

Percepción Computacional

Introducción a los sistemas de percepción

Ideas clave

1.1. ¿Cómo estudiar este tema?

Para estudiar este tema deberás leer con atención las ideas clave que se desarrollan a continuación.

En el presente tema veremos con detalle los sistemas de percepción del ser humano: el sistema auditivo y el sistema visual.

Sistemas de percepción

Sistema auditivo

- ▶ Definición y medición del sonido como fenómeno físico.
- ▶ Partes del oído humano y su función:
 - Oído interno.
 - Oído medio.
 - Oído externo.
- ▶ Respuesta en frecuencia del oído.
- ▶ Características de la voz como señal sonora.

Sistema visual

- ▶ Células específicas del sistema visual humano: características físicas y respuesta en frecuencia.
- ▶ Concepto de diferencia mínima perceptible.

- ▶ Etapas de procesado en el sistema visual:
 - Respuesta logarítmica.
 - Inhibición lateral (filtro paso alto).
 - Muestreo o filtrado temporal.
- ▶ Síntesis de color.

1.2. Objetivos

El objetivo de este tema es comprender los **mecanismos de captación de señales auditivas y visuales** por parte del ser humano. Estos sistemas sirven de referencia en la implementación de técnicas artificiales de percepción, pues su arquitectura y funcionamiento trata, en muchos casos, de emular el funcionamiento de los sistemas biológicos que encontramos en la naturaleza.

Asimismo, el conocimiento de los procesos de recepción y transformación de estímulos en el ser humano, como la luz o el sonido, facilitará la comprensión de las técnicas de tratamiento de información descritas en temas posteriores de la asignatura.

1.3. Funcionamiento del sistema auditivo

El sonido

El sonido se define como una **perturbación mecánica del medio** (elástico) por el que se propaga en forma de onda. El sonido resulta de la vibración de ida y vuelta de las partículas del medio a través del cual se mueve dicha onda. Si una onda de sonido se mueve de izquierda a derecha a través del aire, las partículas de aire se desplazarán

hacia la derecha y hacia la izquierda a medida que la energía de la onda de sonido lo atraviesa.

La **caracterización física** del sonido permite su medición y evaluación cuantitativa ante diferentes fenómenos:

- ▶ La intensidad acústica de una onda sonora viene dada por la rapidez promedio del flujo de energía acústica a través de un área unitaria, midiéndose en W/m^2 .
- ▶ Los niveles sonoros son la medida de la energía sonora en dB. Se define el nivel de intensidad (NI) de la siguiente forma:

$$NI = 10 \log(I/I_{ref})$$

Donde I_{ref} es la **intensidad de referencia**. Esta representa el umbral de audibilidad a 1 KHz en el aire libre y viene dada por:

$$I_{ref} = 10^{-12} \text{ W/m}^2$$

En este apartado se estudiará cómo el cuerpo humano es capaz de captar y procesar estas ondas sonoras. Para ello, en primer lugar, se lleva a cabo una descripción del oído como órgano fundamental en la captación y transducción de estas señales.

El oído

El oído es el órgano encargado de la transmisión de sonidos hasta el cerebro. En él, se distinguen tres partes diferentes: **oído externo, oído medio y oído interno**.

La siguiente imagen muestra la estructura del oído, observándose cada uno de los elementos que lo componen. Además de ser el responsable del sentido del equilibrio, el oído detecta y convierte las señales sonoras en impulsos eléctricos capaces de ser conducidos a través del sistema nervioso.

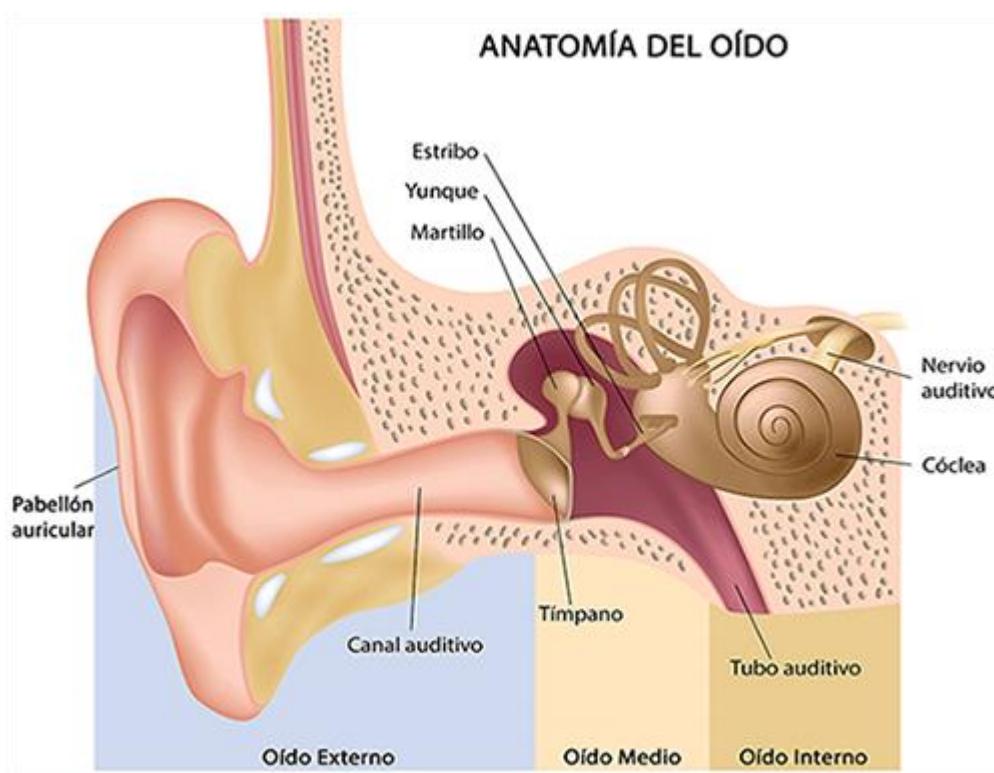


Figura 1. Estructura del oído. Detalle de los elementos que componen oído externo, medio e interno.

Fuente: <https://www.medicaloptica.es/salud-auditiva/como-funciona-el-oido>

Oído externo

La **función** del oído externo es recoger las ondas sonoras procedentes del entorno. Hace las veces de una antena; de hecho, su forma contribuye a reducir la impedancia o resistencia para las ondas, que capta las señales sonoras y las redirige hacia el interior.

Los **elementos** que componen el oído externo son el pabellón auditivo, el conducto auditivo y el tímpano. A continuación, se indica cada una de sus partes y sus correspondientes funciones en el proceso de captación de sonidos:

- ▶ Pabellón auditivo: El pabellón auditivo o, como se conoce comúnmente, **oreja**, se trata de la única parte externa y visible del oído y se caracteriza por su forma helicoidal. Funciona como una antena receptora de señales sonoras, que son captadas y dirigidas hacia el interior del oído.

La **presión en el interior** del oído es mayor que en el exterior, debido a la mayor compresión del aire contenido, por lo que las ondas no penetrarían si no estuviéramos dotados del pabellón auditivo. Este actúa como una interfaz entre el interior y el exterior, disminuyendo la resistencia que la parte interior del oído ofrece a las ondas sonoras y suavizando la diferencia de presión existente. Como resultado, un mayor flujo de sonido es capaz de penetrar en el conducto auditivo.

- ▶ **Conducto auditivo:** Las ondas captadas por el pabellón auditivo son conducidas hacia el tímpano mediante el conducto auditivo. La longitud de este conducto es de dos a tres centímetros. La forma curvada en su extremo permite proteger el tímpano de, por ejemplo, insectos o impurezas que pudieran llegar hasta él. Además, el conducto auditivo se encuentra recubierto de cera (cerumen), que contribuye a impedir la entrada de materias perjudiciales como polvo o suciedad.

Sin embargo, la protección del tímpano no es su función principal. El conducto auditivo amplifica los sonidos bajos y menos penetrantes de la voz humana, actuando como un **audífono**. De este modo, el oído subsana parte de la debilidad de la señal de voz, facilitando la escucha y comprensión de una conversación normal.

- ▶ **Tímpano:** La membrana timpánica o tímpano es una fina membrana situada en el extremo final del conducto auditivo. Supone el final del oído externo y el comienzo del oído medio. El tímpano posee una **gran sensibilidad**, de forma que las ondas de sonido que penetran en el oído hacen que vibre. Estas vibraciones, debidas a la presión de la señal entrante, son transmitidas hacia el interior del oído para su posterior transformación y procesado.

Oído medio

El oído medio es la parte del oído comprendida entre el tímpano y la ventana oval. Lleva a cabo la **transmisión de señales de sonido** desde el oído externo hasta el oído

interno. Se compone de tres huesos conectados entre sí: martillo, yunque y estribo; además de la ventana oval, la ventana redonda y la trompa de Eustaquio.

- ▶ Los huesos del oído medio: las vibraciones del tímpano se transmiten al interior del oído a través de los tres huesos del oído medio: **martillo, yunque y estribo**. Estos transmiten las vibraciones del tímpano como consecuencia de la onda sonora entrante hacia el interior del oído. El estribo es el último de esta cadena de huesos y está conectado con la ventana oval, donde tiene lugar la siguiente etapa de percepción del sonido.
- ▶ La ventana oval: es una fina membrana que recubre el extremo de la cóclea expuesto hacia los huesos del oído medio. Actúa como un **amplificador acústico** previo al oído interno para las señales sonoras transmitidas desde el tímpano a la ventana oval. De hecho, se estima que la presión de la señal sonora es aproximadamente veinte veces superior en la ventana oval respecto al tímpano. Este aumento se debe al mayor tamaño de la superficie de la membrana timpánica en comparación con la de la ventana oval.

Una misma fuerza, en este caso, ejercida por la perturbación sonora sobre las partículas del medio, conlleva un aumento de presión cuando se aplica sobre una superficie menor.

- ▶ La trompa de Eustaquio: es parte del oído medio y supone una **vía de comunicación entre el oído y la parte posterior del paladar**. Es la encargada de equiparar el nivel de presión entre las dos caras de la membrana timpánica, igualando la presión del aire en el interior y exterior del oído. Este conducto se abre cuando realizamos la acción de tragarn, permitiendo la entrada de aire.

Si en el oído se acumula, es decir, el interior del tímpano está sometido a una mayor presión del aire respecto al exterior, se impide que esta membrana vibre adecuadamente. Como consecuencia, se reduce la capacidad auditiva. En escenarios en los que la presión ambiental cambia con rapidez es habitual padecer

esta diferencia de presión entre el oído y el exterior. Este es el caso de los vuelos en avión o la travesía por montañas elevadas.

Oído interno

Es la última parte del oído y supone la conexión de este con el cerebro; contiene la cóclea como elemento principal.

- ▶ La cóclea: esta lleva a cabo la transformación de las ondas sonoras en **impulsos eléctricos capaces** de ser transmitidos al cerebro a través de las conexiones nerviosas. Posteriormente, el cerebro traduce estos impulsos en estímulos que podemos reconocer y entender.
 - La cóclea, con forma de concha de caracol, contiene dos membranas cercanas entre sí y está rellena por un fluido conocido como perilinfa. Estas membranas forman la membrana basilar, una especie de pared de separación en la cóclea.
 - La membrana basilar tiene un pequeño orificio, helicotrema, que permite el movimiento del fluido de un lado al otro de la membrana basilar. El helicotrema garantiza que las vibraciones de la ventana oval se transmitan a todo el fluido que rellena la cóclea.
 - A su vez, el interior de la membrana contiene miles de microscópicas fibras pilosas (24 000 fibras aproximadamente). Estas se mueven como consecuencia del movimiento del fluido en el interior de la cóclea. Existen diferentes tipos de fibras, todas ellas conectadas al nervio auditivo. En función de la naturaleza de los movimientos de la perilinfa, diferentes tipos de fibras se ponen en movimiento y generan impulsos eléctricos en el nervio auditivo.
- ▶ El nervio auditivo es el canal de comunicación que conecta la cóclea con el cerebro. Cuando las células pilosas que se mencionaban previamente se mueven, el nervio transmite la señal eléctrica resultante hasta la zona del cerebro responsable de la audición. Allí, los sonidos captados se interpretan y asimilan. Por tanto, el rol de estas células pilosas es esencial para nuestra capacidad de audición. Si estas

fibras resultaran dañadas, nuestra capacidad auditiva se vería severamente deteriorada.

El oído humano es capaz de captar ondas sonoras en el rango entre los 20 Hz y los 20 KHz de frecuencia. Las ondas por debajo y por encima de estos límites se denominan **infrasonidos** y **ultrasonidos**, respectivamente.

Los sonidos pueden distinguirse gracias a las propiedades de resonancia de la **membrana basilar**, que permite identificar ondas sonoras de diferente frecuencia. Esta membrana tiene propiedades mecánicas que varían desde su base hasta el vértice:

- ▶ La base es más fina, tiene menor masa y elasticidad.
- ▶ En el vértice, al contrario, tanto la masa como la elasticidad de la membrana son mayores. Así, la amplitud del movimiento de la membrana ante la excitación producida por una onda sonora está más cerca de la base para frecuencias de onda mayores.
- ▶ Esta amplitud de pico se encuentra más próxima al vértice conforme disminuye la frecuencia de la onda entrante. Por tanto, esta membrana basilar actúa como un banco de filtros paso-banda que dotan al oído de su capacidad selectiva en frecuencia.

Características físicas del oído humano

Por último, el oído está caracterizado por diferentes umbrales de funcionamiento. El **umbral de audibilidad** es el nivel de intensidad mínimo en espacio libre que el oído es capaz de oír. La máxima sensibilidad se tiene a 4 KHz.

A 120 dB, respecto al umbral de referencia, tenemos el **umbral de sensación**. A este nivel de intensidad se produce cierto cosquilleo en el oído. A partir de 140 dB se puede llegar a producir dolor.

Una de las ondas de mayor interés es la propia voz humana. Físicamente, se caracteriza por un rango frecuencial entre 300 Hz y 3.4 KHz. Por tanto, el ancho de banda asignado en telecomunicaciones para un solo canal de transmisión de voz suele ser de 4 kHz. Como se verá más adelante en la asignatura, este ancho de banda permite utilizar una frecuencia de muestreo de 8 kHz como base del sistema de modulación digital.

1.4. Percepción visual

Funcionamiento del sistema visual humano

La **visión** es el fenómeno resultante de la percepción del color, forma y distancia de los objetos en un espacio. La visión se produce como resultado de la incidencia de la luz, caracterizada como una onda electromagnética, sobre la retina del ojo.

Esta onda debe estar comprendida en un rango frecuencial (luz visible) para producir impresión visual. El color percibido depende de la frecuencia (longitud de onda) de la luz que incide sobre el objeto observado. Asimismo, el estado del propio ojo también influye en la percepción de los colores, como ocurre en el caso del daltonismo.

El comportamiento del ojo humano se asemeja al de una cámara fotográfica. La luz reflejada en los objetos se proyecta sobre un material fotosensible (**retina**) tras pasar a través de una lente (**cristalino**), tal y como se esquematiza en la siguiente imagen.

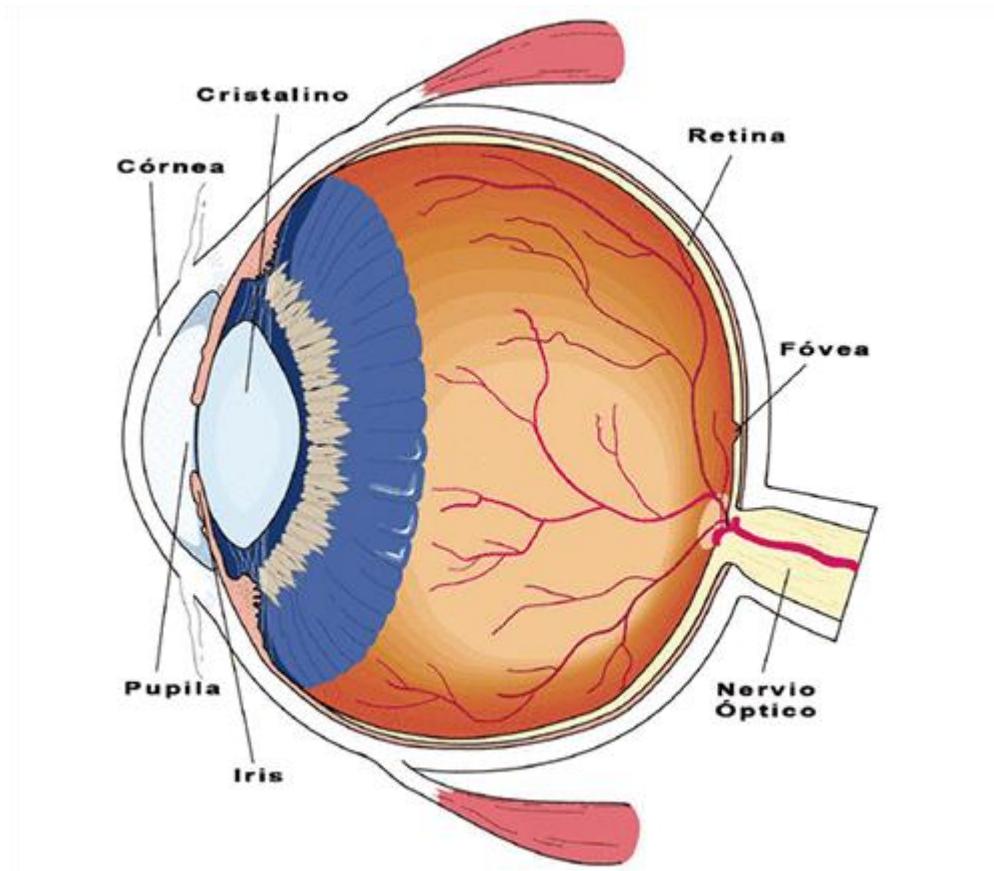


Figura 2. Diagrama simplificado de un corte de ojo humano.

Fuente: <http://www.lucescei.com/estudios-y-eficiencia/extractos-libro-blanco-de-iluminacion/el-sistema-visual-humano/>

La zona visible del cristalino es la **pupila**, que regula el flujo de luz que entra en el ojo. El cristalino refracta los rayos de luz para que estos se focalicen en la zona adecuada de la retina. Gracias al cambio de tamaño de la pupila (cristalino), permite enfocar con precisión objetos cercanos y lejanos. Es decir, se adapta al entorno para producir el resultado de una visión satisfactoria.

La **retina** es una membrana interior y en ella se encuentran las células fotosensibles. Estas células son de dos tipos: conos y bastones.

Conos	Son menos numerosos y muy poco sensibles a la luz. Son las células encargadas de la visión diurna .
Bastones	Precisan de menor cantidad de luz que los conos para su excitación. Por tanto, son las células sensoriales responsables de la visión nocturna .

Figura 3. Tipos de células fotosensibles.

La siguiente gráfica muestra la distribución espacial de estas células en la retina, tomando como referencia el punto central de la retina o **fóvea**. Como puede apreciarse, los bastones se encuentran lejos de la fóvea, principalmente. Mientras, los conos se concentran en torno a esta.

Además, se aprecia la existencia de un **punto ciego en la retina**, carente de células fotosensibles, y que es el lugar desde el que parten los nervios ópticos hacia el cerebro.

La fóvea recibe la luz desde el centro del campo visual, es decir, desde el punto del espacio al que miramos directamente. Los conos, principalmente concentrados en esta zona, nos captan la luz procedente del punto focalizado y son responsables de la visión directa o central. En cambio, los bastones, más alejados de la fóvea, proporcionan visión periférica.

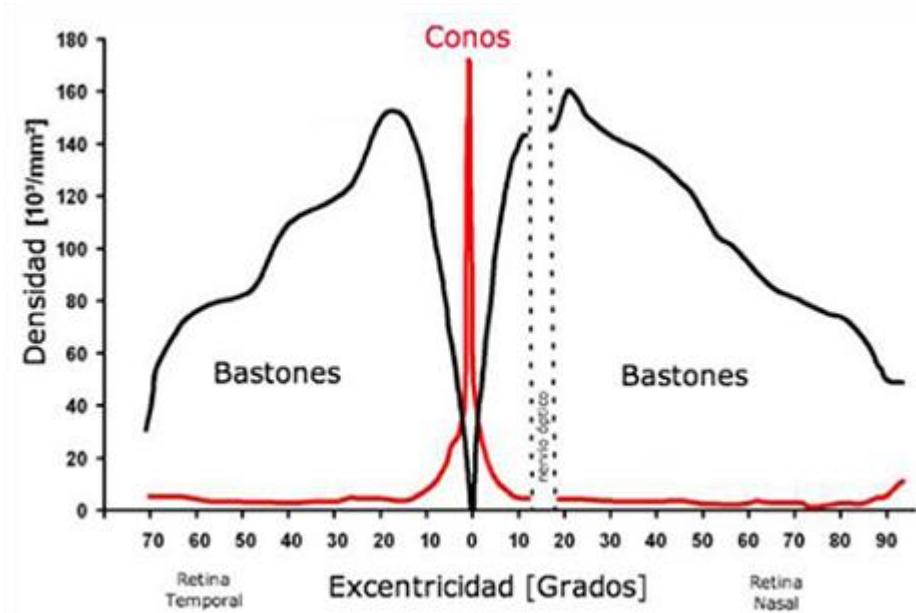


Figura 4. Distribución espacial de conos y bastones en la retina del ojo humano.

Fuente: <https://idus.us.es/xmlui/>

Conos

Los conos actúan con filtros sobre la luz incidente en ellos. Así, los humanos tienen **tres tipos** de conos:

- ▶ El primero responde más a la luz de longitudes de onda largas, alcanzando un máximo a aproximadamente 560 nm. Este tipo de cono se designa mediante la letra **L (long)**.
- ▶ El segundo tipo de cono filtra la luz de longitud de onda media, alcanzando un máximo a 530 nm, y se abrevia mediante la letra **M (medium)**.
- ▶ El tercer tipo responde más a la luz de longitud de onda corta, alcanzando un máximo de 420 nm, y se designa mediante **S (short)**.

Estas longitudes de onda (L, M y S) se corresponden de forma aproximada con los colores **rojo, verde y azul** respectivamente. En el siguiente gráfico vemos la respuesta en frecuencia de los tres tipos de células cono a la luz. Los diferentes tipos de conos permiten percibir distintos colores. Estos se forman a partir de la diferente excitación de los tres tipos de células receptoras.

El color amarillo, por ejemplo, se percibe cuando los conos L se estimulan ligeramente más que los conos M, y el color rojo se percibe cuando los conos L se estimulan significativamente más que los conos M. De manera similar, los tonos azules y violetas se perciben cuando el receptor S se estimula más. Así, mediante la mezcla ponderada de los colores rojo, azul y verde somos capaces de construir la gama cromática perceptible por el ser humano.

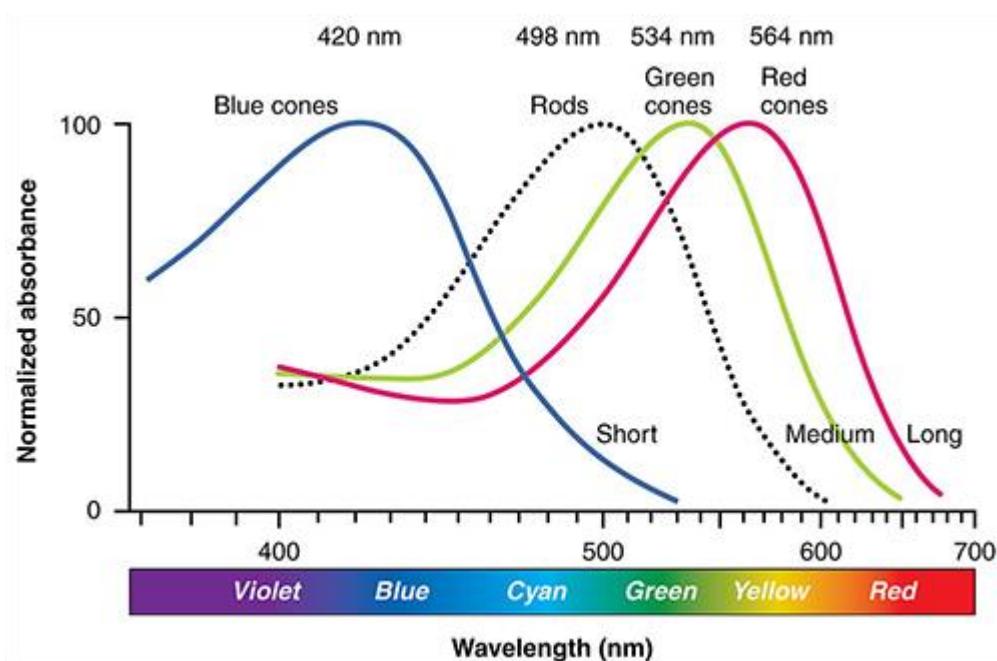


Figura 5. Respuesta en frecuencia de los tres tipos de conos: S (longitud de onda corta, azul), M (longitud de onda media, verde) y L (longitud de onda larga, rojo).

Fuente: https://commons.wikimedia.org/wiki/File:1416_Color_Sensitivity.jpg.

Bastones

Por otro lado, observamos que los bastones **no son sensibles al color** (frecuencia), dado que solo hay un tipo de bastón. Así es que estas células no permiten la formación de colores. Este es el motivo por el que en la oscuridad no somos capaces de distinguir ningún color.

Fenómenos visuales: respuesta logarítmica e inhibición lateral

Respuesta logarítmica

El concepto de *just noticeable difference* (JND) representa, en el ámbito de la psicofisiología, la mínima cantidad de variación ΔI en la magnitud de un estímulo I para que esta sea apreciada.

Por ejemplo, se observa como la siguiente imagen contiene dos áreas de intensidades diferentes. En este caso, dada la clara apreciación de ambas áreas, el incremento de intensidad aplicado sobre el área exterior es mayor a aquel dado por la JND.

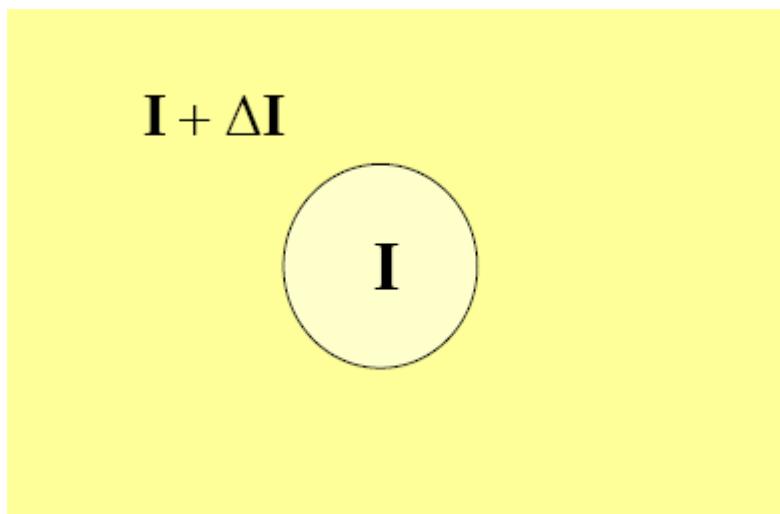


Figura 6. Ilustración del experimento llevado a cabo para la determinación de la JND en una imagen.

Fuente: www.tel.uva.es.

En el siglo XIX, el psicólogo Ernst Weber determinó que existe una relación constante entre ΔI e I en un amplio rango de I . Es decir, la mínima variación necesaria para percibir un cambio en un estímulo se incrementa conforme el estímulo se hace más intenso. Matemáticamente, esta **Ley de Weber** se establece de la siguiente forma:

$$\Delta I/I = \lambda$$

En el marco de la percepción visual, en el que I representaría la cantidad de luz o intensidad percibida, conforme dicha intensidad es mayor, es preciso una variación ΔI más significativa de la misma para que sea apreciable. Por tanto, se concluye que esta **JND es diferente en zonas claras u oscuras de una imagen.**

Así, consideremos:

- ▶ Una imagen Im_0 oscura y su JND asociada, representada por ΔI_0 .
- ▶ Una imagen más clara, Im_1 con una JND representada por ΔI_1 .

De acuerdo a la Ley de Weber, podemos comparar ambas JND y deducir que:

$$\Delta I_0 < \Delta I_1$$

Es decir, los cambios de intensidad en una imagen se aprecian mejor en zonas oscuras. Por tanto, estas zonas serán más sensibles a la presencia de ruido o artefactos de naturaleza aleatoria.

La Ley de Weber muestra, por tanto, que el incremento necesario en la intensidad del estímulo para resultar en una respuesta perceptiva similar aumenta para mayores valores de intensidad. Además, la relación entre la percepción del estímulo y la intensidad de este sigue una relación logarítmica.

Obsérvese el siguiente gráfico. Se ilustra cómo para obtener un incremento similar en la respuesta del ojo humano (en la percepción de la intensidad luminosa recibida), es necesario un mayor incremento de la intensidad de partida para intensidades mayores.

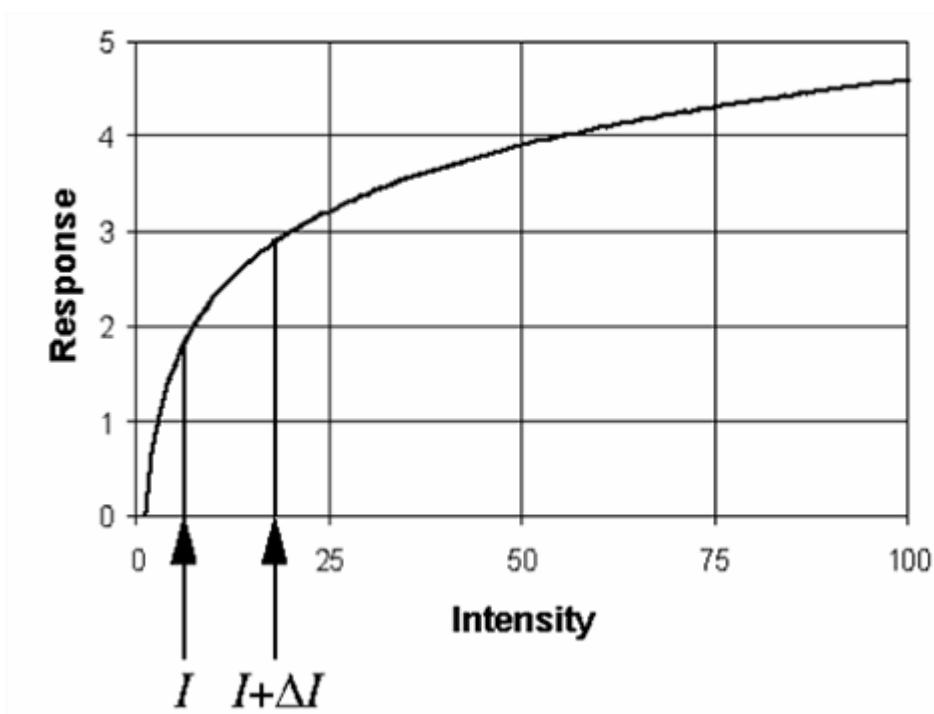


Figura 7. Respuesta logarítmica de la percepción de intensidad de acuerdo a la Ley de Weber.

Fuente: <https://sites.google.com/site/udctopsicologia1/tema-01-introduccion-a-la-percepcion>.

La demostración matemática de este comportamiento puede obtenerse de la siguiente forma. Para una intensidad I , tenemos que el valor de la intensidad mínimamente apreciable es el siguiente:

$$I_{JND} = I + \Delta I = I + \lambda I = (1 + \lambda)I$$

Por tanto, la diferencia entre I_{JND} e I viene dada por:

$$I_{JND} - I = \lambda I$$

Si calculamos esta diferencia en el dominio logarítmico, tenemos la siguiente relación:

$$c_{JND} = \log(1 + \lambda) + \log I$$

$$c = \log I$$

$$c_{JND} - c = \log(1 + \lambda)$$

Es decir, en el dominio logarítmico, los incrementos de intensidad para JND son constantes.

Inhibición lateral

Además de esta transformación logarítmica, el sistema visual humano lleva a cabo un filtrado espacial de la señal luminosa recibida que, como resultado, conlleva el **realce del contraste entre zonas de diferente intensidad**. Las áreas de la imagen donde la intensidad luminosa cambia bruscamente de claro a oscuro, o viceversa, denotan la presencia de bordes.

La conexión de conos y bastones con las células de la retina permite capturar y realizar estos cambios. Tanto conos como bastones están conectados con dos tipos de células (en la segunda y tercera capa de la retina, respectivamente). Estas llevan a cabo un procesado de la señal visual equivalente al que produce el operador laplaciano (diferencial de segundo orden que magnifica las zonas de la señal en los que se aprecian cambios bruscos). Concretamente, obviando la zona central de la retina:

- ▶ Las señales de una gran proporción de conos y bastones vecinos llegan a las células de la tercera capa de la retina, constituyéndose un **agregado** (signo +) de la intensidad recibida.
- ▶ Mientras, las células de la segunda capa retiniana son responsables del envío de **impulsos inhibitorios** (signo -) equivalentes a la substracción del número de impulsos bioeléctricos de otros puntos próximos de la imagen (píxeles).

Este fenómeno constituye la **inhibición lateral**, que, con un comportamiento similar a un filtro de frecuencias altas, nos ayuda a percibir el contraste, facilitando la posterior identificación de límites o fronteras entre regiones de diferente intensidad, así como de contornos o bordes. Este fenómeno fue descrito por Mach, tal y como refleja el experimento mediante bandas de diferente intensidad.

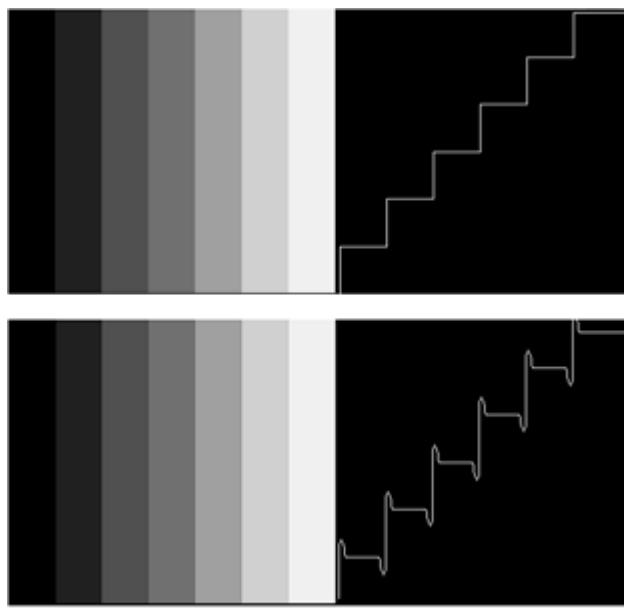


Figura 8. Bandas de Mach y efecto resultante del filtrado espacial llevado a cabo por el ojo humano.

Fuente: <http://www.ub.edu/pa1/node/99>.

Muestreo o filtrado temporal

Por último, el ojo humano también realiza un filtrado temporal de la señal capturada. De hecho, su respuesta temporal es relativamente lenta. Considérese, por ejemplo, el caso en el que se tiene una única fuente de luz que se enciende de forma intermitente. Si la frecuencia de la fuente es baja, se aprecian los períodos en los que no se emite luz, el tiempo entre emisiones de luz consecutivas ha de ser superior a 30 ms. Sin embargo, para frecuencias superiores, no se percibirán los períodos sin emisión de luz, dando la apariencia de una luz continua.

La frecuencia a la que no se percibe la intermitencia de la fuente de luz se denomina **frecuencia de fusión**. Está en torno a 30 Hz, dependiendo del tamaño y del brillo de la fuente. La aplicación práctica de este fenómeno es la **definición de estándares de codificación de vídeo**. Estos no recogen imágenes de forma continua (señal continua), sino que definen la tasa de imágenes necesaria por segundo (frecuencia de muestreo) para que el espectador no perciba las discontinuidades existentes. Así, los antiguos sistemas PAL y NTSC definían una tasa 25 y 30 imágenes por segundo, respectivamente.

Otro ejemplo que refleja la respuesta temporal del ojo humano es el de dos fuentes de luz intermitentes, alternantes (cada una de ellas emitiendo durante 1 ms) y con una separación angular de 1 grado respecto del observador.

- ▶ Si la frecuencia de la intermitencia es superior a 1 s, se perciben como fuentes de iluminación que se alternan.
- ▶ Si la frecuencia se encuentra en torno a 10 ms, la sensación resultante es la de un punto emisor de luz que se traslada desde una fuente a otra.
- ▶ Por último, si la diferencia temporal de encendidos es de aproximadamente 1 ms, la percepción será una iluminación simultánea.

El experimento demuestra que, por tanto, la visión humana está caracterizada por una **frecuencia de rendición del movimiento**. Esta permite crear sensación continua de movimiento a partir de un conjunto de instantáneas. Pero es necesario que la frecuencia a la que se presentan estas instantáneas sea suficientemente alta para que la percepción no refleje discontinuidades.

A partir de los mecanismos descritos, el proceso de captación de imágenes por parte del sistema de visión humana puede esquematizarse en los siguientes pasos:

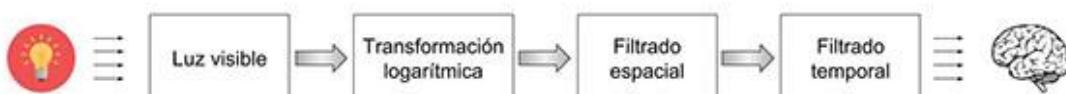


Figura 9. Representación esquemática de las etapas de procesado realizadas por el sistema visual humano en la percepción de imágenes.

- ▶ Filtrado en frecuencia para seleccionar la parte de la radiación luminosa correspondiente al espectro de **luz visible**.
- ▶ Transformación logarítmica (**Ley de Weber**) del estímulo percibido.
- ▶ Filtrado espacial (realce de bordes y fronteras) de acuerdo al mecanismo de **inhibición lateral**.

- ▶ Filtrado temporal (**muestreo de la señal**) que se refleja en la frecuencia crítica de fusión y la frecuencia de rendición del movimiento.

Síntesis de color

Por último, comentaremos en este tema algunas nociones sobre el color. El color de un objeto viene definido por el **contenido espectral de una determinada radiación**, esto es, por $R(f)$. Así, la variación de color de una señal luminosa respecto a otra está asociada a la diferente frecuencia de ambas señales.

Sin embargo, se puede tener dos radiaciones distintas, $R_1(f)$ y $R_2(f)$; con espectros distintos: $R_1(f) \neq R_2(f)$, pero que se traduzcan en la misma percepción de color. Este resultado se debe a que el color percibido es **función de tres canales no independientes**. Como se ha comentado previamente, existen tres tipos de conos, las células fotorreceptoras responsables de la percepción del color. Y cada uno de ellos actúa como un filtro en frecuencia $S_i(f)$ que selecciona parte de los colores de la radiación entrante.

Matemáticamente, la **potencia de la señal de salida** de cada uno de estos canales α_i puede expresarse de la siguiente forma:

$$\alpha_i(R) = \int_{f_{min}}^{f_{max}} R(f)S_i(f)df$$

$$i = 1, 2, 3$$

Por tanto, dos colores $R_1(f)$ y $R_2(f)$ se percibirán de la misma forma si:

$$\alpha_i[R_1(f)] = \alpha_i[R_2(f)] \text{ para } i = 1, 2, 3$$

Así, un determinado color $R(f)$ puede sintetizarse a partir de la superposición de tres primarios $P_k(f)$ encontrando los coeficientes β_k apropiados en la mezcla, tal y como se esquematiza en la siguiente figura.

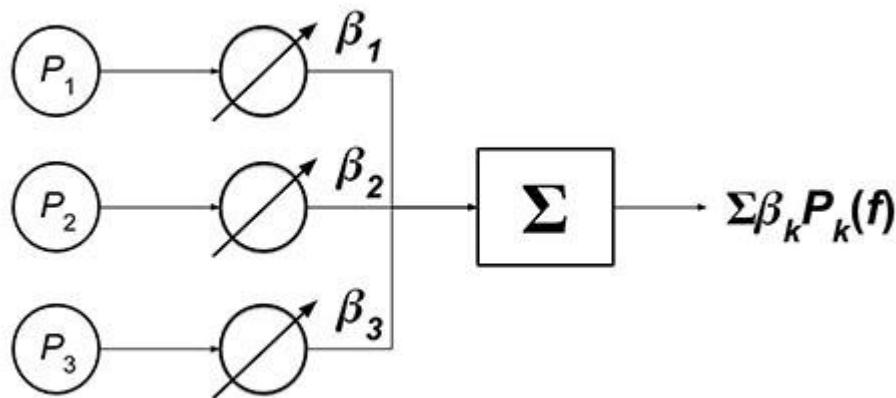


Figura 10. Esquema del proceso de síntesis de color.

Para que el resultado de la síntesis produzca la sensación de color esperada, habrá de cumplirse lo siguiente:

$$\begin{aligned} \alpha_i(R) &= \int_{f_{min}}^{f_{max}} R(f) S_i(f) df = \int_{f_{min}}^{f_{max}} \left[\sum_{k=1}^3 \beta_k P_k(f) \right] S_i(f) df \\ &= \sum_{k=1}^3 \beta_k \int_{f_{min}}^{f_{max}} P_k(f) S_i(f) df = \sum_{k=1}^3 \alpha_{ik} \beta_k \end{aligned}$$

Donde α_{ik} es la respuesta de los conos de tipo i al color primario $P_k(f)$. Este viene dado por la siguiente expresión:

$$\alpha_{ik} = \int_{f_{min}}^{f_{max}} P_k(f) S_i(f) df$$

Por tanto, la igualdad anterior demuestra que los coeficientes de mezcla β_k vienen dados por la solución a un sistema de tres ecuaciones y tres incógnitas, dada la respuesta de los tres filtros de conos a los tres colores primarios inicialmente considerados $P_k(f)$.

Por ejemplo, uno de los **sistemas de color** más relevantes es el **RGB** (*red, green, blue*), comúnmente utilizado en pantallas, y que toma el rojo, el verde y el azul como colores primarios de la mezcla.

Moore, J. K. y Linthicum F. H. Jr. (2007). The human auditory system: a timeline of development. *International journal of audiology*, 46(9), 460-478.

Thorpe, S., Fize, D. y Marlot, C. (1996). Speed of processing in the human visual system. *Nature: International Journal of Science*, 381, 520-522.
doi:10.1038/381520a0

Percepción Computacional

Elementos de un sistema de percepción

Ideas clave

2.1. ¿Cómo estudiar este tema?

Para estudiar este tema deberás leer con atención las ideas clave que se desarrollan a continuación.

2.2. Objetivos

El objetivo principal de este tema es entender los elementos más importantes de un sistema de percepción y cómo, desde el punto de vista computacional, se desarrolla cada módulo de dicho sistema.

En definitiva, se busca dar una visión general de aquellos ingredientes necesarios para que un sistema de percepción computacional tenga sentido y pueda resolver problemas reales.

En este tema no se busca entender los métodos concretos de cada módulo que integra el sistema de percepción (puesto que esos métodos serán material para los siguientes temas).

2.3. Descubriendo los elementos esenciales de un sistema de percepción

Imitando al mundo animal

Existen múltiples maneras de describir los sistemas de percepción. En este tema seguiremos un enfoque que irá del aspecto más genérico al más concreto. Es necesario entender que, en todo momento, los sistemas de percepción intentan imitar a la naturaleza y es, por tanto, lógico pensar que si queremos modelar adecuadamente un sistema, tenemos que empezar por considerar ciertas simplificaciones.

Por ello, sin dejar de imitar a la naturaleza, intentaremos **modelar sistemas que reflejen el comportamiento de organismos sencillos**, como puede ser el de individuos que ni siquiera poseen ojos o elementos concretos para la captura de información visual, como el caso de los moluscos. Estos seres realizan tres funciones de percepción muy sencillas:

Captura de información

- ▶ Este proceso consiste en obtener estímulos del exterior que reflejen qué actividad (movimientos, cambios de temperatura, amenazas, etc.) está teniendo lugar.
- ▶ La captura de información puede considerar múltiples estímulos principalmente físicos, mecánicos o químicos.
- ▶ La información recogida es normalmente mayor que la necesitada para entender el exterior. La adaptación de ese exceso de sensibilidad a las necesidades del entorno es parte de la propia evolución.

Procesamiento

- ▶ La información capturada necesita de un procesamiento posterior para eliminar los datos innecesarios y redundantes; entender si está siendo suficiente y, en caso contrario, redirigir la captura de información hacia otro punto del espacio o bien combinar la información proveniente de una determinada fuente con otras fuentes o con informaciones pasadas.
- ▶ Sin este procesamiento, un ser vivo podría tomar conclusiones erróneas, confundir eventos o incluso conducirlo a su extinción.
- ▶ Una de las características esenciales de este procesamiento es que debe, en todo momento, realizarse a la misma **velocidad** con la que los datos iniciales se están capturando. De esa manera, se asegura de que la información que está llegando al cerebro es únicamente la necesaria.
- ▶ Por el contrario, un animal que almacene, por ejemplo, todo lo que ha oído durante el día y lo analice durante la noche, no podrá esquivar el ataque instantáneo de un depredador.

Toma de decisión y aprendizaje

- ▶ El fin principal de la percepción del exterior no es otra que la toma de decisiones.
- ▶ Una mejor toma de decisiones (basada en más información, contexto o diversidad de fuentes) conducirá, sin lugar a dudas, a una ventaja competitiva mayor.
- ▶ Esa toma de decisiones tiene una consecuencia directa también: el aprendizaje. El ser vivo aprenderá que un determinado estímulo asociado a una determinada decisión tendrá unas consecuencias. Esas consecuencias serán almacenadas para optimizar tanto la toma de decisiones como la captura de información.
- ▶ La **percepción del exterior** es clave para el aprendizaje.

El siguiente esquema capture los puntos anteriormente vistos, enfatizando el hecho de que es un proceso iterativo y constante.

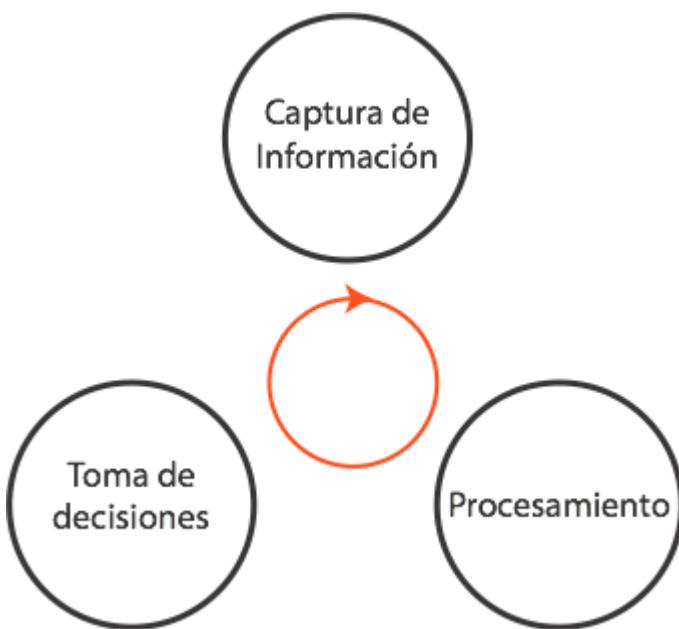


Figura 1. Esquema reducido de percepción del exterior.

Limitaciones de imitar el mundo animal

Los tres puntos anteriormente vistos (captura de información, procesamiento y toma de decisiones) proporcionan una visión muy general de lo que puede ser un sistema de percepción computacional.

Si sustituimos:

- ▶ El módulo de captura de información por una cámara de vídeo.
- ▶ El módulo de procesamiento por un filtrado muy sencillo en color, quedándonos con las componentes en azul, por ejemplo.
- ▶ Y en la toma de decisiones aplicamos una regla que diga si la imagen contiene componentes en azul en más de un 70 % del total.

Podríamos tener un detector de mares o un detector de cielos.

Es un ejemplo muy sencillo, pero muy limitado a una funcionalidad concreta. De hecho, con estos tres módulos es muy complicado hacer funcionalidades más complejas como detectar movimiento, seguir un objeto, analizar la textura de una foto o incluso analizar la forma de un objeto.

Las simplificaciones heredadas de la naturaleza nos ayudan a entender los pasos más esenciales que suceden delante de nuestros ojos, pero el adaptarlo a la resolución de un problema concreto es labor de la inteligencia humana.

Esto es lo que haremos en este tema, entender qué elementos adicionales son necesarios en el esquema de percepción del exterior para desarrollar módulos avanzados de percepción computacional y dar respuesta a los problemas que se plantean actualmente.

2.4. Captura de información

Parámetros que definen una adecuada captura de información

En este apartado nos centraremos en todo lo relativo a la captura de información del exterior. Existen numerosos parámetros que definen cómo es dicha la captura y puede discutirse cuáles son los más importantes. Sin embargo, los más esenciales son los siguientes:

Especificidad

La capacidad de un sistema de captura de información para recoger fielmente los eventos sucedidos es clave. Un ejemplo de especificidad es la **temperatura**.

Un determinado sensor de vídeo, como puede ser una cámara, puede estar en el exterior capturando imágenes de tráfico. Al estar en la calle, dicho dispositivo posee una determinada temperatura. Sin embargo, a pesar de que está sufriendo cambios de temperatura, no puede determinarla ya que no tiene la sensibilidad necesaria para medirla. Hace falta, claramente, un termómetro que sea sensible a esos cambios. Los sensores han de ser creados específicamente para un propósito.

Precisión

El **error de medida** de un determinado evento es también importante. Continuando con el ejemplo del termómetro, la precisión indicará cómo es el error de la medida proporcionada por el dispositivo. Si el termómetro solo puede indicar frío o calor, tendrá baja precisión. Si es capaz de marcar décimas de grado, será un termómetro muy preciso.

Sensibilidad

Esta propiedad refleja la capacidad que tiene un determinado sensor de recoger las fluctuaciones o cambios del evento que se mide. Si la luz en un determinado punto varía hasta treinta veces al día, pero la cámara no es capaz de detectar más que dos de esos cambios, es porque la cámara no tiene la suficiente sensibilidad.

Es importante indicar que **sensibilidad y precisión van de la mano**, pero la precisión está más orientada a ver cómo difiere la realidad de lo que se ha medido, y la sensibilidad está más orientada a evaluar cómo se adapta a los cambios de lo que se mide.

Consumo y tamaño

La mayoría de dispositivos de captura de información necesitan de una fuente de alimentación para funcionar. En función de la sensibilidad y la precisión (y otros factores como la cantidad de información capturada) requerirá un consumo de potencia eléctrica mayor o menor y poseerá un tamaño diferente.

Es necesario considerar el consumo y el tamaño en el diseño de los sistemas de percepción computacional pues suele ser cierto que consumo y tamaño son **inversamente proporcionales a sensibilidad y precisión**.

Existen algunos factores más como pueden ser:

- ▶ El precio del sensor.
- ▶ La usabilidad.
- ▶ La resistencia a agentes externos: si un sensor de temperatura es resistente a la lluvia o una cámara lo es a temperaturas extremas.
- ▶ Amplitud en el rango de la medida.
- ▶ La reparabilidad: si es fácilmente reparable...

Tipos de sensores para la captura de información

Existe una gran cantidad de tipología de sensores. Esto es fruto de que el mercado está dando respuesta a la gran demanda de **especificidad** que se ha visto anteriormente (es decir, cada sensor se está haciendo cada vez más específico) e intenta resolver una casuística concreta.

A continuación presentamos una clasificación genérica para que tengas una visión panorámica de los sensores que existen a día de hoy:

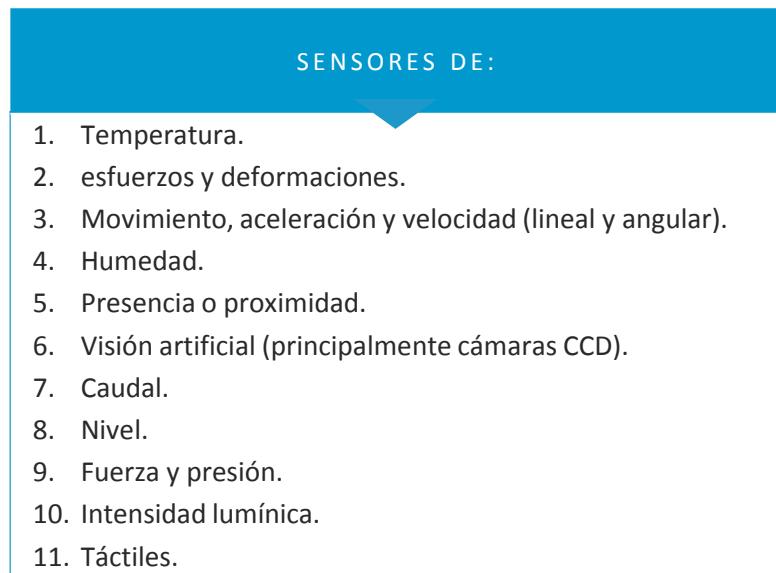


Figura 2. Tipos de sensores que puedes encontrar en el mercado.

Conversión analógico-digital (A/D)

La conversión analógico-digital es uno de los pasos esenciales para poder procesar cualquier información. De por sí, dicha conversión daría para una asignatura en sí misma (incluso para un curso académico), no obstante, en este apartado resumiremos dos puntos:

- ▶ En qué consiste la conversión analógico-digital y cómo funciona.
- ▶ Qué tipos de ventajas aporta dicha conversión.

Conversión

En primer lugar, lo que hace es **tomar fotografías de la realidad que se está midiendo**. Es decir, si estamos midiendo temperaturas, la conversión analógico-digital no estaría midiendo constantemente la temperatura en sí, sino que la mediría a intervalos de tiempos. El valor tomado en cada intervalo, además, se discretizaría, entendiendo por «discretizar» el reducirlo a un formato concreto de datos.

Pongamos un ejemplo más concreto y aplicable a un caso real de nuestro día a día. Estamos en un centro comercial y escuchamos la música que nos gusta por los

altavoces. Queremos guardar esa canción en nuestro móvil y utilizamos la aplicación micrófono para guardar el sonido del exterior.

- ▶ Nuestro móvil tiene un sensor de sonido: un micrófono.
- ▶ Posteriormente un conversor analógico-digital. Dicho conversor A/D muestrea la señal de audio: toma fotos de ella a periodos concretos.
- ▶ Convierte cada medida (que es una intensidad) en un dígito binario (por ejemplo, 2.1 voltios).

Esa conversión de intensidades físicas a dígitos binarios es lo que se conoce como conversor A/D.

Tras dicha conversión, la información almacenada (las fotografías de la señal de audio) ocupa una cantidad mucho menor que el sonido original y permite la realización de un **pre-procesado** (se verá más en detalle en qué consiste esta fase) que, entre otras tareas, filtra la señal del ruido existente y la protege frente a posibles errores; todo esto mediante software sin la necesidad de dispositivos hardware adicionales.

En este gráfico se aprecia cómo la intensidad de una determinada señal a lo largo del tiempo (valores enteros como 3, 5, 7, 7, 7, etc.) se convierte en formato digital/binario.

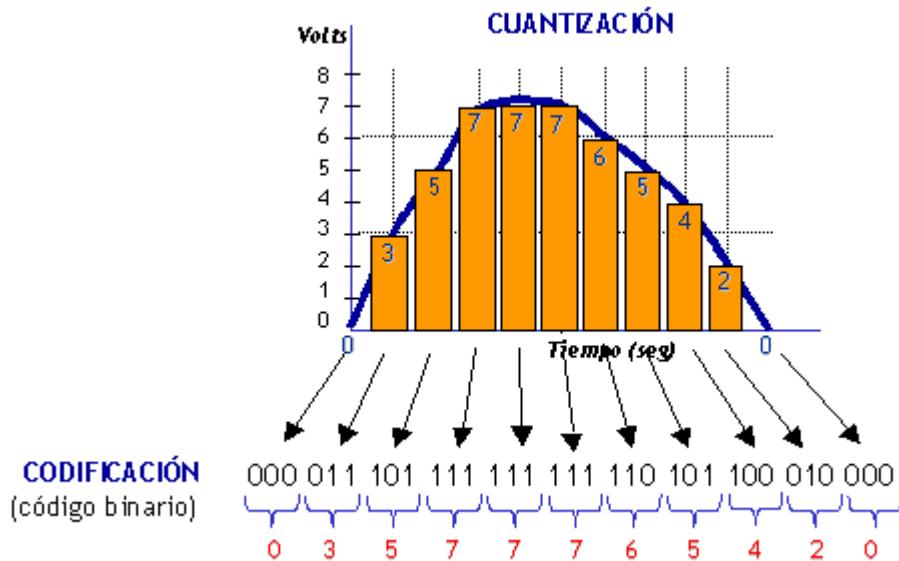


Figura 3. Ejemplo de conversión analógico-digital.

Fuente: <http://panamahitek.com/escalando-unidades-de-conversion-analogica-digitales/>.

Ventajas

Las ventajas de un conversor A/D son numerosas:

- ▶ Reduce enormemente la lógica de procesamiento lo que le permite ser configurable por medio de software.
- ▶ Permite la integración de multitud de sensores en dispositivos corrientes sin necesidad de una integración compleja.
- ▶ La información almacenada es infinitamente menor, lo que ahorra también espacio de almacenamiento.
- ▶ Se puede realizar una transmisión de una información determinada por radio o infrarrojo tras usar un conversor A/D sin necesitar gran ancho de banda ni perder información.
- ▶ Permite y facilita la compresión de información, puesto que la información ha sido convertida a formato digital.

¿Existen desventajas? Por supuesto. La principal es la **pérdida de información**, ya que la señal se está muestreando.

Para solventar esta desventaja, el **teorema de Nyquist-Shannon** propone la frecuencia mínima para muestrear una determinada señal. Este teorema queda fuera del alcance de este temario, sin embargo, hacemos referencia a la existencia de dicha frecuencia mínima dada su relevancia.

Preprocesamiento de la información

Antes de empezar con el apartado dedicado al procesamiento de la información capturada, es necesario añadir una sección dedicada al preprocesamiento. Esta fase podría tener una entidad propia *per se*. De hecho, algunos de los temas de esta asignatura están centrados en el preprocesamiento.

En muchos libros, además, preprocesamiento y procesamiento son sinónimos, siendo en algunos casos uno sustitutivo de otro.

En esta asignatura entenderemos el preprocesamiento como el tratamiento realizado inmediatamente después de la captura de la información y que será común para cualquier procesado posterior que se haga.

Por ejemplo, imaginemos un sistema de percepción computacional que se encarga de seguir un objeto en pantalla y detectar su forma. En este caso, son dos tareas las que se realizan para una misma captura de información: vídeo e imagen.

El preprocesamiento del vídeo y de la imagen consistirá en eliminar ruido en la imagen; asegurar que la iluminación es adecuada; eliminar componentes borrosos del vídeo o incluso en equilibrar las componentes de color.

Por otro lado, el **procesamiento** de dicha captura será doble:

- ▶ Un procesamiento encargado de facilitar la detección de objetos (eliminando detalles del fondo, seleccionando los objetos principales, etc.).
- ▶ Otro procesamiento encargado de analizar la forma y contorno de los objetos.

¿Podrían incluirse las fases de preprocesamiento dentro del procesamiento? Por supuesto, pero entonces habría que realizarla dos veces, una para cada tipo de procesamiento.

**Por tanto, el criterio de esta asignatura es definir el preprocesamiento como
el conjunto de tareas que acomodan la captura de información al
procesamiento de la misma.**

Adicionalmente, el preprocesamiento puede realizarse, bien sobre la información en analógico, o bien después de la conversión a digital.

- ▶ En el caso de realizarlo sobre la información en **analógico**, suele tratarse de un preprocesamiento que involucra hardware como filtros basados en circuitos, mejores sensores, etc.
- ▶ En el caso de realizar el preprocesamiento sobre la información ya **digitalizada** (discretizada), este será eminentemente realizado por software o código.

En general, el preprocesamiento realizado de forma digital suele ser más potente, pero trabaja sobre información ya deteriorada no solo por la captura de la señal, sino también por la propia conversión A/D que ha propagado el error de medida.

Por el contrario, el preprocesamiento realizado de forma analógica suele ser muy eficaz, pero menos flexible y poco configurable. Es decir, al ser muy específico, no puede adaptarse a variaciones en los errores observados.

En general, puede decirse que una combinación de ambos es lo más adecuado, aunque actualmente el avance de la tecnología en sensorización y en computación han desplazado la balanza hacia el preprocesamiento realizado de forma digital, ya que es el elegido en numerosos casos por coste, eficiencia y flexibilidad, dado que los sensores proporcionan una medida de la realidad muy precisa.

Las tareas más comunes de preprocesamiento suelen cubrir los siguientes aspectos:

Eliminación de ruido

El concepto de ruido es muy amplio y genérico, pero puede entenderse como **pérdidas** en la calidad de la información o bien **alteraciones** de la información. El ruido más común es el que se produce en una televisión, donde la imagen no es nítida y se aprecia como algunos píxeles de la pantalla no concuerdan con el contexto de la imagen.

Actualmente, la mayoría de los sensores ya poseen integrados mecanismos de reducción de ruido, pero en muchos casos, es necesario una eliminación de ruido adecuada para asegurar que la señal preprocesada cumple con los requisitos correctos para una posterior extracción de características y toma de decisiones.



Figura 4. Ejemplos de ruido en una determinada imagen.

Fuente: <http://www.clickblog.it/post/194357/cose-il-rumore-dellimmagine-e-come-avertirlo>.

Detección de anomalías

La detección de anomalías se produce cuando una de las medidas realizadas carece de sentido o está fuera de un rango determinado.

Aunque principalmente es algo a evitar, la detección de anomalías puede conducir al estudio de casos concretos que deriven en un conocimiento más amplio de la

problemática. Por ejemplo, en banca, la detección de anomalías suele estar asociada con la detección de comportamientos fuera de lo habitual como pueden ser el fraude o el blanqueo de capitales.

En la siguiente imagen se puede apreciar una anomalía en uno de los latidos de un determinado momento de la señal.

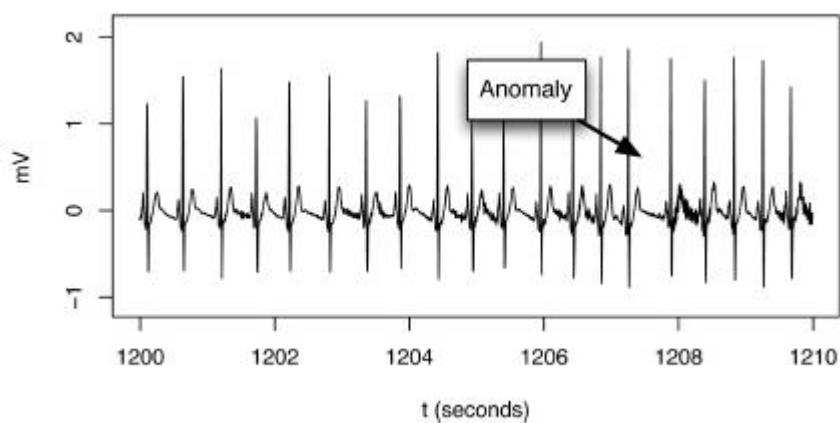


Figura 5. Detección de anomalías en una señal de Electrocardiograma (ECG).

Fuente: Dunning y Friedman (2014).

Corrección de errores

Aunque puede parecer propio de la codificación o transmisión de señales, la corrección de errores permite subsanar rápidamente pequeñas perturbaciones en la captura de información que son estadísticamente corregibles.

Por ejemplo, en una imagen: si en una región todos los píxeles son oscuros y un píxel es completamente blanco, se puede corregir mediante el análisis de los píxeles vecinos.

La corrección de errores no suele ser una de las tareas que se realice por defecto en un preprocesamiento. De hecho, suele venir de una etapa del procesado de información que puede automatizarse y es común al resto del procesamiento.

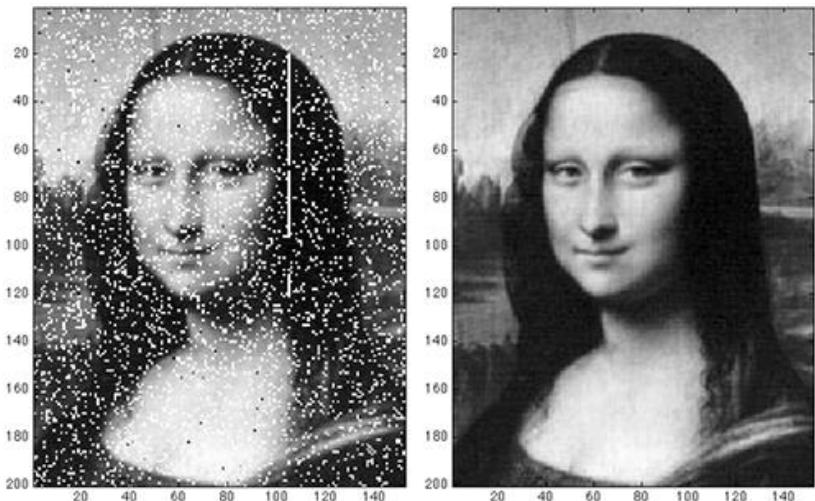


Figura 6. Ejemplo de corrección de errores: (Izq.) Imagen donde se han aplicado errores forzando que algunos píxeles sean blancos. (Dcha.) Imagen reconstruida.

Fuente: <http://cnnespanol.cnn.com/2013/01/18/la-mona-lisa-va-a-la-luna-en-la-primer-transmision-laser-interplanetaria/>.

En definitiva, la etapa de preprocessamiento tiene como fin dotar al sistema de percepción computacional de las herramientas necesarias para:

- ▶ Saber si un determinado sensor está funcionando correctamente: un número elevado de anomalías, errores o ruido puede conducir a la conclusión de que el sensor no está funcionando correctamente.
- ▶ Corregir las posibles desviaciones de los sensores a la hora de realizar una determinada medida.
- ▶ Evitar que los errores en la medida se propaguen hasta la toma de la decisión o, en otras palabras, asegurar en todo momento que la decisión se está tomando con la información más limpia posible.

2.5. Procesamiento de la información

Llegados a este punto, la información ya está depurada, limpia y disponible para realizar el procesamiento de la información propiamente dicho. Este depende completamente de la finalidad que se persiga (detectar caras en un

aeropuerto, leer matrículas de forma automática, detectar enfermedades en la voz, etc.) y es el núcleo de esta asignatura.

Por ello, en este apartado vamos a dar una visión global de los diferentes **métodos de procesamiento** de información existentes y las familias en las que se divide. Todos estos pueden aplicarse de forma secuencial y no suelen ser excluyentes. De hecho, lo habitual en problemas reales es aplicar un conjunto de ellos para encontrar la solución adecuada.

Filtros y suavizado de la información

En general, cuando se habla de filtros se hace referencia a la operación matemática, bien sobre la señal unidimensional, bien sobre la imagen, que permite eliminar o potenciar los detalles de la información a procesar. Un filtro puede eliminar los detalles de una imagen (contrastos, contornos, bordes, etc.) para permitir, por ejemplo, detectar qué objetos existen en dicha imagen.

El **filtrado** suele ser una operación computacionalmente costosa pues requiere recorrer toda la señal unidimensional o toda la imagen, que es una señal bidimensional.

Las operaciones de filtrado normalmente poseen una **operación inversa teórica**, pero puesto que el filtrado y el suavizado de la información conllevan de por sí pérdida de información, volver a la información original es, aparte de costoso, no muy satisfactorio. Además, y no menos importante, se puede decir que la mayoría de filtros son no lineales.

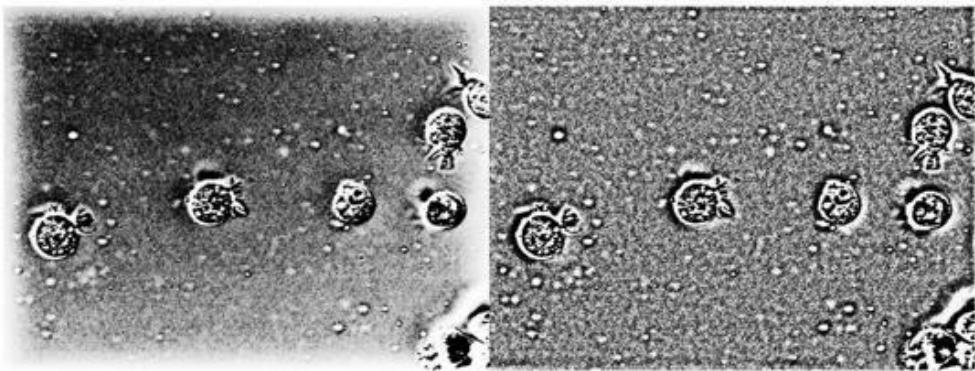


Figura 7. Ejemplo de filtrado de imagen para bajas frecuencias (respetando las altas frecuencias que se corresponden con los detalles de la imagen).

Fuente: <https://blogs.mathworks.com/steve/2013/07/10/homomorphic-filtering-part-2/>.

Segmentación y detección de regiones

Por segmentación se entiende la división de la información en conjuntos de datos con propiedades similares entre ellos.

Un ejemplo claro se aprecia en la segmentación de imágenes. Si consideramos la foto de un elefante en la sabana, la segmentación deberá distinguir las zonas de la imagen más generales como son el suelo, el elefante, el cielo y la vegetación. Se podría haber resumido también en una segmentación de dos elementos: elefante y resto del paisaje. Esta elección es dependiente de la aplicación que se persiga.

La segmentación y detección de regiones es también una de las **operaciones computacionalmente más costosas** que existe en el procesamiento de información.

La dificultad de la segmentación (y que aún se encuentra bajo investigación) es la **detección automática del número de segmentos** dentro de una imagen. Es decir, ¿cuántas subimágenes o grupo de píxeles existen dentro de una imagen?

Para realizar una buena segmentación es necesario que la comparación se realice considerando **características extraídas** de la imagen como pueda ser el color, la textura o incluso la coherencia de las regiones.

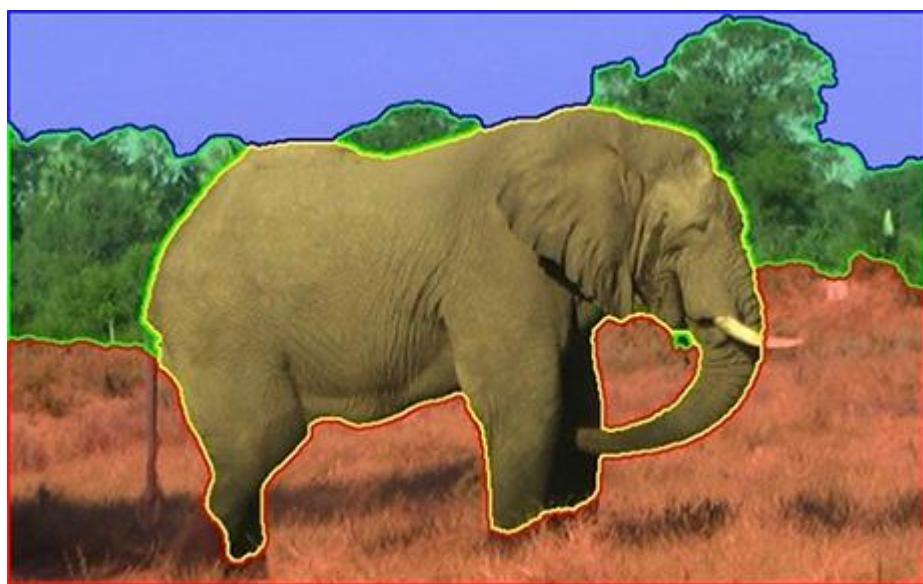


Figura 8. Ejemplo de filtrado de imagen para bajas frecuencias (respetando las altas frecuencias que se corresponden con los detalles de la imagen).

Fuente: <http://ieeexplore.ieee.org/document/7789625/>.

Extracción de características

El siguiente paso lógico a la segmentación es la extracción de características, entendida como un resumen desde el punto de vista analítico de cada región obtenida.

En otras palabras, la extracción de características hace referencia a la capacidad para **distinguir regiones** donde hay diferentes texturas, intensidades o, en el caso de señales unidimensionales, diferentes **patrones de comportamiento**.

La extracción de características puede aplicarse inicialmente sobre la señal unidimensional o sobre la imagen directamente, pero se obtendrá ruido debido a que los extractores de características considerarán el **ruido como parte de la información**. Por eso es recomendable, aunque dependerá del problema en concreto a resolver, aplicar la extracción de características cuando la información esté completamente limpia y preprocesada lo máximo posible.

Ejemplos de características que se pueden extraer:

- ▶ Componentes en frecuencia de diferentes secciones de una imagen.
- ▶ Coeficientes de transformaciones no lineales como puede ser la transformada de Fourier o de Laplace.
- ▶ Suavidad del contorno de un objeto detectado.
- ▶ Área del segmento encontrado después de la segmentación.
- ▶ Descriptores de texturas de una determinada región.

De todas las fases del procesado de la información, es la que más se está desarrollando últimamente y donde hay más foco de investigación. Muchos de los resultados han ayudado a avanzar en la encriptación y la compresión de datos, entre otros.

En definitiva, la extracción de características permite **facilitar la comparación** entre dos regiones dentro de una imagen, entre dos objetos previamente segmentados o incluso entre dos señales unidimensionales.

En la siguiente imagen vemos como la extracción de características acaba con un vector de características que facilita su posterior comparación con otro vector.

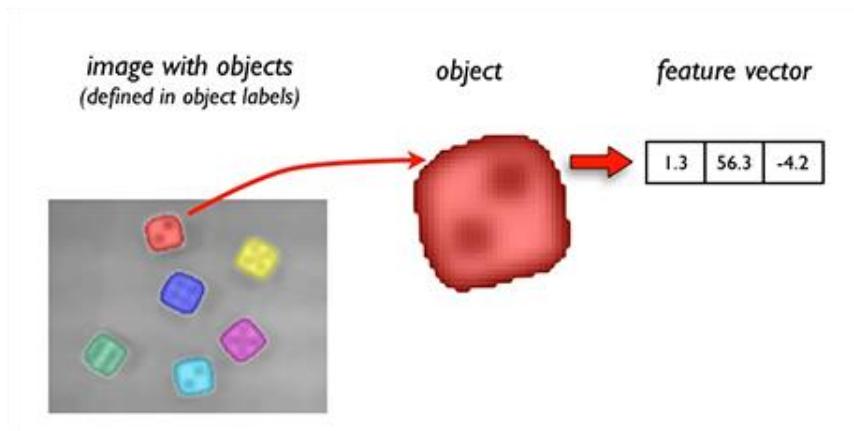


Figura 9. Ejemplo de filtrado de imagen para bajas frecuencias (respetando las altas frecuencias que se corresponden con los detalles de la imagen).

Fuente: http://perclass.com/doc/guide/feature_extraction/intro.html

2.6. Toma de decisión

Para recapitular, cuando se llega a la toma de decisiones se han realizado los siguientes pasos que vemos con un ejemplo sobre la lectura de matrículas de forma automática:

Se ha definido un conjunto de fuentes de información que pueden resultar de interés para resolver un determinado problema.

Ejemplo: leer de forma automática las matrículas de un coche.

Una vez fijadas las fuentes de información, se han decidido qué sensores emplear para capturar dicha información.

Ejemplo: cámaras estándar que capturan fotografías cada segundo.

El preprocesamiento consiste en corregir bien si existe desenfoque en la imagen capturada.

Ejemplo: algo normal, debido a la velocidad a la que se captura la imagen.

El procesamiento consiste en un filtrado para eliminar los detalles de menor importancia, una segmentación en color y una detección de bordes.

Posteriormente se ha hecho una extracción de características.

Ejemplo: detectar dónde está ubicado cada número y letra de la matrícula.

Figura 10. Resumen y ejemplo de las fases anteriores de la de toma de decisiones.

Llegados a este punto tenemos, en el caso de España, cuatro números y tres letras. De todos los métodos presentados en este apartado, ninguno de ellos nos ayuda en:

- ▶ Asegurar que en efecto hay cuatro números y tres letras.
- ▶ Asegurar que estoy leyendo realmente una matrícula y no la marca del coche.
- ▶ Asociar la región correspondiente a un número con un número concreto. Es decir, ¿cómo asociar un determinado conjunto de píxeles a un número?
- ▶ Detectar que no es una matrícula española.

Todas estas tareas deben realizarse en el módulo de toma de decisiones. Dicho módulo es el encargado de aplicar la lógica final para o bien tomar una decisión: ¿Es una matrícula registrada?; o bien dar soporte en la toma de decisiones: marcar en una determinada imagen qué zonas corresponden con paisajes, por ejemplo.

Esta fase debe realizarse principalmente tras la extracción de características. Considerarla basándonos únicamente en filtrados, segmentaciones o incluso en el preprocesado no suele ser común y, en muchos casos complica, enormemente la toma de decisiones. Sin embargo, existen casos donde los módulos de toma de decisiones pudieran no existir.

Por ejemplo, el coche que es capaz de conducir de forma automática (sin supervisión); estos coches poseen un módulo capaz de detectar las líneas de las carreteras. Dicho módulo tiene todos los ingredientes de un sistema de percepción computacional: captura de la información, preprocesamiento y extracción de características.

No obstante, dicho módulo no toma ninguna decisión, sino que proporciona la propia extracción de características a otro módulo que aglutina la información de otras fuentes y toma la decisión. Con lo cual, puede decirse que dicho módulo de percepción computacional no toma una decisión, sino que ayuda a otros módulos tanto a aprender como a dar soporte en la toma de decisiones.

2.7. Referencias bibliográficas

Dunning, T. y Friedman, E. (2014). *Time Series Databases: New Ways to Store and Access Data*. Sebastopol: O'Reilly Media.

González, R. C. y Woods R. E. (2007). *Digital image processing*. Upper Saddle River: Prentice-Hall, Inc.

Percepción Computacional

Captura y digitalización de señales

3.1. ¿Cómo estudiar este tema?

Para estudiar este tema deberás leer con atención las ideas clave que se desarrollan a continuación.

3.2. Objetivos

El muestreo de señales es una parte esencial en la captura de información. En él se basa el resto de procesamiento y la información que se pierda en esta fase no podrá estar disponible.

Es por eso que este tema pretende proporcionar las bases para entender correctamente el funcionamiento del muestreo y la posterior cuantificación de señales.

3.3. Introducción

La teoría de muestreo y cuantificación es uno de los temas más áridos en toda asignatura de procesamiento de señales. Reflejar sobre el papel qué significa muestrear una señal requiere de una complejidad matemática muy alta.

Sin embargo, la mejor manera de entenderlo es verlo con un ejemplo, y para ello recurrimos a la cuantificación más antigua que se conoce: los mosaicos de la antigua Roma.



Figura 1. Ejemplo de discretización o cuantificación: detalle del mosaico de Medusa .y las estaciones.

Fuente: <http://www.man.es/man/dms/man/estudio/publicaciones/boletin-man/MAN-Bol-2015/2015-33-Jabaloyas.pdf>.

El ejercicio de crear un mosaico es claro: con un conjunto «limitado» de colores (como mucho entre cien o doscientos colores) que se corresponden con diferentes tipos de piedra, elaborar un dibujo que refleje la complejidad de la realidad como las sombras, claroscuros, tonalidad, movimiento e incluso la expresión.

Intuitivamente podemos llegar a comprender que cuanto mayor sea el número de **colores** disponibles, mayor la gama cromática que podemos alcanzar y, por lo tanto, mayor el número de matices y expresiones o detalles que se pueden mostrar. Lo mismo sucede con el **tamaño** de la tesela. Parece intuitivo pensar que cuanto más pequeña sea la tesela, mejor nos aproximaremos a la realidad.

Pero si tuviésemos que hacerlo ahora, ¿qué cantidad de colores elegirías?, ¿qué tamaño de tesela?, ¿serían todas del mismo tamaño?, ¿cuál sería el mínimo número de colores y cuál el tamaño de la tesela para que el mosaico fuera lo más barato y bello a la vez? Estas preguntas no son tan obvias de responder. Y los romanos seguro que tampoco tenían las respuestas.

La teoría de muestreo y cuantificación nos ayuda a decidir esos mínimos valores.

Nos da los umbrales a partir de los cuales podemos tener una calidad suficiente de señal o de imagen, sin necesidad de emplear mucho almacenamiento (teselas en el ejemplo del mosaico romano).

Dicha teoría de muestreo funciona de la siguiente manera: imaginemos que queremos entender el funcionamiento de la noche y el día. Vamos a asumir que tenemos siempre los ojos cerrados y que los abrimos en un momento concreto. En este ejercicio, abrir los ojos es una metáfora para **muestrear una realidad**; en definitiva, no es más que observar.

Pongamos que la **primera muestra**, la primera vez que abrimos los ojos, es de día. Si abrimos los ojos una vez al mes, siempre a la misma hora, solo veremos que es de día pudiendo llegar a la conclusión de que en nuestro planeta solo está el día y no hay noche.

Como sabemos teoría de muestreo, decidimos acelerar la **frecuencia de muestreo** a una vez al día, cada 24h. Lamentablemente observamos el mismo efecto: es de día. Decidimos hacer el experimento cada hora y vemos cada día doce muestras de luz y doce muestras de oscuridad (asumiendo que el día y la noche tienen igual número de horas). Evidentemente, con este muestreo se ha observado el fenómeno de la noche. Lo mismo hubiera pasado si miro cada segundo, evidentemente.

Luego, hasta este punto tenemos que si se observa:

- ▶ Cada 24h, siempre hay luz.
- ▶ Cada hora, hay doce observaciones de luz y doce de oscuridad.

Si ahora realizamos el experimento midiendo cada doce horas, realizaremos dos medidas al día: una de noche y una de día. Hemos dado con la **frecuencia mínima**,

también llamada frecuencia de muestreo, para observar el fenómeno de la noche y el día. Parece casualidad que:

La **frecuencia del fenómeno** sea una vez cada 24 horas, es decir, la secuencia luz-noche se repite a las 24h, mientras que la **frecuencia mínima** para observar el fenómeno es justo el doble, una vez cada 12h.

Sería aventurarse mucho decir que:

$$f_{muestreo} \geq 2 f_{observación}$$

Esta igualdad se conoce como **teorema de Nyquist**, a dicha frecuencia de muestreo se la denomina también **Frecuencia de Nyquist** y es la idea principal sobre la que trata este tema. En otras palabras, para procesar información digital hay que perder calidad de la señal e información. La cuestión es, ¿cuál es la mínima información que puedo perder?

3.4. Sistemas de conversión analógico-discreto

En esta sección contaremos en qué consiste un conversor analógico-discreto más en detalle, así como su utilidad y funcionamiento. Antes de nada, es importante entender dos conceptos: analógico y discreto.

Analógico

Por analógico se entiende toda señal o imagen que es continua y que procede de una fuente o fenómeno físicos como puede ser un sonido, un objeto en movimiento, etc. Si continuamos con el ejemplo anterior del mosaico, lo analógico sería el dibujo original del que partió el artista que diseño el mosaico.

- ▶ Este tipo de señales se caracterizan por tener todo el detalle de la realidad, **no pierden nada de información** cuando se transmiten de un punto a otro.
- ▶ El almacenamiento de una señal analógica es complicado y muy poco eficiente (sirva como ejemplo los discos de vinilo o los *cassettes* antiguos).

Discreto

Por discreto (también digital, son sinónimos en este contexto) se entiende una señal o imagen que **solo puede alcanzar unos determinados valores**, por lo que está discretizada. Normalmente, una señal discreta/digital viene de muestrear y discretizar la señal analógica.

- ▶ La mayor ventaja es la reducción de información que supone.
- ▶ La desventaja, si el muestreo no se realiza correctamente, puede ser que la señal pierda la información más importante.

Esquema de un conversor A/D

A continuación se presenta un esquema general de un conversor A/D. Este conversor consta de las siguientes partes:

- ▶ En primer lugar, el **muestreador**, que es el encargado de convertir la señal analógica a discreto tomando fotos de la señal a una frecuencia determinada.
- ▶ En segundo lugar, un **cuantificador**, encargado de transformar la foto tomada en el muestreo y convertirla en un conjunto limitado de valores.
- ▶ Finalmente, un **codificador** que convierte el valor del cuantificador en binario o en un sistema numérico que pueda transmitirse.

En algunos ejemplos de la literatura, cuantificador y codificador van de la mano y suelen ser una misma pieza. No obstante, pueden separarse con vistas a que se apliquen diferentes tipos de codificación (junto con su redundancia para evitar errores).

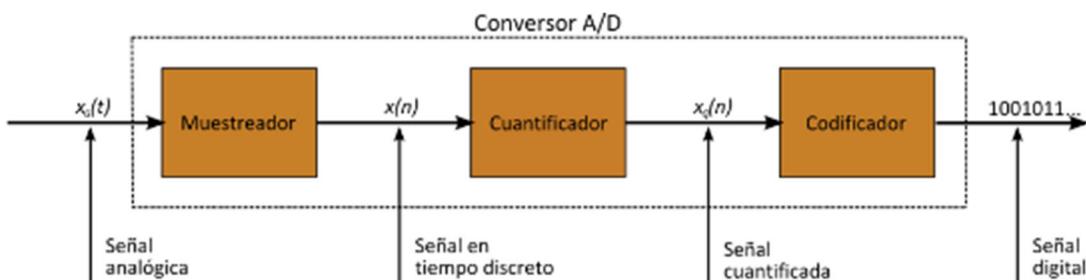


Figura 2. Esquema general de un conversor analógico-discreto.

Fuente: <http://elastixtech.com/fundamentos-de-telefonia/transmision-de-la-voz/>.

Las ventajas principales de hacer este proceso son las siguientes:

- ▶ Permite, en primer lugar, reducir la complejidad de la señal de entrada y reducirla únicamente a un determinado número de posibilidades (**cuantificador**).
- ▶ El filtrado de ruidos, anomalías y otros efectos producidos por una baja calidad son mucho más fácilmente procesables con un filtrado digital que con un filtrado analógico.
- ▶ Los filtros digitales pueden modificarse mediante software, a diferencia de los analógicos, que deben modificarse por hardware, lo que los encarece y hace difícil su mantenimiento.
- ▶ La señal analógica, por el mero hecho de ser convertida a señal digital, ya reduce la cantidad de información a procesar posteriormente, pero otra características es que dicha señal digital puede comprimirse y encriptarse.
- ▶ El almacenamiento de una señal digital es más sencillo y barato que el almacenamiento analógico, así como más perdurable en el tiempo.

Sin embargo, las señales digitales (un archivo MP3 por ejemplo) no pueden oírse directamente, lo único que pueden oírse son las señales analógicas. El proceso de conversión de analógico a digital o discreto debe permitir el proceso inverso,

asumiendo que por la propia definición de este proceso se perderá información, pero la mínima como para que la señal analógica reconstruida suene muy similar a la original.

Para ello, deben seguirse los siguientes pasos que se presentan a continuación:

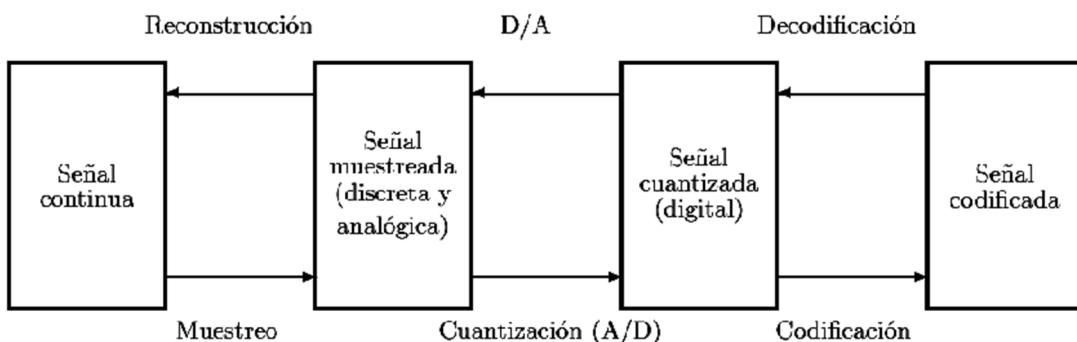


Figura 3. Reconstrucción de una señal analógica.

Fuente: <http://www.lip.uns.edu.ar/pdi/index31.htm>.

En este tema solo veremos el proceso de conversión A/D. El proceso inverso, (conversión D/A), aunque similar en estructura y pasos a dar, contiene otras peculiaridades que se desvían del objetivo de esta asignatura.

Para entender el diagrama anterior, vamos a explicarlo con un ejemplo: vamos a un concierto de nuestro cantante favorito y nos llevamos nuestro móvil. En el momento en que suena nuestra canción favorita, sacamos el móvil y grabamos, ¿qué sucede en ese instante?

- ▶ Primero, el micrófono del móvil captura toda la información analógica proveniente del exterior (no solo la voz del cantante, sino también el ruido de ambiente).
- ▶ Dicho micrófono, aunque puede capturar toda la información, lleva asociado un sistema de muestreo que «observa» la señal de forma periódica, es decir, mide la intensidad de voz cada cierto periodo de tiempo.
- ▶ La señal, una vez muestreada y cuantificada, ya puede almacenarse en el dispositivo móvil. Es la única manera de almacenar información en un dispositivo de este tipo.

- ▶ Para que la información ocupe lo menos posible, la información se comprime y se filtra por defecto para que quede lo más limpia posible.

A continuación puede verse un esquema de los pasos seguidos para guardar un sonido de voz en un dispositivo móvil:

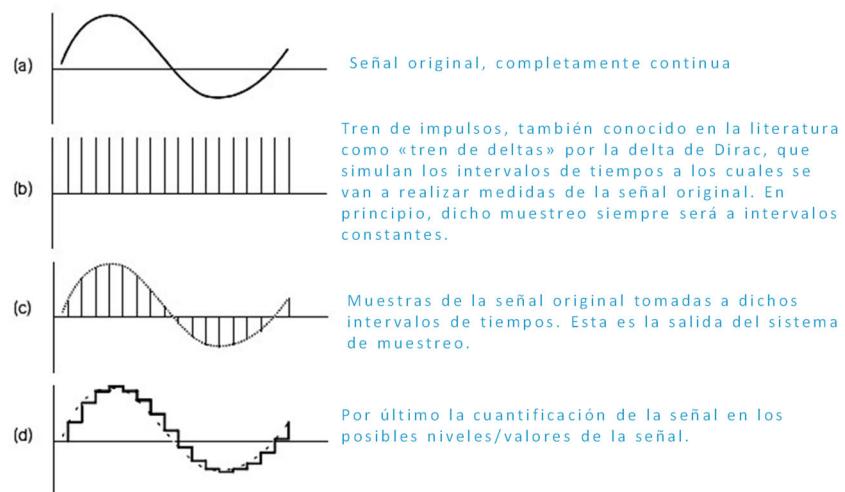


Figura 4. Secuencia de pasos para la conversión analógico-discreto.
Fuente: Adaptado de <http://www.sapiensman.com/tecnoficio/docs/doc39.php>.

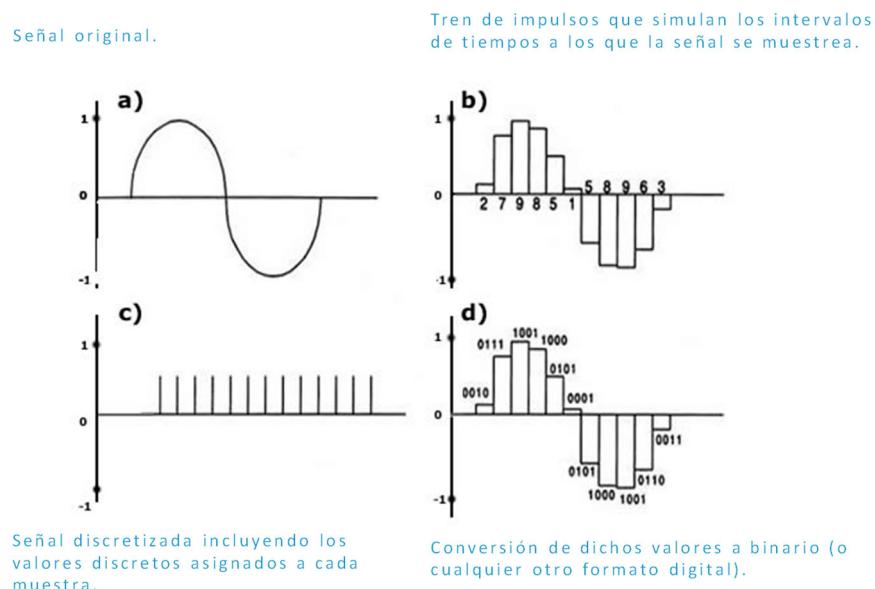


Figura 5. Secuencia para la conversión analógico-discreto incluyendo también la digitalización de la señal.
Fuente: Adaptado de <http://acusticaysonido.com/?p=315>.

3.4. Muestreo y cuantificación

En el apartado anterior, se ha presentado una visión general de los sistemas de conversión analógico-discreto. Ahora, en este apartado, se describen los componentes individuales y la manera en la que se implementan.

Muestreo

Como hemos dicho antes, el muestreo consiste en tomar fotos de la realidad (ya sea una señal o una imagen) a una determinada frecuencia (posteriormente veremos qué criterio) de tal manera que la información manejada se reduce, pero no la información necesaria para comprender el mensaje.

A continuación se presenta un esquema general de muestreo. Dicho sistema consta de:

- ▶ Una señal de muestreo, conocida como **tren de pulsos**, que define la frecuencia y el tiempo de muestreo, es decir, durante cuánto tiempo se va a observar la señal.
- ▶ Un **interruptor** que, en función de la señal de muestreo, se abrirá o cerrará permitiendo tomar las fotos de la señal original.
- ▶ Por último, la **señal muestrada**.

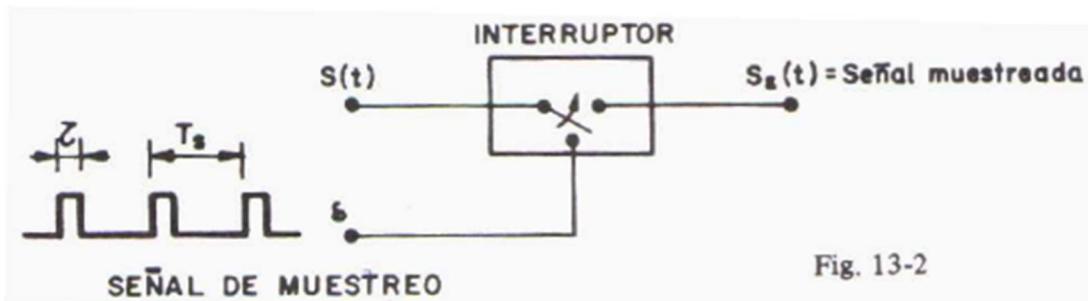


Fig. 13-2

Figura 6. Sistema de muestreo basado en tren de pulsos.

Fuente: http://www.ifent.org/lecciones/digitales/secuenciales/teorema_muestreo.htm.

El mecanismo es sencillo en su visión reducida, pero presenta ciertas **limitaciones**:

- ▶ El muestreo dependerá de la señal de muestreo, tanto de la anchura de los pulsos como de la separación entre ellos.
- ▶ El interruptor en sí mismo presenta discontinuidades muy pronunciadas cuando conmuta tan rápidamente. Esto puede ocasionar (y de hecho los produce) picos de señal en la señal muestreada.
- ▶ El conmutador puede estropearse debido al propio mecanismo y su uso continuado, cualquier solución que no involucre un sistema mecánico ofrecerá ventajas considerables. Es por ello que muchos muestreadores multiplican la señal por dicha señal de muestreo.

Ahora bien, de todos estos parámetros anteriores, solo hay uno que sí que puede modificarse: la frecuencia de muestreo.

Dicha frecuencia viene definida por el **teorema de Nyquist-Shannon** (profundizaremos más sobre esto en la lección magistral de este tema) que indica que si una señal tiene un ancho de banda f , la frecuencia de muestreo deberá de ser:

$$f_{muestreo} \geq 2f$$

Sin embargo, existen dos aspectos importantes a tener en cuenta si no se cumple dicho teorema.

¿Qué sucede si muestreo por debajo de la frecuencia de Nyquist?

Lo que sucederá es lo que se ve en la siguiente imagen. Partiendo de la señal sinusoidal original, si se muestrea a menor frecuencia se verá por tanto una variación de señal que no corresponde con la verdadera. Es decir, veremos que la señal va más despacio.

Este efecto sucede cuando vemos un coche en marcha y sus ruedas parecen girar en el sentido contrario de la marcha.

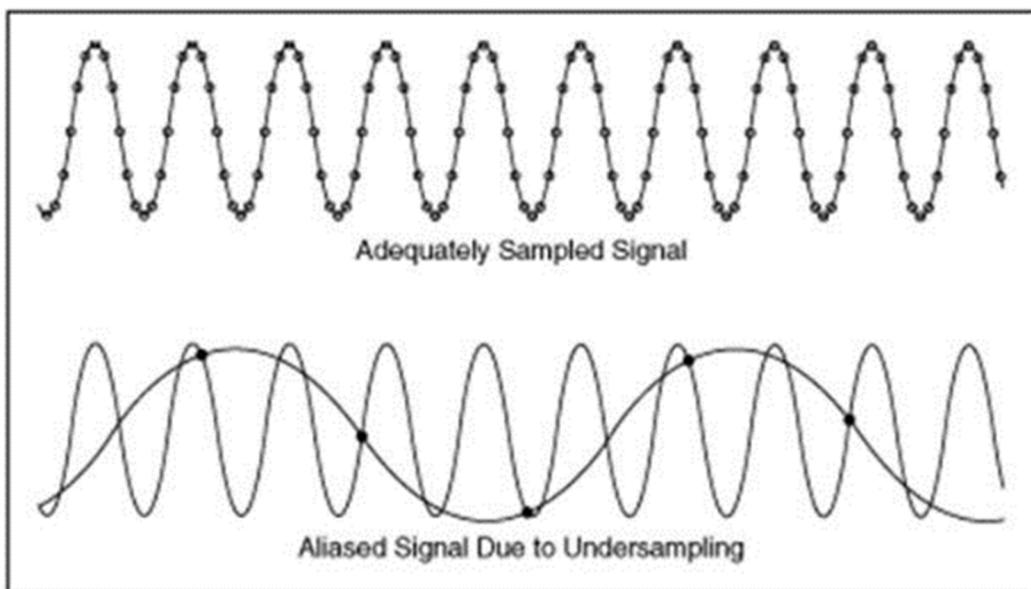


Figura 7. Ejemplo visual de submuestreo o *undersampling*.

Fuente: <https://www.masterhacks.net/blog-ingenero/que-es-el-aliasing/>

¿Qué sucede si muestreamos a mucha más frecuencia?

Entonces, estaremos capturando información de más, lo que produce una reconstrucción errónea de la señal, tal y como se ve en la imagen siguiente. Es decir, consideramos puntos consecutivos, entre los cuales no hay apenas variación, como puntos que poseen una gran variabilidad.

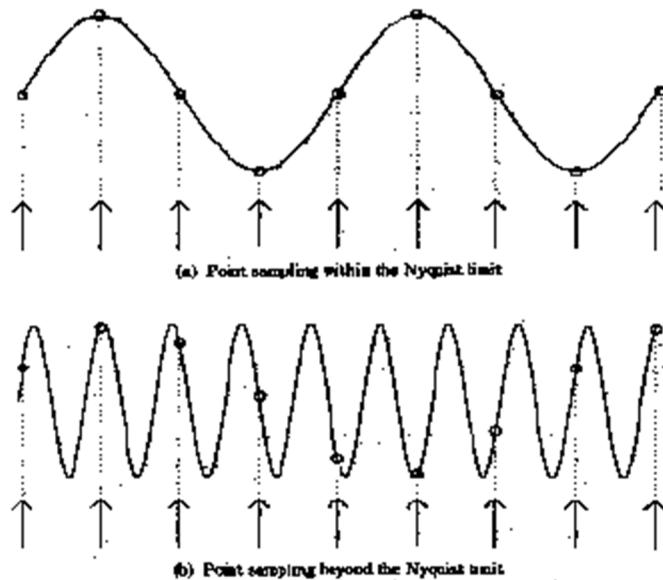


Figura 8. Ejemplo de sobremuestreo de señales u *oversampling*.

Fuente: <https://www.siggraph.org/education/materials/HyperGraph/aliasing/alias1.htm>.

Cuantificación

La cuantificación es el proceso de convertir las muestras tomadas de la señal original en **valores numéricos**.

Dichos valores numéricos poseen una resolución que dependerá del número de bits con los que cuente el procesador. Como se aprecia en la imagen siguiente, a mayor número de bits, mayor parecido con la señal original.

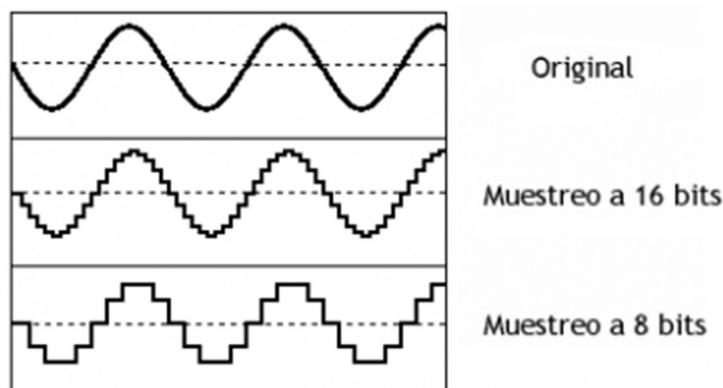


Figura 9. Ejemplo de cuantificación con diferentes niveles de calidad (bits).

Fuente: <http://www.fotoseimagenes.net/frecuencia-de-muestreo>.

En general, los sistemas de cuantificación proporcionan una **cuantificación uniforme**, pero en determinados casos, donde la riqueza de valores es más amplia, suele aumentarse el número de valores en una determinada región de la señal para que así se puede alcanzar mayor resolución donde más se necesita.

La elección del número de bits, así como de la uniformidad del rango es algo que depende de la aplicación y, al contrario del teorema de Nyquist, no existe un teorema concreto que definan ambos parámetros simultáneamente. Es cierto que el teorema de Shannon sí que da una idea de cuántos números de bits harían falta para codificar una determinada señal. Esto lo veremos en detalle en la sesión magistral.

Esta cuantificación realmente está implícita en la propia captura de imágenes. Por ejemplo, las cámaras CCD (ver imagen a continuación) poseen un cuantificador por fotodiodo, con lo cual una imagen nunca es continua, ya es discreta de por sí, y lo mismo sucede con las señales digitales.

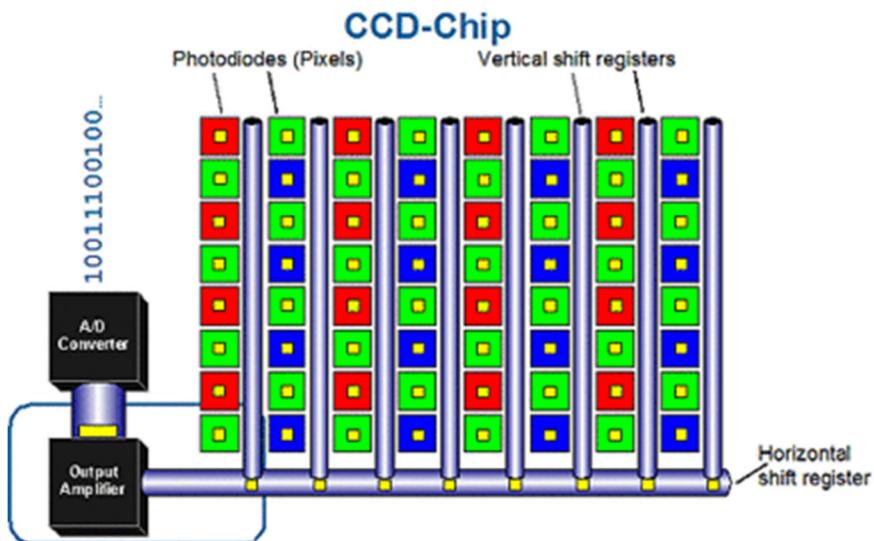


Figura 10. Esquema de funcionamiento de captura de imágenes basada en CCD.

Fuente: <http://distanciafocal.com.ar/%C2%BFcomo-elegir-el-tipo-de-sensor/>.

La cuantificación puede verse como un tema de **resolución o calidad de la señal**. De hecho, en función del número de bits empleados para cuantificar una señal, así será su propia calidad. Normalmente, en imagen, se trabaja con 256 niveles de color, siendo 0 el color negro y 255 el color blanco en escala de grises.

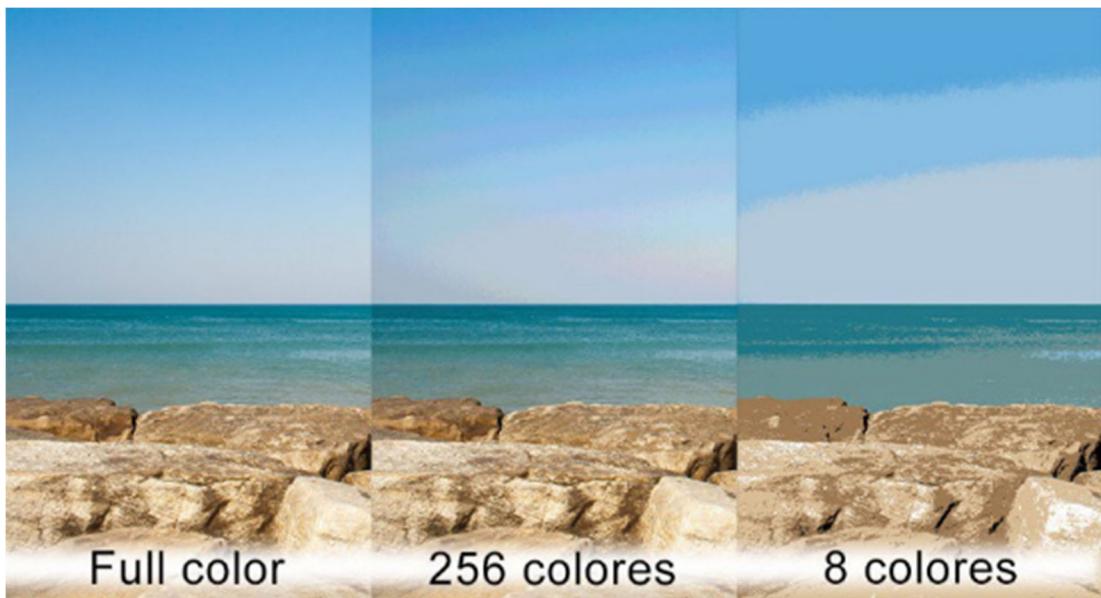


Figura 11. Cuantificación con diferentes niveles de resolución: Full color, 256 colores, 8 colores.

Fuente: <http://www.comercialgumara.es/blog/pixel-vs-imagenes-vectoriales/>.

Así como el muestreo puede volver a recomponerse, siempre y cuando se cumpla el teorema de Nyquist (como veremos en la clase magistral), la información difícilmente puede recuperarse en el caso de la resolución baja. Existen algunos algoritmos, basados en la transformada del Coseno, similar a la transformada de Fourier, pero que su funcionalidad se basa en inferir qué rango de valores posibles podrían haber tomado antes de la cuantificación.

Por último, podemos indicar que hoy día muestreo y cuantificación están tan relacionados que se vinculan incluso a nivel hardware y de dispositivo.

Percepción Computacional

Fuentes y tipos de ruido

4.1. ¿Cómo estudiar este tema?

Para estudiar este tema deberás leer con atención las ideas clave que se desarrollan a continuación.

En el siguiente tema se analiza el **componente aleatorio de las señales**. Como veremos, toda señal estará afectada por algún tipo de perturbación o ruido en un escenario real. El conocimiento de los diferentes tipos de ruido nos permitirá poder diseñar técnicas adecuadas para su eliminación completa o parcial, reduciendo su efecto en la fuente de información.

Con este objetivo, en el tema se presenta, en primer lugar, el concepto de **entropía**. Este hace referencia al grado de incertidumbre e irregularidad en la señal, de forma que nos permite cuantificar el impacto del ruido. Posteriormente, se describen los diferentes **tipos** de ruido. Y finalmente, se presentan las herramientas matemáticas que permiten su modelado: los **procesos estocásticos**.

4.2. Entropía: concepto y estimación

El **ruido** es toda señal no deseada y de naturaleza aleatoria que modifica la intensidad de la señal original a percibir.

En los sistemas de captación de estímulos, como luz o sonido, en primer lugar se lleva a cabo la transformación de estos a señales eléctricas mediante sensores. Por ejemplo, considérese el micrófono o la cámara fotográfica, que captan el sonido o la

luz y los convierten en una señal eléctrica de naturaleza analógica o digital. Posteriormente, estas señales pueden almacenarse en memorias o ser transmitidas por medio de un sistema de comunicación.

Caracterización del ruido

En el mundo real, cualquier señal estará afectada por elementos incontrolables que generen ruido. Generalmente, este se superpone a la señal de interés en forma de **ruido aditivo**. Es decir, nuestra imagen o señal acústica recibida tendrá la siguiente forma:

$$S(t) = f(t) + r(t)$$

Donde:

- ▶ S denota la señal recibida.
- ▶ f denota la señal original.
- ▶ r denota el componente de ruido.

Por tanto, la primera etapa en el tratamiento de una señal debe centrarse en la **identificación y eliminación** del mayor número de artefactos ruidosos posible; en la mayoría de los casos no será viable eliminarlos por completo de la señal capturada.

La **caracterización matemática** del ruido permite abordar el problema de forma analítica. Como se verá más adelante en este tema, las señales que manejamos en la vida real incluyen cierto nivel de ruido, por lo que pueden verse como el resultado de un proceso aleatorio, tal y como sucede cuando lanzamos un dado. La naturaleza aleatoria del ruido provoca que la medición de una misma señal no produzca siempre el mismo resultado.

No estamos ante señales deterministas sino aleatorias: procesos estocásticos.

Como ejemplo, puede considerarse el escenario en el que se capta una señal sonora. En este caso, el resultado de la medición variará dependiendo del nivel de intensidad de otras fuentes sonoras sobre las que no tengamos control (ruido de ambiente).

Definición de entropía

Como se ha comentado anteriormente, las señales contienen información y, además, están afectadas por diferentes fuentes de ruido. En este contexto, surge el concepto de entropía. Al igual que en física, el término hace referencia a la **complejidad de la señal**. La adición de ruido contribuye a incrementar el grado de complejidad de una señal, resultando en una mayor entropía.

En teoría de la información, la entropía se define como la cantidad de información de una fuente aleatoria (en promedio). Por tanto, el concepto de entropía sirve para caracterizar una **variable aleatoria**. En el caso de las señales, estas pueden modelarse como una sucesión de realizaciones de una variable aleatoria en el tiempo (proceso estocástico), por lo que veremos cómo extender la definición de entropía a elementos aleatorios de esta naturaleza.

Dada una variable aleatoria discreta X , que toma valores del conjunto $\{X_1, X_2 \dots, X_M\}$ y de acuerdo a una probabilidad $P(X_i)$, **Shannon** definió la entropía de la siguiente forma:

$$H(X) = E\{-\log_2[P(X)]\} = \sum_{i=1}^M -\log_2[P(X_i)]P(X_i)$$

Donde $-\log_2[P(X_i)]$ se interpreta como la cantidad de información.

Así, valores poco probables de la variable X están asociados con una mayor cantidad de información (efecto sorpresa) frente a valores de mayor probabilidad.

En el caso de que la variable X posea únicamente dos posibles resultados, $\{X_1, X_2\}$, es decir, cuando se trata de una variable caracterizada por una distribución de Bernoulli, puede observarse como la entropía de X , en función de la probabilidad $P(X = X_1)$, alcanza su máximo cuando esta es 0.5. En este caso, ambos sucesos poseen la misma probabilidad y en promedio obtendremos siempre la misma cantidad de información de X .

El siguiente gráfico muestra la variación de la entropía de X como función de la probabilidad asociada al suceso X_1 .

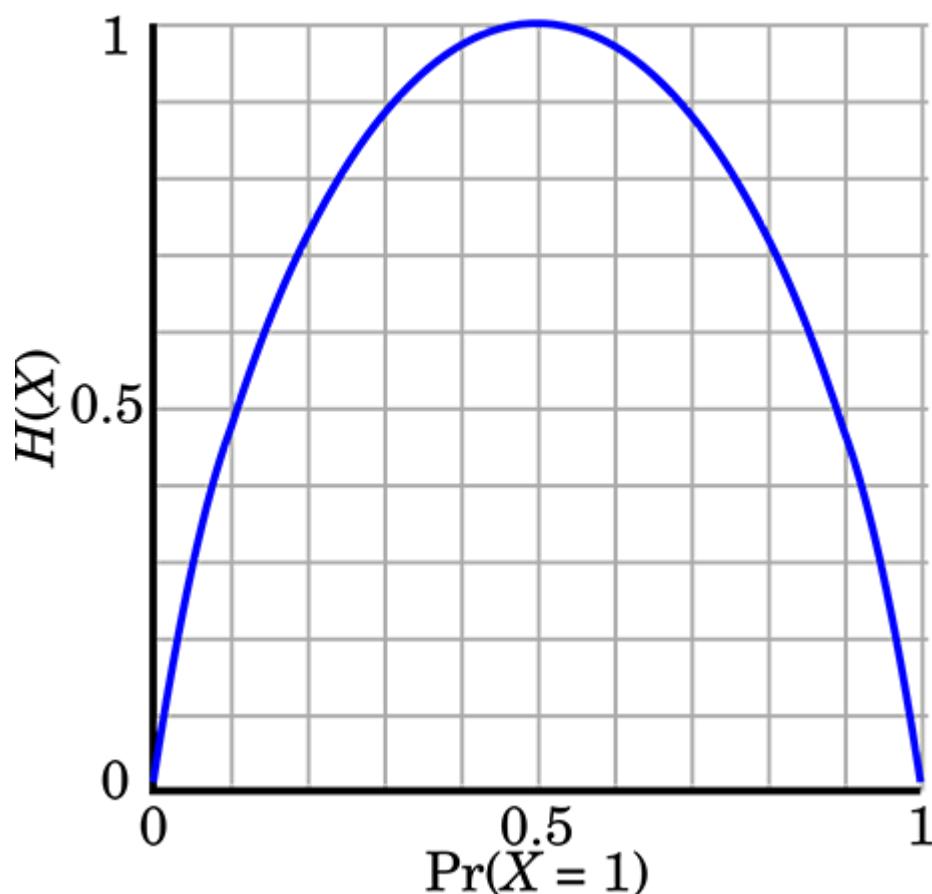


Figura 1. Dependencia de la entropía con la probabilidad de observación del suceso X_1 para una variable aleatoria X con una distribución de Bernoulli.

Fuente: [https://es.wikipedia.org/wiki/Entrop%C3%ADa_\(informaci%C3%B3n\)](https://es.wikipedia.org/wiki/Entrop%C3%ADa_(informaci%C3%B3n)).

Las señales pueden modelarse, desde un punto de vista matemático, como un conjunto de variables aleatorias (proceso estocástico). Por ejemplo: una señal de voz

de una duración determinada puede verse como una serie temporal finita, de forma que cada una de sus muestras representa la realización de una variable aleatoria.

La inclusión de nuevas muestras en la serie contribuirá a aumentar el contenido de información en esta, lo que muestra la dependencia de la entropía del proceso con la longitud de este. Por tanto, tiene sentido medir la variación de la entropía de la señal debida a la inclusión de una nueva muestra, hablamos de **la tasa de entropía o entropía diferencial**.

Considérese una señal de longitud N dada por $x = [x_1, x_2 \dots, x_N]$ y compuesta por la secuencia de la realización de N variables aleatorias. La entropía del proceso estocástico viene dada por:

$$\begin{aligned} H_N &= E\{-\log_2[p(x_1, x_2 \dots, x_N)]\} \\ &= - \int_{-\infty}^{\infty} \log_2[p(x_1, x_2 \dots, x_N)] p(x_1, x_2 \dots, x_N) dx_1 \dots dx_N \end{aligned}$$

Donde $p(x_1, x_2 \dots, x_N)$ es la función de densidad de probabilidad conjunta de las variables que componen el proceso estocástico.

A partir de esta expresión, el **ratio de entropía**, o simplemente entropía, de la señal se define de la siguiente forma:

$$E_N = \lim_{N \rightarrow \infty} H_{N+1} - H_N$$

Existen diferentes métodos para la estimación de la entropía de una señal. En este caso, propondremos la **entropía aproximada (ApEn)** como procedimiento de estimación.

Este algoritmo consiste en estimar la entropía de subseries de longitud m y $m+1$, de forma que la diferencia de ambas estimaciones proporciona el valor final de entropía.

A continuación, se describe cómo obtener la estimación de la entropía para subseries de **longitud** m . El mismo procedimiento se aplicaría para longitud $m + 1$.

Para nuestra serie original $x = [x_1, x_2 \dots, x_N]$:

- ▶ Se extraen todas las subseries de longitud m representadas por $x_i^{(m)}$.
- ▶ Posteriormente, se define una tolerancia r que permite identificar el número $N^{(m)}(i)$ de las subseries $x_j^{(m)}$ que se asemejan a $x_i^{(m)}$, de acuerdo a una métrica de la distancia entre ambas $d[x_i^{(m)}, x_j^{(m)}] \leq r$.
- ▶ Por tanto, la probabilidad de encontrar en la serie original una subserie similar a $x_i^{(m)}$ viene dada por la siguiente expresión:

$$C^{(m)}(i) = N^{(m)}(i)/(N - m + 1)$$

Donde $N - m + 1$ es el número total de subseries de longitud m que podemos encontrar en la serie original.

El término $C^{(m)}(i)$ proporciona una estimación discreta de la función de densidad de probabilidad $f(x^{(m)})$. Esta puede emplearse para aproximar la entropía de la variable aleatoria $x^{(m)}$, que vendrá dada por el promedio de la cantidad de información de esta fuente aleatoria, empleando la definición de entropía de Shannon. Así, la entropía de los procesos representados por $x^{(m)}$ se calcula como:

$$H_N^{(m)} = -1/(N - m + 1) \sum_{i=1}^{N-m+1} C^{(m)}(i)$$

El **ratio de entropía de la serie** (proceso estocástico) inicial, es decir, el incremento de entropía por la inclusión de una nueva muestra se obtiene como:

$$E_N = H_N^{(m+1)} - H_N^{(m)}$$

El valor de la entropía de una señal puede interpretarse como el grado de incertidumbre de la misma. De forma equivalente, refleja la capacidad para predecir un estado o valor posterior a partir del conocimiento u observación de los valores que previamente ha tomado la señal. Un mayor valor de entropía reflejará, por tanto, mayor complejidad y caos en la señal bajo estudio.

Como resultado, la entropía nos da una idea del nivel de impacto del ruido en una señal.

Si tomamos una muestra de una misma señal en las mismas condiciones pero en instantes de tiempo diferentes, la señal con mayor entropía será aquella con un mayor nivel de ruido.

Hasta ahora, se ha presentado el concepto de entropía tomando como contexto la influencia del ruido en una señal, y se ha descrito cómo estimar su valor en **series temporales**, señales dependientes de la variable tiempo de una única dimensión.

La naturaleza y el modelado matemático de las **imágenes** son diferentes al de este tipo de señales. Una imagen no lleva implícita la variable tiempo, como sí ocurre en una señal de voz o un electrocardiograma, sino que representa la **luz capturada** en cada posición del espacio. Además, la información de una imagen se representa en dos dimensiones.

Por tanto, en imágenes, de la misma forma que en series se caracterizó el ratio de incremento de la entropía respecto a nuevas muestras, podría pensarse en un ratio de entropía respecto a la unidad de área representada. Para la **estimación de la entropía en una imagen** se toma el histograma de los niveles de intensidad de esta. La estimación final se obtiene como la entropía de la variable aleatoria caracterizada por dicho histograma.

Al igual que sucedía en señales unidimensionales, la entropía tenderá a incrementarse, o al menos permanecerá igual, si el área de la imagen considerada

para la estimación se amplía. Así, menores valores de entropía se asociarán con patrones repetitivos en la imagen que conduzcan a un histograma con picos marcados (textura). En cambio, la entropía aumenta si existe una mayor variabilidad en los valores de intensidad observados en la imagen, no hay patrones marcados que produzcan un histograma más plano. En este sentido, se deduce que el ruido contribuye a aumentar la entropía de la imagen, ya que provoca que la variabilidad de los niveles de intensidad observados aumente.

4.3. Fuentes perturbadoras de las señales

Clasificación de los tipos de ruido

Las señales eléctricas se ven afectadas por diferentes tipos de ruido:

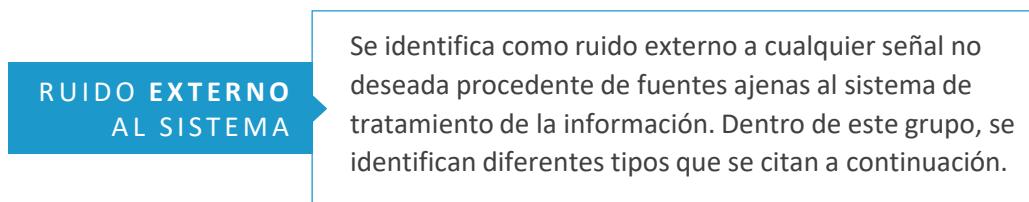


Figura 2. Definición de ruido externo al sistema.

Ruido atmosférico

El ruido atmosférico viene dado por las señales eléctricas derivadas de las descargas naturales que tienen lugar bajo la ionosfera. Las tormentas o las cargas eléctricas de las nubes son fuentes de este tipo de ruido que, generalmente, afecta en mayor medida a sistemas de comunicaciones que utilizan el espectro radioeléctrico como canal.

Aproximadamente, **la potencia del ruido atmosférico es inversamente proporcional a la frecuencia**. Así, el ruido atmosférico tiene mayor impacto en bandas de baja y

media frecuencia, mientras que el ruido de menor potencia afecta a las bandas VHF y UHF. Como resultado, el ruido atmosférico afecta a la banda de comunicaciones AM y decrece significativamente en frecuencias de TV y FM. Más allá de los 30 MHz, el ruido atmosférico tiene menor impacto negativo que el propio ruido del receptor.

Producido por el hombre

Se refiere a los artefactos eléctricos generados por fuentes como automóviles, motores eléctricos, interruptores, líneas de alto voltaje... Se conoce también como **ruido industrial**. La intensidad de estas señales ruidosas cobra mayor magnitud en grandes núcleos urbanos y zonas industriales. En estas áreas, el ruido de esta naturaleza prevalece sobre otras fuentes de ruido en el rango de frecuencias entre 1 MHz y 600 MHz.

Ruido impulsivo o *shot*

Este tipo de ruido origina la aparición de valores anómalos (*outliers*) en la señal. Se caracteriza por un aumento brusco de la intensidad durante un corto periodo de tiempo. Por lo general, su origen es un agente externo al sistema de información: un rayo o una interferencia derivada de una chispa de un motor. Sin embargo, no debe confundirse con el ruido atmosférico o producido por el hombre, ya que la duración de estos es más prolongada en el tiempo.

Galáctico

Tiene su origen en las perturbaciones producidas más allá de la atmósfera terrestre. Como principales fuentes de ruido galáctico aparecen el sol y otras estrellas.

- ▶ Solar: el sol es una gran fuente de emisión de energía en forma de radiación electromagnética. Por tanto, estas señales afectan a los sistemas de telecomunicaciones. El **rango frecuencial** de estas emisiones es muy amplio, incluyendo las bandas comúnmente empleadas para los sistemas de comunicación

por radio. La intensidad de la emisión producida por el sol varía de forma cíclica, con un período de unos once años aproximadamente. En los niveles más altos, esta radiación puede llegar a hacer inutilizables algunas bandas frecuenciales.

- ▶ Cómico: de la misma forma que el sol, otras estrellas próximas a nuestro planeta emiten energía en forma de radiación electromagnética que puede llegar a afectar a nuestras señales y sistemas de comunicaciones.

RUIDO INTERNO AL SISTEMA

El ruido interno o inherente a un sistema incluye al conjunto de señales aleatorias generadas en los dispositivos y su circuitería electrónica como resultado de sus propiedades físicas. Existen diferentes tipos de ruido interno que veremos a continuación.

Figura 3. Definición de ruido externo al sistema.

Ruido térmico

Esta fuente de ruido se debe a la agitación aleatoria de los electrones en los elementos de un circuito electrónico. Este movimiento solo podría cancelarse en condiciones de cero absoluto de temperatura. Por tanto, se trata de una fuente de **ruido inevitable** que siempre va estar presente en un sistema de captación y procesado de señales. El movimiento de los electrones es mayor conforme la temperatura del conductor aumenta, dando lugar a pequeñas corrientes eléctricas. Esta señal ruidosa se distribuye en una amplio rango de frecuencias, de forma que siempre afectará en cierto grado al sistema, a pesar de llevar a cabo diferentes etapas de filtrado.

Ruido de parpadeo (*flicker*) o 1/f

Se denomina 1/f debido a que su potencia decrece por debajo de 1 KHz cuando la frecuencia aumenta. Por lo que tiene mayor impacto sobre las bajas frecuencias. Las causas físicas de este tipo de ruido no están del todo claras. Se origina en elementos

como transistores o resistencias, y se hipotetiza que se debe a procesos de intermodulación en estos elementos.

Relación señal a ruido (SNR)

Ante una fuente de información afectada por artefactos ruidosos, el valor de la relación señal a ruido (SNR, *signal to noise ratio*) indica cuantitativamente **la calidad de la señal** de interés. Esta relación viene dada por el cociente entre la potencia de la señal recibida y la potencia estimada del ruido. Así, un valor mayor que la unidad indica una mayor presencia de señal frente al ruido. Generalmente, la relación entre ambos términos de potencia viene expresada en decibelios (dB), por lo que la SNR se calcularía como:

$$SNR = 10 \log_{10}(P_S/P_N)$$

Donde:

- ▶ P_S se corresponde con la potencia de la señal.
- ▶ P_N se corresponde con la potencia del ruido.

4.4. Caracterización matemática del ruido: procesos estocásticos

El término proceso estocástico ha sido previamente empleado en este tema para referirnos a una **señal aleatoria**. En nuestro caso, cualquier señal será el resultado de la combinación de la señal de interés y una señal no deseada, de naturaleza aleatoria y caótica, que contribuye a aumentar la entropía. Esta señal no deseada es el ruido.

Por lo que la señal resultante es, en sí misma, una señal aleatoria. Al igual que sucede con una variable aleatoria, de la que tomamos una muestra y obtenemos valores de acuerdo a una función de densidad de probabilidad, las señales que manejamos son realizaciones de un proceso estocástico. Cada vez que extraemos una muestra de la fuente de información, obtenemos una señal diferente.

En esta sección, se proporciona una definición formal de proceso estocástico que permita comprender el modelado y caracterización del ruido en el tratamiento de señales.

Una variable aleatoria está caracterizada por los siguientes tres elementos:

Espacio muestral. Es el conjunto de todos los resultados que pueden observarse en la realización de un experimento.

Conjunto de sucesos. Subconjunto del espacio muestral.

Ley de probabilidad. Asignación de probabilidad a cada uno de los sucesos observables.

Un proceso estocástico puede verse como una variable aleatoria para la que el resultado de un experimento viene dado en forma de señal. De la misma forma que una variable aleatoria, está caracterizado por la terna mencionada: espacio muestral, conjunto de sucesos y ley de asignación de probabilidades.

En la siguiente imagen vemos la interpretación del concepto de proceso estocástico X :

- ▶ Para diferentes realizaciones del proceso (a_1 , a_2 y a_3), se obtienen diferentes señales como resultado: $x(t, a_1)$, $x(t, a_2)$ y $x(t, a_3)$.
- ▶ Si nos fijamos en un único instante de tiempo t_0 , obtenemos muestras de la variable aleatoria $X(t_0)$.

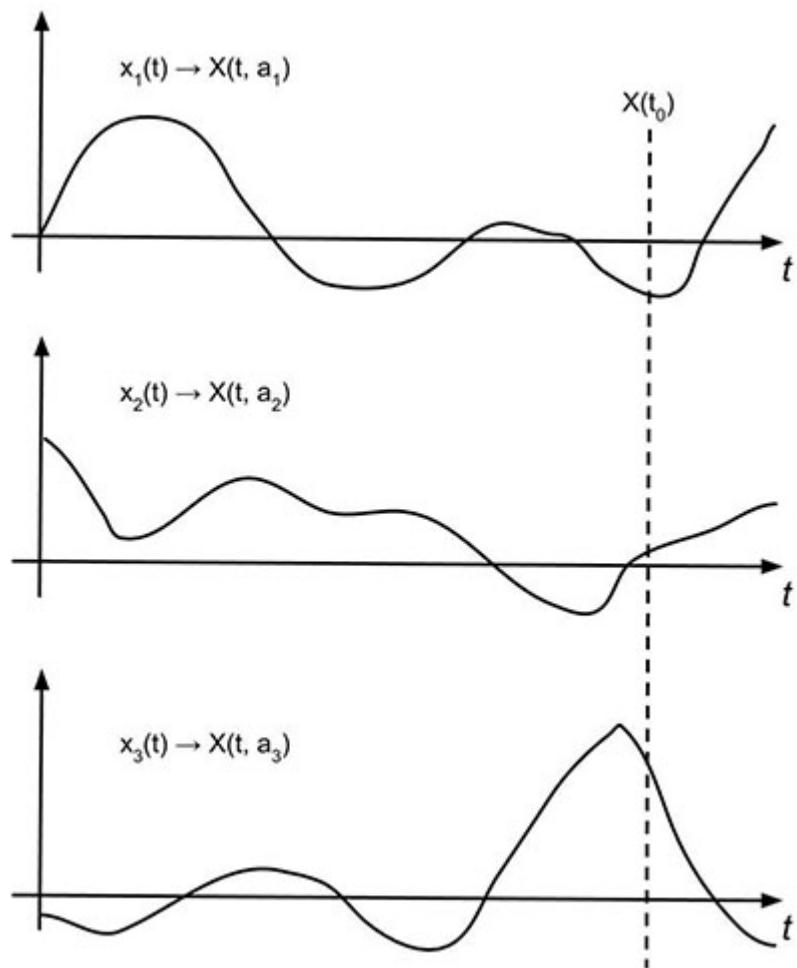


Figura 4. Interpretación del concepto de proceso estocástico.

En la práctica, como se ha mencionado previamente en este tema, tendremos señales ruidosas que, desde un punto de vista matemático, serán modeladas como un proceso estocástico. La componente de ruido se supondrá aditiva, por lo que la señal capturada $y(t)$ tendrá la siguiente forma:

$$y(t) = x(t) + \varepsilon(t)$$

Donde:

- ▶ $x(t)$ refleja la señal de interés.
- ▶ Y el término $\varepsilon(t)$ corresponde al ruido.

Considérese, por ejemplo, que la señal de interés corresponde a un tono de frecuencia f y que la componente ruidosa obedece a una distribución gaussiana de media nula y varianza s .

En el siguiente gráfico, podemos ver esta señal objetivo (parte superior) y una realización del proceso estocástico que corresponde a la señal observada (parte inferior).

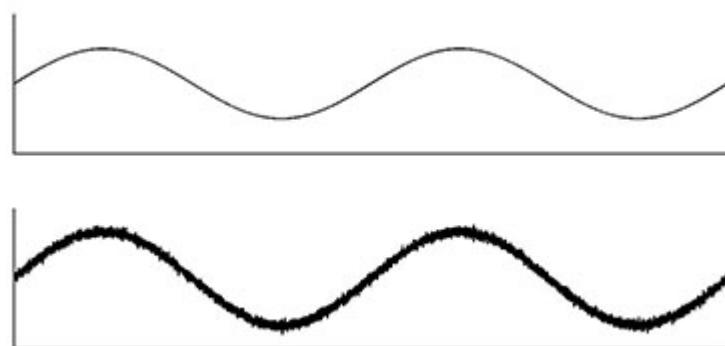


Figura 5. Ejemplo de una señal ruidosa modelada como un proceso estocástico.

Como puede apreciarse, el componente de ruido dota a la señal de naturaleza aleatoria que nos impide conocer con exactitud el valor de la misma en un instante t . A fin de caracterizar el proceso estocástico, el objetivo será conocer las propiedades estadísticas del mismo. Las funciones de distribución y densidad de probabilidad permiten modelar estadísticamente el proceso. Estas funciones vendrían dadas de la siguiente forma:

- ▶ Función de distribución: $F_X(x, t) = P(X(t) \leq x)$
- ▶ Función de densidad de probabilidad: $f_X(x, t) = dF_X(x, t)/dx$

A partir de estas funciones, puede definirse la estacionariedad del proceso:

- ▶ Un proceso es estacionario en **sentido estricto** si la función de densidad de probabilidad que caracteriza al proceso no varía con el tiempo. Es decir, para c constante tal que $c > 0$, se tendrá lo siguiente: $F_X(x, t) = f_X(x, t + c)$
- ▶ Un proceso es estacionario en **sentido amplio** si los momentos estadísticos que lo caracterizan (media, varianza, etc.) no varían respecto al tiempo.

Retomemos el ejemplo anterior. En esta ocasión, la señal capturada muestra, además del ruido gaussiano, otro componente que provoca una clara tendencia a lo largo del tiempo. El siguiente gráfico recoge este nuevo ejemplo. Como resultado de esta tendencia, las propiedades estadísticas de la señal no se mantienen a lo largo del eje temporal, por lo que no puede considerarse una señal estacionaria. Será necesario eliminar el componente de ruido que provoca esta tendencia para eliminar la no estacionariedad presente en nuestra información.

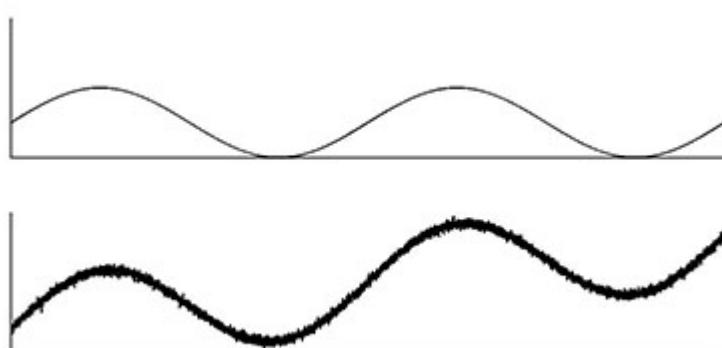


Figura 6. Proceso estocástico no estacionario.

Percepción Computacional

Detección y cancelación de anomalías

5.1. ¿Cómo estudiar este tema?

Para estudiar este tema deberás leer con atención las ideas clave que se desarrollan a continuación.

En el siguiente tema se aborda, como una etapa más en la preparación de las señales capturadas, la detección y eliminación de muestras de datos atípicas conocidas como anomalías (*outliers*).

El tema se organiza de la siguiente forma:

- ▶ Definición de anomalía. En primer lugar, se proporciona una definición de anomalía en el contexto del procesado de datos y señales. Así, se indican los **diferentes tipos de anomalías** que podremos observar en un conjunto de datos y las características de cada uno de ellos.
- ▶ Métodos de identificación. Se indica la taxonomía de los métodos que pueden emplearse para la **identificación** de anomalías.
- ▶ Eliminación de anomalías. Por último, se profundiza en aquellos métodos que son comúnmente empleados para la **detección y eliminación** de anomalías en señales, dado que estas son la fuente de información con la que se trabaja en la asignatura.

El siguiente esquema permite memorizar de forma sencilla los elementos de este tema.

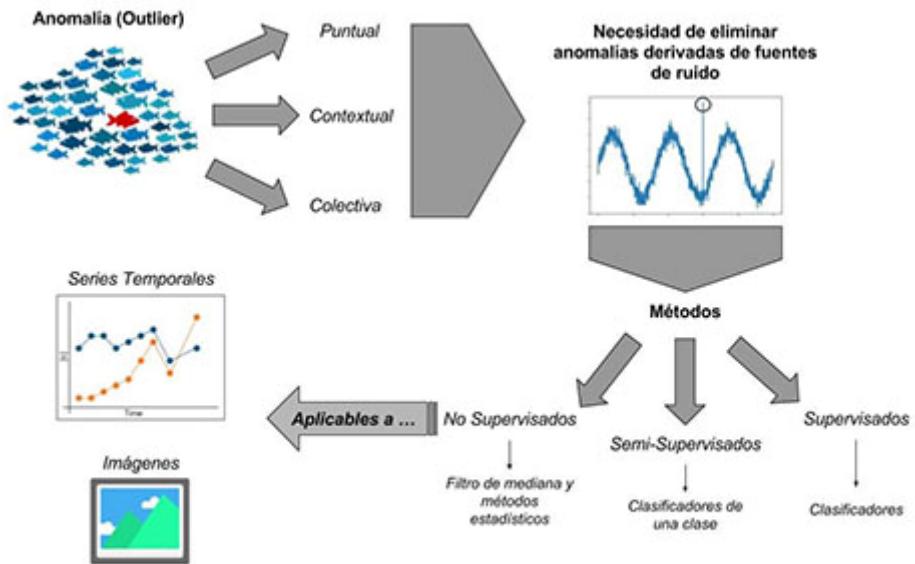


Figura 1. Conceptos relacionados con la presencia e identificación de anomalías en señales como series temporales e imágenes.

5.2. Definición de anomalía

La detección de anomalías tiene como objetivo la **identificación de valores atípicos en la fuente de información**, comúnmente conocidos por su vocablo en inglés, *outliers*. Estos se definen como patrones inusuales que no se ajustan al comportamiento esperado. La aparición de *outliers* en una señal o imagen refleja la existencia de ruido, generalmente de tipo impulsivo motivado, por ejemplo: por un valor de pico en un campo eléctrico cercano, o de inestabilidades en el procedimiento de captura, por ejemplo: el movimiento brusco de una cámara.

La detección de anomalías tiene aplicación directa en diferentes escenarios prácticos.

A continuación se citan algunas:

- ▶ Detección de intrusos en una red. Identificación de patrones atípicos en el tráfico de red que pueden indicar un ataque.

- ▶ Diagnóstico médico. Reconocimiento de lesiones con una bajo índice en la población que pueden indicar la existencia de alguna patología.
- ▶ Detección de transacciones fraudulentas. La gran mayoría de las transacciones que se realizan son lícitas y solo una pequeña proporción se corresponden con actividades fraudulentas.
- ▶ Predicción de fuga de clientes en grandes compañías. En los sectores de banca, seguros y telecomunicaciones, una pequeña parte de los clientes abandona la compañía, por lo que la identificación de estos comportamientos puede realizarse mediante técnicas de detección de anomalías.

Existen diferentes tipos de anomalías. Vamos a definir cada uno de ellos:

Anomalías puntuales

En el caso de que una muestra individual pueda considerarse notablemente diferente respecto al resto de los datos, esta puede ser tomada como un *outlier*. Este tipo de anomalía es el más simple y el foco de la mayoría de los trabajos de investigación sobre este tema.

Un claro ejemplo correspondiente a un escenario real sería el fraude cometido con una tarjeta de crédito. Si nos fijamos en una variable como la cuantía de la transacción, aquellas transacciones para las que la cantidad sea muy alta en comparación con el promedio del resto de transacciones previas son susceptibles de ser anomalías puntuales y, por tanto, sospechosas de fraude. Así, una anomalía puntual se expresa mediante la aparición de **valores pico** que se alejan excesivamente del conjunto de valores que encontramos.

En la siguiente figura podemos ver una señal en la que una de las muestras toma un valor que no observamos en ninguna otra. Se trata, claramente, de una muestra candidata a ser una anomalía.

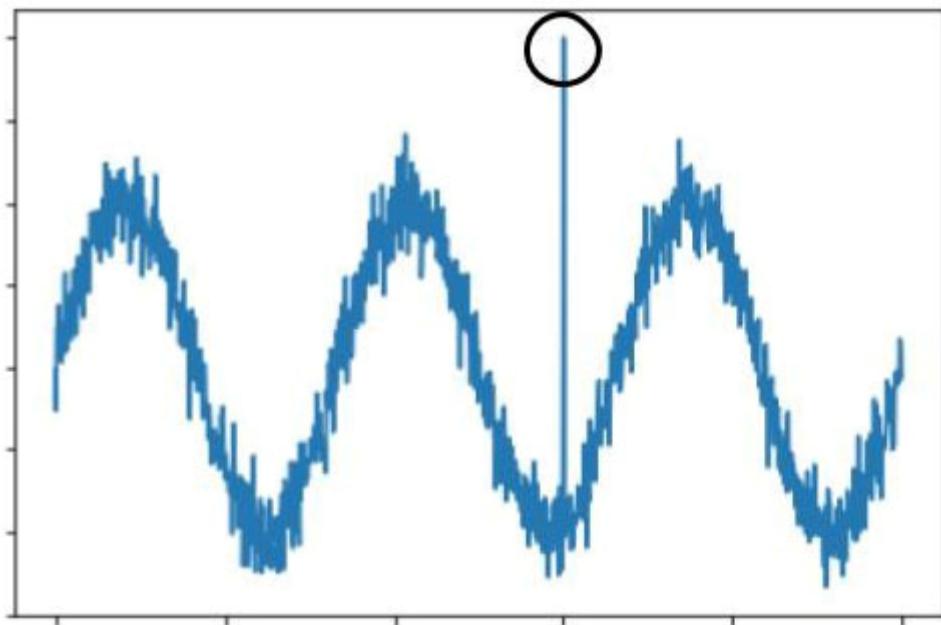


Figura 2. Ejemplo de anomalía puntual en una señal temporal.

Anomalías contextuales

Si una muestra de datos es anómala en un contexto específico (pero no de otro modo), se denomina anomalía contextual. La noción de contexto viene dada por la naturaleza de los datos. Cada muestra de datos se define teniendo en cuenta los siguientes atributos.

Atributos contextuales: estos se usan para determinar el contexto (o vecindad) para esa muestra. Vienen dados por la **naturaleza de la fuente de datos**. Por ejemplo, en conjuntos de datos espaciales, la longitud y la latitud de una ubicación son los atributos contextuales. En una serie temporal, el tiempo es un atributo contextual que determina la posición de una muestra en toda la secuencia.

Atributos de comportamiento: estos definen el carácter no contextual de una instancia. Es decir, representa el **valor de la muestra**. Siguiendo con el ejemplo de los datos espaciales, si se trata de cuantificar la precipitación promedio en cualquier punto de la superficie mundial, la cantidad de lluvia en cualquier lugar es un atributo de comportamiento.

El comportamiento anómalo se determina usando los valores de los atributos de comportamiento dentro de un contexto específico.

Una instancia de datos podría ser una anomalía contextual en un determinado contexto, pero una instancia de datos idéntica (en términos de atributos de comportamiento, es decir, de su valor) podría ser considerada normal en un contexto diferente. Esta propiedad es clave para identificar atributos contextuales y de comportamiento para una técnica de detección de anomalía contextual.

A diferencia de las anomalías puntuales definidas previamente, en las que únicamente se lleva a cabo una comparación de las muestras de datos disponibles para identificar un valor atípico, en señales temporales (series temporales) e imágenes se tiene en cuenta el contexto para definir un valor anormal.

Por ejemplo, en una imagen, es posible identificar un píxel anómalo si su intensidad es muy diferente a la de los píxeles vecinos. De la misma forma, en una serie temporal también será el entorno de un punto el que nos dé la información contextual necesaria para identificar un valor anómalo, tal y como se ejemplifica en la siguiente imagen.

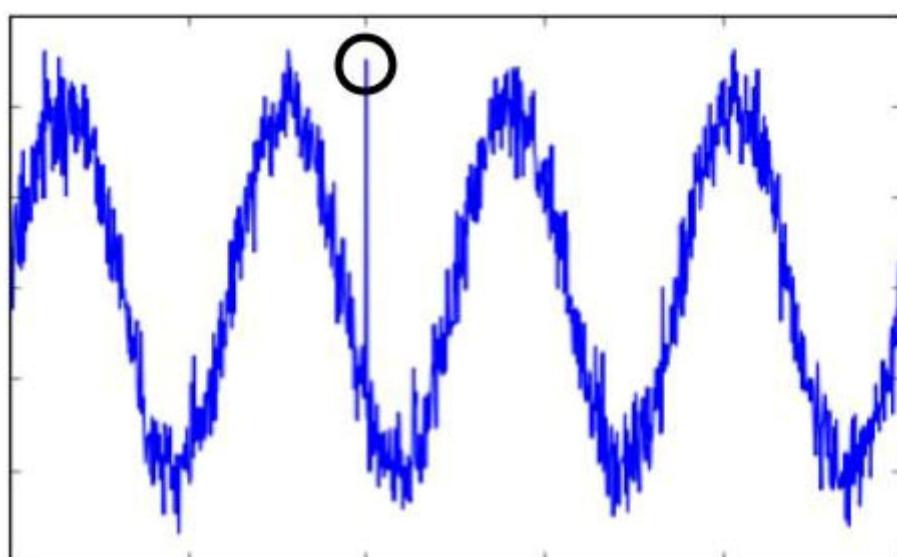


Figura 3. Ejemplo de anomalía contextual identificada en una serie temporal.

En ella vemos que la serie toma en algún momento valores similares al de la anomalía, pero el contexto nos indica que en este caso es una muestra atípica.

Anomalías colectivas

Si una **colección de instancias de datos** relacionadas es anómala con respecto a todo el conjunto de datos, se denomina anomalía colectiva. Las instancias de datos individuales en una anomalía colectiva pueden no ser anomalías por sí mismas, pero su ocurrencia conjunta como colección es anómala.

El siguiente gráfico ilustra un ejemplo de anomalía colectiva en una señal electrocardiográfica. La región resaltada denota una anomalía porque la señal toma aproximadamente el mismo valor durante un tiempo inusualmente largo. Sin embargo, ese valor no es una anomalía por sí mismo.

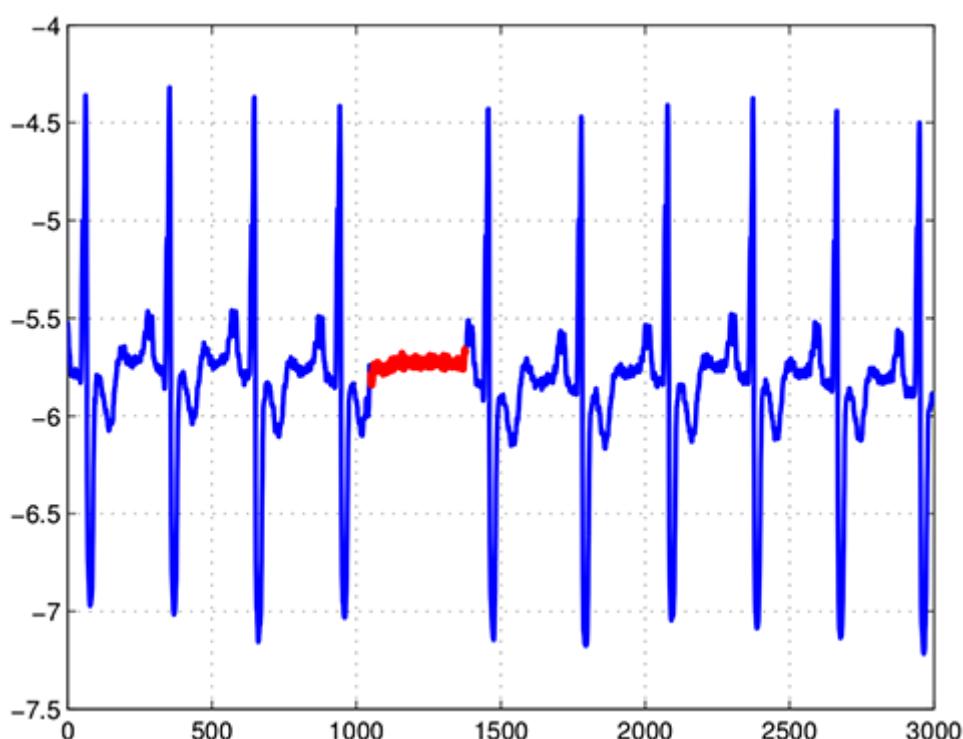


Figura 4. Ejemplo de anomalía colectiva en una señal de ECG.

Fuente: <https://www.datasciencecentral.com/profiles/blogs/anomaly-detection-for-the-oxford-data-science-for-iot-course>.

En este tema se tratará de identificar la presencia de anomalías en nuestras fuentes de información como una etapa más en la limpieza y preparación de la señal. Nos ceñiremos al escenario en el que la anomalía se corresponde con un artefacto ruidoso que debería ser eliminado a fin de preservar la calidad de la información.

En este sentido, los tres tipos de anomalías descritos previamente pueden tener como origen un elemento perturbador de la señal. Generalmente, las anomalías o artefactos a los que nos enfrentaremos vendrán dados por valores de pico inusuales derivados de la presencia de **fuentes ruidosas de tipo impulsivo**.

5.3. Métodos de identificación de anomalías

Adiferencia de los problemas convencionales de clasificación, donde se cuenta con un conjunto de datos de entrenamiento, previamente etiquetados con su correspondiente clase, y un conjunto de test que permite la estimación objetiva del rendimiento del modelo, hay múltiples configuraciones posibles cuando se habla de detección de anomalías.

Básicamente, la configuración de detección de anomalías que se utilizará depende de las etiquetas disponibles en el conjunto de datos. Así, podemos distinguir entre tres tipos principales:

Métodos supervisados

En este caso, se dispone de un conjunto de datos en los que cada muestra está asociada a una etiqueta que indica si es o no una anomalía. A partir de estos datos, se puede **entrenar un clasificador que capture el patrón característico de la anomalía**. Posteriormente, este clasificador se emplearía como identificador automático de estos valores atípicos.

Este escenario es muy similar al reconocimiento tradicional de patrones en el que las dos clases (no anomalías vs. anomalía) están fuertemente desequilibradas. Como consecuencia, debe tenerse en cuenta que no todos los algoritmos de clasificación se adaptan perfectamente a esta tarea, pues depende de la función de error que trate de optimizarse durante su entrenamiento.

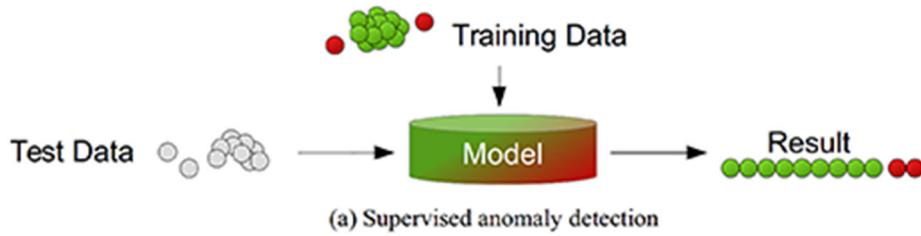


Figura 5. Métodos y escenarios para la implementación de técnicas de detección de anomalías: supervisado.
Fuente: Goldstein y Uchida, 2016.

Por ejemplo: los árboles de decisión, que buscan divisiones de los datos en las que se maximice la pureza de las categorías, no responden bien a conjuntos desequilibrados, mientras que las máquinas de vector de soporte (SVM, *Support Vector Machines*) o las redes neuronales artificiales (ANN, *Artificial Neuronal Network*) proporcionarán un resultado más plausible.

Sin embargo, la identificación de anomalías basada en técnicas supervisadas **no está muy extendida**, pues el punto de partida es la suposición de que las anomalías son conocidas y etiquetadas correctamente. Para muchas aplicaciones, las anomalías no se conocen de antemano o pueden ocurrir espontáneamente como novedades durante la fase de test, por lo que estas técnicas no tendrían cabida.

La utilización de modelos derivados del aprendizaje supervisado como herramientas de identificación de anomalías es común en aplicaciones prácticas como la detección de fraude o el diagnóstico médico, donde el número de muestras positivas es notablemente menor en proporción y, por tanto, se consideran una anomalía.

Métodos semisupervisados

Se trata de un procedimiento similar al anterior, pues también se emplea un conjunto de datos para entrenar un modelo. Sin embargo, en este caso, los datos de entrenamiento **contienen únicamente muestras no anómalas**. La idea fundamental es que el modelo aprenda la clase normal, de forma que se detectaría la anomalía al identificarse una desviación del patrón aprendido.

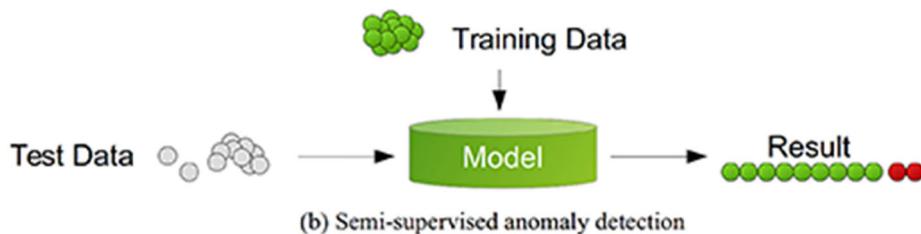


Figura 6. Métodos y escenarios para la implementación de técnicas de detección de anomalías: semisupervisado.

Fuente: Goldstein y Uchida, 2016.

Esta aproximación se conoce como clasificadores de una sola clase (*one-class classifiers*). Usualmente, suelen emplearse para tal fin los algoritmos SVM de una clase y los *autoencoders*. Además, cualquier método de estimación de densidad puede usarse para modelar la función de densidad de probabilidad de las muestras normales. Por ejemplo, los modelos de mezcla gaussiana o la estimación basada en funciones núcleo.

Métodos no supervisados

Representa el procedimiento **más flexible** para la implementación de un método de identificación de anomalías pues no requiere de ningún tipo de conocimiento previo sobre los datos (etiquetas). La idea es que un algoritmo de detección de anomalías no supervisado califique los datos únicamente en función de las propiedades intrínsecas de estos. Normalmente, las **distancias o densidades** se utilizan para dar una estimación de lo que es normal y lo que es un valor atípico.

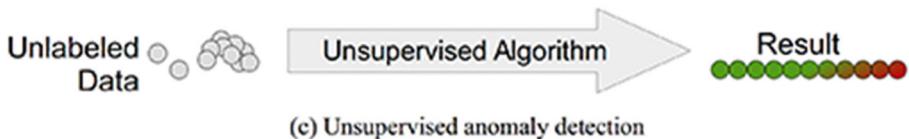


Figura 7. Métodos y escenarios para la implementación de técnicas de detección de anomalías: no supervisado.

Fuente: Goldstein y Uchida, 2016.

La salida devuelta por un algoritmo de detección de anomalías puede ser de dos tipos diferentes. Por un lado, puede emplearse **una etiqueta para cada muestra**. Esta indica de forma categórica si una instancia es anómala o no. Este es el caso, por lo general, de los métodos basados en aprendizaje supervisado.

En segundo lugar, puede obtenerse un **valor continuo** que refleje el grado de anormalidad de la muestra. Los algoritmos de detección de anomalías semisupervisados y no supervisados producen una salida de este tipo. Esto se debe a razones prácticas, pues el procedimiento de identificación solo informa de los casos más sospechosos, por lo que es necesario establecer un orden en las instancias a partir del grado de anomalía asignado.

El objetivo de la asignatura es el manejo de señales que, generalmente, se corresponden con:

- ▶ Series temporales: señales unidimensionales con el tiempo como variable independiente.
- ▶ Imágenes: señales bidimensionales con las coordenadas espaciales como variables independientes.

En este contexto, como se indicó previamente, las anomalías son el resultado de una fuente ruidosa que conlleva la aparición de elementos espurios en la señal. Por tanto, el propósito de la etapa de detección de anomalías es eliminarlas para mejorar la calidad de la señal en etapas posteriores. Es decir, no formaría parte de esta etapa la identificación de patrones anómalos de naturaleza no ruidosa en la señal.

Por ejemplo: el resultado de una arritmia reflejada en una señal de ECG o la presencia de determinados objetos en una imagen. De acuerdo a este contexto, el procedimiento habitual para la detección y eliminación de anomalías será la utilización de **métodos no supervisados** dado que, inicialmente, no disponemos de ninguna descripción de las posibles anomalías que pudiera haber en la señal.

5.4. Eliminación de anomalías

A

continuación se explican los procedimientos no supervisados más habituales para la eliminación de anomalías en señales.

Filtro de mediana

El filtro de mediana ha sido comúnmente empleado sobre señales 1D y 2D para la eliminación de ruido impulsivo. Estos artefactos se reconocen fácilmente mediante la inspección visual de la señal, pues están asociados a valores de pico que destacan notablemente sobre el resto de la señal (anomalía puntual) o sobre el entorno más cercano (anomalía de tipo contextual, pues el valor atípico iría en discordancia con aquellos de su entorno).

En imágenes, este tipo de anomalías se conoce como ruido «sal y pimienta» (*salt & pepper*), ya que el efecto que genera es el de píxeles colocados de forma aleatoria que toman valores extremos de intensidad (1 o 0). Aquí vemos una imagen afectada por este tipo de ruido.



Figura 8. Imagen con ruido de tipo «sal y pimienta» reflejado en píxeles con un valor extremo de intensidad.

Fuente: http://in.mathworks.com/matlabcentral/mlc-downloads/downloads/submissions/16201/versions/3/previews/toolbox_image/html/content.html.

El filtro de mediana es una operación que se aplica punto a punto mediante una ventana deslizante. El tamaño de esta viene determinado por el usuario.

En el caso de señales unidimensionales como series temporales, se trata de una ventana de longitud N , mientras que en imágenes la ventana se define en ambas coordenadas y es de tamaño $N \times N$. El valor de N es impar, ya que la ventana se centra en el punto de la señal que se pretende filtrar. Así, el valor resultante en este punto viene dado por la mediana de los puntos considerados por la ventana.

El filtro, tal y como se puede apreciar en su definición, no crea nuevos valores de la señal, sino que **selecciona como salida uno de los valores entrantes**. A continuación veremos el resultado de aplicar el filtro de mediana a una serie temporal y a una imagen.

En la mitad izquierda del siguiente gráfico tenemos una serie temporal donde pueden observarse dos puntos que representan sendas anomalías frente al resto de valores de la señal. Para eliminar estos *outliers*, se aplica un filtro de mediana con una ventana de longitud 3. El resultado del filtro aparece en la parte derecha del gráfico. Tal y como se aprecia, el filtro permite suprimir los valores atípicos, estimando su valor como la mediana de los puntos de su vecindad.

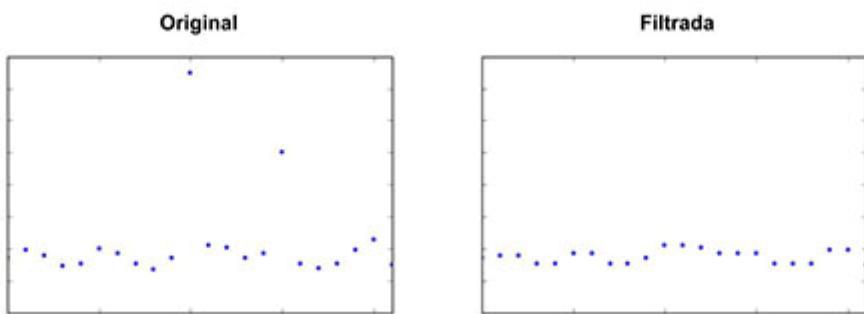


Figura 9. Aplicación de un filtro de mediana sobre una serie temporal afectada por la presencia de *outliers*.

Y a continuación vemos el resultado de utilizar el filtro de mediana para la eliminación de ruido de tipo «sal y pimienta» en la imagen anterior (figura 8). El tamaño de ventana empleado ha sido 3×3 . Como puede verse en la imagen filtrada, la aplicación del filtro conlleva cierta distorsión de la imagen, pues se modifica el valor de intensidad de los píxeles que no están afectados por este tipo de ruido.

El filtro de mediana es muy similar a un filtro de promedio que obtendría, para cada ventana, el valor medio de los píxeles o puntos considerados. Esta operación es equivalente a la utilización de un filtro paso-bajo en frecuencia, por lo que las variaciones rápidas de la señal, reflejadas como contrastes significativos en una imagen, quedan suavizados por el filtro.

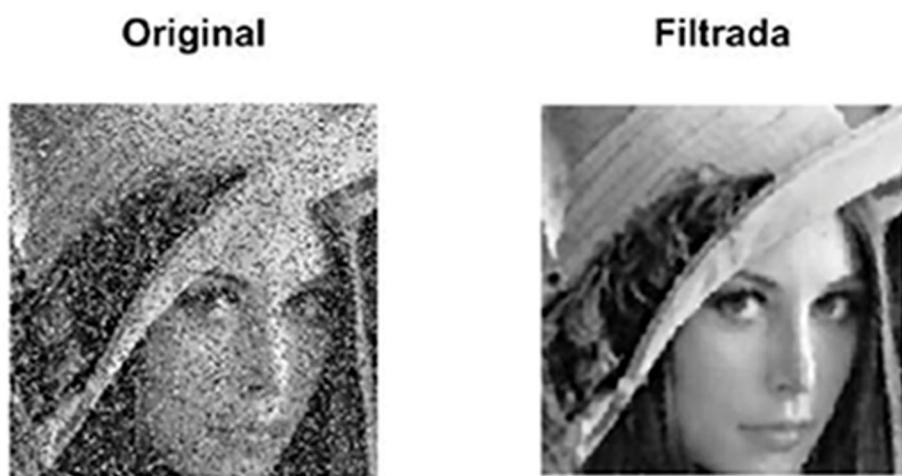


Figura 10. Efecto del filtro de mediana sobre una imagen con ruido impulsivo «sal y pimienta».

Fuente: Adaptado de http://in.mathworks.com/matlabcentral/mlc-downloads/downloads/submissions/16201/versions/3/previews/toolbox_image/html/content.html#17 .

Técnicas estadísticas

Otra técnica de detección y corrección de anomalías empleada habitualmente se basa en la utilización de la **función de densidad de probabilidad de los datos**. Dada la función de densidad $f(x)$, donde x es uno de los valores que puede tomar la variable aleatoria correspondiente, puede obtenerse una medida que cuantifica el grado de anomalía para la muestra x_1 como la inversa de $f(x_1)$.

Aquellos valores muy poco probables tenderán a ser identificados como atípicos y, por tanto, debe emplearse una estrategia adecuada para su tratamiento, por ejemplo: eliminarlos o estimar su valor como la media de los puntos vecinos.

Otra estrategia para la identificación de valores atípicos a partir de la función $f(x)$ sería considerar como tales aquellos **que estén en los extremos** del dominio de $f(x)$. Por ejemplo, a partir de $f(x)$ puede identificarse los valores x_a y x_b , tales que:

$$P(x \leq x_a) = P(x \geq x_b) = P_{min}$$

Si la muestra x_1 es menor que x_a o mayor que x_b , sería considerada una anomalía.

La utilización de estos procedimientos puede llevarse a cabo de forma global o local:

- ▶ En el primer caso, la función $f(x)$ corresponde al conjunto total de muestras disponibles. Es decir, todos los puntos de la serie temporal o todos los píxeles de la imagen. En este caso, podríamos detectar **anomalías de tipo puntual**.
- ▶ A fin de identificar **anomalías contextuales**, sería necesario aplicar este método en un entorno local del punto a estudiar. Para ello, la función $f(x)$ se correspondería únicamente con la vecindad del punto que se pretende evaluar como anomalía. Tal y como sucede con el filtro de mediana explicado previamente, sería necesario emplear una ventana deslizante centrada en el punto objetivo para definir dicha vecindad.

Además, ambos métodos basados en $f(x)$ requieren fijar inicialmente un umbral de decisión: para comparar el *score* obtenido con el primero de los métodos y para definir el valor P_{min} que identifica los valores extremos de una distribución en el caso del segundo procedimiento.

Este umbral determina la definición de anomalía en nuestro conjunto de datos y que ha de ser establecido por el usuario.

En la definición de los métodos basados en el empleo de la función de densidad de probabilidad $f(x)$ se ha asumido el conocimiento de esta. Sin embargo, esta función es desconocida, por lo que será necesario aplicar técnicas de estimación para obtener una aproximación a la misma.

Estas son algunas técnicas que pueden ser utilizadas para la estimación de esta función.

Histograma

Representa la técnica más sencilla. A partir de las muestras de una variable, esta se discretiza mediante la división de su dominio en un número limitado de intervalos de igual tamaño e identificados por su punto medio. Estos puntos representan los valores discretos que la variable puede tomar. Así, se obtiene la frecuencia (probabilidad) asociada a cada posible valor a partir del conjunto total de datos inicial contando el número de muestras de la variable que caen en cada intervalo.

La **elección del número de intervalos** empleado para la discretización de la variable tiene una influencia muy significativa en la aproximación obtenida. Un número demasiado pequeño de intervalos resultará en una aproximación excesivamente simple que no captura las particularidades de la distribución objetivo. Sin embargo, un número excesivo de intervalos conlleva que la estimación resultante presente discontinuidades (valores nulos) y cambios bruscos en su perfil.

Existen diferentes reglas para obtener el número óptimo de intervalos a considerar. Entre ellas, una de las más comunes es la **regla de Freedman-Diaconis**, que viene dada por la siguiente expresión:

$$T = 2IQR(x)/\sqrt[3]{n}$$

Donde:

- ▶ x hace referencia a la muestra disponible de la variable.
- ▶ n es el número de muestras.
- ▶ IQR es el rango intercuartil. Este se obtiene de la diferencia entre los percentiles 75 % y 25 %, y representa el grado de dispersión de la muestra, es decir, si la distribución está extendida en el dominio de x o si, por el contrario, está concentrada en un rango de valores.

Estimación paramétrica

Se asume que la función de densidad de probabilidad que caracteriza estadísticamente a la variable es de tipo normal. Por tanto, la media y varianza de esta distribución son los parámetros a obtener. Para ello, se emplean las estimaciones derivadas de la muestra disponible:

$$\mu_x = 1/n \sum_{i=1}^n x_i$$

$$\sigma_x = \sqrt{1/n \sum (x_i - \mu_x)^2}$$

Obviamente, la principal limitación de este método viene dada por la suposición inicial sobre la forma de la distribución. El error en la estimación será más representativo, por tanto, cuanto más difiera la distribución real de la variable del perfil normal inicialmente supuesto.

Funciones núcleo (método de Parzen)

Se trata de un procedimiento híbrido entre la estimación basada en histogramas y la estimación paramétrica. En este caso, la estimación de la función de densidad de probabilidad viene dada por la superposición de funciones núcleo centradas en cada uno de las muestras x_i observadas inicialmente. La expresión de la función estimada se obtiene de la siguiente forma:

$$\hat{f}_x = 1/n \sum_{i=1}^n g(x - x_i, \theta),$$

Donde:

- ▶ $g(x, \theta)$ es la función núcleo.
- ▶ θ representa el conjunto de parámetros de esta función.

Comúnmente, se emplea una normal Gaussiana como función núcleo, de forma que el conjunto de parámetros θ viene dado únicamente por la varianza de la normal, dado que cada función núcleo se centra en la muestra x_i correspondiente.

Es habitual emplear el mismo valor de varianza para el conjunto de funciones núcleo, por lo que la función de densidad de probabilidad estimada se obtendría como:

$$\hat{f}_x = 1/n \sum_{i=1}^n 1/\sqrt{2\pi\sigma^2} \exp[-(x - x_i)^2/(2\sigma^2)]$$

Como se observa, el efecto de la varianza de las funciones núcleo normales es similar al del tamaño del intervalo para el cálculo del histograma. De hecho, el histograma puede verse como un caso particular de estimación basada en funciones núcleo, en el que estas funciones vendrían dadas por pulsos uniformes de altura igual a la unidad y centrados en el punto medio del intervalo.

Una regla extendida para la obtención de un valor adecuado de la varianza de las funciones núcleo es fijar esta al siguiente valor:

$$\sigma = 1.06\sigma_x n^{-1/5}$$

5.5. Referencias bibliográficas

Goldstein, M. y Uchida, S. (2016). A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data. *PLOS ONE*, 11(4): e0152173.
Recuperado de <https://doi.org/10.1371/journal.pone.0152173>

Percepción Computacional

Procesamiento de imagen. Operaciones elementales

6.1. ¿Cómo estudiar este tema?

Para estudiar este tema deberás leer con atención las ideas clave que se desarrollan a continuación.

Dentro del marco de la asignatura, este tema se centra en el procesado de imagen y, concretamente, se ubica en las primeras etapas de tratamiento de una imagen. En él se describen técnicas básicas para comenzar el procesado y que persiguen realzar el contenido de la imagen bajo estudio. Es decir, en un contexto o aplicación determinados, se busca **capturar la información relevante de la imagen** frente a la más prescindible.

Formalmente, las operaciones de realce de imagen consisten en la utilización de una o varias técnicas para resaltar en ella determinadas propiedades. Estas características a magnificar dependen del propósito final del escenario práctico en el que nos encontramos. Por ejemplo, lesiones concretas en imágenes sobre tejidos corporales para la identificación de una patología o los bordes de un objeto en un sistema de visión automática.

A modo ilustrativo, resumimos las siguientes operaciones que pertenecen al ámbito del realce de imagen:

- ▶ La supresión del ruido de fondo derivado del proceso de captura de una imagen. En este caso, se pretende **remarcar el contenido** de la imagen en detrimento de la textura en el fondo de esta. El objetivo es que el fondo de la imagen presente una intensidad lo más constante posible, ya que el detalle de interés no está en esa parte de la señal.

- ▶ El ajuste de intensidad que persigue **aumentar el nivel de contraste** de la imagen, de forma que puedan distinguirse con mayor nitidez los elementos presentes en ella.
- ▶ El realce de bordes, cuyo objetivo es **enfatizar y delimitar las transiciones** entre objetos diferentes a fin de poder ser distinguidos posteriormente con mayor facilidad.

El propósito final de las operaciones de realce de imagen es facilitar las etapas posteriores destinadas a la extracción de información. Así, estas etapas pueden ser llevadas a cabo por:

- ▶ Un observador humano. El resultado de las operaciones de realce debe hacer que la imagen sea más fácilmente «digerible», es decir, que puedan distinguirse con mayor nitidez las estructuras y elementos implicados en el caso de uso a tratar.
- ▶ Una máquina. Las operaciones de realce representan, en este caso, una etapa de preprocesado, cuyo objetivo es generar inputs adecuados para etapas posteriores destinadas a extraer información de la imagen. Así, el indicar cuáles son los bordes existentes entre diferentes estructuras puede no generar una imagen fácilmente interpretable para un humano, pero sí para una etapa posterior del sistema automático de procesado.

Existen diferentes operaciones para el realce de imagen. En este tema, se pone el foco en las **operaciones elementales de tipo punto a punto**. Estas operaciones se caracterizan por que el valor de un píxel en la imagen procesada es función únicamente del valor de ese mismo píxel en la imagen original. Matemáticamente, se expresa de la siguiente forma:

$$B(x, y) = T(A(x, y))$$

Donde:

- ▶ B es la imagen procesada.
- ▶ A es la imagen original.

- ▶ La coordenada (x, y) identifica al píxel sobre el que se aplica la operación.
- ▶ T es la función u operador de transformación.

La siguiente figura esquematiza la aplicación de este tipo de operadores.

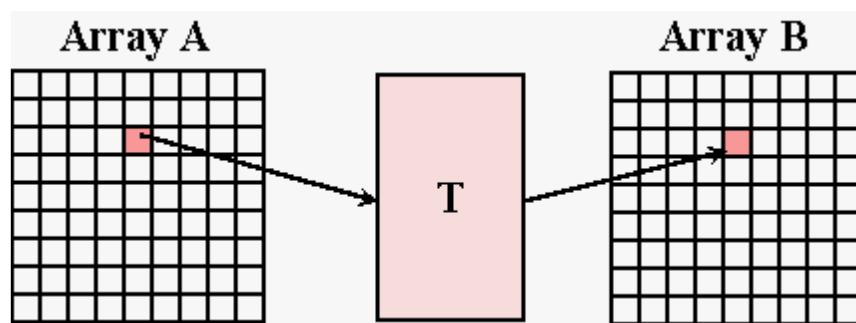


Figura 1. Aplicación de un operador punto a punto T sobre una imagen A y dando como resultado una imagen B.

Fuente: www.cis.rit.edu

Con el fin de sintetizar la información proporcionada en este tema, en el siguiente esquema se reflejan los principales conceptos.

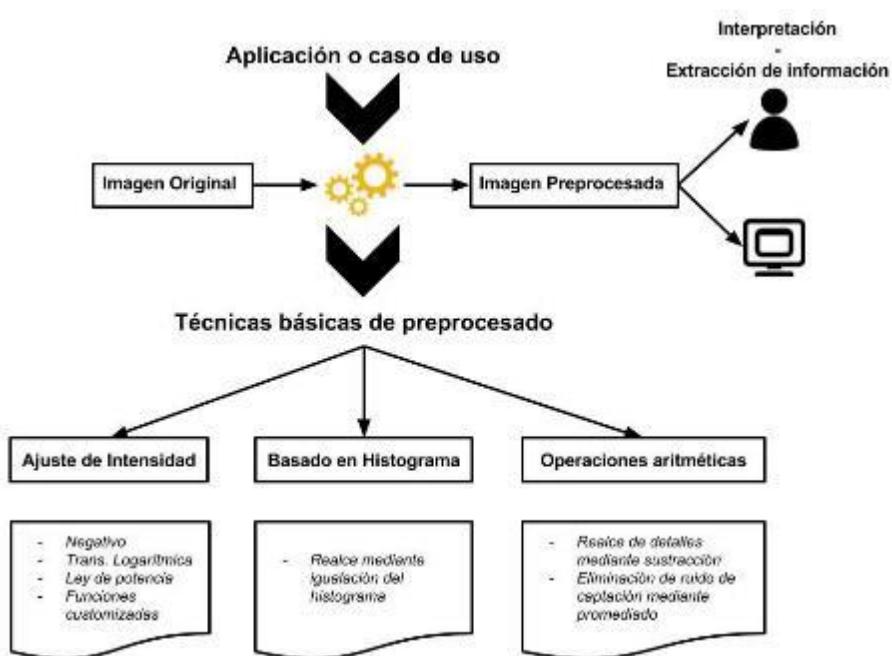


Figura 2. Síntesis de los conceptos elementales descritos en el tema: preprocessado de imagen mediante operaciones elementales.

6.2. Ajuste de intensidad

Las operaciones de ajuste de intensidad consisten en diferentes expresiones del operador T de la ecuación anterior. La figura 3 recoge algunas de las operaciones de transformación más comunes para una imagen con L niveles de intensidad diferentes.

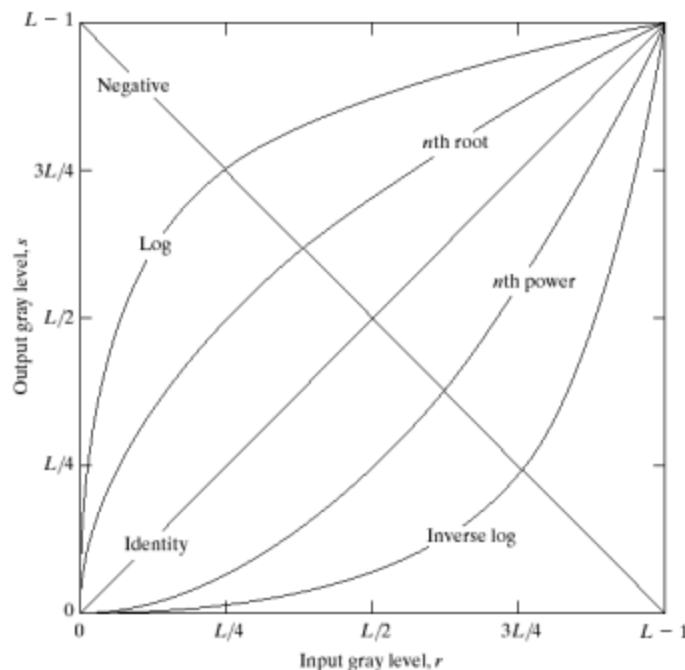


Figura 3. Perfil de las funciones de transformación punto a punto más comúnmente empleadas.

Fuente: González y Woods, 2008.

Negativo de una imagen

El negativo de una imagen representa la imagen invertida respecto a la original. La expresión matemática de la transformación viene dada por la siguiente ecuación:

$$T(u) = L - u$$

Donde L es el mayor nivel de intensidad que puede tomar un píxel.

Como puede apreciarse, se trata de una **transformación lineal**. La figura 4 muestra un ejemplo de la aplicación de esta transformación a una imagen médica con el objetivo de resaltar un detalle de interés que, en este caso, viene dado por una lesión. Tal y como se aprecia, la lesión destaca con mayor nitidez en la imagen transformada, a pesar de que la transformación no varía el contenido visual de la señal al ser de naturaleza lineal.

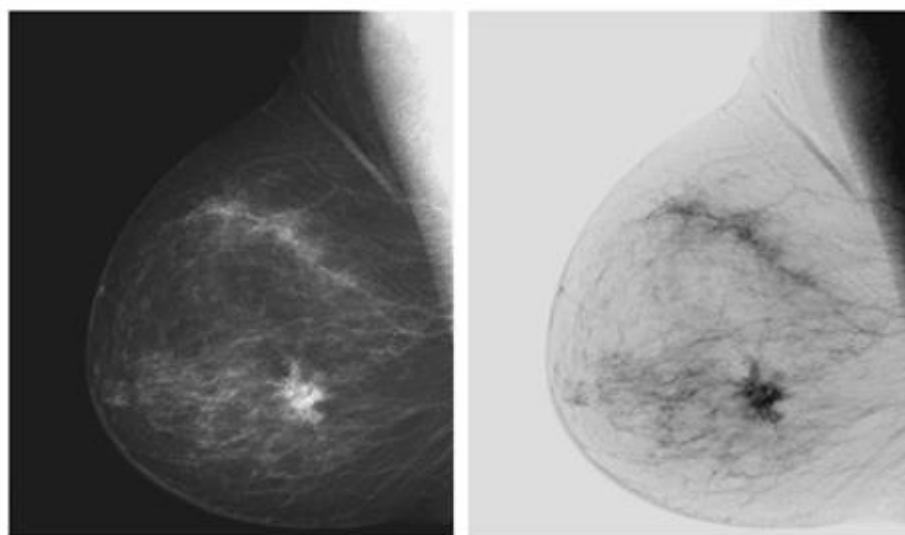


Figura 4. Ejemplo de aplicación del operador de transformación para la obtención del negativo en una imagen médica.

Fuente: González y Woods, 2008.

Esta transformación tiene especial utilidad para realzar estructuras de mayor intensidad, es decir, de tonos más próximos al color blanco (que representa el valor máximo de intensidad) encerradas en áreas oscuras de la imagen.

Transformaciones logarítmicas

La expresión de los operadores logarítmicos es la siguiente:

$$T(u) = C \log(1 + u)$$

Donde C es una constante.

El **perfil del operador** puede apreciarse en la figura 3. Como puede observarse, esta transformación mapea un rango pequeño de valores menores de intensidad en la imagen de entrada en un rango amplio de salida. Sin embargo, sucede lo contrario para los valores de intensidad mayores, que tienden a ser concentrados en un rango estrecho de valores de salida.

Por tanto, esta transformación se emplea cuando se pretende expandir el rango de intensidad de los píxeles oscuros al mismo tiempo que se equiparan entre sí los píxeles más claros.

La utilidad práctica de este operador puede apreciarse en aquellas situaciones en las que la imagen de entrada posee un rango dinámico muy amplio y su representación, por tanto, nos llevará a distinguir únicamente píxeles negros y blancos.

Esta situación suele observarse con la representación de imágenes en el dominio de la frecuencia. El espectro de la imagen típicamente tiene valores significativamente mayores que el resto en el entorno de la frecuencia origen: gran parte de la potencia de la señal se concentra en las frecuencias más bajas. Por tanto, a la hora de representarlo gráficamente, no es posible apreciar la evolución de la señal hacia frecuencias mayores.

La figura 5 muestra un ejemplo de **cómo emplear la transformación logarítmica** para obtener una representación más informativa del espectro de una imagen. En este caso, se han realizado los píxeles en niveles menores de intensidad, manteniéndose prácticamente intactos aquellos píxeles cercanos al máximo valor de intensidad.

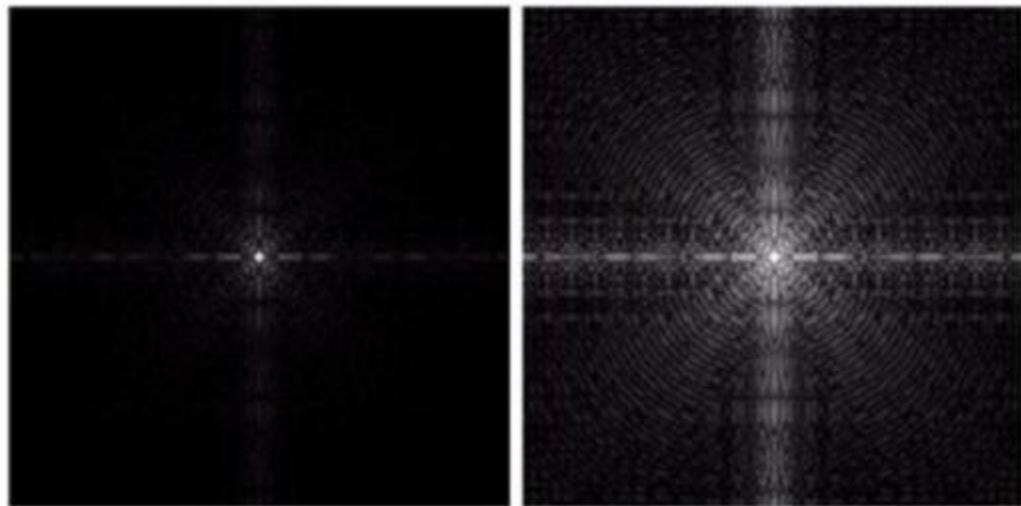


Figura 5. Ejemplo de utilización de una transformación logarítmica para el realce de las componentes frecuenciales más altas en el espectro de una imagen.

Fuente: González y Woods, 2008.

Ley de potencia

La función de transformación de acuerdo a la ley de potencia viene dada por la siguiente expresión:

$$T(u) = Cu^\gamma$$

Donde C y γ son constantes positivas. El perfil de las funciones obtenidas de acuerdo a esta expresión se muestra en la siguiente imagen.

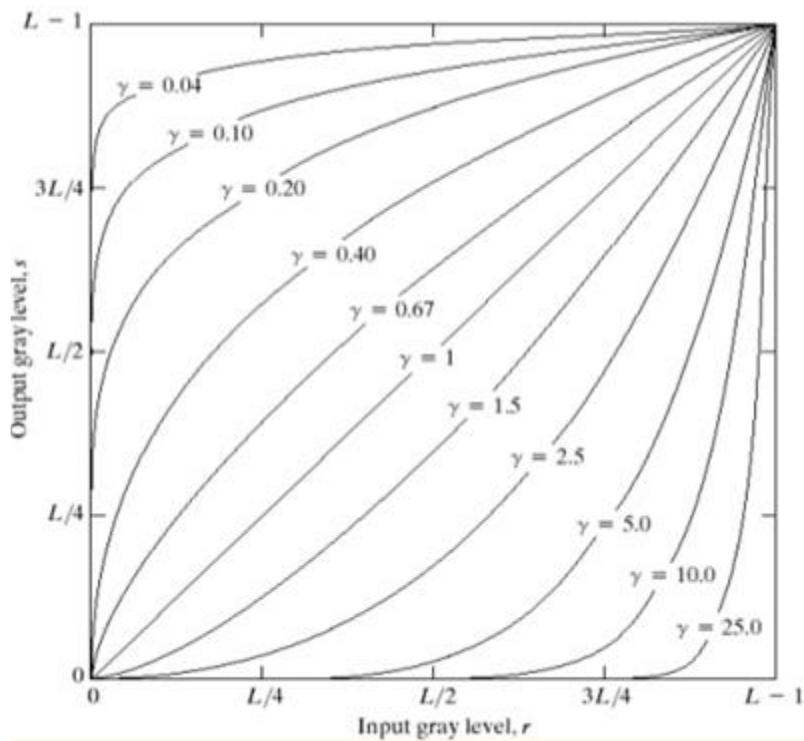


Figura 6. Funciones de transformación punto a punto obtenidas de acuerdo a la ley de potencias.

Fuente: González y Woods, 2008.

Como en el caso de las transformaciones logarítmicas, valores de γ menores que la unidad tienden a expandir el rango de intensidad de los píxeles más oscuros, mientras que este se comprime para los píxeles más claros.

La ventaja de estas funciones respecto a la transformación logarítmica es la posibilidad de generar una amplia familia de transformaciones simplemente mediante la variación del parámetro γ .

Así, las curvas generadas con $\gamma > 1$ tienen exactamente un efecto opuesto a las correspondientes a $\gamma < 1$.

El origen de este tipo de transformación se encuentra en la respuesta observada en algunos dispositivos antiguos destinados a la captura de imagen o la impresión. Esta respuesta obedece a una función de este tipo, por lo que es necesario aplicar una transformación opuesta para restaurar la imagen. Este procedimiento se conoce como **corrección gamma**, haciendo referencia al exponente de la función de transformación.

Funciones definidas a trozos

Las funciones anteriores tienen una expresión matemática para todo el dominio de aplicación, es decir, para cualquier valor de intensidad de entrada. Sin embargo, puede ser interesante aplicar diferentes transformaciones dependiendo del rango de intensidad de los píxeles sobre los que se opera. En esta situación, se emplean operadores definidos a trozos o tramos.

La **principal ventaja** de estas funciones es que su forma o expresión puede ser arbitrariamente compleja. Por ejemplo, puede dividirse el rango de intensidades de entrada en tantos tramos como se desee, definiéndose una transformación específica para cada uno de ellos.

Sin embargo, el **inconveniente más relevante** es que requieren un alto grado de participación del usuario para su definición. Generalmente, un humano define la función mediante la inspección visual del efecto producido por cada una de las transformaciones que se aplican sobre los diferentes tramos.

Una de las aplicaciones más comunes de este tipo de funciones es el **realce del contraste en una imagen**. En algunas situaciones, la imagen puede tener un rango dinámico pequeño debido a una pobre iluminación de la escena, el sensor empleado o la configuración de la lente empleada en la captura. Por tanto, la idea del realce del contraste es aumentar el rango dinámico de la imagen de entrada. Para ello, se trata de remarcar las diferencias entre píxeles cuyos valores de intensidad están concentrados en un rango intermedio. El perfil de la transformación será, en consecuencia, similar al mostrado en la figura 7.

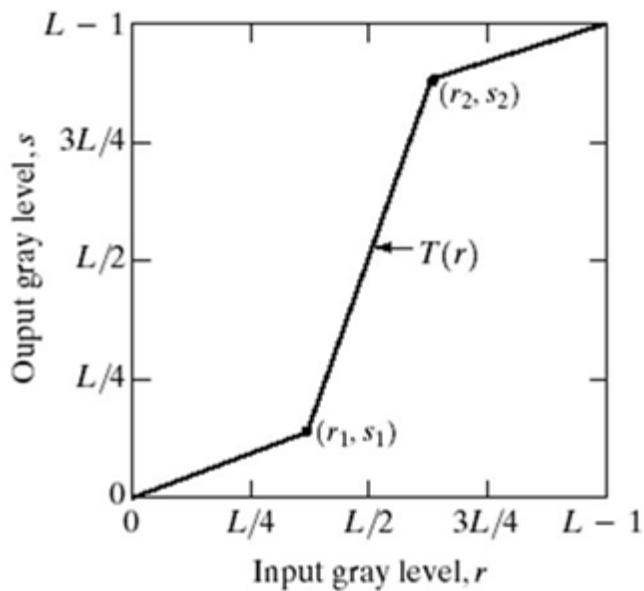


Figura 7. Perfil de la función de transformación definida a tramos para conseguir el realce de contraste.

Fuente: González y Woods, 2008.

Por otro lado, la figura 8 muestra el efecto de la aplicación de una transformación de este tipo sobre una imagen con un rango dinámico reducido. Como puede observarse, el resultado obtenido es un realce significativo respecto al contraste inicial entre los objetos y el fondo de la imagen.

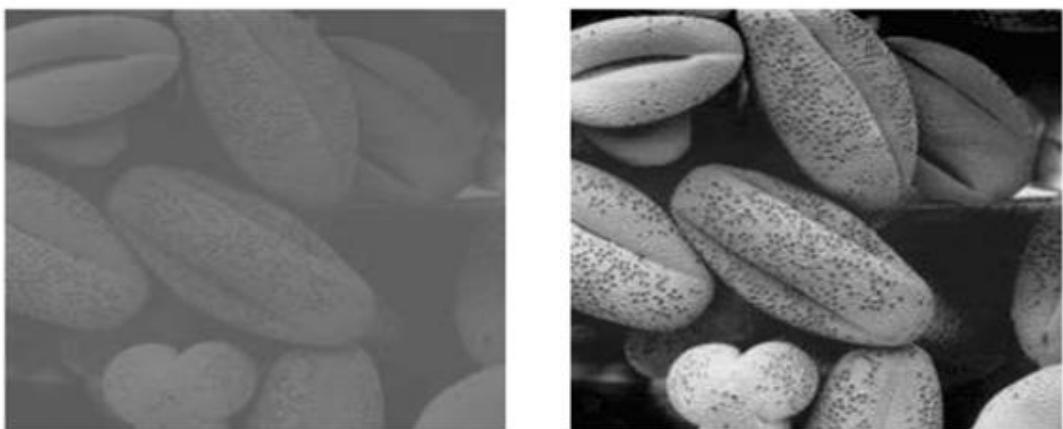


Figura 8. Ejemplo de realce de una imagen mediante la utilización de una función de transformación definida a tramos.

Fuente: <https://www.mathworks.com/>.

6.3. Procesado sistemático del histograma

Como se indicaba previamente, la expresión de la función de transformación definida a trozos para el realce de contraste se obtiene de forma manual, tras diferentes ciclos de prueba. Este modo de proceder conlleva demasiado y puede llegar a convertirse en una tarea tediosa. Una opción alternativa a este método para el realce de contraste es el procesado de la imagen basado en las propiedades de su histograma.

El histograma de una imagen proporciona una estimación de la función de densidad de probabilidad de la variable, la cual representa el valor de intensidad de los píxeles.

Para una imagen cuyos píxeles toman un nivel de intensidad de entre L niveles diferentes, el histograma se obtiene mediante el conteo del número de píxeles en cada nivel. Para obtener una versión normalizada en la que los conteos puedan interpretarse como probabilidades, el número de píxeles asociados a cada nivel se divide entre el número total de puntos en la imagen.

La **operación de igualación del histograma de una imagen** permite realizar el contraste de esta, ensalzando las estructuras presentes en la imagen de forma automática sin intervención externa del usuario.

Esta operación parte de la modelización matemática de la imagen de tamaño $N = n_r \times n_c$ como la realización de N variables aleatorias, una por cada píxel Y_i , que toma valores en el rango $L' - 1$. Se denota por f_{Y_i} a la función de densidad de probabilidad de cualquiera de estas variables aleatorias.

El objetivo de la transformación es convertir cada una de estas variables aleatorias en variables uniformes X_i en el rango $[0, \dots, L - 1]$.

Para ello, como es sabido, una variable aleatoria puede convertirse en uniforme transformándola a través de su **función de distribución**. Por tanto, basta con obtener la variable X_i como la transformación de Y_i a partir de su función de distribución:

$$X_i = F_Y(Y_i) = \int_{-\infty}^{Y_i} f_Y(\tau) d\tau$$

Sin embargo, la función de distribución de Y no es conocida, por lo que debe emplearse una **estimación** de la misma. Dicha estimación se obtiene a partir del histograma.

El valor de este para el nivel de intensidad l puede interpretarse como la probabilidad de que la variable Y tome dicho valor. Es decir:

$$P(Y = l) = p_l \cong \frac{n_l}{N}$$

Donde n_l es el número de píxeles con un nivel de intensidad l en la imagen. Por tanto, la función de distribución puede aproximarse mediante el histograma acumulado:

$$F_Y(l) = \sum_{k=0}^l p_k, \quad l = 0, \dots, L - 1$$

La figura 9 muestra un ejemplo en el que se emplea la igualación de histograma para realizar las estructuras de una imagen. Además del efecto logrado en la propia imagen, puede apreciarse el impacto de la transformación en el histograma. Este se caracteriza por un perfil más uniforme tras la transformación. En otras palabras, hay un mayor grado de dispersión en los valores de intensidad que toman los píxeles de la imagen transformada. En el caso de la original, estos tienden a estar concentrados en un rango menor de valores de intensidad.

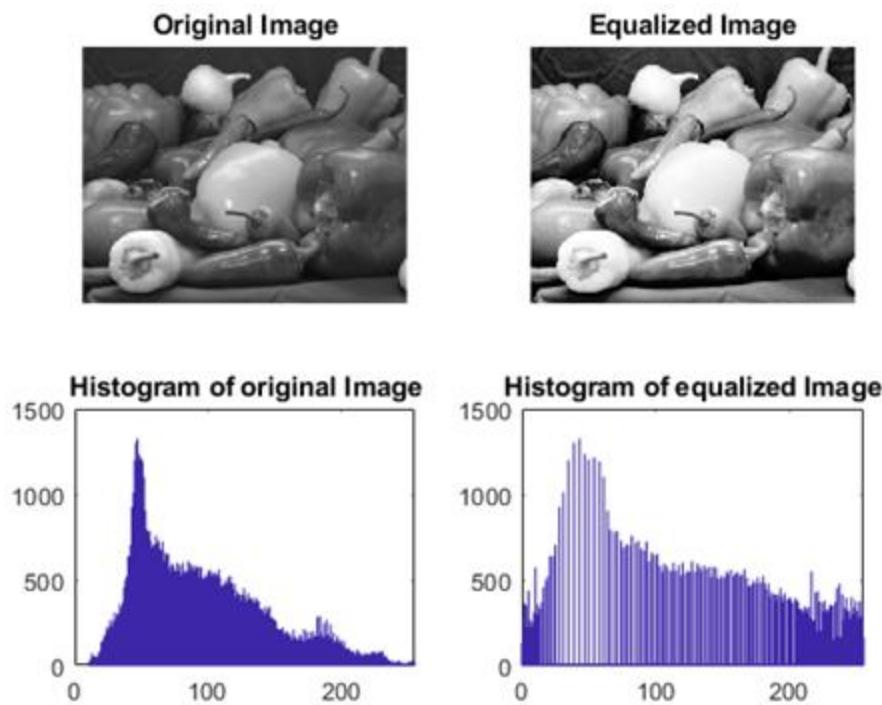


Figura 9. Ejemplo del realce de contraste mediante igualación del histograma.

Fuente: <https://www.mathworks.com/>.

6.4. Suavizado y realce mediante operadores aritméticos

De las cuatro operaciones aritméticas posibles (suma, resta, multiplicación y división), la **resta** y la **suma** son los dos operadores más útiles en las etapas de preprocesado para conseguir el realce de una imagen.

Realce de imagen mediante el operador resta

La diferencia $C(x,y)$ entre dos imágenes $A(x,y)$ y $B(x,y)$ se obtiene mediante la resta entre cada par de píxeles en ambas:

$$C(x,y) = B(x,y) - A(x,y)$$

Esta operación permite **realizar las diferencias entre dos imágenes**. Sirva como ejemplo la figura 10.

- ▶ En la esquina superior izquierda se tiene la imagen original en la que cada píxel se codifica mediante 8 bits: hay 256 niveles diferentes de intensidad.
- ▶ La imagen de la esquina superior derecha se obtiene asignando un valor 0 al cuarto bit de cada píxel. Aparentemente, no se aprecia ninguna diferencia entre ambas imágenes.
- ▶ Sin embargo, esta hipótesis puede ser contrastada mediante la resta de las dos señales píxel a píxel. Este resultado se muestra en la imagen de la esquina inferior izquierda, que refleja cierta diferencia entre ambas imágenes.
- ▶ Para apreciarse mejor, en la esquina inferior derecha se ensalzan estas diferencias mediante una operación de igualación de histograma sobre la imagen anterior.

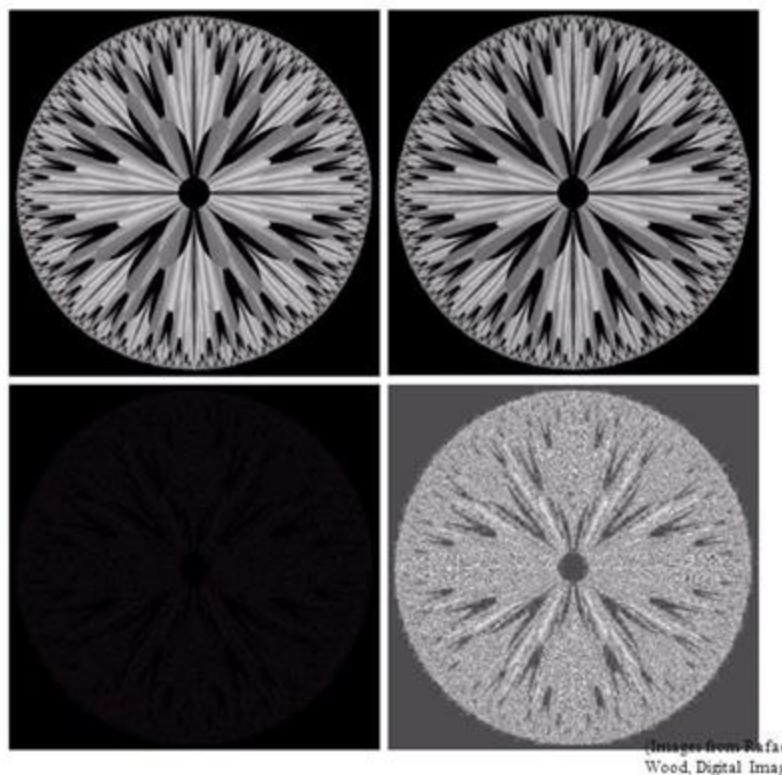


Figura 10. Utilización del operador resta entre dos imágenes para el realce de diferencias entre ambas.
Fuente: González y Woods, 2008.

Una de las aplicaciones prácticas más comunes de la sustracción de imágenes puede encontrarse en el **ámbito médico**. Partiendo de una imagen de referencia de un tejido vascular, en el que puede apreciarse una determinada sustancia, se toma a continuación una secuencia de imágenes de este tejido. Calculando la diferencia entre cada una de estas imágenes y la imagen de referencia, puede apreciarse el movimiento de la sustancia de interés, obteniéndose una especie de película a partir de las sucesivas diferencias.

Suavizado de imagen mediante el operador suma

De forma similar a la resta, el operador suma aplicado a dos o más imágenes consiste en la adición de los valores de intensidad de los píxeles correspondientes en cada una de ellas. El interés de este operador radica en su utilización para la **obtención de la imagen promedio** a partir de un conjunto, lo que permite reducir el ruido de captación.

Así, para un conjunto de M imágenes $A_i(x, y)$ ($i = 1, \dots, M$), el promedio de estas se obtiene de la siguiente forma:

$$C(x, y) = 1/M \sum_{i=1}^M A_i(x, y)$$

Cada una de las instantáneas $A_i(x, y)$ puede verse como la realización de un **proceso estocástico espacial**, una variable aleatoria cuyas observaciones suponen imágenes. Por tanto, estas instantáneas serían expresadas matemáticamente de la siguiente forma:

$$A_i(x, y) = F(x, y) + \eta(x, y)$$

Donde:

- $F(x, y)$ representa la imagen ideal libre de ruido.

- ▶ El proceso estocástico $\eta(x, y)$ se asume estacionario y gaussiano, con media nula y varianza σ^2 .

Si el ruido aditivo $\eta(x, y)$ se asume incorrelado entre dos instantáneas cualesquiera de nuestro conjunto, puede demostrarse que el promedio de todas ellas tiende a la imagen $F(x, y)$:

$$\lim_{M \rightarrow \infty} C(x, y) = F(x, y)$$

La potencia del ruido de captación se atenúa de acuerdo a un factor M en la imagen promedio resultante:

$$\sigma_C^2 = \sigma^2/M$$

Por último, reseñar que esta estrategia de realce basada en la atenuación del ruido de captación asume que las instantáneas de partida se encuentran perfectamente alineadas. De no ser así, el resultado final mostraría una clara difuminación de los bordes de las estructuras en la imagen.

En el siguiente ejemplo (figura 11) se muestra en la parte izquierda la imagen libre de ruido, en el centro una de las instantáneas capturadas afectada por ruido aditivo gaussiano y, en la derecha, el resultado de aplicar el promediado sobre un conjunto de instantáneas ruidosas.



Figura 11. Ejemplo de utilización del promediado de imágenes para la atenuación del ruido de captación.
Fuente: González y Woods, 2008.

6.5. Referencias bibliográficas

González, R. C. y Woods, R. E. (2008). *Digital image processing*. New Jersey: Pearson Education.

Percepción Computacional

Procesamiento de imagen. Operaciones espaciales

7.1. ¿Cómo estudiar este tema?

Para estudiar este tema deberás leer con atención las ideas clave que se desarrollan a continuación.

En el tema anterior se describieron un conjunto de técnicas de realce de imagen. Estas persiguen ensalzar las estructuras presente en la imagen, de forma que, en etapas posteriores, pueda obtenerse información de interés. Los operadores analizados fueron calificados como elementales, dada la naturaleza de las operaciones que, en ese caso, correspondía a funciones de transformación punto a punto.

Operadores espaciales

En el presente tema se aborda la utilización de operadores más complejos para el realce de imagen. Concretamente, se estudiarán diferentes operaciones espaciales. Para este tipo de operadores, el valor del píxel resultante es **función del píxel** correspondiente en la imagen original **y de la vecindad de este**. Además, se emplea una subimagen de tamaño similar a esta vecindad y compuesta por elementos denominados coeficientes que se conoce como **máscara**. Así, el operador puede expresarse de la siguiente forma:

$$v = T(u, u_1, u_2, \dots, u_{P^2-1}; \lambda, \lambda_1, \lambda_2, \dots, \lambda_{P^2-1})$$

Donde:

- ▶ v denota al valor de intensidad del píxel resultante.
- ▶ u es el píxel correspondiente en la imagen original.

- ▶ $u_i (i = 1, \dots, P^2 - 1)$ son los píxeles vecinos ($P^2 - 1$) para una vecindad de tamaño $P \times P$ centrada en u .
- ▶ $\lambda, \lambda_i (i = 1, \dots, P^2 - 1)$ son los coeficientes de la máscara que define el operador espacial.

Si el operador es de **tipo convolutivo**, el valor del píxel resultante será una función lineal de los valores de intensidad de los píxeles de entrada, expresándose el operador T de la siguiente forma:

$$v = \lambda u + \sum_{i=1}^{P^2-1} \lambda_i u_i$$

Como se ha indicado, el operador espacial queda definido por los coeficientes λ_i de la máscara. Por tanto, es común proporcionar una descripción del operador mediante la representación de esta máscara.

Por ejemplo, para un operador espacial 3×3 , la máscara tendría el aspecto mostrado en la siguiente figura:

λ_1	λ_2	λ_3
λ_4	λ	λ_5
λ_6	λ_7	λ_8

Figura 1. Máscara de un operador espacial de dimensión 3×3 .

Típicamente, se toma P impar, de forma que la máscara se centra en el píxel bajo estudio. Mediante esta representación, puede obtenerse una interpretación del operador a través de una inspección visual de su máscara.

En los siguientes epígrafes de este tema se presentan algunas implementaciones de filtros espaciales comúnmente empleados en las etapas de preprocesado de una imagen. La figura 2 recoge los principales conceptos expuestos en el tema:

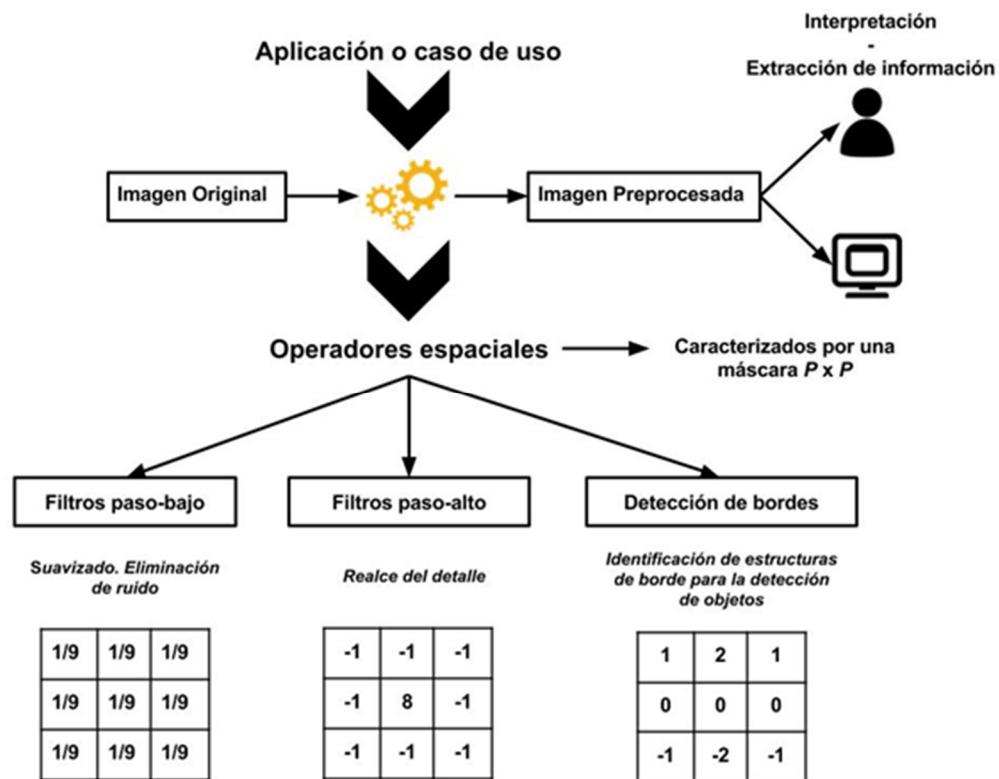


Figura 2. Esquema de los conceptos y descripción de los principales operadores espaciales sobre imágenes.

El **concepto de filtro** tiene su origen en las funciones de transferencia de los sistemas diseñados íntegramente en el dominio de la frecuencia. Sin embargo, en este tema, se emplea el término filtro para referirnos a las operaciones llevadas a cabo directamente sobre los píxeles de la imagen. Así, el término filtrado espacial se emplea para denominar a este tipo de operaciones frente a los filtros diseñados y aplicados en el dominio frecuencial.

En este punto, cabe recordar el filtro de mediana descrito en anteriormente. Se trata, al igual que los operadores descritos en este tema, de un filtro espacial, pues su respuesta se calcula directamente sobre el píxel de interés y su vecindad. Sin embargo, el filtro de mediana supone una dependencia no lineal de la respuesta y el conjunto de píxeles de entrada.

Una cuestión a considerar cuando se implementan filtros espaciales es la estrategia a seguir para aquellos píxeles cercanos al borde de la imagen. Cuando la máscara se centra en uno de estos píxeles, parte de ella queda fuera de la imagen. Concretamente, para aquellos **píxeles alejados del borde** una distancia inferior a $(P - 1)/2$ píxeles, parte de la máscara quedará fuera de la imagen. En este caso, existen diferentes estrategias para manejar esta situación:

- ▶ La más simple de todas ellas es aplicar la máscara solo a aquellos píxeles que estén situados a una **distancia mayor o igual a $(P - 1)/2$ píxeles** del borde.
 - La principal limitación de esta estrategia radica en que la imagen de salida será de tamaño menor que la imagen original.
- ▶ Otra opción es **calcular la respuesta del filtro** para aquellos puntos cercanos al borde, teniendo en cuenta únicamente los píxeles de la imagen que caen dentro de la máscara. Esta opción es tomada comúnmente cuando se requiere que la imagen de salida conserve el tamaño de la imagen original.
- ▶ Por último, cabe la posibilidad de **añadir filas/columnas a la imagen** con un valor determinado. Habitualmente, estas filas/columnas añadidas contienen píxeles con valor 0, 1 o el promedio de la intensidad de la imagen. También es posible replicar la última fila/columna de la imagen a fin de añadir estos nuevos píxeles.
 - La limitación fundamental de esta estrategia es que la respuesta del filtro en los píxeles cercanos al borde está influenciada por el valor de intensidad de los píxeles añadidos de forma artificial.

7.2. Filtros paso bajo y paso alto

A continuación se exponen las características de los filtros espaciales empleados habitualmente para la selección de frecuencias bajas y altas.

Tal y como se ha mencionado, estos filtros se definen y aplican sobre los píxeles originales de la imagen y, a pesar de su nombre, no hacen referencia a filtros diseñados y aplicados en el dominio de la frecuencia.

Filtros espaciales paso bajo (suavizado)

El efecto del filtro paso bajo se obtiene mediante el **cálculo del promedio** a partir del conjunto de píxeles definidos por la máscara, por tanto, los coeficientes de esta serán mayores que cero. Además, a fin de preservar la componente continua de la señal original, estos coeficientes son normalizados de forma que la suma de todos ellos sea la unidad.

La idea detrás de estos filtros de promediado es simple. Si se reemplaza el valor de intensidad de un píxel por el promedio de los píxeles vecinos, **se atenúan las variaciones bruscas de intensidad**. Por definición, un cambio significativo de intensidad en píxeles cercanos está asociado a una componente frecuencial alta que, de esta manera, sería eliminada por el filtro. Este suele ser el caso del ruido de captación, que representa una componente aleatoria de la señal de entrada caracterizada por transiciones bruscas entre píxeles contiguos.

En el caso de no contar con un conjunto de instantáneas de la misma escena para, tal y como se describió en el tema anterior, obtener el promedio de todas ellas y reducir el ruido, la aplicación de un **filtro espacial de promediado** es una solución plausible a fin de mejorar la calidad de la imagen y ensalzar sus estructuras. La principal desventaja de este filtro es el efecto de suavizado o desenfoque sobre algunas de las estructuras caracterizadas por componentes frecuenciales elevadas como, por ejemplo, los bordes entre objetos. La figura 3 muestra las máscaras correspondientes a sendos filtros espaciales de promediado (paso bajo).

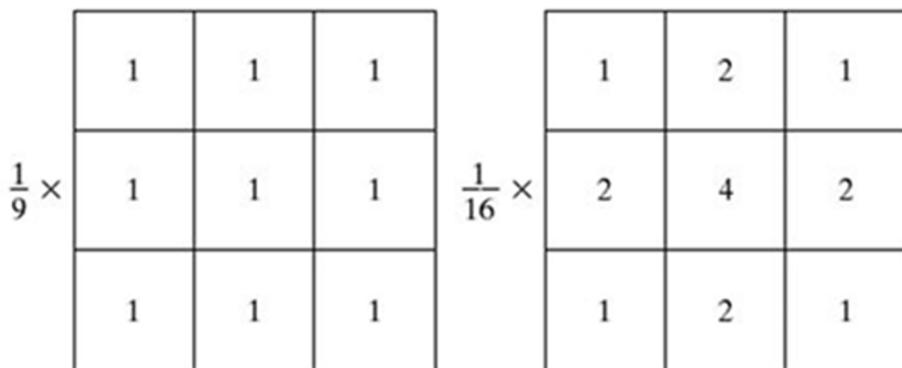


Figura 3. Máscaras correspondientes a dos filtros espaciales de promediado.

La máscara de la izquierda posee el mismo valor para todos sus coeficientes, mientras que la máscara de la derecha, al tener coeficientes de diferente valor, lleva a cabo el **cálculo del promedio ponderado**.

La máscara de la izquierda refleja un filtro en el que se lleva a cabo el promediado simple de los píxeles de la vecindad, pues todos los coeficientes tienen el mismo valor.

Un filtro de promediado en el que todos sus coeficientes son iguales se denomina *box filter*.

En el caso de la máscara de la derecha, como puede observarse, hay coeficientes que tienen más peso que otros. En este caso, el promediado está ponderado, dando mayor relevancia a unos píxeles frente a otros. Concretamente, en el ejemplo de la figura, el píxel central es el de mayor peso, seguido de los píxeles contiguos.

Los filtros mostrados en el ejemplo tienen un tamaño 3×3 . Sin embargo, puede considerarse cualquier **tamaño del filtro**. Este factor influye, a su vez, en el resultado del filtrado y debe escogerse de acuerdo al nivel de ruido presente en la imagen original y al nivel de calidad exigido por el escenario práctico en el que nos encontramos. Por lo general, un filtro de mayor dimensión tenderá a eliminar el ruido de forma más eficiente, pero a costa de producir una imagen más desenfocada, ya que el promediado se computa a partir de un mayor número de píxeles.

Filtros espaciales paso-alto (afilado)

El principal objetivo de los filtros espaciales para el afilado de la imagen es realzar los detalles más finos de la imagen o, también, corregir el desenfoque producido como consecuencia del proceso de adquisición.

En el epígrafe anterior, se describían los filtros espaciales enfocados en el suavizado de la imagen para la supresión de ruido. Las máscaras que caracterizan estos filtros implementan una operación de promediado que, matemáticamente, se corresponde con la operación de integración o suma de elementos. Los filtros de afilado o ensalzado del detalle persiguen un objetivo opuesto. Es decir, tratan de **magnificar las variaciones o contrastes** entre estructuras de la imagen. Por tanto, las máscaras correspondientes a estos filtros incluirán coeficientes tales que lleven a cabo cálculos de diferenciación. El **cálculo de diferencias** es necesario para la obtención de la derivada de una función. A su vez, la derivada representa la operación inversa a la integración desde un punto de vista matemático.

La derivada de una función se hace más pronunciada en aquellos puntos en los que hay cambios más significativos.

Así, los operadores espaciales paso alto que aproximan la derivada de la función imagen en las coordenadas (x, y) ensalzan los bordes y otras discontinuidades (por ejemplo, el ruido), atenuando aquellas áreas de la imagen en las que la variación es poco relevante o nula. En definitiva, actúan como un filtro de característica paso alto que retiene las componentes frecuenciales más elevadas de la imagen.

Los filtros espaciales paso alto se basan en la aproximación de las derivadas de primer y segundo orden. Nótese que estas operaciones matemáticas se llevan a cabo sobre funciones discretas, como son las imágenes. Los valores de intensidad que pueden observarse en estas son finitos y, por tanto, las variaciones entre píxeles así lo serán. Además, la menor distancia entre dos puntos de la función entre los que cuantificar una variación vendrá dada por dos píxeles adyacentes.

Para la definición de la aproximación de las derivadas de primer y segundo orden en un contexto digital como el descrito, emplearemos por simplicidad **señales unidimensionales**; el procedimiento es extensible a señales bidimensionales sin pérdida alguna de generalidad.

La **derivada de primer orden** de una señal discreta $f[n]$ puede estimarse como:

$$\frac{\partial f}{\partial n} = f[n + 1] - f[n]$$

De forma similar, la **derivada de segundo orden** puede computarse de la siguiente forma:

$$\frac{\partial^2 f}{\partial^2 n} = f[n + 1] + f[n - 1] - 2f[n]$$

Para definir las propiedades de ambas operaciones, podemos observar la siguiente imagen donde se muestra el cálculo de ambas derivadas sobre una secuencia numérica que representa diferentes tipos de discontinuidades.

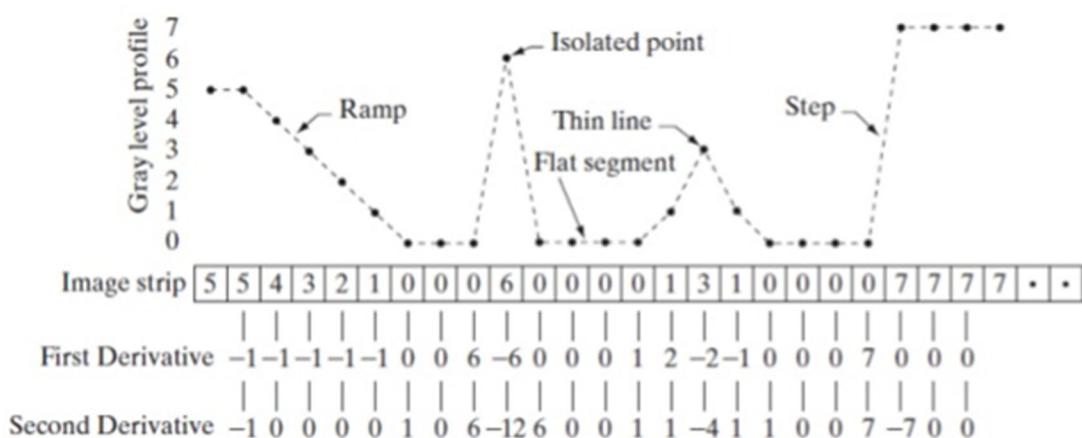


Figura 4. Estimación de derivadas de primer y segundo orden sobre una secuencia numérica. Análisis de la capacidad de estos operadores para la identificación de diferentes tipos de discontinuidades.

Fuente: González y Woods, 2008.

En primer lugar, se aprecia como la primera derivada es distinta de cero durante toda la rampa inicial de la secuencia, mientras que la segunda derivada solo toma valores distintos de cero durante el comienzo y final de la rampa. Dado que los **bordes** en una imagen vienen dados por este tipo de perfil en forma de rampa, se puede extraer que la primera derivada produce bordes gruesos, mientras que los identificados por la segunda derivada son mucho más finos.

Posteriormente, se aprecia un **punto aislado en la secuencia**. Para este, la respuesta es mucho más significativa en el caso de la segunda derivada. Esto hecho no es sorprendente, ya que la derivada de segundo orden magnifica los cambios de forma más pronunciada que la derivada de primer orden. Por tanto, este será el resultado con este tipo de puntos aislados y, por consiguiente, con elementos espurios como el ruido.

La línea delgada que encontramos a continuación en nuestra secuencia produce un resultado similar. La respuesta de la segunda derivada sería aún más pronunciada si esta línea fina alcanzara el nivel de intensidad del punto aislado que se ha visto previamente. Finalmente, la secuencia muestra una discontinuidad brusca que es capturada de forma muy parecida por ambos operadores de derivada.

A continuación, exponemos un resumen de las **conclusiones** extraídas de la comparación del resultado producido por las derivadas de primer y segundo orden.

COMPARACIÓN ENTRE DERIVADAS DE PRIMER Y SEGUNDO ORDEN	
DERIVADA DE PRIMER ORDEN	DERIVADA DE SEGUNDO ORDEN
Produce bordes más gruesos en la imagen resultante.	Tiene una respuesta más pronunciada a detalles más finos como líneas delgadas o puntos aislados.
Presenta una respuesta más acusada a cambios con forma de escalón en el nivel de intensidad de la imagen.	Tiende a producir una doble respuesta en cambios de tipo escalón. Además, se aprecia que su respuesta es más pronunciada en puntos que en líneas delgadas, y en estas que en un escalón.

Figura 5. Conclusiones de la comparación entre derivadas de primer y segundo orden.

De acuerdo a las propiedades descritas, la siguiente máscara (figura 6) corresponde a un **filtro espacial paso alto** basado en el cálculo de la derivada de primer orden. En este caso, el filtro está diseñado para la detección de bordes en posición vertical, tal y como reflejan los coeficientes de la máscara. Este filtro es equivalente al cómputo del **operador gradiente** en la dirección del eje horizontal de la imagen.

1	0	-1
1	0	-1
1	0	-1

Figura 6. Ejemplo de un filtro espacial paso alto basado en el cómputo de la derivada de primer orden (gradiente) de la señal.

En el caso de un filtro de segundo orden, este vendría dado por una máscara similar a la mostrada en la figura 7. El operador equivalente a este filtro es el **laplaciano** de la señal. En este caso, el filtro es de carácter isotrópico, pues es capaz de identificar discontinuidades en cualquiera de las direcciones de la imagen.

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

Figura 7. Ejemplo de un filtro espacial paso alto basado en el cómputo de la derivada de segundo orden (laplaciano) de la señal.

Finalmente, la siguiente imagen (figura 8) ilustra el efecto de aplicar el operador laplaciano sobre la imagen de la izquierda. Como se puede ver, la imagen resultante (derecha) contiene los bordes de las estructuras identificadas en la imagen original.

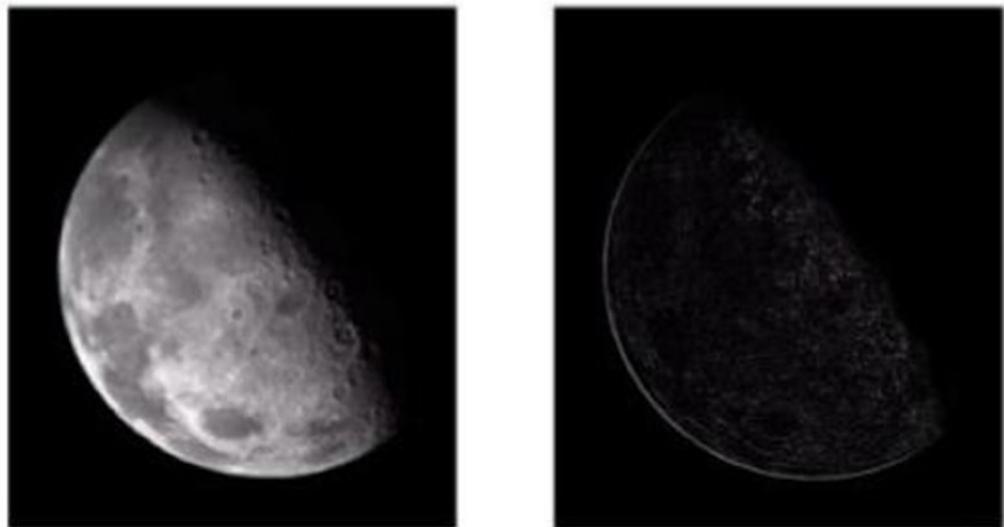


Figura 8. Aplicación del operador laplaciano para el realce de detalles en una imagen.
Fuente: González y Woods, 2008.

7.3. Detección de bordes

Ya se ha visto cómo pueden ser identificados los bordes de las estructuras presentes en una imagen mediante la aplicación de filtros espaciales paso alto. En este punto, se profundiza en el conocimiento de operadores destinados a la detección de bordes. Esta denominación se emplea para identificar a los operadores basados en el **cálculo del gradiente**. Como se ha visto, se corresponden con los filtros espaciales que aproximan el cómputo de la derivada de primer orden en direcciones determinadas de la imagen.

A continuación se detallan los filtros espaciales más comunes para la detección de estructuras de borde en una imagen.

Operadores de Roberts

Estos operadores se caracterizan por tener un tamaño de máscara 2×2 , que supone una novedad con las máscaras vistas hasta el momento. Los coeficientes de las máscaras se muestran en la siguiente figura, indicándose en negrita el píxel sobre el que se calcula la operación.

1	0
0	-1

0	1
-1	0

Figura 9. Máscaras que definen los operadores de Roberts para la detección de bordes.

La máscara de la izquierda aproxima el resultado del operador gradiente en la dirección diagonal dada por el ángulo 135° . En el caso de la máscara de la derecha, como puede deducirse por sus coeficientes, lleva a cabo el cálculo del gradiente en la diagonal de 45° .

Operadores de Prewitt

Estos operadores vienen definidos por máscaras de tamaño 3×3 que definen el cálculo del operador gradiente tanto en la dirección vertical como en la dirección horizontal de la imagen. Estas serían las máscaras correspondientes a este operador.

1	1	1
0	0	0
-1	-1	-1

-1	0	1
-1	0	1
-1	0	1

Figura 10. Máscaras correspondientes a los operadores de Prewitt para la identificación de bordes con orientación horizontal y vertical.

Operadores de Sobel

Los operadores de Sobel son similares a los de Prewitt. La diferencia estriba en que los primeros dan más peso a los píxeles cercanos al píxel de interés (centro de la máscara) para el cálculo de diferencias.

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

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

Figura 11. Máscaras correspondientes a los operadores de Sobel para la identificación de bordes con orientación horizontal y vertical.

Algoritmo de Canny

Por último, se incluye en este epígrafe del tema uno de las transformaciones más ampliamente utilizadas y efectivas para la detección de bordes como es el filtro de Canny.

Este operador se basa en los siguientes **criterios**:

- ▶ El filtro debe identificar el mayor número de bordes de entre todos los que hay en la imagen. Es decir, el operador, a fin de funcionar correctamente, debe minimizar el número de falsos negativos y falsos positivos en la identificación de los bordes.
- ▶ El punto identificado como borde por parte del operador debe corresponderse con el centro de la estructura de borde identificada en la imagen.
- ▶ Un borde de la imagen debe ser identificado una única vez e, idealmente, el ruido presente en la imagen no debe dar lugar a falsos bordes.

Así, los **pasos** que sigue el algoritmo de detección de bordes de Canny son los siguientes:

- ▶ Se aplica, en primer lugar, un filtro paso bajo de tipo gaussiano que suaviza la imagen y reduce el posible ruido en la misma.
- ▶ Se calcula el gradiente de la imagen en las diferentes direcciones.
- ▶ Se aplica una transformación sobre los píxeles no máximos que consiste en poner a 0 el valor de intensidad de los píxeles que no forman parte de un máximo local.
- ▶ Se aplica un umbral sobre la intensidad de los píxeles para binarizar la imagen y quedarnos con las estructuras que potencialmente pueden corresponderse con bordes.
- ▶ Se eliminan aquellas estructuras más débiles que no están conectadas con bordes mayores.

A continuación mostramos un ejemplo en el que se aplica el operador de Canny para la detección de los bordes en una imagen.

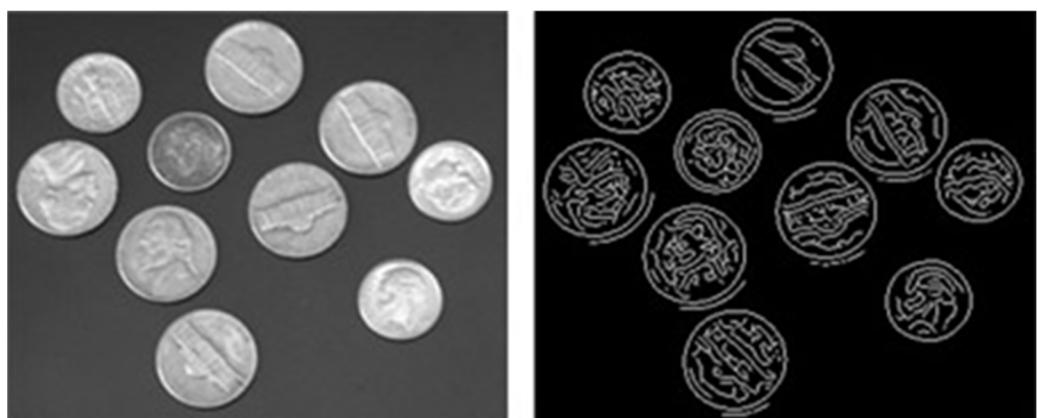


Figura 12. Resultado de la aplicación del operador de Canny sobre una imagen.

Fuente: <https://www.mathworks.com/>.

7.4. Referencias bibliográficas

González, R. C. y Woods, R. E. (2008). *Digital image processing*. New Jersey: Pearson Education.