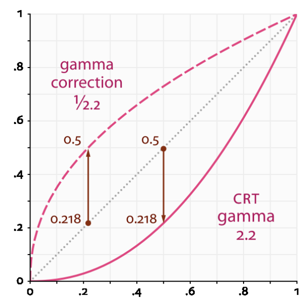
Video en multimedia

- Tiempo de respuesta: tiempo que tarda un elemento LCD en ir desde su mínima transparencia a la máxima, y otra vez a la mínima. Es importante para la reproducción de video. Suele estar entre 20 y 35 ms.
- Espacio de color: cuántos (y qué) colores diferentes puede reproducir.
- Tasa de refresco: número de veces por segundo que puede cambiar la imagen. Está limitada por el tiempo de respuesta, y suele ir entre 60 y 75 Hz.
- Consumo eneregético: energía consumida por unidad de tiempo, en Watios.

En general, la retroiluminación LED mejora el contraste, amplía el espacio de color y disminuye el consumo.

Respecto a la linealidad del monitor (intensidad luminosa más o menos lineal con el valor de la señal que se le envía), la curva de respuesta suele ser de la forma $I = A \cdot s^\gamma$ donde s es el valor de la señal, normalizado en $[0..1]$, y γ es una constante, que para los monitores CRT suele estar entre 2.35 y 2.55 y para los LCD sobre 2.2.



Digitalización de señales de audio y video

Una **señal** es una magnitud física variable en el tiempo y/o en el espacio. En el caso que nos ocupa, las variables físicas de interés son la presión del aire para el sonido, y la luminancia y cada componente del color para la imagen. Como vimos, en ambos casos éstas se convertían a señal eléctrica mediante micrófonos y cámaras de vídeo.

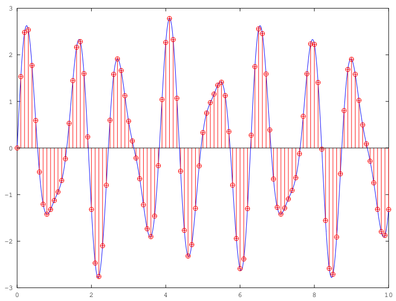
Una señal **analógica** es aquella que toma valores en todos los instantes del tiempo, y que puede tomar cualquier valor comprendido en un intervalo conocido (p. ej., 0 a 5 Voltios).

Una señal **digital** es aquella que toma valores sólo en ciertos instantes del tiempo, llamados instantes de muestreo, normalmente separados entre sí por un tiempo constante (el periodo de muestreo, T) y además el valor es uno de entre un conjunto discreto conocido (p. ej. los enteros entre 0 y 255).

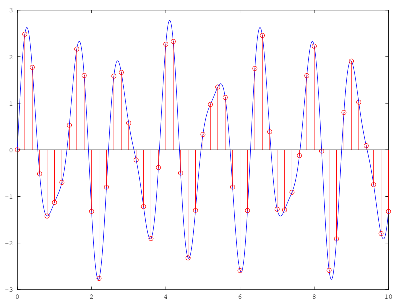
Digitalización es el proceso por el cual una señal analógica es convertida a una señal digital perdiendo información. Para ello son necesarios dos procesos: el **muestreo** y la **cuantización**.

Digitalización de señales de audio y video

Ejemplos de muestreo de una señal unidimensional

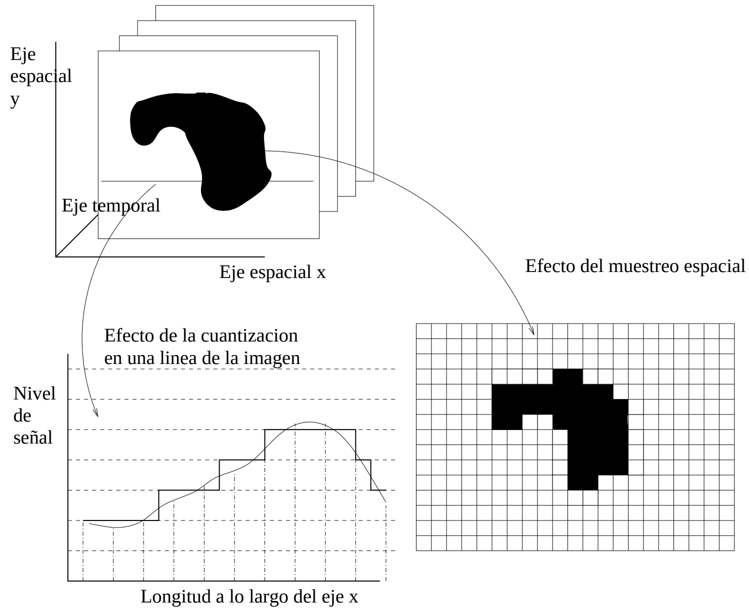


Periodo apropiado



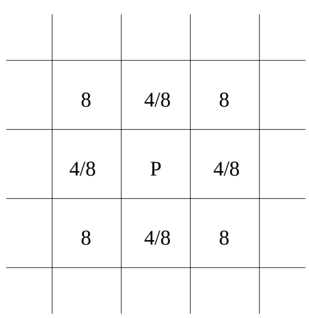
Periodo inapropiado

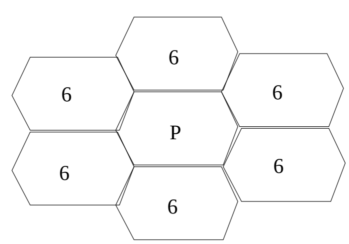
Acción del muestreo y la cuantización sobre una secuencia de imágenes:



Digitalización de señales de audio y video

El concepto de muestreo lleva a los de pixel y voxel, que son las unidades discretas de información en imágenes 2D ó 3D respectivamente.
La forma de muestrear genera la trama, que puede ser rectangular o hexagonal.

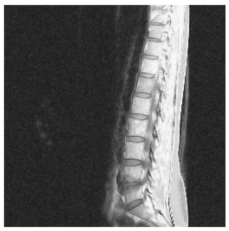
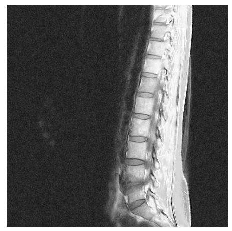
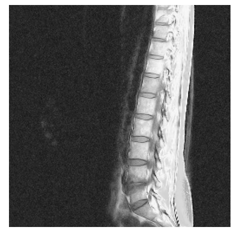
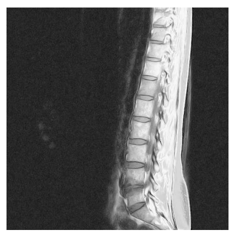
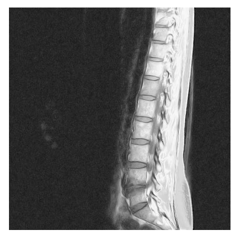




La conectividad es el número y disposición espacial de los pixels que se consideran vecinos de uno dado. En trama rectangular suele usarse conectividad a 4 ó a 8. En trama hexagonal, conectividad a 6.

Digitalización de señales de audio y video

Diferencias al usar distintos valores de muestreo y cuantización:

	6 bpp	8 bpp	12 bpp
128×128			
256×256			
512×512			

Digitalización de señales de audio y video

En audio, la frecuencia de muestreo más común es 44100 Hz (se verá el motivo cuando se entiendan la transformada de Fourier y el teorema del muestreo). La cuantización se suele realizar a 16 bits, es decir, las muestras resultan ser enteros en el rango $[0..65535]$; esto es la calidad de los CD.

Es posible también muestrear a frecuencias desde 8192 Hz y cuantizar a 8, 10 ó 12 bits, por supuesto perdiendo calidad. En el otro extremo, el DVD-Audio muestrea hasta 192000 Hz y cuantiza hasta 24 bits. El formato de audio más simple (wav) consiste sólo en una cabecera indicando periodo de muestreo, nivel de cuantización, número de canales (en wav, sólo mono o estéreo), longitud en número de muestras y otros detalles, seguido de las propias muestras, en formato `unsigned`. Si se trata de dos canales se alternan las muestras de ambos.

Un CD contiene simplemente las muestras, a 44100 Hz, 16 bits, con una duración estándar de 74 minutos, lo cual da 746 Mbytes:
44100muestras/seg$\cdot 2$bytes/muestra$\cdot 2$ canales$\cdot 60$seg/min$\cdot 74$minutos=783216000 bytes=746 Mbytes

Transformadas integrales discretas

Una vez una señal ha sido digitalizada, se puede ver como un vector de N muestras. Incluso si se trata de una imagen de resolución $M \times N$, pueden concatenarse las filas (o las columnas) para generar un vector de $M \times N$ componentes.
Como sabemos, en todo espacio vectorial (E.V.) de dimensión finita N se puede encontrar una base, que se define como cualquier conjunto de N vectores, $B = \{\mathbf{x}_0, \ldots, \mathbf{x}_{N-1}\}$ tal que:

- Sean linealmente independientes, es decir, si $\alpha_0 \cdot \mathbf{x}_0 + \ldots + \alpha_{N-1} \cdot \mathbf{x}_{N-1} = \mathbf{0}$, entonces tienen que ser $\alpha_i = 0, \forall i = 0 \ldots N-1$, siendo $\mathbf{0}$ el vector nulo.
- Sean generadores de todo vector del E.V., es decir, $\forall \mathbf{y}, \exists \beta_0 \ldots \beta_{N-1}$ tal que $\mathbf{y} = \beta_0 \cdot \mathbf{x}_0 + \ldots + \beta_{N-1} \cdot \mathbf{x}_{N-1}$

La **base canónica** es el conjunto de vectores $\mathbf{x}_j = [0, .., 0, 1, 0, ..., 0]$ con el 1 en la posición j-ésima, para cada $j = 0..N-1$. La descomposición trivial de un vector es aquella que lo expresa en la base canónica.
Una **transformada integral discreta** no es más que la descomposición de un vector conteniendo las muestras de una señal en una cierta base, distinta de la base canónica.

Transformadas integrales discretas

La **transformada discreta de Fourier** (DFT) de un vector **y** de N componentes en 1D es la descomposición en la base cuyos vectores básicos (con componentes complejas) son

$$\mathbf{x}_k = \frac{1}{\sqrt{N}} \left(e^{\frac{2\pi \cdot 0 \cdot k}{N} j}, e^{\frac{2\pi \cdot 1 \cdot k}{N} j}, \ldots, e^{\frac{2\pi \cdot (N-1) \cdot k}{N} j} \right) \quad k = 0..(N-1) \text{m es decir, } N \text{ vectores c}$$

siendo j la unidad imaginaria, $j = \sqrt{-1}$. Así, la descomposición de un vector cualquiera **y** sería

$$\mathbf{y} = \sum_{k=0}^{N-1} c_k \cdot \mathbf{x}_k$$

y al vector de números complejos $\mathbf{Y} = \{c_0, \ldots, c_{N-1}\}$ formado por los coeficientes de dicha descomposición lo llamamos la **transformada de Fourier discreta** de $\mathbf{y}$. Cada una de las c_k se puede hallar como

$$c_k = \frac{1}{\sqrt{N}} \sum_{l=0}^{N-1} y_l e^{-\frac{2\pi j \cdot kl}{N}} = \frac{1}{\sqrt{N}} \sum_{l=0}^{N-1} y_l \left(\cos \frac{2\pi \cdot kl}{N} - j \cdot \sin \frac{2\pi \cdot kl}{N} \right)$$

Transformadas integrales discretas

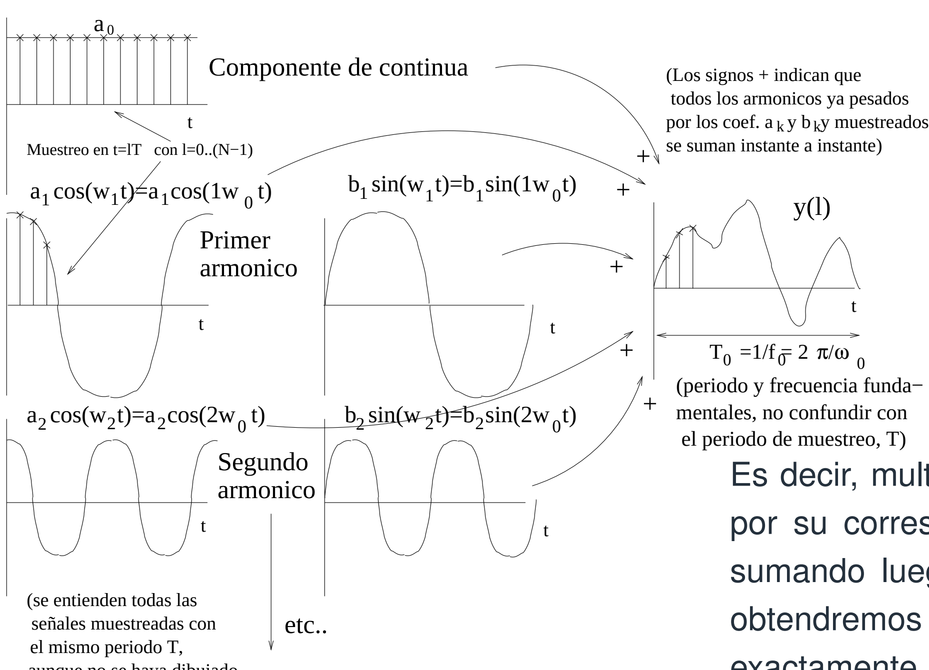
Los números complejos c_k tendrán partes reales e imaginarias, que llamaremos $c_k = a_k - b_k \cdot j$.

Como todas las transformadas integrales útiles, la DFT es invertible, lo que significa que se pueden recuperar las componentes de $\mathbf{y}$ a partir de las de $\mathbf{Y}$:

$$y_l = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} c_k e^{\frac{2\pi j \cdot kl}{N}} = \frac{1}{\sqrt{N}} \left(a_0 + \sum_{k=1}^{N-1} \left(a_k \cdot \cos \frac{2\pi kl}{N} + b_k \cdot \sin \frac{2\pi kl}{N} \right) \right)$$

Es muy importante observar que, debido a la invertibilidad, la **misma** información contenida en el vector de muestras original, $\mathbf{y}$, está igualmente contenida en $\mathbf{Y}$, aunque expresada de distinta manera. ¿De qué manera?. En la expresión de las componentes y_l podemos observar que son una suma de funciones seno y coseno tales como $\sin(\omega_k t)$, muestreadas en $t = l$, y para las que la frecuencia angular ω_k vale $\omega_k \equiv \frac{2\pi k}{N} \equiv k \cdot \omega_0$ donde a ω_0 se le llama **frecuencia fundamental** y vale $\omega_0 = \frac{2\pi}{N}$. Las otras son **múltiplos enteros** de ésta, y se llaman **armónicos**. El armónico de frecuencia 0 es $\cos(0l) = 1$, una constante, y se llama **término de contínua**.

Transformadas integrales discretas



Es decir, multiplicando cada armónico por su correspondiente coeficiente, y sumando luego todos los resultados, obtendremos $\mathbf{y}$. Los armónicos son exactamente N (contando al término de contínua) y el de más alta frecuencia es el de $\omega_{N-1} = \frac{2\pi(N-1)}{N}$.

Transformadas integrales discretas

Esto significa que a_k y b_k (o si usamos los coeficientes complejos, $\mid c_k \mid$) dan el "peso" o importancia relativa que la onda de frecuencia ω_k (el armónico k-ésimo) tiene para la construcción de la onda total. Para medir esto se define la **potencia** del armónico k-ésimo como

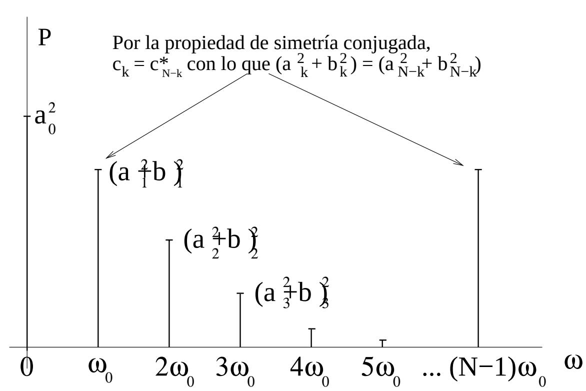
$$P_k = \mid c_k \mid^2 = (a_k^2 + b_k^2) \quad \text{si } k \neq 0 \text{ y } \; P_0 = a_0^2 \quad \text{para } k = 0$$

La potencia total de una señal periódica se define como la suma de las potencias de cada uno de los armónicos que la componen, es decir

$$P = a_0^2 + \sum_{k=1}^{N-1} \mid c_k \mid^2 = a_0^2 + \sum_{k=1}^{N-1} (a_k^2 + b_k^2)$$

Transformadas integrales discretas

Se puede representar esto en un diagrama llamado espectro de potencias, en el que cada barra representa la potencia que la onda de frecuencia ω_k aporta a la señal periódica.



Una observación importante: para reconstruir $\mathbf{y}$ <u>no</u> hemos necesitado todas las frecuencias posibles, sino sólo los armónicos de la frecuencia fundamental. P. ej., si el periodo de la onda original hubiese sido $T_0 = 10$ seg., entonces $\omega_0 = \frac{2\pi}{10}$ rad/s., y en la reconstrucción de $\mathbf{y}$ sólo intervendrán armónicos de frecuencias $\frac{2\pi}{10}, \frac{4\pi}{10}, \frac{6\pi}{10}, \frac{8\pi}{10}, \ldots$ pero no, p. ej., una onda de frecuencia $\frac{3\pi}{10}$.

Transformadas integrales discretas

Algunas propiedades de la DFT (siendo $\mathbf{X} = DFT(\mathbf{x})$ e $\mathbf{Y} = DFT(\mathbf{y})$):

- Linealidad: $DTF(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha\mathbf{X} + \beta\mathbf{Y}, \forall\alpha, \beta \in R$.
- Simetría conjugada: $X(k) = X^*(N-k)$, siendo N el tamaño de la señal $\mathbf{X}$.
- Convolución circular: Se define la convolución circular de dos señales de longitud N, $\mathbf{x}$ e $\mathbf{y}$ como

$$(\mathbf{x} * \mathbf{y})(k) = \sum_{l=0}^{N-1} x_l \cdot y_{[(l-k)\,\%N]} = \sum_{l=0}^{N-1} x_{[(l-k)\,\%N]} \cdot y_l$$

siendo $a\,\%b$ el módulo (resto de la división entera de a entre b). Entonces,

$$DFT(\mathbf{x} * \mathbf{y})_k = DFT(\mathbf{x})_k \cdot DFT(\mathbf{y})_k = \mathbf{X_k} \cdot \mathbf{Y_k}$$

(Producto de las señales elemento a elemento).

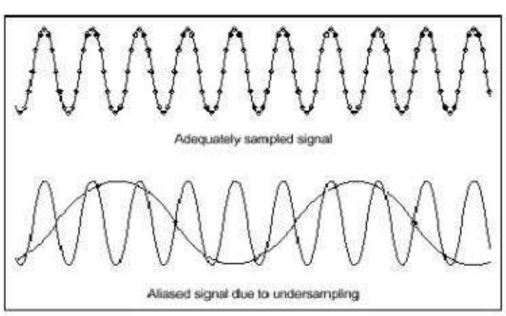
- Identidad de Parseval:

$$\sum_{k=0}^{N-1} x_k^2 = \sum_{k=0}^{N-1} \mid X_k \mid^2 \quad \text{(la potencia total se conserva)}$$

Transformadas integrales discretas

Para elegir la frecuencia de muestreo apropiada para cada señal usamos el **teorema del muestreo** o de Shannon. Para entenderlo definimos:
Una señal de tiempo contínuo (antes de la digitalización) $y(t)$ es de **banda limitada** si existe una frecuencia angular $\omega_0 = 2\pi f_0$ tal que
$\mid Y(\omega) \mid = 0, \forall\omega / \mid \omega \mid > \omega_0$, siendo $Y(\omega)$ la transformada de Fourier de $y(t)$.
Entonces,

Toda señal $y(t)$ de banda limitada es reconstruíble sin pérdidas a partir de las muestras $\{y_k\}$ tomadas con periodo T ($y_k = y(kT)$) si $T \leq \frac{\pi}{\omega_0}$ (o, equivalentemente, si $f_s = \frac{1}{T}$, debe ser $f_s \geq 2f_0$).



Transformadas integrales discretas

Las consecuencias del muestreo en señal unidimensional dependiente del tiempo (audio) son:

- El oído humano es sensible a frecuencias de hasta aprox. 20000 Hz
- Todo sonido es de banda limitada (ningún medio material puede vibrar a frecuencia infinita) aunque algunos sonidos tienen un ancho de banda mayor de 20000 Hz
- No obstante, si eliminamos de ellos las frecuencias superiores a 20000 Hz el oído humano no debería ser capaz de distinguirlos del sonido original
- Ello hace que se pueda reconstruir una onda sonora sin pérdidas perceptuales muestreando con frecuencia de al menos $2 \cdot 20000Hz = 40000Hz$. Por ello para el CD se escogieron 44100 Hz.

Transformadas integrales discretas

Transformadas integrales en dos dimensiones

Todo lo dicho para una dimensión es directamente aplicable a dos, si la imagen se aplana por filas (o por columnas). No obstante, hay una formulación alternativa que aclara la comprensión del significado físico. Dada una imagen $f(m,n)$ con $m = 1..(M-1)$ y $n = 0..(N-1)$, y en la que $f(m,n)$ representaría el nivel de gris en el píxel de la fila m, columna n, se define la transformada integral de f como otra imagen $F(u,v)$, también $M \times N$,

$$F(u,v) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} f(i,j) K(i,j,u,v)$$

done $K(i,j,u,v)$ es el llamado kernel de la transformada, una función real o compleja de cuatro variables. Nótese que se suma sobre las variables espaciales (i y j) de la función, que por tanto desaparecen, quedando la transformada como función de las nuevas variables u y v. La idea es buscar kernels tales que los valores de la transformada en el plano (u,v) tengan sentido físico y utilidad.

Transformadas integrales discretas

La fórmula anterior se suele expresar en forma matricial; si la imagen es la matriz f de $M \times N$ la transformada es otra imagen $M \times N$:

$$F \equiv T(f) = P \cdot f \cdot Q$$

donde P es una matriz $M \times M$ y Q una matriz $N \times N$. En forma explícita,

$$\begin{aligned} F(u,v) &= \sum_{m=0}^{M-1} P(u,m) \cdot \left(\sum_{n=0}^{N-1} f(m,n) Q(n,v) \right) = \\ &= \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} P(u,m) f(m,n) Q(n,v) \end{aligned}$$

En el caso de que las matrices cuadradas P y Q fuesen invertibles, la imagen original se obtiene a partir de la transformada con la transformación inversa:

$$f = P^{-1} \cdot F \cdot Q^{-1} \quad \text{sin pérdida de información}$$

Transformadas integrales discretas

Definiciones sobre matrices

Dada una matriz de números reales M, decimos que es:
- Simétrica: si $M = M^T$ siendo M^T la matriz traspuesta
- Ortogonal: si $M \cdot M^T = M^T \cdot M = I \Rightarrow M^T = M^{-1}$

Si es a la vez simétrica y ortogonal se cumple que $M = M^{-1}$

Dada una matriz de números complejos C, decimos que es:
- Hermítica: si $C = C^{T^*}$ siendo C^{T^*} la matriz traspuesta conjugada
- Unitaria: si $C \cdot C^{T^*} = C^{T^*} \cdot C = I \Rightarrow C^{T^*} = C^{-1}$

Si es a la vez hermítica y unitaria se cumple que $C = C^{-1}$

Si P y Q son reales, simétricas y ortogonales, o bien complejas, hermíticas y unitarias,

$$F = P \cdot f \cdot Q \quad \text{y} \quad f = P \cdot F \cdot Q$$

lo que hace que se pueda usar el mismo algoritmo y programa para la transformada directa y para la inversa. Se dice entonces que la transformada es ortogonal.

Transformadas integrales discretas

Entre las transformadas que se estudiarán sólo la de Walsh-Hadamard es ortogonal. No obstante, la de Fourier usa matrices complejas hermíticas, con lo que

$$F = P \cdot f \cdot Q \quad \text{y} \quad f = P^{T^*} \cdot F \cdot Q^{T^*}$$

y la transformada discreta del coseno usa matrices reales ortogonales, con lo que

$$F = P \cdot f \cdot Q \quad \text{y} \quad f = P^{T} \cdot F \cdot Q^{T}$$

Transformadas integrales discretas

Transformada Discreta de Fourier (DFT)

Sea Φ_{JJ} la matriz $J \times J$ cuyos elementos son

$$\Phi_{JJ}(k,l) = \frac{1}{\sqrt{J}} e^{-\frac{2\pi i}{J}kl}$$

Se define la DFT de una imagen f de tamaño $M \times N$ como la imagen F, también de tamaño $M \times N$ como

$$\mathcal{F}\left\{f(m,n)\right\} \equiv F = \Phi_{MM} \cdot f \cdot \Phi_{NN}$$

o sea,

$$F(u,v) = \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m,n) e^{-2\pi i\left(\frac{mu}{M} + \frac{nv}{N}\right)} \quad \text{con} \quad \begin{cases} 0 \leq u \leq M-1 \\ 0 \leq v \leq N-1 \end{cases}$$

Def.:- Espectro de frecuencias, o simplemente espectro: la propia imagen compleja $F(u,v)$

- Espectro de potencia: la imagen $\| \; F(u,v) \; \|^2$
- Espectro de fase: la imagen $\angle F(u,v)$

Implantación de la DFT

La matrices Φ_{JJ} cumplen que $\Phi_{JJ}^{T*} = \Phi_{JJ}^{-1}$. Por ello,

$$f(m,n) = \Phi_{MM}^{T*} \cdot F \cdot \Phi_{NN}^{T*} = \frac{1}{\sqrt{MN}} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u,v) e^{2\pi i \left(\frac{mu}{M} + \frac{nv}{N}\right)}$$

(la misma fórmula que para la transf. directa, excepto el signo del exponente).

Ahora, para programar la transformada (p. ej., directa) se usa

$$F(u,v) = \frac{1}{\sqrt{M}} \sum_{m=0}^{M-1} \left[\frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} e^{-\frac{2\pi i}{N} nv} \cdot f(m,n) \right] \cdot e^{-\frac{2\pi i}{M} mu}$$

El contenido del corchete es la DFT unidimensional de la fila m-ésima de la imagen. Una vez hecho esto para cada fila y almacenado en una matriz temporal T, se realiza la DFT unidimensional de cada columna de T. Se puede probar que cada transformada 1-D de un vector de p elementos tiene un coste $\mathcal{O}(p \log p)$ si p es potencia de 2. Así, el coste total de la DFT 2-D sería del orden de $\mathcal{O}(p^2 \log p)$ siendo $p = max(M,N)$.

Propiedades de la DFT

Llamando $F(u,v) = \mathcal{F}\{f(m,n)\}$ y $G(u,v) = \mathcal{F}\{g(m,n)\}$ se cumplen:

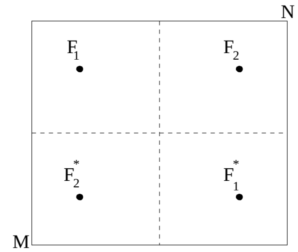
- Linealidad: $\mathcal{F}\{\alpha f(m,n) + \beta g(m,n)\} = \alpha F(u,v) + \beta G(u,v)$
- Similaridad: $\mathcal{F}\{f(am,bn)\} = \frac{1}{|ab|} F(\frac{u}{a}, \frac{v}{b})$
- Desplazamiento circular: $\mathcal{F}\{f(m-a, n-b)\} = e^{-2\pi j(au+bv)} F(u,v)$
- Convolución circular: $\mathcal{F}\{f(m,n) * g(m,n)\} = F(u,v) \cdot G(u,v)$
- Diferenciación: $\mathcal{F}\left\{\left(\frac{\partial^p}{\partial m^p}\right)\left(\frac{\partial^q}{\partial n^q}\right) f(m,n)\right\} = (2\pi j)^p (2\pi j)^q F(u,v)$

  En particular,
  - $\mathcal{F}\left\{\frac{\partial f}{\partial m}\right\} = 2\pi j u F(u,v)$
  - $\mathcal{F}\left\{\frac{\partial f}{\partial n}\right\} = 2\pi j v F(u,v)$
  - $\mathcal{F}\{\nabla^2 f\} = \mathcal{F}\left\{\frac{\partial^2 f}{\partial m^2} + \frac{\partial^2 f}{\partial n^2}\right\} = -4\pi^2(u^2+v^2) F(u,v)$
- Teorema de Rayleigh (identidad de Parseval):
  $\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \mid f(m,n) \mid^2 = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \parallel F(u,v) \parallel^2$

- Simetría conjugada:

  Si $f(m,n)$ es real, entonces su TF cumple que

  $$F\left(\tfrac{M}{2}\pm u, \tfrac{N}{2}\pm v\right) = F^*\left(\tfrac{M}{2}\mp u, \tfrac{N}{2}\mp v\right)$$



- Rotación: rotar el plano de la imagen implica rotar igualmente el de Fourier
  $\mathcal{F}\left\{f(m\cos\theta - n\sin\theta, m\sin\theta + n\cos\theta)\right\} =$
  $F(u\cos\theta - v\sin\theta, u\sin\theta + v\cos\theta)$
- Proyección: Se define la proyección, p. ej. por filas de la imagen $f(m,n)$ como el vector 1-D $p(m) = \sum_{n=0}^{N-1} f(m,n)$. La TF de este vector unidimensional sería

$$\begin{aligned} P(u) &= \frac{1}{\sqrt{M}}\sum_{m=0}^{M-1} p(m)e^{-\frac{2\pi jmu}{M}}dm = \frac{1}{\sqrt{M}}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} f(m,n)e^{-2\pi j\left(\frac{mu}{M}+\frac{n0}{N}\right)} = \\ &= \sqrt{N}F(u,v)\mid_{v=0} = \sqrt{N}F(u,0) \end{aligned}$$

Significado físico de la DFT
Recordemos que el valor de la TF en un punto del plano de Fourier (PF) (u,v) y el de la imagen en un punto del plano image (PI) vienen dados por

$F(u,v) = \frac{1}{\sqrt{MN}}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} f(m,n)e^{-2\pi i\left(\frac{mu}{M}+\frac{nv}{N}\right)}$ y
$f(m,n) = \frac{1}{\sqrt{MN}}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1} F(u,v)e^{2\pi i\left(\frac{mu}{M}+\frac{nv}{N}\right)}$

Nótese que <u>todos</u> los puntos del PI $f(m,n)$ intervienen en el cálculo de <u>cada uno</u> de los puntos del PF. Para entender el significado de cada punto del PF consideremos cuál sería la imagen cuya TF fuese un punto de valor K en el centro del plano ($(u,v) = (0,0)$).

$$f(m,n) = \frac{1}{\sqrt{MN}}Ke^{2\pi i\left(\frac{m0}{M}+\frac{n0}{N}\right)} = \frac{K}{\sqrt{MN}}e^0 = \frac{K}{\sqrt{MN}}$$

es decir, una imagen constante. Al punto $(0,0)$ del plano de Fourier se le llama el término de contínua, y su valor es proporcional al valor medio de la imagen.

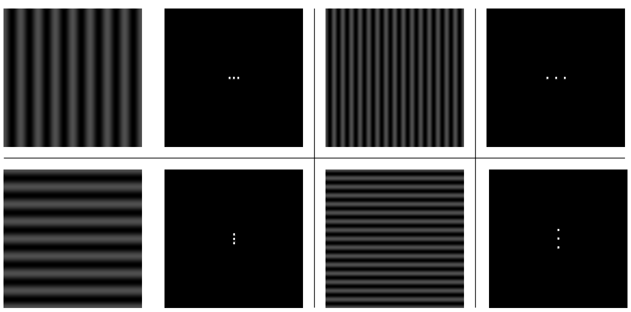
Transformadas integrales discretas

Sea ahora una imagen cuya TF fuesen dos puntos de valor K situados en $(0, v_0)$ y $(0, -v_0)$. En ese caso,

$$f(m,n) = \frac{1}{\sqrt{MN}} K \left(e^{2\pi i(\frac{m0}{M}+\frac{nv_0}{N})} + e^{2\pi i(\frac{m0}{M}-\frac{nv_0}{N})} \right) = \frac{2K}{\sqrt{MN}} \cos(\frac{nv_0}{N})$$

Análogamente, consideremos cuál sería la imagen cuya TF fuesen dos puntos de valor K situados en $(u_0, 0)$ y $(-u_0, 0)$.

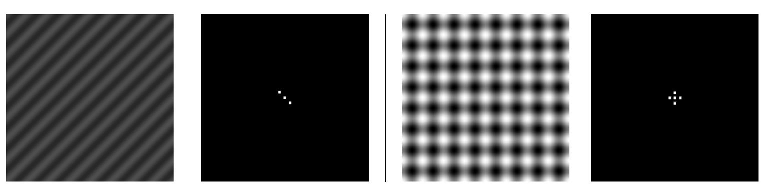
$$f(m,n) = \frac{1}{\sqrt{MN}} K \left(e^{2\pi i(\frac{mu_0}{M}+\frac{n0}{N})} + e^{2\pi i(-\frac{mu_0}{M}+\frac{n0}{N})} \right) = \frac{2K}{\sqrt{MN}} \cos(\frac{mu_0}{M})$$



Cuanto más separados están los puntos del centro del PF, mayor es la frecuencia de repetición de las ondas. u_0 y v_0 dan las frecuencias espaciales (horiz. y vert. respect.), medidas en repeticiones por imagen (aquí, 8 y 16).

Transformadas integrales discretas

Del mismo modo, por la propiedad de rotación obtenemos que la TF de ondas diagonales son los dos puntos más el central anteriormente vistos, girados. Y por la linealidad, la TF de la suma de ondas horizontales y verticales es la suma de las TF de cada caso.



Ahora, hay que entender la imagen f del PI como la suma sopesada de ondas elementales, tanto cosenos como senos, orientadas horizontal, verticalmente, y en todas las otras posibles direcciones. El valor del punto (u, v) del FP (es decir, $F(u, v)$) nos dará el peso con que la onda de frecuencias espaciales (u, v) contribuye en la imagen f. Concretamente, Re(F(u,v)) es el coef. de la onda coseno, y Im(F(u,v)) el de la onda seno.

Las frecuencias espaciales bajas se agrupan al centro del plano de Fourier, y las altas, en la periferia. Además, hay una frecuencia espacial límite: $N/2$ repeticiones/imagen en horizontal y $M/2$ en vertical. Esto serían imágenes con una línea blanca y una negra de anchura un pixel cada una, alternadas.

Transformadas integrales discretas

Transformada discreta del coseno (DCT)

Como vimos, una transformada integral se define por $F = P \cdot f \cdot Q$. En el caso de la DCT, las matrices P y Q se definen a partir de la matriz de $J \times J$ C_{JJ} cuyos elementos son:

$$C_{JJ}(k,l) = \begin{cases} \frac{1}{\sqrt{J}} & \text{si } l = 0 \\ \sqrt{\frac{2}{J}} \cos\left(\frac{(2k+1)l\pi}{2N}\right) & \text{si } l \neq 0 \end{cases}$$

Se define la DCT de una imagen real f de $M \times N$ como la imagen real F de $M \times N$

$$F(u,v) = C_{NN} \cdot f \cdot C_{MM}$$

En el caso de imagen cuadrada de $N \times N$ las fórmulas explícitas de la DCT e IDCT son

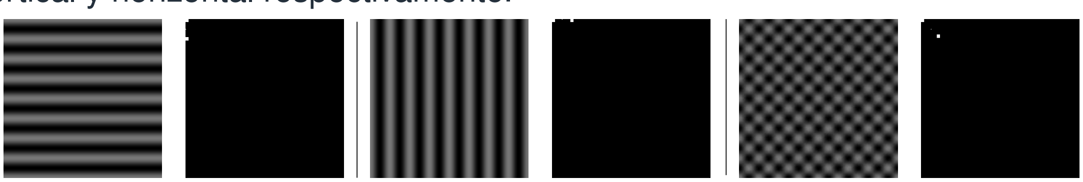
$$\begin{aligned} F(u,v) &= \frac{2c(u)c(v)}{N} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} f(m,n) \cos\left(\frac{(2m+1)\pi u}{2N}\right) \cos\left(\frac{(2n+1)\pi v}{2N}\right) \\ f(m,n) &= \frac{2}{N} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} c(u)c(v)F(u,v) \cos\left(\frac{(2m+1)\pi u}{2N}\right) \cos\left(\frac{(2n+1)\pi v}{2N}\right) \end{aligned}$$

siendo $c(k) = \frac{1}{\sqrt{2}}$ si $k = 0$ y 1 si no

Transformadas integrales discretas

Significado físico:

Es similar al de la TF, aunque aquí sólo tienen sentido valores positivos de u y de v. La transformada inversa de una imagen con un único punto de valor K en (u_0, v_0) es una imagen que representa el producto de dos ondas coseno de frecuencias espaciales u_0 y v_0 en vertical y horizontal respectivamente.



Las diferencias con la TF son que la DCT de una imagen real resulta siempre en una imagen real (por tanto no hay propiedad de simetría conjugada) y que aquí las bajas frecuencias espaciales se encuentran en la esquina superior izquierda. A medida que nos alejamos de esta esquina los puntos representan frecuencias espaciales crecientes. Además, se puede mostrar que, para un amplio rango de imágenes naturales y artificiales, la DCT es la transformada que más concentra la energía, es decir, los valores de unos pocos puntos cerca de la esquina superior izquierda del plano de la DCT tienden a ser muy altos, y los valores de la inmensa mayoría de puntos restantes tienden a ser muy pequeños.

Transformadas integrales discretas

Transformada de Walsh-Hadamard (WHT)

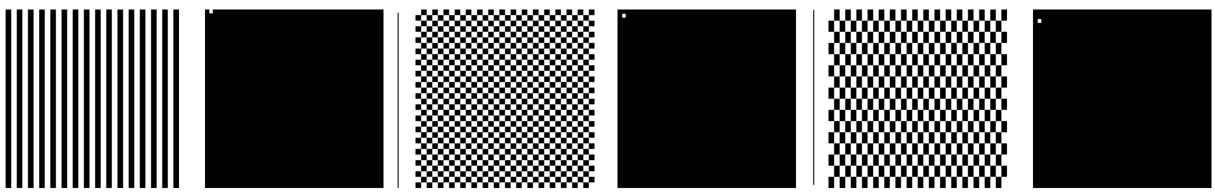
Se define del mismo modo que las demás, ($F = P \cdot f \cdot Q$) pero aquí P y Q deben ser cuadradas y de dimensiones potencia de 2. Se construyen como sigue:

$$H_{22} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \text{ y para otro tamaño potencia de 2, } H_{2J,2J} = \begin{pmatrix} H_{JJ} & H_{JJ} \\ H_{JJ} & -H_{JJ} \end{pmatrix}$$

La inversa de una matriz WH es $H_{JJ}^{-1} = \frac{1}{J} H_{JJ}$, con lo cual las definiciones de la transformada de Walsh-Hadamard directa e inversa de una imagen f de $M \times N$ son:

$$F = H_{MM} \cdot f \cdot H_{NN} \quad \text{y} \quad f = \frac{1}{MN} H_{MM} \cdot F \cdot H_{NN}$$

Las funciones de base de esta transformada son ondas cuadradas:



La TWH es real, y las frecuencias espaciales se distribuyen en el plano WH igual que en el caso de la DCT. Su cálculo es muy ligero computacionalmente: sólo involucra sumas (las multiplicaciones serían por +1 ó -1). Es útil en compresión y codificación de imágenes.

Transformadas integrales discretas

Las consecuencias del muestreo en señal bidimensional dependiente del espacio (imagen) son:

- No hay una frecuencia de muestreo límite; aun cuando el ojo humano tenga un límite de resolución, éste es dependiente de la distancia, y puede aumentarse por medios ópticos (magnificación); por tanto, no está unívocamente definido.
- El límite viene marcado por la propia digitalización que se aplique. En una imagen digital la máxima frecuencia espacial representable es de 0.5 lineas/pixel (la línea representa un ciclo espacial (onda completa) en cierta dirección
- De todos modos, el ojo sí presenta distinta agudeza visual para ciertas frecuencias espaciales específicas, y además ésta depende de la dirección de las líneas (ondas).

Formatos de audio sin compresión

Codificación es la asignación de un valor binario a un cierto valor discreto que resulta del muestreo y la cuantización de una señal en un instante del tiempo (o en un punto del espacio, si se trata de imagen). En el caso de audio esto se llama DPCM. Hay diferentes subtipos:

LPCM, en el que a cada muestra se le asigna el valor digital más cercano a su valor analógico.

A-Law y μ-Law, usados en transmisión de señal telefónica en Europa y EEUU/Japón respectivamente, y en los que a cada muestra se le asigna el valor digital más cercano al logaritmo de su valor analógico.

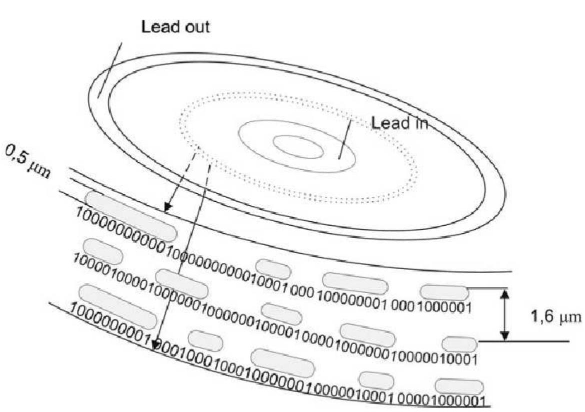
DPCM en la que se codifica de modo lineal la diferencia entre muestras adyacentes o la diferencia entre la predicción generada por muestras previas y el valor actual. Esto es una forma de compresión sin pérdidas que veremos después.

ADPCM, es igual que el DPCM pero el filtro de predicción cambia con el tiempo para ir mejorando (reduciendo el residuo).

Formatos de audio sin compresión

El formato de almacenamiento más simple es el Wave (.WAV), que contiene una cabecera indicando número de pistas, frecuencia de muestreo, número de bits por muestra, duración y metadatos seguida de los propios datos, normalmente en LPCM.

El medio de almacenamiento más común es el CD-Audio. En él los datos binarios se graban como marcas láser (se quema una superficie fotosensible) cuya longitud codifica una cadena de bits.



Pits length		Size in nm
T3 = 10001		833
T4 = 100001		1,111
T5 = 1000001		1,388
...	...	...
T11 = 1000000000001		3,054

Formatos de audio sin compresión

Además de los datos en bruto se graba también un código corrector de errores (CRC) obtenido como el resto de dividir el polinomio asociado a cada 16 bits de datos entre un polinomio fijado, lo cual se puede hacer usando sólo operaciones de OR-exclusivo y rotaciones de bits, y también se escriben los datos intercalados usando un código Reed-Solomon. Ver páginas 109-120 de este libro.

Para extraer los datos de formato WAV de un CD el probablemente mejor programa es cdparanoia.

Formatos de video sin compresión

Es muy inusual almacenar video no comprimido, dado el enorme volumen que ocupa. Como ejemplos, con los valores en GBytes por hora:

		Resolución			
fps	bps	720×486	720×576	1280×720	1920×1080
29.97^*	8	NTSC, 70 GB/hr			
	10	NTSC, 94 GB/hr			
25.00^*	8		PAL, 70 GB/hr		
	10		PAL, 93 GB/hr		
59.94	8			HDTV, 370 GB/hr	
	10			HDTV, 494 GB/hr	
25.00^*	8				HDTV, 348 GB/hr
	10				HDTV, 463 GB/hr
60.00	10			HDTVp, 742 GB/hr	
60.00^*	10				HDTVi, 834 GB/hr

fps: frames por segundo. Un frame es la transmisión de las líneas pares seguidas de las impares, para formato entrelazado, y todas las líneas en orden consecutivo para formato no entrelazado (pregresivo).
bps: bits por muestra (por cada componente de color de cada píxel).
(*): indica formato entrelazado.

Formatos de video sin compresión

Prácticamente el único formato de vídeo sin compresión usado es el HDMI (High Definition Multimedia Interface), que además encapsula también el audio digital asociado. Se transmite por cables especiales que contienen varios canales, concretamente el TDMS que es el que lleva el audio y vídeo, el CEC que sirve para enviar señales de control al monitor o dispositivo reproductor (conocido por varios nombres comerciales, el más común $Anynet^{©}$) y el infame HDCP, para protección anticopia.

El canal TMDS transmite los datos digitales con una frecuencia de píxeles de vídeo entre 25 y 330 MHz. Los formatos de vídeo con tasa binaria inferior a 25MHz (p. ej. los 13.5 MHz del NTSC a 480i) se transmiten usando un esquema de repetición de píxeles.
Se pueden transmitir hasta 24 bits por píxel, independientemente de la frecuencia. Los píxeles se pueden codifican en RGB 4:4:4, en YCbCr 4:2:2 ó en YCbCr 4:4:4.
Las frecuencias de muestreo permitidas para el audio son 32 kHz, 44.1 kHz, 48 kHz, 88.2 kHz, 96 kHz, 176.4 kHz y 192 kHz, con un total de hasta 8 canales de audio.

Compresión de señales digitales

Definición
La **compresión** es el proceso por el que una señal digital, o una señal similar a ella, es almacenada ocupando menos espacio (menor cantidad de bits) del que ocupaba inicialmente.

La expresión ”similar a ella” resulta ambigua, y se desarrollará en detalle posteriormente. En cualquier caso, las ideas fundamentales de la compresión son:

- Aprovechar las redundancias (valores de la señal que se repiten)
- Reducir aquella información que hace que la señal reconstruída difiera poco de la original.

También será necesario indicar cómo se puede medir esa diferencia, y qué significa ”diferir poco”.

Compresión de señales digitales

Conceptos previos

El **histograma** de una señal es un vector unidimensional en el que el índice de cada componente indica un valor de la señal, de entre el conjunto de valores cuantizados que ésta puede tomar, y el valor de la componente indica cuántas muestras de la señal tienen dicho valor. Si se normaliza dividiendo por el número total de muestras el histograma es una aproximación a la densidad de probabilidad del proceso que generó la señal.

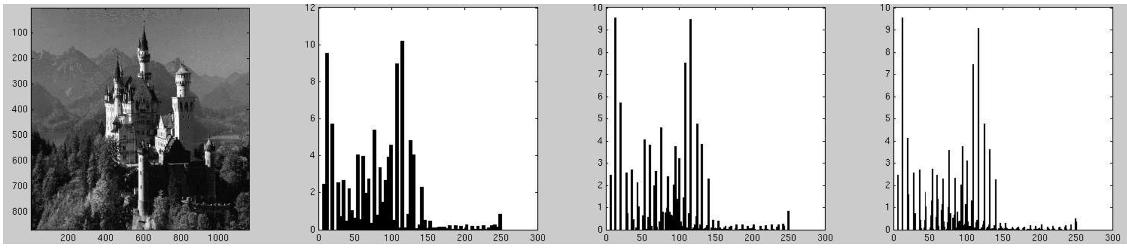


Imagen en grises y sus histogramas con 64, 128 y 256 bins respectivamente

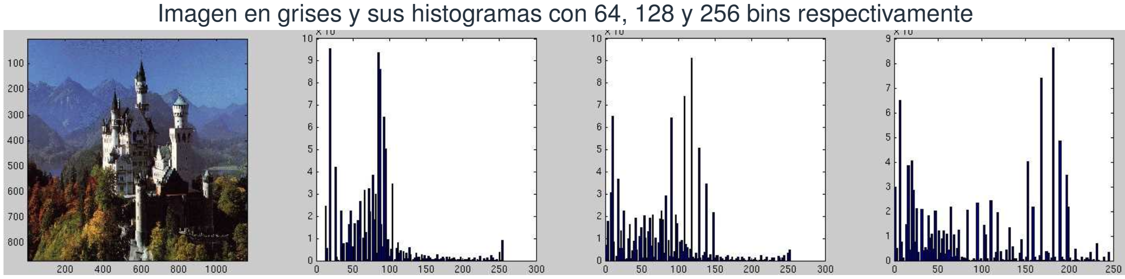


Imagen en color e histogramas de sus bandas R,G y B

Compresión de señales digitales

La **entropía** de una señal es la suma para todos los valores posibles de la señal del producto de la probabilidad de que el valor aparezca en la señal por el logaritmo en base 2 de dicha probabilidad, cambiada de signo.

$$S(\mathbf{x}) = -\sum_{i=0}^{N_v} p(v_i)\log(p(v_i)) \quad \text{Si } p(v_i)=0, \text{ se toma } p(v_i)\log(p(v_i))=0$$

La entropía mide el opuesto al contenido informativo de una señal, es decir, su desorden. La señal con menor contenido informativo es la que contiene un sólo valor, y en este caso $S = 1 \cdot \log(1) = 0$. Por contra, la señal más ”desordenada” es aquella en la que cada valor aparece con igual probabilidad $\frac{1}{V}$, un ejemplo de lo cual es el **ruido blanco**, y en este caso, $S = -\sum_{v=0}^{V} \frac{1}{V}\log(\frac{1}{V}) = \log(V)$. Esta no es la única definición posible de entropía, y no es la más adecuada para imágenes porque no tiene en cuenta la distribución espacial.

Ejercicio: pensar un ejemplo en el que esta definición de entropía no de cuenta de la noción intuitiva de orden.

Ayuda: la entropía se mide a partir del histograma, pero el histograma **no** es una representación invertible.

Métricas entre imágenes

A la hora de evaluar la calidad de un algortimo de compresión el procedimiento usual es tomar una muestra representativa de señales o imágenes de prueba, bien de diversos tipos o bien de un sólo tipo, si el algoritmo está orientado a comprimir específicamente ésas. Para cada una de ellas se realizará la compresión y la descompresión y se comparará el resultado con la señal o imagen original. Existen varias métricas posibles para la comparación (I_o e I_r son la original y resultante del proceso, y N el número total de muestras):

$$L_2(I_o, I_r) = \frac{1}{N}\sqrt{\sum_{i=1}^{N} \mid I_o(i) - I_r(i) \mid^2} \quad \text{(métrica Euclídea)}$$

$$L_1(I_o, I_r) = \frac{1}{N}\sum_{i=1}^{N} \mid I_o(i) - I_r(i) \mid \quad \text{(métrica } L_1\text{)}$$

$$L_\infty(I_o, I_r) = \max_i(\mid I_o(i) - I_r(i) \mid) \quad \text{(diferencia máxima)}$$

$$SNR = 10\log_{10}\frac{P_s}{P_n}(dB) \quad \text{(razón señal/ruido, con } P_s \text{ pot. de la señal y } P_n \text{ pot. del ruido)}$$

Ruido es la señal que resulta al restar la original y la reconstruída: $Ru(i) = I_o(i) - I_r(i)$.

$$PSNR = 10\log_{10}\frac{\max(I_o^2)}{\frac{1}{N}P_n}(dB)$$

(razón señal/ruido de pico). Ahora en lugar de tomar la potencia de la señal se toma el cuadrado del máximo valor que puede tomar (p. ej., si usamos 8 bps, 255).

Tanto la SNR como la PSNR se suelen expresar en dB.

En una imagen en color tanto la P_s como la P_n se calculan como la potencia media de las tres bandas. En el numerador de la PSNR se usa el mismo valor que para el caso monobanda.

Se considera que un incremento de 0.5 dB o más de PSNR se puede percibir visualmente como una mejora.

Compresión de señales digitales

Tasa de compresión es el cociente entre la longitud en bits de la señal original y la comprimida. En igualdad de otros factores, un método es mejor cuanto mayor sea su tasa de compresión.

Una primera división de los métodos de compresión es la que los separa en métodos **sin pérdidas** y métodos **con pérdidas**.
En los métodos sin pérdidas la señal reconstruída es binariamente idéntica al original, con lo que la distancia entre ambas usando cualquiera de las métricas vistas es 0, y las razones señal/ruido y señal/ruido de pico son ambas infinitas.
En los métodos con pérdidas esto no ocurre, y se intentará que la distancia en la señal original y la reconstruída sea lo más pequeña posible. En general, hay un balance entre tasa de compresión y distancia (o razón señal/ruido): a mayor tasa de compresión con pérdidas, menor razón señal/ruido. No obstante, dos métodos de compresión con la misma razón señal/ruido pueden generar una señal reconstruída muy distinta (y mucho mejor en uno de ellos) desde el punto de vista humano. Por ello, para caracterizar los algoritmos de compresión de audio, imagen y video se suelen emplear métricas **perceptuales**.

Compresión de señales digitales

Codificación es el proceso por el cual a un valor de la señal se le asigna otro valor (su **código**), de tal modo que la correspondencia sea suprayectiva (un valor del código puede corresponder a varios valores originales, pero un valor original tiene asignado uno y sólo un código). Lógicamente, la codificación no puede ser invertible a menos que sea biyectiva (cada valor original se corresponde con exactamente un código diferente). La manera normal de implementar las codificaciones es en forma de tabla (array) de posibles valores de la señal.
Codificación uniforme es aquella que asigna a cada valor de la señal el propio valor, o uno proporcional a él.
Codificación no uniforme es aquella que asigna a cada valor de la señal otro valor cualquiera, a voluntad del usuario, que puede escoger cualquier tabla de codificación que tenga buenas propiedades.
Al almacenar la señal en un ordenador es preciso hacerlo como una secuencia de valores que son números binarios, cada uno de los cuales necesitará una representación de cierto número de bits, dependiendo de su valor. El número de bits usados es la guía para construir un sistema de codificación no uniforme; hay dos casos importantes de codificación no uniforme: la óptima y la de Huffmann.

Compresión de señales digitales

Codificación óptima para una cierta señal:
Es la asignación que da los códigos con menor número de bits a los valores más frecuentes, de modo que la longitud en bits de la señal codificada sea mínima. Esta es la forma más simple de compresión sin pérdidas. El problema importante de la codificación óptima es que, cuando se envían los valores en serie, como un torrente de bits (*bitstream*), al no tener todos los símbolos la misma longitud no se puede saber dónde acaba uno y empieza el siguiente.

Codificación de Huffmann
Es el método más usado de codificación no uniforme. Asigna símbolos de longitud inversamente proporcional a la frecuencia de aparición de cada valor (aproximadamente), pero construyéndolos en forma de árbol, lo que elimina el problema de la separación entre símbolos. Se basa en ordenar los símbolos por su probabilidad de aparición, y proceder por pasos, agrupando en cada paso los dos símbolos menos probables. Hecho esto, se vuelve atrás asignando a cada grupo de símbolos un bit distinto.

Compresión de señales digitales

Ejemplo de codificación de Huffmann

Fuente original		Reduccion de la fuente			
Simbolo	Probabilidad	1	2	3	4
a_2	0.4	0.4	0.4	0.4	0.6
a_6	0.3	0.3	0.3	0.3	0.4
a_1	0.1	0.1	0.2	0.3	
a_4	0.1	0.1	0.1		
a_3	0.06	0.1			
a_5	0.04				

Fuente original			Reduccion de la fuente							
Simbolo	Probabilidad	Codigo	1		2		3		4	
a_2	0.4	1	0.4	1	0.4	1	0.4	1	0.6	0
a_6	0.3	00	0.3	00	0.3	00	0.3	00	0.4	1
a_1	0.1	011	0.1	011	0.2	010	0.3	01		
a_4	0.1	0100	0.1	0100	0.1	011				
a_3	0.06	01010	0.1	0101						
a_5	0.04	01011								

Al decodificar un *bitstream* de Huffmann, si encontramos un 1 sabemos que se trata del símbolo a_2, y el siguiente bit es ya parte de otro símbolo. Si es un 0, deberemos mirar el siguiente bit; si es 0, estamos ante a_6 y el símbolo ha terminado. Si no, miraremos el siguiente bit, y así sucesivamente.
Nótese que en una codificación uniforme el número de bits a usar sería de 3 (hay 6 símbolos) con lo que la longitud media es 3; en cambio en ésta la longitud media sería $\bar{l} = \sum_{i=1}^{6} p_i l_i = 0.4 + 0.3 * 2 + 0.1 * 3 + 0.1 * 4 + 0.06 * 5 + 0.04 * 5 = 2.20.$
La reducción en este caso sería pues de un 25.6 %

Compresión de señales digitales

Fases de la compresión

Señal original $\longrightarrow$ compresión $\longrightarrow$ señal comprimida

Transformación a una representación invertible más adecuada (normalmente, una transformada integral) En el caso de video, transformación del movimiento (cambios entre frames) a una representación más compacta.	$\longrightarrow$	Alteración de la representación en el espacio transformado, normalmente por cuantización no uniforme, para obtener valores repetidos próximos. Esta alteración es el único paso que puede ser **no invertible**, y por tanto provocar pérdidas.	$\longrightarrow$	Codificación de los valores obtenidos, mediante Huffmann, códigos *run-length* o alguna otra codificación no uniforme.

Señal original alterada $\longleftarrow$ descompresión $\longleftarrow$ señal comprimida

Formatos de audio comprimido

Compresión de audio sin pérdidas: FLAC

FLAC (free looseless audio compression) es un formato libre bajo el patrocinio de la fundación Xiph. Su funcionamiento se basa en:

- Dividir la señal en bloques que se comprimirán independientemente. Por defecto son de 4096 muestras consecutivas.
- Decorrelar las señales de los canales izquierdo y derecho construyendo con ellas otras dos señales, $mid = \frac{left+right}{2}$ y $side = left - right$.
- Buscar cuál es el predictor autoregresivo óptimo para la señal. Un predictor autoregresivo es una forma de obtener el valor de la señal en un instante del tiempo t a partir de valores anteriores:

$$\hat{s}(t) = \sum_{i=1}^{p} a_i s(t-i)$$

con coeficientes $a(i)$ cuantizados a 7 bits.