca nula sobre E^S , $VC(\pi)$ el potencial canónico de π , V un \mathcal{G}^0 -potencial no necesariamente normalizado con respecto al punto ϖ tal que π es una distribución de Gibbs con respecto a \mathcal{G}^0 y $V, \Lambda \subset S$ no vacío $x \in E^S$. Entonces:

$$VC(\pi)(x, \Lambda) = \sum_{\Delta \subset \Lambda} \sum_{\Lambda \subset \Delta^* \subset S} (-1)^{\#(\Lambda \setminus \Delta)} V(x^\Delta, \Delta^*).$$

Notación 26 Sean: π una probabilidad nunca nula sobre E^S , $s \in S, \Delta \subset S, x \in E^S$. Pondremos:

$$\pi_{s, \Delta}(x) = \pi (\{y \in E^S / y(s) = x^\Delta(s)\} | \{y \in E^S / y(t) = x^\Delta(t), t \neq s\}).$$

Proposición 27 Sean: π una probabilidad nunca nula sobre E^S , $VC(\pi)$ el potencial canónico de $\pi, s \in S, \Lambda \subset S$ no vacío, $x \in E^S$. Entonces

$$VC(\pi)(x, \Lambda) = \sum_{\Delta \subset \Lambda} (-1)^{\#(\Lambda \setminus \Delta)} \ln (\pi_{s, \Delta}(x)).$$

Antes de enunciar el resultado más importante de esta Sección, veamos algunas notaciones útiles.

Notación 28 a) Sea $\mathcal{V} = \{V_s / s \in S\}$ un sistema de vecindades sobre S . Para cada $\Lambda \subset S$ sea

$$\mathcal{V}(\Lambda) = \bigcup_{t \in \Lambda} (V_t \setminus \Lambda).$$

b) Para cada $\Lambda \subset S$ y cada $x \in E^S$ sea

$$\{x, \Lambda\} = \{y \in E^S / y(s) = x(s), s \in \Lambda\}.$$

Para cada $z \in E^\Lambda$ sea

$$I_\Lambda(z, x) = \begin{cases} z(t) & \text{si } t \in \Lambda \\ x(t) & \text{si } t \notin \Lambda \end{cases}.$$

c) Sean: $\mathcal{G}=(S, \mathcal{V})$ un grafo; V un \mathcal{G} -potencial; $\Lambda \subset S$ no vacío; $x \in E^S$.
 Pondremos:

$$H_{V \cap \Lambda}(x) = \sum_{\Delta \in \mathcal{C}(\mathcal{G}) \cap * } V(x, \Delta),$$

donde $\mathcal{C}(\mathcal{G}) \cap * = \{ \cdot \in \mathcal{C}(\mathcal{G}) / \Delta \cap \Lambda \neq \emptyset \}$.

Teorema 29 Hammersley y Clifford. Sean: $\mathcal{G}=(S, \mathcal{V})$ un grafo y π una probabilidad nunca nula sobre E^S .

a) Sea V un \mathcal{G} -potencial tal que π es de Gibbs con respecto a \mathcal{G} y V . Entonces π es \mathcal{G} -markoviana.

b) Si π es \mathcal{G} -markoviana entonces $VC(\pi)$ es un \mathcal{G} -potencial y π es de Gibbs con respecto a \mathcal{G} y $VC(\pi)$. Además

$$\pi_s(x) = \frac{\exp(-H_{VC(\pi) \cap \{s\}}(x))}{\sum_{z \in E} \exp(-H_{VC(\pi) \cap \{s\}}(I_{\{s\}}(z, x)))}$$

Proposición 30 Sean: $\mathcal{G}=(S, \mathcal{V})$ un grafo; π una probabilidad nunca nula sobre E^S ; V un \mathcal{G} -potencial tal que π es de Gibbs con respecto a \mathcal{G} y V ; $\Lambda \subset S$ no vacío; $\Delta \subset S$ tal que $\mathcal{V}(\Lambda) \subset \Delta$ con $\Lambda \cap \Delta = \emptyset$; $x \in E^S$. Entonces:

$$\pi(\{x, \Lambda\} | \{x, \Delta\}) = \pi(\{x, \Lambda\} | \{x, \mathcal{V}(\Lambda)\}) = \frac{\exp(-H_{V \cap \Lambda}(x))}{\sum_{z \in E^\Lambda} \exp(-H_{V \cap \Lambda}(I_\Lambda(z, x)))}$$

5 La Estadística en el procesamiento de imágenes

La idea central subyacente al procesamiento de imágenes es la de transformar una imagen por medio de una ecuación o una serie de ecuaciones (usando la computadora), para obtener una nueva imagen digital o pictórica, que a su vez pueda ser manipulada por otros programas computacionales o mostrada en pantalla, por impresora, etc.

A las transformaciones más usuales que se realizan sobre las imágenes obtenidas de los sensores antes de su utilización con diversos objetivos las podemos agrupar en:

- 1- Rectificación y restauración de imágenes.
- 2- "Mejoramiento" de imágenes.
- 3- Clasificación de imágenes.
- 4- Combinación de imágenes e integración con S.I.G. (sistemas de información geográfica).
- 5- Construcción de modelos biofísicos.

Ahora bien, toda imagen de teledetección presenta perturbaciones no deseadas, que son debidas a diversas causas provenientes de deficiencias en los procesos de percepción, digitalización de señales y/o registro de datos.

El ruido puede o bien degradar o bien ocultar totalmente el contenido de información radiométrica de una imagen digital. Por ello la remoción del ruido habitualmente debe preceder a cualquier técnica de mejoramiento o clasificación de una imagen. El objetivo en la remoción de ruidos es restaurar una imagen para obtener una lo más aproximada posible a la escena real focalizada en la imagen.

Veamos a seguir cómo incorporar el mencionado ruido en un modelo matemático adecuado. Lo haremos en el caso muy simple de imágenes binarias.

Ejemplo 31 Imágenes binarias y ruido. Consideremos el Modelo de Ising 2-D con los parámetros: $k = T = 1, H = m = 0$ y $J \neq 0$. Luego, el potencial V está dado por:

$$V(x, \Lambda) = \begin{cases} 0 & \text{si } \Lambda \notin \mathcal{C} \\ Jx(s)x(t) & \text{si } \Lambda = \{s, t\}, \text{ con } \Lambda \in \mathcal{C} \end{cases},$$

y su función de energía es

$$H_V(x) = \frac{J}{2} \sum_{\|s-t\|=1} x(s)x(t).$$

Sean: (Ω, \mathcal{A}, P) ; $X : \Omega \rightarrow E^S$ un campo aleatorio con valores en E^S tal que P_X es de Gibbs con respecto a \mathcal{G} y V , esto es:

$$P(X = x) = \frac{1}{Z_J} \exp \left(-\frac{J}{2} \sum_{\|s-t\|=1} x(s)x(t) \right),$$

con

$$Z_J = \sum_{z \in E^S} \exp \left(-\frac{J}{2} \sum_{\|s-t\|=1} z(s)z(t) \right).$$

Sea $0 < \varepsilon < \frac{1}{2}$.

Sea $W : \Omega \rightarrow E^S$ un campo aleatorio tal que $\{W(s) / s \in S\}$ es una familia de variables independientes con

$$P(W(s) = -1) = \varepsilon = 1 - P(W(s) = 1).$$

No es difícil ver que se cumple:

$$P(W = w) = K_N(\varepsilon) \left(\frac{1-\varepsilon}{\varepsilon} \right)^{\frac{1}{2} \sum_{s \in S} w(s)},$$

con

$$K_N(\varepsilon) = (\varepsilon(1-\varepsilon))^{N^2/2}.$$

Suponemos ahora que el proceso realmente observado no sea X sino otro proceso $Y : \Omega \rightarrow E^S$ dado por:

$$Y(\omega)(s) = X(\omega)(s) \cdot W(\omega)(s),$$

para todo $\omega \in \Omega$ y todo $s \in S$, y que X y W son independientes.

En este contexto, el problema de rectificación y restauración de imágenes se puede plantear en términos formales como: **estimar X conocido Y** .

Sea $\Pi : E^S \times E^S \rightarrow (0, 1]$ dada por:

$$\Pi(x, y) = P(X = x | Y = y).$$

Se puede probar que en este caso particular se cumple:

$$\Pi(x, y) = \frac{1}{Z_{J,y}} \exp \left(-\frac{J}{2} \sum_{\|s-t\|=1} x(s)x(t) - \frac{1}{2} \sum_{s \in S} \frac{y(s)}{x(s)} \ln \left(\frac{1-\varepsilon}{\varepsilon} \right) \right),$$

donde

$$Z_{J,y} = \sum_{z \in E^S} \exp \left(-\frac{J}{2} \sum_{\|s-t\|=1} z(s)z(t) - \frac{1}{2} \sum_{s \in S} \frac{y(s)}{z(s)} \ln \left(\frac{1-\varepsilon}{\varepsilon} \right) \right).$$

Como en general, J y ε no son conocidos, debemos sustituirlos por estimadores (basados en Y). Este problema de estimación puede verse por ejemplo en ([1]).

Por un abuso de notación pongamos J y ε para los respectivos estimadores.

Hecho esto, debemos encarar el problema de estimar X conocido Y .

Dos estimadores "naturales" son:

1) Estimador MR o "misclassification rate" Está definido para cada $\omega \in \Omega$ por:

$$\hat{X}_{MR}(\omega)(s) = \begin{cases} 1 & \text{si } \sum_{z \in E^S \setminus \{s\}} \Pi((1, z)_s, Y(\omega)) \geq \frac{1}{2} \\ -1 & \text{caso contrario} \end{cases},$$

para todo $s \in S$. $(1, z)_s$ es el elemento de E^S definido por

$$(1, z)_s(t) = \begin{cases} 1 & \text{si } t = s \\ z(t) & \text{si } t \neq s \end{cases}.$$

2) Estimador MAP o "maximum a posteriori" Está dado para cada $\omega \in \Omega$ por $\hat{X}_{MAP}(\omega)$ tal que

$$\Pi(\hat{X}_{MAP}(\omega), Y(\omega)) \geq \Pi(x, Y(\omega)),$$

para todo $x \in E^S$.

Ejemplo 32 Un modelo general para imágenes con distorsión y ruido ([2]). Sean: S un subconjunto finito de $\mathbb{Z} \times \mathbb{Z}$; E un conjunto finito; $\mathcal{V} = \{V_s / s \in S\}$ un sistema de vecindades sobre S ; $\mathcal{G} = (S, \mathcal{V})$ un grafo; (Ω, \mathcal{A}, P) un espacio de probabilidades; $X : \Omega \rightarrow E^S$ un \mathcal{G} -MRF con $P_X(\{x\}) = P(X = x) > 0$, para todo $x \in E^S$; F y G dos conjuntos finitos, que representan los posibles valores de las distorsiones y ruido respectivamente que podrían afectar a una imagen;

$\mathcal{B} = \{B_s / s \in S\}$ otro sistema de vecindades sobre S .

Para cada $s \in S$ sean: $B_s^* = B_s \cup \{s\}$; $H_s : E^{B_s^*} \rightarrow F$ suryectiva. Con H_s denotamos la posible distorsión en la señal de retorno desde s causada por la influencia de los retornos desde los píxeles situados en una cierta vecindad de s . Finalmente, sean: $\phi : F \rightarrow G$ suryectiva; $\Theta : G \times G \rightarrow G$, $\Phi : G \times G \rightarrow G$ suryectivas tales que

$$\Phi(a, \Theta(a, b)) = b = \Theta(a, \Phi(a, b)),$$

para todo $(a, b) \in G \times G$.

Sea $\tilde{\varepsilon} : \Omega \rightarrow G^S$ un campo markoviano con respecto a (S, \mathcal{B}) , con $\mathcal{B} = \{B_s / s \in S\}$ tal que con $P(\tilde{\varepsilon} = g) > 0$, para todo $g \in G^S$. Este campo $\tilde{\varepsilon}$ representa el “ruido” aleatorio. Supongamos que X y $\tilde{\varepsilon}$ son independientes.

Finalmente, sea $Y : \Omega \rightarrow G^S$ el campo aleatorio que representa las observaciones dado por

$$Y(\omega)(s) = \Theta\left(\left(\phi\left(H_s\left(X(\omega)_{B_s^*}\right)\right), \tilde{\varepsilon}(\omega)(s)\right)\right).$$

En el ya mencionado trabajo ([2]) se prueba con detalle la propiedad enunciada en ([3]) de que el campo aleatorio X condicional a la imagen observada Y sigue siendo markoviano con respecto a un cierto sistema de vecindades que comprende a los sistemas \mathcal{V} y \mathcal{B} .

Ejemplo 33 Caso particular del modelo general para imágenes con distorsión y ruido. Sean: $N \geq 2$; S ; $\mathcal{V} = \{V_s / s \in S\}$; $\mathcal{G} = (S, \mathcal{V})$ y $E = \{-1, +1\}$ como en el Modelo de Ising 2-D. Para cada $s = (s_1, s_2)$ en S , sea

$$B_s = \left\{t = (t_1, t_2) / 0 < (s_1 - t_1)^2 + (s_2 - t_2)^2 \leq 2\right\},$$

y $\mathcal{B} = \{B_s / s \in S\}$.

Para facilitar la notación, pongamos

$$\begin{aligned} \text{Int} &= \{(s_1, s_2) / 1 < s_i < N, i = 1, 2\}, \\ \text{Esq} &= \{(1, 1), (1, N), (N, 1), (N, N)\}, \\ \text{Lad} &= S \setminus (\text{Int} \cup \text{Esq}). \end{aligned}$$

Se puede verificar fácilmente que

$$\#(B_s^*) = \begin{cases} 4 & \text{si } s \in \text{Esq} \\ 6 & \text{si } s \in \text{Lad} \\ 9 & \text{si } s \in \text{Int} \end{cases}.$$

Definimos la distorsión por

$$H_s(x_{B_s^*}) = \frac{1}{\#(B_s^*)} \sum_{t \in B_s^*} x_{B_s^*}(t),$$

para todo $s \in S$.

Sea

$$F = \left\{0, \pm \frac{1}{9}, \pm \frac{3}{9}, \pm \frac{1}{2}, \pm \frac{5}{9}, \pm \frac{6}{9}, \pm \frac{7}{9}, \pm 1\right\},$$

y

$$G = E.$$

Luego

$$G \times G = \{(-1, -1), (-1, 1), (1, -1), (1, 1)\}.$$

Sea $\Theta : G \times G \rightarrow G$ dada por

$$\Theta((i, j)) = \begin{cases} -1 & \text{si } i = j \\ 1 & \text{si } i = -j \end{cases}$$

Sea $\Phi : G \times G \rightarrow G$ igual a Θ . Esta operación se puede interpretar de la siguiente manera: supongamos que los píxeles que toman el valor -1 y $+1$ indican, respectivamente, ausencia y presencia de señal, tanto para la imagen verdadera como para el ruido. Θ significa que la superposición de no-señal y no-ruido o de señal y ruido dan no-señal, mientras que la de señal y no-ruido o no-señal y ruido producen señal.

La función de distorsión $\phi : F \rightarrow G$ podría ser

$$\phi(z) = \begin{cases} -1 & \text{si } z \leq 0 \\ 1 & \text{si } z > 0 \end{cases}.$$

A su vez, esta función puede interpretarse como una función “umbral de respuesta”, que enfatiza el efecto de “borroneo”: aquellos píxeles que toman valores grises oscuros (no positivos) son asociados con el color negro (valor -1) y los grises claros (positivos) son asociados con el blanco ($+1$).

Suponemos ahora que las características locales correspondientes al campo aleatorio X que representa la imagen verdadera están dadas por:

$$\pi_s(x) = P(X(s) = x(s) | X_{V_s} = x_{V_s}) = \frac{\exp\left(\frac{1+x(s)}{2} (\alpha\Upsilon_s + \beta\Psi_s)\right)}{1 + \exp(\alpha\Upsilon_s + \beta\Psi_s)},$$

donde, suponiendo que $s = (i, j) \in \text{Int}$ definimos:

$$\begin{aligned} \Upsilon_s &= x(i-1, j) + x(i+1, j), \\ \Psi_s &= x(i, j-1) + x(i, j+1). \end{aligned}$$

Con modificaciones evidentes para los casos en que $s = (i, j) \in \text{Esq}$ y $s = (i, j) \in \text{Lad}$. Esto es, suponemos que la familia de probabilidades de transición asociadas con X es una paramétrica con parámetro (α, β) .

Finalmente, suponemos que el campo representativo del ruido ε , es markoviano con respecto al sistema de vecindades \mathcal{B} y que la familia de sus características locales también es paramétrica dada por:

$$\tilde{\varepsilon}_s(g) = P(\tilde{\varepsilon}(s) = g(s) | \tilde{\varepsilon}_{B_s} = g_{B_s}) = \frac{\exp\left(\frac{1+g(s)}{2} (\gamma A_s + \delta C_s)\right)}{1 + \exp(\gamma A_s + \delta C_s)},$$

donde, suponiendo que $s = (i, j) \in \text{Int}$ definimos:

$$\begin{aligned} A_s &= g(i-1, j) + g(i+1, j) + g(i, j-1) + g(i, j+1), \\ C_s &= g(i-1, j-1) + g(i+1, j-1) + g(i-1, j+1) + g(i+1, j+1). \end{aligned}$$

Con modificaciones evidentes para los casos en que $s = (i, j) \in \text{Esq}$ y $s = (i, j) \in \text{Lad}$.

6 Algoritmos para simulación de distribuciones de Gibbs

6.1 Preliminares.

Sean: S un subconjunto finito de $\mathbb{Z} \times \mathbb{Z}$; E un conjunto finito; $\mathcal{V} = \{V_s / s \in S\}$ un sistema de vecindades sobre S ; $\mathcal{G} = (S, \mathcal{V})$ un grafo; \mathcal{C} la familia de todos los subconjuntos \mathcal{G} -completos de S ; $V : E^S \times \mathcal{P}(S) \rightarrow \mathbb{R}$ un \mathcal{G} -potencial; $H_V : E^S \rightarrow \mathbb{R}$ la función de energía asociada a V ; (Ω, \mathcal{A}, P) un espacio de probabilidades; $X : \Omega \rightarrow E^S$ un campo aleatorio que es de Gibbs con respecto a \mathcal{G} y V . Esto es:

$$P_X(x) = \frac{1}{Z_V} \exp(-H_V(x)),$$

donde

$$H_V(x) = \sum_{\Lambda \in \mathcal{C}} V(x, \Lambda).$$

y

$$Z_V = \sum_{y \in E^S} \exp(-H_V(y)).$$

Para fijar ideas, veamos el siguiente

Ejemplo 34 Campo aleatorio de Gibbs con ruido aditivo gaussiano.

Consideremos el Ejemplo 31 con las siguientes modificaciones:

$E = \{0, \pm 1, \dots, \pm L\}$. El potencial V está dado por:

$$V(x, \Lambda) = \begin{cases} 0 & \text{si } \Lambda \notin \mathcal{C} \\ J \mathbf{1}_{\{x(s)\}}(x(t)) & \text{si } \Lambda = \{s, t\}, \text{ con } \Lambda \in \mathcal{C} \end{cases},$$

y su función de energía es

$$H_V(x) = -\frac{J}{2} \sum_{\|s-t\|=1} \mathbf{1}_{\{x(s)\}}(x(t)).$$

Sean: (Ω, \mathcal{A}, P) ; $X : \Omega \rightarrow E^S$ un campo aleatorio con valores en E^S tal que P_X es de Gibbs con respecto a \mathcal{G} y V , esto es:

$$P(X = x) = \frac{1}{Z_J} \exp\left(-\frac{J}{2} \sum_{\|s-t\|=1} \mathbf{1}_{\{x(s)\}}(x(t))\right),$$

con

$$Z_J = \sum_{z \in E^S} \exp \left(-\frac{J}{2} \sum_{\|s-t\|=1} \mathbf{1}_{\{z(s)\}}(z(t)) \right).$$

Sea $W : \Omega \rightarrow E^S$ un campo aleatorio tal que $\{W(s) / s \in S\}$ es una familia de variables independientes con

$$P(W(s) = e) = K_\sigma^{-1} \exp \left(-\frac{1}{2} \left(\frac{e}{\sigma} \right)^{\frac{1}{2}} \right),$$

con

$$K_\sigma = \sum_{d \in E} \exp \left(-\frac{1}{2} \left(\frac{d}{\sigma} \right)^{\frac{1}{2}} \right).$$

Es decir, W es “gaussiano discreto”.

Suponemos ahora que el proceso realmente observado no es X sino otro proceso $Y : \Omega \rightarrow E^S$ dado para todo $\omega \in \Omega$ y todo $s \in S$ por:

$$Y(\omega)(s) = X(\omega)(s) \oplus W(\omega)(s),$$

donde \oplus entre dos elementos de E a y b significa

$$a \oplus b = (a + b) \text{ mod } L.$$

para todo $s \in S$ y que X y W son independientes.

Sea $\Pi : E^S \times E^S \rightarrow (0, 1]$ dada por:

$$\Pi(x, y) = P(X = x | Y = y).$$

En este caso particular se cumple:

$$\Pi(x, y) = \frac{1}{Z_{J,\sigma,y}} \exp \left(-\frac{J}{2} \sum_{\|s-t\|=1} \mathbf{1}_{\{x(s)\}}(x(t)) - \frac{1}{2\sigma^2} \sum_{s \in S} (y(s) - x(s))^2 \right),$$

donde

$$Z_{J,\sigma,y} = \sum_{z \in E^S} \exp \left(-\frac{J}{2} \sum_{\|s-t\|=1} \mathbf{1}_{\{z(s)\}}(z(t)) - \frac{1}{2\sigma^2} \sum_{s \in S} (y(s) - z(s))^2 \right).$$

Como vemos entonces, la probabilidad condicional de la “verdadera” imagen, X , dada la imagen observada, $y = Y$, es un campo de Gibbs. Luego, una vez estimados los parámetros J y σ , resta estimar X observada la imagen dada $y = Y$. Para esto último es que debemos saber cómo generar campos de Gibbs (a partir de los valores estimados de J y σ y a partir de la imagen y).

En este breve curso veremos en forma muy simplificada dos algoritmos para simular campos de Gibbs: los algoritmos de Metrópolis y el así llamado “muestreo de una distribución de Gibbs”.

Pero antes veamos algunos conceptos y resultados que nos serán de utilidad tanto para enunciar los algoritmos en cuestión como para justificar su exactitud.

Definición 35 Se dice que una función $q : E^S \times E^S \rightarrow [0, 1]$ es una **matriz de transición** si

$$\begin{aligned} q(x, y) &\geq 0, \text{ para todo } (x, y) \in E^S \times E^S, \\ \sum_{y \in E^S} q(x, y) &= 1, \text{ para todo } x \in E^S. \end{aligned}$$

Se dice que una tal q es **irreducible** si para todo $(x, y) \in E^S \times E^S$ existen $n \geq 1$, x_0, x_1, \dots, x_n en E^S con $x_0 = x$, $x_n = y$ y $q(x_{k-1}, x_k) > 0$ para $k = 0, \dots, n-1$.

Notemos que para cada $x \in E^S$, se tiene que la función $q_x : E^S \rightarrow [0, 1]$ dada por

$$q_x(y) = q(x, y),$$

es una función de probabilidad sobre E^S .

Proposición 36 Sean: q una matriz de transición; ρ una probabilidad nunca nula sobre E^S ; $(x, y) \in E^S \times E^S$. Definimos $p(\cdot, \cdot; \rho, q)$ sobre $E^S \times E^S$ por:

$$p(x, y; \rho, q) = \begin{cases} q(x, y) \frac{\rho(\{y\})}{\rho(\{x\})} & \text{si } \rho(\{y\}) < \rho(\{x\}), \\ q(x, y) & \text{si } \rho(\{y\}) \geq \rho(\{x\}), x \neq y, \\ 1 - \sum_{z \neq x} p(x, z; \rho, q) & \text{si } x = y. \end{cases}$$

Entonces $p(\cdot, \cdot; \rho, q)$ es una matriz de transición.

Notación 37 Sea

$$\mathbb{I} = \{0, 1, 2, \dots\}.$$

Consideremos a $(E^S)^\mathbb{I}$ como espacio medible con la σ -álgebra producto que denotaremos por \mathcal{B} .

Para cada probabilidad, μ , nunca nula sobre E^S y cada sucesión de matrices de transición $(p_n)_{n \geq 1}$ sea $P(\cdot; \mu, (p_n)_{n \geq 1})$ la única probabilidad sobre $((E^S)^\mathbb{I}, \mathcal{B})$ tal que:

$$\begin{aligned} &P\left(\left\{\tilde{x} \in (E^S)^\mathbb{I} / \tilde{x}(0) = x_0, \tilde{x}(1) = x_1, \dots, \tilde{x}(n) = x_n\right\}; \mu, (p_n)_{n \geq 1}\right) \\ &= \mu(\{x_0\}) \cdot p_1(x_0, x_1) \cdot \dots \cdot p_n(x_{n-1}, x_n), \end{aligned}$$

para todo $n \geq 0, x_0, x_1, \dots, x_n$ en E^S . Si $x \in E^S$ y μ es la probabilidad sobre E^S con masa 1 en x , pondremos

$$P(\cdot; x, (p_n)_{n \geq 1}) = P(\cdot; \mu, (p_n)_{n \geq 1}),$$

y si para todo n , $p_n = p$ (matriz de transición) ponemos

$$P(\cdot; \mu, p) = P(\cdot; \mu, (p_n)_{n \geq 1}).$$

Notemos que decir $(X_n)_{n \geq 0}$ es una cadena de Markov con respecto a $P(\cdot; \mu, (p_n)_{n \geq 1})$ sobre $((E^S)^\mathbb{I}, \mathcal{B})$ con valores en E^S significa decir que $(X_n)_{n \geq 1}$ es una sucesión de funciones medibles con dominio en $(E^S)^\mathbb{I}$ y rango en E^S tal que:

$$P(\{X_0 = x\}; \mu, (p_n)_{n \geq 1}) = \mu(\{x\}),$$

para todo $x \in E^S$ y

$$P(\{X_k = y | X_{k-1} = x\}; \mu, (p_n)_{n \geq 1}) = p_k(x, y),$$

para todo $k \geq 1$ y todo $(x, y) \in E^S \times E^S$.

Supongamos que q es simétrica, esto es:

$$q(x, y) = q(y, x),$$

para todo $(x, y) \in E^S \times E^S$.

Proposición 38 Sea ρ una probabilidad nunca nula sobre E^S . Se cumplen:

a)

$$\rho(\{x\})p(x, y; \rho, q) = \rho(\{y\})p(y, x; \rho, q),$$

para todo $(x, y) \in E^S \times E^S$.

b) ρ es $p(\cdot, \cdot; \rho, q)$ -estacionaria. Es decir:

$$\rho(\{y\}) = \sum_{x \in E^S} \rho(\{x\})p(x, y; \rho, q)$$

para todo $y \in E^S$.

c) $p(\cdot, \cdot; \rho, q)$ es irreducible si y sólo si q es irreducible.

Teorema 39 Sean: μ , probabilidad nunca nula sobre E^S , $x \in E^S$. Si $(X_n)_{n \geq 0}$ es una cadena de Markov con respecto a $P(\cdot; x, p)$ sobre $((E^S)^\mathbb{I}, \mathcal{B})$ con valores en E^S , donde p es la matriz de transición $p(x, y; \mu, q)$, entonces

$$\lim_{n \rightarrow \infty} P(\{X_n = y\}; x, p) = \mu(\{y\})$$

para todo $y \in E^S$.

6.2 Dinámica de Metropolis

Sea μ una probabilidad nunca nula y no idénticamente constante sobre E^S . El algoritmo de Metropolis consiste en generar una “muestra” de la distribución μ , valiéndose del Teorema del final de la Subsección anterior.

Supongamos que q es una matriz de transición irreducible.

Sea $x \in E^S$ cualquiera.

Definiremos por inducción una sucesión $(X_n)_{n \geq 0}$ de funciones con dominio en $(E^S)^\mathbb{I}$ de manera tal que esta sucesión resulte ser una cadena de Markov con respecto a $P(\cdot; x, p)$ sobre $\left((E^S)^\mathbb{I}, \mathcal{B}\right)$ con valores en E^S , siendo p la matriz de transición $p(\cdot, \cdot; \mu, q)$.

Algoritmo 40 Metropolis. Sea

$$X_0 = x.$$

Sea $k \geq 0$ y supongamos definidas X_0, \dots, X_k .

Definiremos a seguir X_{k+1} .

Para simplificar la notación pongamos

$$X_k = v.$$

Paso 1: Generar una muestra “y” de la probabilidad sobre $(E^S, \mathcal{P}(E^S))$ dada por la función de probabilidad q_v .

Paso 2: Si $y = v$, definir

$$X_{k+1} = v.$$

Paso 3: Si $y \neq v$ y $\mu(\{y\}) \geq \mu(\{v\})$, definir

$$X_{k+1} = y.$$

Paso 4: Si $y \neq v$ y $\mu(\{y\}) < \mu(\{v\})$, entonces:

4.1) Generar una muestra “u” de la distribución uniforme sobre $(0, 1)$.

4.2) Si

$$u < \frac{\mu(\{y\})}{\mu(\{v\})},$$

definimos

$$X_{k+1} = y.$$

Si

$$u \geq \frac{\mu(\{y\})}{\mu(\{v\})},$$

definimos

$$X_{k+1} = v.$$

Veamos ahora un resultado que justifica el algoritmo de Metropolis.

Sea $x \in E^S$ cualquiera. Sea X_0 la función medible definida sobre $(E^S)^\mathbb{I}$ por

$$X_0(\tilde{y}) = x,$$

para todo $\tilde{y} \in (E^S)^\mathbb{I}$.

Sea $k \geq 0$ y supongamos definidas X_0, \dots, X_k .

Sean:

$$A = \{(v, z) \in E^S \times E^S \mid \mu(\{v\}) < \mu(\{z\})\};$$

ξ y U funciones medibles definidas sobre $(E^S)^\mathbb{I}$ con valores en E^S y $[0, 1]$ respectivamente que son $P(\cdot; x, p)$ -independientes y tales que:

1)

$$P(\xi = z \mid X_k = v) = q(v, z),$$

para todo z y v en E^S ;

2) U y (X_0, \dots, X_k) son $P(\cdot; x, p)$ -independientes;

3) Bajo $P(\cdot; x, p)$, U tiene distribución uniforme sobre $[0, 1]$. Esto es:

$$P(U \leq t; x, p) = \begin{cases} 0 & \text{si } t \leq 0, \\ t & \text{si } 0 < t < 1, \\ 1 & \text{si } t \geq 1. \end{cases}$$

Sea W función medible con valores en $\{0, 1\}$ definida sobre $(E^S)^\mathbb{I}$ por

$$W(\tilde{y}) = \mathbf{1}_A(\xi(\tilde{y}), X_k(\tilde{y})) \cdot \mathbf{1}_{[i(\tilde{y}), 1)}(U(\tilde{y})),$$

donde

$$i(\tilde{y}) = \frac{\mu(\{\xi(\tilde{y})\})}{\mu(\{X_k(\tilde{y})\})}.$$

Finalmente, sea X_{k+1} definida sobre $(E^S)^\mathbb{I}$ por

$$X_{k+1}(\tilde{y}) = \xi(\tilde{y}) + W(\tilde{y})(X_k(\tilde{y}) - \xi(\tilde{y})).$$

Proposición 41 *Se cumple*

$$P(\{X_{k+1} = y \mid X_k = v\}; x, p) = p(v, y),$$

para todo y y v en E^S .

Luego $(X_n)_{n \geq 0}$ es una $P(\cdot; x, p)$ -cadena de Markov.

Notemos que este algoritmo puede ser aplicado a cualquier probabilidad, μ , nunca nula sobre E^S . Esto es, si queremos generar una muestra de una distribución que sabemos es de Gibbs, este algoritmo no usa este mayor conocimiento. Por lo tanto, es razonable preguntarse si no será posible encontrar un algoritmo que sí use esa información y que sea más eficiente en algún sentido que el algoritmo de Metropolis. Una respuesta a esta cuestión la veremos a seguir.

6.3 Muestreo de una distribución de Gibbs propiamente dicha (Gibbs sampler)

Como antes, μ una probabilidad nunca nula y no idénticamente constante sobre E^S . $\mathcal{V} = \{V_s \mid s \in S\}$ un sistema de vecindades sobre S ; $\mathcal{G} = (S, \mathcal{V})$ un grafo; \mathcal{C} la familia de todos los subconjuntos \mathcal{G} -completos de S ; $V : E^S \times \mathcal{P}(S) \rightarrow \mathbb{R}$ un

\mathcal{G} -potencial; $H_V : E^S \rightarrow \mathbb{R}$ la función de energía asociada a V ; μ una distribución de Gibbs con respecto a \mathcal{G} y V . Esto es:

$$\mu(\{x\}) = \frac{1}{Z_V} \exp(-H_V(x)),$$

donde

$$H_V(x) = \sum_{\Lambda \in \mathcal{C}} V(x, \Lambda).$$

y

$$Z_V = \sum_{y \in E^S} \exp(-H_V(y)).$$

Este algoritmo toma en consideración la familia de características locales asociadas a μ , $\mu_s : E^S \rightarrow [0, 1]$, definidas para todo $s \in S$, según vimos en la Definición 1 y que en este caso están dadas por:

$$\mu_s(x) = \frac{\exp(-H_{V \cap \{s\}}(x))}{\sum_{z \in E} \exp(-H_{V \cap \{s\}}(I_{\{s\}}(z, x)))},$$

con

$$H_{V \cap \{s\}}(x) = \sum_{s \in \Lambda \in \mathcal{C}} V(x, \Lambda),$$

y

$$I_{\{s\}}(z, x)(t) = \begin{cases} z & \text{si } t = s, \\ x(t) & \text{si } t \neq s. \end{cases}$$

Sea $(s_k)_{k \geq 1}$ una sucesión en S tal que

$$\#(\{k / s_k = s\}) = \infty,$$

para todo $s \in S$.

Para cada $k = 1, 2, \dots$, sea $q_k : E^S \times E^S \rightarrow [0, 1]$ definida por

$$q_k(v, y) = \begin{cases} \mu_{s_k}(y) & \text{si } y(t) = v(t) \text{ para todo } t \neq s_k, \\ 0 & \text{caso contrario.} \end{cases}$$

Sea $x_0 \in E^S$ cualquiera.

Definiremos por inducción una sucesión $(X_n)_{n \geq 0}$ de funciones con dominio en $(E^S)^{\mathbb{I}}$ de manera tal que esta sucesión resulte ser una cadena de Markov con respecto a $P(\cdot; x_0, (q_k)_{k \geq 1})$ sobre $((E^S)^{\mathbb{I}}, \mathcal{B})$ con valores en E^S .

Algoritmo 42 Sea

$$X_0 = x_0.$$

Sea $k \geq 0$ y supongamos definidas X_0, \dots, X_k .

Definiremos a seguir X_{k+1} .

Para cada $s \in S$ sea $\xi_s : (E^S)^{\mathbb{I}} \rightarrow E^S$ una función medible tal que:

$$P(\xi_s = z | X_k = v, X_{k-1} = w_2, \dots, X_1 = w_k, X_0 = x_0) = \mu_s(I_{\{s\}}(z, v)),$$

cualesquiera sean $z \in E$, v, w_2, \dots, w_k en E^S .

Definiremos a seguir X_{k+1} sobre $(E^S)^{\mathbb{I}}$ por

$$X_{k+1}(\tilde{y})(t) = \begin{cases} X_{k+1}(\tilde{y})(t) & \text{si } t \neq s_{k+1}, \\ \xi_{s_{k+1}}(\tilde{y}) & \text{si } t = s_{k+1}. \end{cases}$$

Se puede ver entonces que:

$(X_n)_{n \geq 0}$ es una cadena de Markov con respecto a $P(\cdot; x_0, (p_k)_{k \geq 1})$ sobre $((E^S)^{\mathbb{I}}, \mathcal{B})$ con valores en E^S .

También que se cumple el siguiente Teorema que fundamenta el algoritmo propuesto.

Teorema 43 *Teorema de convergencia de Geman y Geman*

$$\lim_{n \rightarrow \infty} P(\{X_n = y\}; x_0, (p_k)_{k \geq 1}) = \mu(\{y\})$$

para todo $y \in E^S$.

Bibliografía

- [1] FRIGESSI, A. y PICCIONI, M. (1988), *Parameter estimation for the two-dimensional Ising fields corrupted by noise*, Quaderno, IAC-CNR, Roma.
- [2] BUSTOS, O. H. y FRERY, A. C. (1992), *A contribution to the study of markovian degraded images: an extension of a theorem by Geman and Geman*, Computational & Applied Mathematics, 11, pags. 17-29.
- [3] GEMAN, S., and GEMAN, D. (1984), *Stochastic relaxation, Gibbs distribution, and the bayesian restoration of images*, IEEE Trans. Pattern Anal. Machine Intell., PAMI 6, pags. 721-741.