

Curso Optativo Procesamiento de Imágenes

Ciencia de la Computación

Ángela León Mecías

Curso 2020-2021

“A picture is worth a thousand words” *

Índice

1. Introducción	2
2. Adquisición de una imagen. Tipos de imágenes	4
2.1. Cómo se forma una imagen digital?	7
2.2. Escala de las imágenes	8
2.3. Dimensión de las imágenes	8
3. Digitalización de las imágenes	9
3.1. Muestreo	9
3.1.1. Algunos fundamentos matemáticos	9
3.1.2. Redimensionamiento de una imagen. Interpolación	15
3.2. Cuantificación	15
3.2.1. Tamaño de una imagen de datos	16
3.3. Cuestiones de notación	16
4. Procesamiento de imágenes. Operaciones básicas	17
4.1. Operaciones básicas en el preprocesamiento de imágenes	17
4.1.1. Operaciones puntuales	18
4.2. Histograma de una imagen	18
4.2.1. Manipulación del Histograma	19
4.2.2. Umbralización	20
4.3. Transformación de una imagen mediante operadores lineales	20
4.3.1. Filtros espaciales lineales	22
5. Procesamiento de imágenes en el dominio de Fourier	23
5.1. Transformada de Fourier en 2D	25
5.2. Filtros en el dominio de frecuencias	27
5.3. Filtros ideales en imágenes	29
6. Métodos basados en Ecuaciones Diferenciales Parciales. Modelos de difusión	29
6.1. Suavizado con filtro Gaussiano. Equivalencia con el filtrado por difusión lineal	30
6.2. Filtrado por difusión no lineal (anisotrópica)	32
6.2.1. Aproximaciones discretas del gradiente	33

* “Una imagen vale por mil palabras”

Bibliografía

1. Suganya, R., Rajaram, S., Sheik Abdullah, A., Big Data in Medical Image Processing, CRC Press Taylor and Francis Group, Boca Raton, London, New York, 2018.
2. Toennis, Klaus D., Guide to Medical Image Analysis: Methods and Algorithms, Second Edition, Springer Verlag London Ltd, 2017.
3. Jirí J., Medical Image Processing, Reconstruction and Restoration. Concepts and Methods, CRC Press Taylor and Francis Group, Boca Raton, FL, 2006.
4. Tony F. Chan and Jianhong Shen, Image Processing and Analysis. Variational, PDE's, wavelet and stochastic methods, SIAM 2005.
5. Giles Aubert and Pierre Kornprobst, Mathematical Problems in Image Processing. PDE's and Calculus of Variations, Second Edition Springer 2006.
6. Image Processing: The Fundamentals, Maria Petrou, Costas Petrou, Second Edition John Wiley and Sons, 2010.

1. Introducción

La Ciencia de las Imágenes o Imagenología integra tres áreas fundamentales:

- Adquisición de la imagen: se estudian los mecanismos físicos mediante los cuales los diferentes dispositivos generan las observaciones como imágenes y se investigan los modelos matemáticos asociados, así como los algoritmos empleados en las computadoras integradas a tales dispositivos.
- Procesamiento de la imagen: tiene que ver con cambiar la naturaleza de una imagen con el objetivo de mejorar su información gráfica para la interpretación humana, ó para hacerla más apropiada para la percepción automática de una máquina. (será el Centro de este curso)
- Interpretación de la imagen: capacidad de "ver", monitorear e interpretar la porción deseada del mundo hecha imagen. Intrínsecamente conectada con la visión basada en inteligencia artificial

Una imagen se puede interpretar como una función (señal) de brillantez (o niveles de grises), definida sobre un dominio espacial bidimensional, que teóricamente puede ser ilimitado, pero que en la práctica se restringe dentro de un área acotada, y los valores que toma la función también se encuentran dentro de un intervalo acotado.

Matemáticamente podemos distinguir entre:

- una función bidimensional continua (imagen analógica): función real que describe una escena del mundo real $u : \mathbb{R}^2 \rightarrow \mathbb{R}^d$,

que, teóricamente puede cubrir todo el plano \mathbb{R}^2 , pero en la práctica, $(x, y) \in \Omega = [x_{min}, x_{max}] \times [y_{min}, y_{max}]$. Si $d=1$ (imágenes en escala de grises) los valores de $u(x, y)$ representan intensidad del brillo en la posición espacial (x, y) . Si $d=3$ imágenes a color, $u(x, y) = [u_R(x, y), u_G(x, y), u_B(x, y)]^T$. Para algunos sistemas de imágenes tomográficas, $u : \Omega \subset \mathbb{R}^3 \rightarrow \mathbb{R}^d$.

- Una función que toma valores discretos, sobre una región discreta (imagen digital): $u : \Omega \subset \mathbb{R}^2 \rightarrow G \subset \mathbb{Z}$ donde Ω es una malla finita de puntos equidistantes, con frecuencia de $2^n \times 2^n$ puntos. $G = [0, 255]$, niveles de grises, 0 representa el negro, 255 el blanco.

El procesamiento y el análisis de las imágenes se ha convertido en un componente importante dentro de las ciencias y la tecnología moderna, con una amplia gama de aplicaciones que van desde la astronomía, la geología, la medicina, la biología, la nanotecnología, la ecología y el medio ambiente, la meteorología, video vigilancia en la seguridad y defensa, la industria del entretenimiento, la comunicación digital, la arqueología, la historia del arte, entre otras. Los conceptos básicos son comunes a todo este espectro de aplicaciones, pero con énfasis diferentes en cada una de ellas.



Figura 1: Imagen, poderoso medio de comunicación

Las imágenes nos permiten desarrollar tareas complejas en la vida diaria, comunicarnos, transmitir información, representar y entender el mundo que nos rodea y siempre que se tiene que interactuar con imágenes pues se cae en el dominio de acción de la visión computacional y del procesamiento de imágenes. Para referirse a distintos aspectos del cálculo con imágenes, se usan términos como procesamiento de imágenes, análisis de imágenes, visión computacional, visión robótica, visión de máquina, y la diferencia radica en la información que se usa y la salida que se obtiene después de realizado algún proceso. En el procesamiento de imágenes la información que mayormente se usa son los valores de intensidad en los píxeles y la salida es propiamente una imagen, ver [3].

En este curso nos enfocaremos en el procesamiento de imágenes, que según Alasdair McAndrew en su libro "An Introduction to Digital Image Processing with MatLab" se relaciona con cambiar la naturaleza de una imagen con el objetivo de:

1. Mejorar su información gráfica para la interpretación humana (mejoramiento y restauración para facilitar la transmisión, interpretación, impresión y almacenamiento de fotos según [7]).
2. Hacerla más apropiada para la percepción automática de una máquina (segmentación y descripción de imágenes como una primera etapa de la visión por computadoras).

En general, cualquier procesador de imágenes se puede formular de forma abstracta como un sistema entrada-salida donde se denota por u_0 la imagen de entrada (captada, sensada), que puede estar degradada a causa de pobres condiciones en la adquisición o problemas durante el almacenamiento y la comunicación, como se muestra en la tabla 1.

La entrada u_0 puede contener más de una imagen. Por ejemplo, imágenes obtenidas por cámaras de vigilancia, videos, películas, etc. En el caso de procesamiento de videos y películas, la variable tiempo puede conducir a nuevas características tales como distribución de velocidad, evolución dinámica de las formas y regiones de interés, etc. Se presentan dos grandes tareas

- Modelación de la imagen
- Modelación de los procesos que actúan sobre las imágenes

Matemáticamente hablando un procesador de imágenes T puede ser cualquier operador lineal o no lineal que actúa en la entrada produciendo determinadas características o patrones objetivos en F . El mayor reto en el diseño y cálculo de T es que la mayoría de los problemas de procesamiento de imágenes son problemas inversos mal planteados.

La tercera etapa de la ciencia de las imágenes según Tony Chan y Jianhong Shen, corresponde a la interpretación de la imagen y la inteligencia visual, es decir, la capacidad de ver, monitorear e interpretar la porción deseada del mundo hecho imagen. La interpretación de la imagen está intrínsecamente conectada con la visión basada en inteligencia

Q_0	proceso	Q
$u_0 = Ku + n$	Desemborronar y eliminar ruido	u Imagen más nítida
u_0	segmentación	Objetos $[u_k, \Omega_k], k = 1, 2, ..$
$u_0 _{\Omega \setminus D}$	Implante ó completamiento	$u_0 _{\Omega}$ Imagen completa
u_0	Espacio escala	Imágenes multi-escalas $u_{\lambda_1}, u_{\lambda_2}$
(u_0^1, u_0^2)	Estimación de movimiento	Flujos ópticos (\vec{v}^1, \vec{v}^2)

artificial. Molina en la página 53 Epígrafe 2.7 de su libro "Introducción al procesamiento y análisis de imágenes digitales", considera el procesamiento de imágenes como una tarea de visión artificial. A veces el procesamiento de imágenes clásico (eliminar ruido, suavizar, interpretar datos pedidos, segmentar, ¿algo más?) se identifica con visión de bajo nivel, mientras que la interpretación se identifica con visión de alto nivel (reconocimiento de rostros o de iris, lectura de terrenos para pilotos automáticos). Un sistema de visión artificial ideal debe ser capaz de realizar la mayoría de las tareas del sistema de visión humana (estimación de movimiento, percepción, estimación de profundidad, reconocimiento y diferenciación de objetos). De ahí que sea necesaria una combinación del procesamiento de imágenes con la interpretación de la imagen.

Elementos de un sistema de procesamiento de imágenes

- Adquisición
- Almacenamiento
- Procesamiento
- Comunicaciones
- Visualización

2. Adquisición de una imagen. Tipos de imágenes

Para entender el proceso de adquisición de una imagen, se estudian los mecanismos físicos mediante los cuales los diferentes dispositivos (sensores) generan las observaciones como imágenes y se investigan modelos matemáticos asociados, así como los algoritmos empleados en las computadoras integradas a tales dispositivos.

El sistema de visión humana es el dispositivo biológico de la madre naturaleza, desarrollado para realizar una percepción visual inteligente y es el patrón que se estudia para desarrollar los dispositivos artificiales. La visión humana está formada por periféricos ópticos que incluyen la córnea, la pupila, el iris, así como dos tipos diferentes de fotorreceptores de la retina: los conos y los bastones, que son responsables de capturar las intensidades de los valores de gris y el color de manera separada. Partículas de luz llamadas fotones rebotan en los objetos a nuestro alrededor y llegan a nuestros ojos. Los fotones son absorbidos por cerca de 126 millones de células sensibles a la luz y nuestro cerebro traduce las diferentes energías y direcciones de los fotones en formas y colores que nos permiten ver el mundo en ténicolor. Las retinas izquierda y derecha forman el extremo frontal de un sistema de visión complejo que transforma luz y fotones en señales electroquímicas y prepara para el siguiente procesamiento e interpretación de la imagen en las cortezas visuales internas a través de las fibras ópticas de neuronas. Existen tres tipos de conos, cada uno puede registrar cerca de 100 tonalidades, se relacionan con la visión en colores y la visión diurna; los bastones se relacionan con la visión nocturna y condiciones de baja luminosidad. En el ojo humano existen más de 100 millones de

bastones y cerca de 4 millones de conos. Relacionada con la visión humana está el concepto de ilusión óptica, que es cualquier ilusión del sentido de la visión que nos lleva a percibir la realidad de varias formas. Puede ser de carácter fisiológico, asociada a los efectos de una estimulación excesiva en los ojos o el cerebro, o de carácter cognitivo, en las que interviene nuestro conocimiento del mundo. Muchas veces nuestros ojos nos dan la información correcta y es el cerebro el que no realiza una correcta interpretación. La adquisición de una imagen se puede realizar además mediante:

- Procesos basados en cómo diseñar un hardware y combinarlo con un potente software para construir las imágenes en el mundo real 3D.
- Gráficos por computadora (*Computer Graphics*): basados únicamente en software. De ahí su importancia en la industria del cine. Con una escena imaginaria en mente un diseñador crea primero las coordenadas 3D G , para formas e información geométrica de objetos imaginarios, así como sus propiedades ópticas tales como grado de transparencia y reflectancia R (cantidad de luz que reflejan los objetos a la escena). El diseñador crea una fuente de luz I para la iluminación, que puede ser puntual o esparcida e irregular. También selecciona un punto de vista Θ de un observador imaginario. Entonces la imagen imaginaria u es sintetizada basada en alguna relación óptica propia

$$u = F(G, R, I, \Theta).$$

Si se incluye movimiento, entonces se incluye el tiempo t

$$u(t) = F(G(t), R(t), I(t), \Theta(t)).$$

Existe una gran diversidad de tipos de imágenes digitales que pueden ser obtenidas de casi todos los tipos de radiación. Sin embargo los sensores así como el ojo humano son capaces de "grabar" las radiaciones en una banda limitada del espectro electromagnético, ver Figura 2. De hecho aunque pensamos que vemos las cosas tal y como son, en realidad se pierde mucha información. Una banda de color es un rango de longitudes de onda del espectro

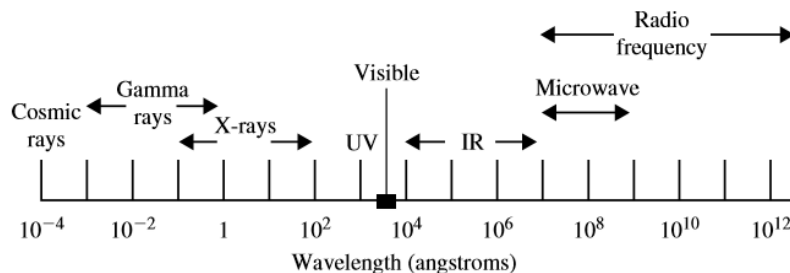


Figura 2: Espectro electromagnético. Tomada de [5]

electromagnético, sobre el cual los sensores que se usan para capturar la imagen tienen una sensibilidad determinada distinta de cero. Las imágenes a color tienen tres bandas de color lo que significa que han sido capturadas por tres conjuntos diferentes de sensores, cada uno con una función de sensibilidad diferente.

Los diferentes tipos de imágenes están relacionados por una parte con la longitud de onda del espectro electromagnético de las radiaciones captadas por los sensores.

El espectro de luz perceptible está entre 400 y 700 nm ($1 \text{ nm} = 10^{-9} \text{ m}$) en términos de longitud de onda.

Algunos tipos de imágenes

- Visión nocturna e imágenes térmicas

La visión humana es sensible a un rango dinámico grande de intensidades de la luz. Las longitudes de onda mayores que 700 nm pertenecen al rango infrarrojo (Madre naturaleza fuerza al ser humano a dormir durante la

noche, haciendo imperceptible la luz infrarroja). Con determinados equipos se puede percibir hasta pocas micras (10^{-6} m). La mayoría de las ondas de luz en el espectro infrarrojo no son reflejadas por los objetos como en el espectro visible, sin embargo, son emitidas por los objetos a través de su actividad térmica intrínseca a niveles atómicos. Los dispositivos de imágenes térmicas pueden recolectar y procesar dichos fotones produciendo los llamados termogramas que brindan un mapeo detallado de la distribución de temperatura de la escena tomada.

- *Imágenes de Radar y SAR*: Un radar es un dispositivo que emplea ondas de radio para detectar y ubicar objetivos. Las señales se encuentran en el rango de microondas entre 1 cm y 1 m de longitud de onda y polarizadas. Las antenas de radar emiten pulsos de ondas de radio polarizadas y reciben ecos reflejados de los objetos. El tiempo de espera entre la emisión y la recepción así como la intensidad, la potencia del eco o disipación de vuelta, revelan información importante como el rango, configuración local y diferenciación del material del objetivo. Aplicados a las imágenes, los radares son una herramienta poderosa para los sensores remotos. Para obtener imágenes básicas, el dispositivo de radar se adjunta a un vehículo volante que emite, recibe y procesa continuamente pulsos de señalizaciones de microondas. La imagen de radar es construida a lo largo de su línea de vuelo. La resolución de la imagen obtenida es proporcional a la dimensión del radar, es decir, antenas grandes resultan en imágenes de alta resolución.
- *Imágenes de Ultrasonido*: Así como la luz ultravioleta, los ultrasonidos se refieren a ondas acústicas cuyas frecuencias son mayores que el rango audible para el oído humano normal, es decir, entre 20 Hz y 20 KHz. Los ultrasonidos médicos con frecuencia están por encima de 1 MHz y por debajo de 20 MHz. Las diferentes respuestas de los tejidos y estructuras a los pulsos de ultrasonido forman la base fisiológica de la imagenología ultrasónica. Midiendo los intervalos de tiempo entre la emisión y recepción de los pulsos sonoros así como su intensidad, se puede inferir y mostrar (usando la computadora) información topológica local de la entidad fisiológica objetivo. Más aún, usando *traslación doppler* (el eco reflejado de un objeto en movimiento tiene mayores o menores frecuencias que la onda incidente, dependiendo de si el movimiento es hacia o se aleja de la fuente de la onda), la imagen de ultrasonido puede ser empleada para el estudio de órganos en movimiento en tiempo real.
- *Rayos X* Los rayos X fueron descubiertos por el físico alemán Wilhelm Roentgen en 1895 (Primer Premio Nobel de Física en 1901). A diferencia del espectro visible, los rayos infrarrojos para visión nocturna y las microondas para radares, los rayos X tienen longitudes de onda en la escala de nanómetros (10^{-9} m) o picómetros (10^{-12} m) los cuales tienen energía suficiente para penetrar la mayoría de los materiales, incluyendo tejidos biológicos, lo que hace a los rayos X ideales para pruebas médicas no invasivas e imagenológicas. Los diferentes tejidos y órganos en el cuerpo humano absorben los rayos X a diferentes tasas. Los huesos, por ejemplo, absorben más rayos X que los tejidos más suaves o blandos. Los distintos patrones de intensidad en el film diferencian los diferentes órganos y estructuras.
- *Tomografía Axial Computarizada (TAC)* El TAC puede ser llamado tecnología de imágenes de rayos X localizada y rotacional ya que consiste en una fuente de rayos X localizadas y un detector, se colocan uno opuesto al otro y rotan alrededor del eje central longitudinal del cuerpo humano. Mejora las imágenes de rayos X convencionales por medio de un diseño delicado de hardware así como un software de reconstrucción de la imagen final. De las señales de rayos X reunidas a lo largo de todas direcciones se sintetiza una imagen bidimensional rebanada (cortes).
- *Imágenes de Resonancia Magnética (IRM)* En 2003 Paul Lauterbur y Peter Mansfield obtuvieron el Premio Nobel de Fisiología o Medicina por su contribución a las IRM. La tecnología de IRM ha sido construida sobre la teoría física de la resonancia magnética nuclear, descubierta y desarrollada por dos físicos estadounidenses: Felix Bloch y Edward Purcell, este último ganó el Premio Nobel en 1952. Esta tecnología cuenta con un importante desarrollo en el mapeo del cerebro y la imagenología del tensor difusión (diffusion tensor imaging) para estudiar fibras nerviosas. Para más información sobre cómo trabaja esta tecnología consultar la página 22 de "Image Processing and Analysis. Variational, PDE's, wavelet and stochastic methods", Tony F. Chan, Jianhong Shen.

Hasta ahora todos los procesos de adquisición de imágenes están basados en el cómo diseñar un hardware y combinarlo con un potente software para construir las imágenes objetivo en el mundo 3D real. En el área de imágenes médicas aparecen cada vez más nuevos sensores que graban imágenes de fuentes de radiación que con anterioridad eran muy poco usadas tales como tomografía por emisión de positrones e imágenes por resonancia magnética; o

que captan las radiaciones de otra forma como tomografía asistida por computadora y rayos X, donde los datos se coleccionan desde múltiples ángulos para formar una imagen agregada más rica en información.

2.1. Cómo se forma una imagen digital?

Qué es una imagen digital?

Son "fotos" que han sido convertidas en un formato que pueda ser leído por la computadora (foto que no cambia con el tiempo).

Las imágenes digitales se obtienen convirtiendo señales continuas en formato digital, aunque los sistemas directamente digitales aparecen cada vez con mayor frecuencia.

Definición: Una **imagen digital** es una imagen $u(x, y)$ donde han sido discretizadas tanto las coordenadas espaciales (x, y) como el valor de la intensidad del brillo $u(x, y)$. La intensidad del brillo discretizada se conoce como niveles de gris. La imagen digital se representa mediante un arreglo bidimensional (imágenes en escalas de gris) o mediante una serie de arreglos bidimensionales, uno para cada banda de color (imágenes a color),

$$f(x, y) = \begin{bmatrix} f(1, 1) & f(1, 2) & \dots & f(1, N) \\ f(2, 1) & f(2, 2) & \dots & f(2, N) \\ \vdots & \vdots & \dots & \vdots \\ f(N, 1) & f(N, 2) & \dots & f(N, N) \end{bmatrix}, \quad (1)$$

donde $0 \leq f(x, y) \leq k$, N vienen expresados como potencias enteras de dos $N = 2^n, k = 2^m$. Cada elemento del arreglo es llamado píxel (o pel), que proviene del término "picture element".

Observación:

- Los índices en la matriz dependen de la elección de la posición del origen de coordenadas en el plano $x - y$. Las matrices de imágenes generalmente consideran el origen en la esquina superior izquierda por lo que el primer elemento de la matriz tendrá índices 00. En MatLab los índices comienzan en (1,1).
- Para ciertos algoritmos del procesamiento de imágenes, la matriz se convierte en un vector columna, leyendo la matriz por columnas

$$\bar{\mathbf{f}} = [f_{11}, f_{21}, \dots, f_{N1}, \dots, f_{1N}, f_{2N}, \dots, f_{NN}]^T.$$

- Una gran parte de los algoritmos de procesamiento de imágenes se refieren a imágenes en escalas de gris y la razón fundamental es que estos se pueden aplicar por separado a las diferentes bandas de color (RGB).

Cómo se forma una imagen digital?

Cada píxel de una imagen corresponde a una parte de un objeto físico en el mundo 3D, entonces los objetos son iluminados por un rayo de luz que proviene de alguna fuente de radiación y según la forma en que la radiación interactúa con el objeto 3D se obtienen diferentes tipos de imágenes. Los sensores ubicados para 'grabar' las imágenes lo que hacen es captar parte de la radiación que es reflejada por el objeto y cada sensor construye un píxel que representa un pequeño parche de la escena irradiada. El valor grabado por cada sensor depende de su curva de sensibilidad. Cuando un photon de una cierta longitud de onda alcanza un sensor, su energía es multiplicada por el valor de la función de sensibilidad de dicho sensor para esa longitud de onda y va siendo acumulada durante todo el tiempo que dura la 'exposición a la radiación'. La energía total acumulada por un sensor es eventualmente usada para calcular el valor de gris del píxel que le corresponde. (Ver en Maria Petrou 2010 significado físico de brillantez de una imagen en un píxel determinado).

De aquí que otro tipo de clasificación, un poco informal, de los tipos de imágenes tiene que ver con la forma en que la radiación interactúa con el objeto, ver Figura 3 tomada de [5]:

- Imágenes reflejadas (por reflexión): información de la superficie del objeto. El sensor capta la energía que ha sido reflejada por la superficie del objeto, la radiación puede provenir del ambiente o puede ser artificial. La mayoría de las imágenes ópticas captadas por nuestros ojos en la experiencia diaria son de este tipo, así como las imágenes de radar, sonar, imágenes láser.
- Imágenes emitidas (por emisión): información interna del objeto. En este caso el objeto que esta siendo captado es luminoso. Se incluyen las imágenes térmicas o infrarrojas comunes en la medicina, la astronomía y aplicaciones

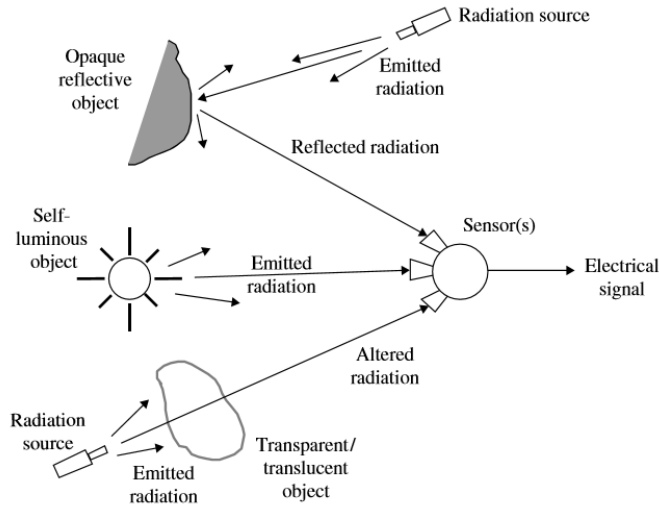


Figura 3: Interacción de la radiación con el objeto

militares. Estas imágenes revelan cómo los objetos crean radiación y entonces parte de su estructura interna es 'imagenada (fotografiada)'. También pueden ser captadas por esta vía imágenes externas por ejemplo cuando se usan cámaras térmicas para fotografiar escenas donde están presentes objetos con calor como las personas.

- Imágenes absorbidas (por absorción): información de la estructura interna del objeto. La radiación pasa a través del objeto y es parcialmente absorbida o atenuada por el material que lo compone. El grado de absorción dicta el nivel de radiación censada en la imagen grabada. Un ejemplo de este tipo de imágenes son los rayos X.

Las imágenes se pueden obtener de fuentes de luz visibles. Sin embargo, los datos que son tomados mediante sensores o dispositivos de grabación son con frecuencia multidimensionales y pueden ser dispuestos en un formato que puede ser visible para los humanos. Estos conjuntos de datos pueden ser considerados imágenes y como tal procesados.

2.2. Escala de las imágenes

Tiene que ver con el orden de magnitud de los objetos que visualizamos. Dentro del campo de las imágenes, la diversidad no solo se manifiesta en la forma de adquirir las imágenes, sino también en la escala de las imágenes. En nuestra experiencia diaria nos encontramos (visualizamos) objetos que tienen 3 o 4 veces la magnitud de 1 m. Sin embargo, los dispositivos que permiten magnificar y amplificar una imagen han hecho posible extender la visión, es decir, tener imágenes de estructuras que son 10^{30} veces 1 m (en el cosmos) o en el microcosmos donde es posible adquirir imágenes de objetos tan pequeños como $10^{-10}m$, es decir, somos capaces de captar imágenes en un rango de 40 órdenes de magnitud.

La escala puede ser también interpretada en el sentido de que una imagen puede contener objetos que existen en diferentes escalas o que un mismo objeto existe en diferentes escalas.

2.3. Dimensión de las imágenes

Un rasgo importante de las imágenes es que son señales (funciones) de dos o quizás tres variables espaciales (cuando se da información de profundidad en cada píxel), es decir son multidimensionales. Un video como función incluye una tercera o cuarta dimensión temporal.

Dimensión: Es el número de coordenadas requeridas para ubicar, indexar un punto dado en la imagen.

3. Digitalización de las imágenes

El medio que nos rodea existe, en una escala de observación razonable, como un continuo espacio-tiempo. De ahí que las imágenes del medio que nos rodea (antes de ser fotografiadas, censadas) sean analógicas. Esto significa que las señales son funciones definidas sobre un dominio espacio/tiempo continuo y que los valores que toma son elegidos también de un continuo de posibilidades.

Sin embargo, para trabajar con las imágenes ellas deben ser leídas por la computadora, es decir solo pueden ser evaluadas en un conjunto finito de puntos, o lo que es lo mismo deben estar definidas en un dominio discreto de tiempo-espacio; esto se llama muestreo, además deben tomar valores en un conjunto discreto de posibilidades. De aquí que el proceso de conversión de analógico a digital (A-D) tiene dos partes:

- Muestreo (discretización de la región espacio/temporal donde están definidas), la tasa de muestreo determina la resolución espacial de la imagen digital.
- Cuantificación (el nivel de cuantificación determina el número de niveles de gris), que deben ser suficientes para que el ojo humano pueda percibir finos detalles de sombreado en la imagen. Cuando la imagen ha sido cuantificada con niveles de brillantez insuficientes aparecen contornos falsos.

3.1. Muestreo

El muestreo espacial "ideal" se puede representar mediante la multiplicación de la señal analógica (continua), por una función de muestreo bidimensional que denotaremos por $s(x, y)$. Esta función de muestreo va a estar formada por un número infinito de impulsos de Dirac repetidos periódicamente $s(x, y) = \sum_{i=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \delta(x - i\Delta x, y - k\Delta y)$. La imagen muestreada de manera uniforme, puede ser representada como el producto entre la imagen analógica y la función de muestreo,

$$u_s(x, y) = u(x, y)s(x, y) = \sum_{i=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} u(i\Delta x, k\Delta y)\delta(x - i\Delta x, y - k\Delta y). \quad (2)$$

Como se observa la expresión anterior no es más que una suma de Dirac con pesos, los pesos son las muestras de la función; esto es, solo se necesitan aquellos valores de la función analógica definidos sobre los nodos de la malla de muestras, con lo cual nos parecería que la información contenida por los valores intermedios de la imagen se pierde; pero esto no tiene que ser exactamente así, si se toman en cuenta ciertas restricciones, de manera tal que la matriz de valores correspondiente a la imagen $\bar{u} = [u_{si,k}]$; $u_{si,k} = u_s(i\Delta x, k\Delta y)$, hasta ahora ilimitada, contenga toda la información de ésta.

Surge entonces la interrogante: Cuánta información sobre una función está contenida en un conjunto finito o infinito de muestras?

Desde el punto de vista matemático, nos estaríamos preguntando:

- ¿Cómo recuperar la función $u(x, y)$ a partir de sus muestras $u(i\Delta x, j\Delta y)$?, o lo que es lo mismo
- ¿Cómo recuperar $u(x, y)$ a partir de la función discreta $u_s(x, y)$?

Un papel importante en este análisis lo juega la fórmula de Suma de Poisson, que es el puente entre la Transformada de Fourier y la Serie de Fourier de ambas funciones, u y u_s .

Veamos, mientras la Transformada de Fourier es fundamental para un análisis abstracto en el procesamiento de señales e imágenes, es la serie de Fourier la que se usa para realizar el trabajo y todo tiene que ver con la simplicidad a la hora de la implementación computacional eficiente. La Transformada de Fourier y su inversa requieren del cálculo de integrales sobre todo el eje real, sin embargo, la serie de Fourier requiere del cálculo de sumas infinitas e integrales sobre intervalos finitos, que en realidad son aproximados por sumas finitas.

3.1.1. Algunos fundamentos matemáticos

Elementos del análisis de Fourier Sea $f(x)$ una función continua real, su transformada de Fourier \hat{f} existe si $f \in L^1(\mathbb{R})$, es decir si $\int_{-\infty}^{\infty} |f(t)| dt < +\infty$ y

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(x)e^{-2i\pi\omega x} dx$$

. Si $\hat{f} \in L^1(\mathbb{R})$, entonces existe la Transformada inversa

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\omega) e^{-2i\pi\omega x} d\omega.$$

Aunque f sea una función real, \hat{f} es en general una función compleja, esto es $\hat{f}(\omega) = R(\omega) + iI(\omega)$, escrita en su forma exponencial

$$\hat{f}(\omega) = |\hat{f}(\omega)| e^{i\phi(\omega)}$$

$|\hat{f}(\omega)| = \sqrt{R^2(\omega) + I^2(\omega)}$ llamado espectro de Fourier de $f(x)$, y $\phi(\omega) = \arctan(I(\omega)/R(\omega))$ la fase, $|\hat{f}(\omega)|^2$ es el espectro de potencias; ω recibe el nombre de frecuencia que proviene de la fórmula de Euler

$$e^{2i\pi\omega x} = \cos(2\pi\omega x) + i\sin(2\pi\omega x),$$

donde ω representa el número de ciclos completos por unidad de x .

Ejemplo 1 Sea

$$f_\varepsilon(x) = \begin{cases} 1/\varepsilon & \text{si } |x| < \varepsilon/2 \\ 0 & \text{e.o.c} \end{cases}$$

$$\begin{aligned} \hat{f}_\varepsilon(\omega) &= \int_{-\varepsilon/2}^{\varepsilon/2} \frac{1}{\varepsilon} e^{-2i\pi\omega x} dx \\ &= \frac{\sin(\pi\omega\varepsilon)}{\pi\omega\varepsilon} \end{aligned}$$

es decir $\hat{f}_\varepsilon(\omega)$ es real. Sin embargo, si consideramos la función $h_\varepsilon(x) = f_\varepsilon(x - x_0)$, su transformada de Fourier es una función compleja aunque el espectro es el mismo

$$\begin{aligned} \hat{h}_\varepsilon(\omega) &= \int_{-\varepsilon/2+x_0}^{\varepsilon/2+x_0} \frac{1}{\varepsilon} e^{-2i\pi\omega x} dx \\ &= \frac{\sin(\pi\omega\varepsilon)}{\pi\omega\varepsilon} e^{-2i\pi\omega x_0} \end{aligned}$$

Para $f(x, y)$, se define la Transformada de Fourier bidimensional

$$\hat{f}(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-2i\pi(ux+vy)} dx dy$$

y la Transformada Inversa de Fourier bidimensional

$$f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}(u, v) e^{2i\pi(ux+vy)} du dv$$

Definición formal de impulso puntual de Dirac en el dominio continuo La función generalizada (o de distribución) delta δ o impulso unitario de Dirac¹ es muy útil para explicar operaciones sobre funciones, pero esta no se define directamente, si no que existen varios modelos para hacerlo, uno de ellos es como límite de una sucesión de funciones: Se define una función rectángulo (rect) como

$$rect(t) = \begin{cases} 1 & |t| < \frac{1}{2} \\ 0 & \text{si no} \end{cases} \quad (3)$$

Se considera entonces una sucesión de funciones

$$\delta_n(t) = n rect(nt)$$

¹Paul Adrien Maurice Dirac, físico inglés 1902 Bristol Reino Unido-1984 Tallahassee, Florida USA

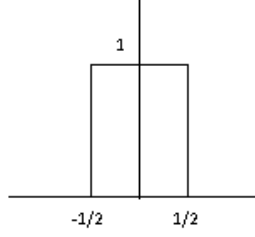


Figura 4: Función rectángulo

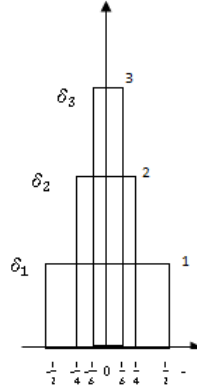


Figura 5: Sucesión de rectángulos

$$\delta_1(t) = \begin{cases} 1 & |t| < \frac{1}{2} \\ 0 & \text{si no} \end{cases} \quad (4)$$

$$\delta_2(t) = \begin{cases} 2 & |t| < \frac{1}{4} \\ 0 & \text{si no} \end{cases} \quad (5)$$

$$\delta_3(t) = \begin{cases} 3 & |t| < \frac{1}{6} \\ 0 & \text{si no} \end{cases} \quad (6)$$

$$: \quad (7)$$

$$\delta_n(t) = \begin{cases} n & |t| < \frac{1}{2n} \\ 0 & \text{si no} \end{cases} \quad (8)$$

$$(9)$$

Como se observa cada función de la sucesión es un pulso corto de duración $\frac{1}{n}$ que tiene la misma área unitaria, pero es de altura creciente que tiende a $+\infty$ cuando $n \rightarrow \infty$. Entonces, en realidad el límite de δ_n cuando $n \rightarrow \infty$ en el sentido usual no existe, pero es conveniente suponer que sí y que es una función de altura infinita, pero de ancho cero, de forma que el área sea siempre la unidad, y el límite se representa con frecuencia como un vector (o punto sobre el eje y) indicando que el área del pulso está concentrada en $t = 0$ y se denota por $\delta(t)$,

$$\delta(t) = \lim_{n \rightarrow \infty} \delta_n(t) \quad (10)$$

y tal que: $\int_{-\infty}^{\infty} \delta(t)dt = 1$. Se cumple además la propiedad

$$\int_{-\infty}^{\infty} f(t)\delta(t-a)dt = f(a). \quad (11)$$

Denotando $\delta_a = \delta(t-a)$, se tiene

$$\widehat{\delta}_a(\omega) = \int_{-\infty}^{\infty} \delta(t-a)e^{-i\omega t}dt = e^{-i\omega a}$$

Teniendo en cuenta que las imágenes son funciones bidimensionales, entonces se definen fuentes (impulsos) de brillantez constante

$$\delta_n(x, y) = n^2 \text{rect}(nx, ny),$$

donde n es una constante positiva y

$$\text{rect}(nx, ny) = \begin{cases} 1 & |nx| \leq \frac{1}{2}, |ny| \leq \frac{1}{2} \\ 0 & \text{si no} \end{cases} \quad (12)$$

La brillantez total de la fuente viene dada por

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta_n(x, y) dx dy = n^2 \underbrace{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{rect}(nx, ny) dx dy}_{\text{área del rectángulo}} = 1 \quad (13)$$

En el límite δ_n se convierte en la función impulso delta de Dirac

$$\delta(x, y) = \begin{cases} \infty & \text{para } x = y = 0 \\ 0 & \text{si no} \end{cases},$$

con la propiedad

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(x, y) dx dy = 1 \quad (14)$$

La integral

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta_n(x, y) g(x, y) dx dy$$

no es más que la promediación de la imagen $g(x, y)$ sobre un cuadrado de lado $1/n$ centrado en $(0, 0)$. En el límite se tiene que

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(x, y) g(x, y) dx dy = g(0, 0),$$

que no es más que el valor de brillantez de la imagen en el origen. Interpretada como una imagen estaríamos en presencia de un plano $x-y$ negro con un punto infinitésimo brillante en el origen, cuya luminosidad total es unitaria. Además esta imagen puede ser considerada como el caso límite de una sucesión de imágenes con cuadrados brillantes centrados en el origen, que mantienen la luminosidad total constante a medida que se hace más pequeña la dimensión del cuadrado. Haciendo una extensión de la propiedad (11) se cumple ver Petrou pag. 35 pdf:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \delta(x-a, y-b) dx dy = g(a, b)$$

Esta última igualdad nos indica que una imagen se puede representar como superposición de fuentes puntuales, que en el caso de imágenes digitales la integral doble se sustituye por doble suma, ver la expresión (2).

Se cumple además que

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-i2\pi(xu+vy)} du dv = \delta(x, y) \quad (15)$$

Sobre Teoremas de muestreo Un modelo básico para obtener una medición es mediante la evaluación de una función en un punto. Un conjunto de puntos $\{x_j\}$ contenidos en un intervalo (a, b) es un conjunto discreto si no existe ninguna subsucesión que converja a un punto en (a, b) . La evaluación de una función en un conjunto discreto de puntos es llamada muestreo. Se asume que la función sólo se evaluará en puntos donde es continua.

Definición 2 Sea f una función definida en el intervalo (a, b) y $\{x_j\}$ un conjunto discreto de puntos en (a, b) . Los puntos $\{x_j\}$ son llamados puntos muestrales y los valores $\{f(x_j)\}$ muestras de f en $\{x_j\}$.

En la mayoría de las aplicaciones el conjunto discreto se escribe como $\{x_0 + jl, j \in \mathbb{Z}\}$, y l es un número positivo fijo. Estas son llamadas muestras igualmente espaciadas; l es el espaciado muestral y l^{-1} es la tasa de muestreo (sampling rate). La teoría de muestreo estudia el problema de reconstruir funciones de variable continua a partir de un conjunto de muestras, y la relación entre esta reconstrucción y los datos idealizados.

Desde el punto de vista matemático existe una cota inferior en la frecuencia de muestreo en el tiempo-espacio, que determina el número mínimo de muestras posibles requeridas para retener la información en la señal (imagen). Esto permite que la señal pueda ser reconstruida a partir de las muestras, [4] [Página 8].

Para trabajar con mediciones, un modelo más realista que la evaluación puntual de una función es la evaluación de la convolución $f * \varphi$, donde $\varphi \in L^1$ es una llamada función de peso que modela todo el aparataje de las mediciones. $f * \varphi$ es una función continua. Adicionalmente si pasamos al espacio de Fourier, para obtener información en frecuencias, pues tenemos que por el Teorema de Convolución

$$\widehat{f * \varphi} = \hat{f} \cdot \hat{\varphi} \quad (16)$$

\hat{f} denota la transformada de Fourier de f . Y como $\varphi \in L^1$ entonces por el Teorema de Riemann-Lebesgue [CITAR], su transformada de Fourier $\hat{\varphi}$ es una función continua que tiende a cero en el infinito, es decir para $\eta \in \mathbb{R}$

$$\lim_{\xi \rightarrow \eta} \hat{f}(\xi) = \hat{f}(\eta)$$

y

$$\lim_{\xi \rightarrow \mp \infty} \hat{f}(\xi) = 0.$$

Esto significa que el aparato de medición φ atenúa la información de las altas frecuencias de f .

Definición 3 Una función $f \in L^2(\mathbb{R})$ se dice que está limitada en una banda de frecuencias, si su transformada de Fourier \hat{f} tiene soporte compacto, es decir si existe una constante $\Omega > 0$ t.q. $\hat{f}(\lambda) = 0$, para $|\lambda| > \Omega$.

En otras palabras, el soporte de \hat{f} está en el intervalo $[-\Omega, \Omega]$ y f es una función limitada a la banda de frecuencias de ancho Ω . Se dice también que f es una función con ancho de banda finito, en este caso Ω .

Por ejemplo el oído humano puede oír sonidos solo con frecuencias menores que $20KHz$, ($1KHz = 1000$ ciclos por segundo). Conversaciones telefónicas son señales limitadas en bandas de frecuencia.

Para las funciones con frecuencias en un ancho de banda limitado se cumple:

- son siempre infinitamente diferenciables
- si $f \in L^1$ o $f \in L^2$, entonces $\hat{f} \in L^1$ y existe la Transformada inversa de Fourier y se tiene que

$$f(x) = \frac{1}{2\pi} \int_{-\Omega}^{\Omega} \hat{f}(\xi) e^{ix\xi} d\xi$$

Y como el soporte de la función del integrando es un intervalo finito, entonces la integral se puede derivar tantas veces como se quiera y con esto f es infinitamente diferenciable

- Si $f \in L^2$, entonces todas las derivadas también
- El teorema de Nyquist postula que f estará determinada por un conjunto de muestras uniformemente espaciadas, siempre y cuando el espacio entre las muestras sea “suficientemente pequeño”.

Teorema 4 (de Nyquist ²) Si $f \in L^2$ y $\hat{f}(\xi) = 0$, para $|\xi| > \Omega$, entonces f está completamente determinada por el conjunto de muestras $\{f(\frac{n\pi}{\Omega}), n \in \mathbb{Z}\}$

Cuando Ω es la menor frecuencia para la cual se cumple la desigualdad anterior entonces la frecuencia natural $\gamma := \frac{\Omega}{2\pi}$ es llamada frecuencia de Nyquist y $2\gamma = \frac{\Omega}{\pi}$ es la tasa de Nyquist.

Por ejemplo las funciones exponenciales $e^{\mp i\Omega x}$ tienen período $(2\pi)/\Omega$ y frecuencia $\Omega/(2\pi)$. Si la función está limitada en una banda de frecuencia de ancho Ω , entonces $\Omega/(2\pi)$ es la frecuencia más alta que aparece en su representación en serie de Fourier. El Teorema de Nyquist postula que debemos muestrear dicha función con una tasa Ω/π , es decir el doble de la frecuencia más alta. Si se muestrea con una tasa menor no obtendremos información suficiente para determinar f completamente y a esto se le llama submuestreo y se se toma una tasa mayor entonces se tiene información redundante y se llama sobremuestreo. El teorema de Nyquist también fue formulado y demostrado formalmente por Claude Shannon, quien usó la fórmula de interpolación de Whittaker,

Teorema 5 (de muestreo de Shannon-Whittaker) ³ Sea $\hat{f}(\lambda)$ continuamente diferenciable por tramos, limitada en una banda de frecuencias para alguna frecuencia fija $\Omega > 0$. Entonces $f = \mathcal{F}^{-1}[\hat{f}]$ está completamente determinada por sus valores en los puntos $t_j = \frac{j\pi}{\Omega}$, $j = 0, \pm 1, \pm 2, \dots$, es decir f se puede representar mediante la serie

$$f(t) = \sum_{j=-\infty}^{\infty} f\left(\frac{j\pi}{\Omega}\right) \frac{\text{sen}(\Omega t - j\pi)}{\Omega t - j\pi} \quad (17)$$

que es uniformemente convergente.

La fórmula (17) es conocida como fórmula de interpolación de Whittaker-Shannon o interpolación “sinc” (seno cardinal), viene de la definición

$$\text{sinc} x = \frac{\text{sen} x}{x}$$

. Además se tiene que La demostración del teorema es simple. Por simplicidad tomemos $\Omega = \pi$. Si \hat{f} tiene soporte compacto en $[-\pi, \pi]$, entonces haciendo una extensión periódica de \hat{f}

$$\hat{f}_p(\xi) = \sum_{k=-\infty}^{\infty} \hat{f}(\xi + 2\pi k)$$

podemos hallar su serie de Fourier

$$\hat{f}_p(\xi) = \sum_{k=-\infty}^{\infty} \hat{f}(\xi + 2\pi k) = \sum_{k=-\infty}^{+\infty} c_k e^{ik\pi\xi/\Omega} \quad (18)$$

con $c_k = \frac{1}{2\Omega} \int_{-\Omega}^{\Omega} \hat{f}(\xi) e^{-ik\pi\xi/\Omega} d\xi$. La expresión (18) se conoce como Sumatoria (dual) de Poisson y se puede asegurar la convergencia según el teorema 8.2.2 ⁴ en [Epstein, Introduction to the Mathematics of Medical Imaging].

Como $\hat{f}(\xi) = 0, |\xi| > \Omega = \pi$;

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{f}(\xi) e^{-ik\xi} d\xi = \frac{1}{\sqrt{2\pi}} f(k),$$

entonces

$$\hat{f}(\xi) = \sum_{k=-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} f(k) e^{ik\xi} \quad (19)$$

²Formulado por primera vez en forma de conjetura en 1928 por Harry Nyquist

³Demostrado por primera vez por Whittaker en 1935 en un libro de Teoría de Interpolación y redescubierto en 1949 por Claude Shannon matemático en los laboratorios Bell, este teorema constituye la base para la tecnología digital y fue descubierto por primera vez de forma independiente por Harry Nyquist en 1928

⁴Si f es tal que $\hat{f} \in L^1$, y $\sum_{-\infty}^{\infty} |f(\frac{\pi n}{L})| < \infty$, entonces en los puntos de continuidad de \hat{f}_p se cumple $\sum_{-\infty}^{\infty} \hat{f}(\omega + 2nL) = \frac{\pi}{L} f(\frac{n\pi}{L}) e^{-\frac{ni\pi\omega}{L}}$

y esta serie converge uniformemente ya que \hat{f} es continua y suave por tramos⁵ con lo cual se puede hallar por la fórmula de la Transformada Inversa

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{f}(\xi) e^{ix\xi} d\xi = \frac{1}{\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} c_k \int_{-\pi}^{\pi} e^{i\xi(x-k)} d\xi = \sum_{k=-\infty}^{\infty} f(k) \frac{\text{sen}\pi(x-k)}{\pi(x-k)}$$

En nuestro escenario, esto significa que una imagen puede ser totalmente reconstruida a partir de sus muestras en tiempos regularmente espaciados. Otra forma de ver el teorema de muestreo es que para retener información vital mediante el muestreo con período de muestreo T , la frecuencia de muestreo $\omega_s = \frac{2\pi}{T}$ debe ser mayor que dos veces la mayor frecuencia ω_x contenida en la señal en tiempo continuo $x_a(t)$ que será muestreada, llamada tasa de Nyquist $2\omega_x$. Si $\omega_s \leq 2\omega_x$ ocurre el llamado efecto aliasing de frecuencia⁶. Ordinariamente una curva continua puede ser caracterizada sólo de manera aproximada si toma un número finito de puntos por donde ella pasa; para tener una información completa sobre la curva se requieren en general una cantidad infinita de puntos. Sin embargo según Shannon si la curva está compuesta por un rango limitado de frecuencias, ésta se puede reproducir de manera exacta tomando un número finito de muestras (consecuencia muy importante para la transmisión y procesamiento de información). Por ejemplo si el rango de frecuencias transmitida por una línea telefónica es de 4000Hrtz (ciclos por segundo), cuando usted habla por teléfono su voz es medida alrededor de 8000 veces por segundo.

En las imágenes es importante que no se pierda información visual, que la imagen o el video se "vean bien".

Las muestras son indexadas por números enteros lo que permite procesar una imagen como un arreglo multidimensional de números.

En las imágenes se seleccionan con frecuencia números de filas y columnas que sean potencias enteras de 2 (lo que es útil para aplicar FFT).

El muestreo reduce la resolución espacial de las imágenes.

Tarea: Escoger una imagen y ver efectos visuales de los diferentes rangos de muestreo ($2^6 \times 2^6$, $2^7 \times 2^7$, $2^8 \times 2^8$)

3.1.2. Redimensionamiento de una imagen. Interpolación

La interpolación en una imagen ocurre cuando se redimensiona o se distorsiona, al pasar de una malla de píxeles a otra. El redimensionamiento se produce cuando es necesario aumentar o disminuir la cantidad total de píxeles, por ejemplo el proceso de "magnificación" (zooming) se refiere al proceso de aumentar la cantidad de píxeles, de manera que se vean mejor los detalles en la imagen. Como se conoce del curso de Matemática Numérica básico la aproximación por interpolación se construye a partir de datos conocidos, asumiendo que están libres de error, para estimar valores en puntos desconocidos. En este contexto será una interpolación bidimensional, donde la aproximación del valor de intensidad se estima a partir de los valores en píxeles vecinos. Los algoritmos de interpolación pueden ser adaptativos o no; con adaptatividad entendemos en este caso que el tratamiento sea diferenciado según la región. Es interesante señalar que muchas cámaras digitales pueden realizar dos tipos de magnificación: óptica y digital. La magnificación óptica se refiere a que el lente de la cámara se mueve para magnificar la luz y el digital degrada la calidad de la imagen con la interpolación.

3.2. Cuantificación

Usualmente las imágenes se describen por el valor de intensidad (es un valor positivo), p.e. en una imagen en blanco y negro se dice que los valores en los píxeles nos dan los niveles de gris. El proceso de cuantificación consiste en convertir la imagen que toma valores continuos en una que toma valores discretos; este proceso usualmente se realiza mediante un proceso de redondeo, truncamiento o algún otro proceso no lineal de destrucción de información irreversible.

La intensidad de la imagen debe ser representada con precisión finita en un procesador digital, p.e. los valores que pueden tomar los niveles de gris se acotan $(1, \dots, k-1)$ donde por conveniencia $k = 2^\beta$; el número de valores de gris permisibles es k y el número de bits asignado a cada nivel de gris por píxel es β (usualmente $\beta = 8$, es decir, cada píxel se representa por un byte). En algunas aplicaciones médicas se pueden usar 12, 16 o más bits por píxel.

⁵Teorema: La serie de Fourier de una función f continua y suave por tramos, 2π -periódica, converge uniformemente a f en $[-\pi, \pi]$

⁶Si el rango de frecuencias de una señal medido en ciclos por segundos es n , entonces la señal puede ser representada con completa precisión midiendo su amplitud $2n$ veces por segundo

$$u_0 \rightarrow \boxed{T} \rightarrow F = T[u_0].$$

Figura 6: Procesamiento de una imagen. Sistema entrada-salida

A diferencia del muestreo, la cuantificación es un tópico difícil de analizar, es un proceso no lineal, **no existe una medida matemática de la pérdida de información por cuantificación, aunque sus efectos son visualmente obvios**. Ver [5][Páginas 17 y 18] para conocer más sobre los efectos.

Se puede concluir que una imagen digital en escalas de gris se representa por una matriz de valores enteros positivos acotados.

3.2.1. Tamaño de una imagen de datos

El volumen de datos es clave en el procesamiento, almacenamiento, transmisión y visualización de la información de una imagen.

El espacio requerido para una imagen digital monocromática de $N \times M$ y B bits de resolución de niveles de gris es NMB bits. La siguiente tabla fue tomada de [5], página 22,

Spatial dimensions	Pixel resolution (bits)	Image type	Data volume (bytes)
128 × 128	1	Monochromatic	2,048
256 × 256	1	Monochromatic	8,192
512 × 512	1	Monochromatic	32,768
1,024 × 1,024	1	Monochromatic	131,072
128 × 128	8	Monochromatic	16,384
256 × 256	8	Monochromatic	65,536
512 × 512	8	Monochromatic	262,144
1,024 × 1,024	8	Monochromatic	1,048,576
128 × 128	3	Trichromatic	6,144
256 × 256	3	Trichromatic	24,576
512 × 512	3	Trichromatic	98,304
1,024 × 1,024	3	Trichromatic	393,216
128 × 128	24	Trichromatic	49,152
256 × 256	24	Trichromatic	196,608
512 × 512	24	Trichromatic	786,432
1,024 × 1,024	24	Trichromatic	3,145,728

3.3. Cuestiones de notación

Supongamos $f(n)$ definida sobre un espacio discreto bidimensional $n = (n_1, n_2)$ o $n = (M, N)$. La imagen tiene soporte finito con dominio $[0, M - 1] \times [0, N - 1]$.

Los datos de la imagen pueden estar contenidos en una matriz o array de dimensiones $M \times N$.

La imagen en espacio discreto se obtiene muestreando una imagen continua $f(x, y)$. Más aún, se asume que la imagen $f(n)$ está cuantizadas en k niveles $0, \dots, k - 1$ y cada pixel tiene uno de esos valores enteros. Como $f(n)$ es resultado del muestreo y la cuantización, es una imagen digital.

4. Procesamiento de imágenes. Operaciones básicas

La imagen de entrada u_0 , previamente adquirida, puede estar degradada debido a pobres condiciones en su adquisición o problemas durante su almacenamiento y comunicación. Entonces, ¿Cuál es el propósito del procesamiento de una imagen?

- **Mejoramiento (en forma subjetiva) (Image Enhancement):** mejorar la calidad de la imagen, usualmente aumentando el contraste, aumento de la nitidez, de la intensidad, eliminando ruido.
- **Compresión (Image Compression):** usar la menor cantidad posible de bits para representar la imagen, sin comprometer su calidad.
- **Restauración (Image Restoration):** mejoramiento de forma objetiva, revertir un daño causado por una fuente conocida; por ejemplo reducir el emborronamiento, eliminar distorsión óptica.
- **Extracción de información (Feature Extraction):** Resaltar (hacer explícitas) ciertas características de la imagen que ayuden a identificar objetos. Segmentar, resaltar bordes son formas de identificar objetos y estructuras presentes en la imagen.

Hay textos [Sungaya2018] que consideran el mejoramiento “subjetivo” como una etapa de Preprocesamiento. Los algoritmos que se usan para cada una de las tareas no representan clases disjuntas, puede que un mismo algoritmo se use por ejemplo en el mejoramiento y la restauración de imágenes.

Los métodos para lograr el mejoramiento (Enhancement) se pueden agrupar en

- Métodos de filtrado lineal: suavizado y filtrado de paso bajo; acentuación (sharpening) o filtrado de paso alto.
- Manipulación del Histograma
- Algoritmos de desemborronamiento genérico
- Algoritmos de eliminación de ruido sin emborronamiento

Toda operación en el procesamiento de imágenes transforma los valores de gris en los píxels, sin embargo, las operaciones se pueden dividir en clases, basadas en la información requerida para realizar la transformación. Por ejemplo si el valor del píxel se transforma sin tener conocimiento de lo que lo rodea, se habla de operaciones puntuales; si para cambiar el nivel de gris del píxel se necesita conocer el valor de los niveles de gris en una pequeña vecindad que encierra el píxel en cuestión se habla de filtrado espacial y si se requiere información de todos los valores de gris en la imagen completa se habla de transformación.

4.1. Operaciones básicas en el preprocesamiento de imágenes

En el nivel más básico se habla de tres clases de **Operaciones** en el procesamiento digital de imágenes en escala de grises, las cuales pueden actuar sobre la reducción del ruido, el incremento de la nitidez, la mejora del contraste por solo mencionar algunos efectos.

1. **Operaciones puntuales:** Son aquellas que manipulan cada píxel de manera individual, independientemente de sus vecinos. Estas operaciones son definidas solamente como funciones de la intensidad de los píxels, en el caso continuo se representa como

$$g(x, y) = h(f(x, y))$$

y en el caso discreto

$$g(i, j) = h(f(i, j)).$$

Las operaciones aritméticas entre imágenes de la misma dimensión son también consideradas operaciones puntuales, se pueden usar por ejemplo para la detección de movimiento (image differencing)

2. **Operaciones de vecindad o de filtrado espacial** El nuevo valor de intensidad en cada píxel depende de una cantidad reducida de valores en una vecindad del píxel.

3. **Operaciones geométricas** Son complementarias a las operaciones puntuales, en el sentido de que no están definidas como función de la intensidad de la imagen, en su lugar son función de la posición espacial. Este tipo de operaciones cambian la apariencia de la imagen cambiando las coordenadas de intensidades (rotación o traslación de la imagen) y produce efectos especiales, pero no modifica el valor de intensidad en los píxels. Los efectos geométricos con tiempo variable son considerados en *Computer Graphics*.

Otras aproximaciones en el procesamiento de imágenes consideran una complejidad mayor en los operadores, en la Figura 6 representado por T , que actúa sobre la imagen de entrada y produce el objetivo deseado. Se pueden mencionar como las aproximaciones más antiguas las basadas en el Análisis Espectral (Fourier) y en conceptos básicos de probabilidades y estadísticas. En los últimos 25 años se han desarrollado aproximaciones basadas en la modelación estocástica, la aproximación con funciones wavelets (análisis espacio-escala), aproximaciones basadas en Ecuaciones Diferenciales Parciales, y las basadas en métodos variacionales, entre otras.

El reto mayor en el diseño y cálculo de T es que la mayoría de los problemas de procesamiento de imágenes son problemas inversos mal planteados. Conocida F (digamos los bordes), se puede conocer la imagen original u_0 . Sin embargo, en la mayoría de los problemas en procesamiento de imágenes se trata de recuperar F a partir de u_0 (Problema inverso). Mal planteado significa problemas en la unicidad y estabilidad de la solución. La salida F puede ser alguna colección de rasgos que visualmente tienen algún significado p.e. localización de esquinas, fronteras de los objetos, regiones disjuntas asociadas con distintos objetos o el orden relativo o profundidad de diferentes objetos en la escena.

4.1.1. Operaciones puntuales

Las operaciones puntuales son particularmente ventajosas en el pre-procesamiento. Por ejemplo, operaciones lineales puntuales

$$g(x, y) = Mf(x, y) + L,$$

donde con $M > 0$ denotamos una acción multiplicativa sobre la intensidad de la escala de gris en el píxel, que a veces es llamada ganancia (gain) y se dice que controla el contraste; con L una acción aditiva al valor de la intensidad, que es llamada sesgo (bias) y se dice que controla la intensidad del brillo, veamos si $M = 1$

$$g(x, y) = f(x, y) + L,$$

entonces si $L > 0$, la imagen g tendrá valores de intensidad más brillantes que la función original. Hay que ser cuidadosos a la hora de elegir L , porque puede suceder que el valor de intensidad resultante se salga del rango definido. El Histograma de ambas imágenes cumplirán con la relación

$$H_f(k) = H_g(k - L),$$

es decir se produce un corrimiento del Histograma hacia la izquierda o hacia la derecha. La herramienta básica para diseñar operaciones punto a punto es el **Histograma** de la imagen.

Controlando la intensidad del brillo y la ganancia se puede mejorar la apariencia de la imagen, sin embargo, cómo saber ¿cuáles serían los mejores valores?. Según se plantea en la literatura, una posibilidad sería buscar el valor más oscuro y el más brillante y hacerle corresponder el negro puro y el blanco puro respectivamente; o encontrar el valor de intensidad promedio en la imagen, ubicarlo en el gris medio y expandir el rango para que estén más cercanos todos los posibles valores. Pero cómo visualizar el conjunto de valores de luminosidad en una imagen para probar tales heurísticas. La respuesta es plotear el Histograma!!

4.2. Histograma de una imagen

El histograma H_f es un gráfico de la frecuencia de aparición de cada nivel de gris en una imagen f , es decir, simplemente el conteo de la cantidad de píxels en la imagen que toman cierto valor de gris. Supongamos una imagen con niveles de gris en el rango $[0, L - 1]$, entonces el histograma es una función discreta

$$H_f(n g_k) = \frac{n_k}{MN}, \quad (20)$$

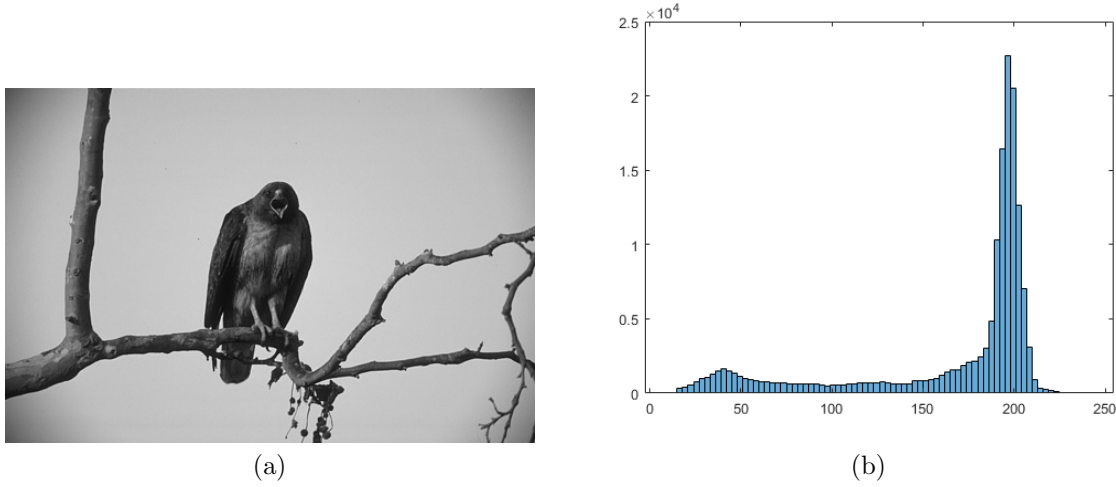


Figura 7: (a) Imagen original, (b) Histograma

donde ng_k es el k -ésimo nivel de gris, n_k es el número de píxeles en la imagen con ese nivel de gris, MN es el número total de píxeles y $k = 0, 1, \dots, L - 1$. H_f es una función unidimensional definida sobre $0, \dots, L - 1$ y el rango se extiende desde 0 hasta el número de píxeles en la imagen MN . Cuando esta función se normaliza, es decir que la suma sobre todos los píxeles de los valores de gris sea igual a 1, entonces puede ser vista como una función de densidad de probabilidad, que expresa la probabilidad de que cierto valor de gris pueda ser encontrado en una imagen. Visto de esta forma el valor de gris de un píxel se convierte en una variable aleatoria que toma valores de acuerdo a la salida de un experimento aleatorio.

Como el histograma representa una reducción de dimensión relativa (se pierde información) a la imagen original, la imagen original no se puede deducir del histograma H_f excepto en casos triviales.

Aunque el histograma no contiene información espacial sobre f , la información es muy rica y de gran utilidad. En una imagen oscura, los niveles de gris estarán agrupados en el extremo inferior de la escala en el histograma. En una imagen con brillo uniforme, los niveles de gris estarán agrupados en el extremo superior de la escala. En una imagen con buen contraste, los niveles de gris estarán bien distribuidos sobre casi todo el rango de niveles.

Muchas operaciones del procesamiento de imágenes se pueden derivar del histograma de la imagen.

Por ejemplo, la densidad óptica promediada (*average optical density* o AOD), que es la medida básica de una imagen, promedia la brillantez o nivel de gris, y se puede calcular directamente de la imagen:

$$AOD(f) = \frac{1}{NM} \sum_{n_1=0}^{N-1} \sum_{n_2=0}^{M-1} f(n_1, n_2) \quad (21)$$

o por medio del Histograma

$$AOD(f) = \frac{1}{NM} \sum_{k=0}^{L-1} k H_f(k) \quad (22)$$

O si en una imagen no se pueden percibir los detalles, la razón puede ser que píxeles que representan diferentes objetos o diferentes partes de un mismo objeto tenga valores de intensidad de gris similares, este sería el caso de un Histograma “comprimido” a un rango pequeño de escalas de gris, que indica una imagen con un pobre contraste visual. Por el contrario un histograma bien distribuido en todo el rango de escalas de gris es signo de una imagen en la que se pueden apreciar bien los detalles, ver páginas 49-50 del pdf en [5]. De aquí que la manipulación del Histograma sea una de las operaciones más simples en el procesamiento de imágenes.

4.2.1. Manipulación del Histograma

La forma más simple es lo que se conoce como “stretching”

Supongamos que se tiene una imagen de bajo contraste, es decir los valores de gris se encuentran en un rango pequeño de valores $[g_{min}, g_{max}]$. Si se quiere ampliar el rango de valores al intervalo $[0, G - 1]$, siendo $G - 1 > g_{max} - g_{min}$, entonces se deben reemplazar los valores de gris según la relación

$$g_{nuevo} = \frac{g_{viejo} - g_{min}}{g_{max} - g_{min}} G + 0,5$$

Con esta expresión no se modifica la cantidad de píxeles en cada nivel de gris, sino que los valores de gris como que se expanden, se dispersan más. El 0,5 se suma para redondear al entero más cercano (Operación $\lfloor \cdot \rfloor$) **Tarea:**

1. Seleccione una imagen y realice las operaciones precedentes y concluya qué sucede con el Histograma
2. Implemente operaciones puntuales para encontrar el negativo de una imagen
3. Dada una imagen con escalas de gris $0, \dots, K - 1$, donde el mayor valor de intensidad es B y el menor valor de intensidad es A . Implemente $g(x, y) = \frac{K-1}{B-A}(f(x, y) - A)$ (FSHS) Full Scale Histogram Stretch y experimente con imágenes con Histogramas con diferentes distribuciones. Concluya.

Estudio independiente: Operaciones puntuales no lineales (Ecualización del Histograma, ver página 391 pdf Maria Petrou)

4.2.2. Umbralización

1. Umbralización simple: Por ejemplo en una imagen en escalas de gris $f : [x_0, x_1, \dots, x_N] \times [y_0, y_1, \dots, y_N] \rightarrow [0, 255]$ se transforma en una imagen binaria,
 - si para el nivel de gris de un píxel, n_i $n_i > T$, transformar a blanco ($n_i = 255$),
 - si para el nivel de gris de un píxel, n_i $n_i < T$, transformar a negro ($n_i = 0$)siendo T un umbral en los niveles de gris. La umbralización es importante en la segmentación de imágenes, donde se quiere aislar un objeto de su entorno, también es importante en la visión robótica.
2. Umbralización doble:
 - si para el nivel de gris de un píxel, n_i $T_1 < n_i < T_2$, transformar a blanco (255),
 - en otro caso, transformar a negro (0)

La umbralización doble se usa en la detección de bordes

La umbralización es de gran utilidad cuando se quieren eliminar detalles innecesarios y resaltar lo esencial; sacar a la luz detalles escondidos; eliminar un fondo de un texto o pintura.

4.3. Transformación de una imagen mediante operadores lineales

Cuando la entrada es una imagen degradada u_0 , afectada por algún tipo de ruido, o con bajo contraste, o emborronada, entonces antes de extraer cualquier tipo de información, se necesita mejorar la calidad de la imagen, esto se conoce como preprocesamiento; la salida F puede ser su versión ideal mejorada, denotada como u .

Para realizar transformaciones en las imágenes, los operadores pueden ser desde el punto de vista matemático, lineales o no lineales. Los operadores lineales tienen muchas características en común independientemente de la tarea para la que se usen, a diferencia de los operadores no lineales, que en general son específicos para determinada aplicación o tarea. Por definición un operador T es lineal si dadas dos imágenes cualesquiera f y g , y valores escalares α y β , se cumple

$$T[\alpha f + \beta g] = \alpha T(f) + \beta T(g)$$

Los operadores lineales se definen en término de lo que se conoce como su función respuesta al impulso puntual, “point spread function” (PSF) en inglés. La PSF de un operador T es la función respuesta que se obtiene si se aplica

el operador T a un impulso puntual matemáticamente definido mediante la función generalizada o función Delta de Dirac, [CITAR])

$$T[\delta(a-x, b-y)] = h(x, a, y, b),$$

donde $\delta(a-x, b-y)$ es una fuente (impulso) puntual de intensidad unitaria, centrada en el punto (x, y) . La PSF h expresa el efecto del operador sobre la fuente puntual. PSF se usa también para caracterizar equipos y dispositivos de captura de imágenes, ver [5], [CITAR otros].

Si el operador es lineal y la fuente puntual es m -veces más brillante (intensa) entonces la PSF del operador será m -veces más alta

$$T[m\delta(a-x, b-y)] = mh(x, a, y, b).$$

Una imagen digital (como se vio más arriba), es una colección de fuentes (impulsos) puntuales (los píxeles), cada una con su propio valor de intensidad (brillantez). Por ejemplo si asumimos que f es una imagen representada por una matriz de 3×3 , entonces

$$f = \begin{pmatrix} f(1,1) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & f(1,2) & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \dots + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & f(3,3) \end{pmatrix}, \quad (23)$$

se puede decir que la imagen es la suma de estas fuentes (o impulsos) puntuales y el efecto de un operador T con PSF $h(x, a, y, b)$ sobre una imagen $f(x, y)$ se puede representar como

$$g(a, b) = \sum_{x=1}^N \sum_{y=1}^N f(x, y)h(x, a, y, b), \quad (24)$$

donde $f(x, y)$ es la imagen de entrada, $g(a, b)$ es la imagen de salida, ambas con dimensión $N \times N$. Veamos qué lectura le damos a la expresión (24); como $f(x, y)$ representa la intensidad del brillo en la fuente puntual en la posición (x, y) ; cuando se aplica el operador T sobre ésta, se obtiene la función *PSF* del operador multiplicada por la intensidad de la fuente y como el operador es lineal, se suma sobre todos los píxeles (fuentes puntuales).

Por otra parte la PSF $h(x, a, y, b)$ expresa cuánto influye el valor de entrada en la posición (x, y) en el valor de salida en la posición (a, b) . Propiedades de la PSF:

- Si la influencia expresada por la PSF (el valor que toma) es independiente de la posición actual, pero depende de la posición relativa entre el píxel que ejerce influencia (el de salida) y el píxel influenciado (el de entrada), entonces se dice que la PSF es invariante a la traslación (Ver ejemplo 1.6, página 37 Petrou)

$$h(x, a, y, b) = h(a-x, b-y),$$

entonces la ecuación (24) se convierte en una convolución

$$g(a, b) = \sum_{x=1}^N \sum_{y=1}^N f(x, y)h(a-x, b-y) \quad (25)$$

- Si la influencia sobre las filas se realiza de manera independiente a la influencia sobre las columnas de la imagen, entonces se dice que la PSF es separable

$$h(x, a, y, b) = h_c(x, a)h_f(y, b).$$

Entonces la ecuación (24) se escribe como una cascada de transformaciones en una dimensión

$$g(a, b) = \sum_{x=1}^N h_c(x, a) \sum_{y=1}^N f(x, y)h_f(y, b) \quad (26)$$

- Si la PSF es invariante a traslaciones y separable, entonces la ecuación (24) se puede escribir como una cascada de convoluciones en una dimensión

$$g(a, b) = \sum_{x=1}^N h_c(a-x) \sum_{y=1}^N f(x, y)h_f(b-y) \quad (27)$$

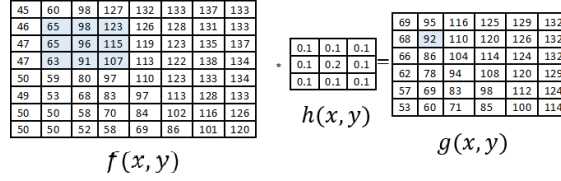


Figura 8: Píxeles sombreados en azul indican los píxeles procesados para obtener el píxel sombreado a la derecha

TAREA: Implementar el ejemplo 1.5 de la página 36 pdf del Petrou.

Resumiendo, si se considera la expresión (24), se puede decir que en la práctica el efecto de un operador lineal sobre una imagen se puede expresar con la ayuda de operaciones matriciales. Reescribiendo la matriz de intensidades que representa a una imagen $\{f(i, j)\}_{i,j=1,..N}$ como vector,

$$\mathbf{f}^T = (f(1, 1), f(2, 1), \dots, f(N, 1), f(1, 2), f(2, 2), \dots, f(N, 2), \dots, f(1, N), f(2, N), \dots, f(N, N))$$

ordenando consecutivamente por las columnas y haciendo lo mismo con la matriz de la PSF

$$h_{a,b}^T = (h(1, a, 1, b), h(2, a, 1, b), \dots, h(N, a, 1, b), h(1, a, 2, b), h(2, a, 2, b), \dots, h(N, a, 2, b), h(1, a, N, b), h(2, a, N, b), \dots, h(N, a, N, b)),$$

la expresión (24) se puede ver como el producto escalar del vector $h_{a,b}^T$ con el vector f^T . Si pensamos lo mismo para el cálculo de cada uno de las componentes de la imagen de salida $g(a, b)$, considerando que cada vector $h_{a,b}^T$ para b fijo y haciendo variar $a = 1, ..N$ es la fila de una matriz H , entonces la ecuación (24) se puede escribir de forma compacta como

$$g = Hf \quad (28)$$

donde H es una matriz de $N^2 \times N^2$, formada de $N \times N$ submatrices cada una de dimensión $N \times N$. En un lenguaje coloquial, los elementos en cada fila de H son matrices de dimensión $N \times N$ y en cada fila se tienen N - matrices, y lo mismo para las columnas.

La ecuación (28) es la ecuación fundamental del procesamiento lineal de imágenes.

4.3.1. Filtros espaciales lineales

En el lenguaje de procesamiento de imágenes es más común oír hablar de **filtros lineales** y no de operadores lineales y como además estaremos hablando de procesamiento de imágenes digitales pues nos referimos a filtros lineales en lugar de operadores de vecindad lineales, en el cual el valor de intensidad de un píxel en la imagen de salida se determina por la suma con pesos aplicada a los valores de intensidad de los píxeles de la imagen de entrada

$$g(i, j) = \sum_{k=1}^N \sum_{l=1}^N f(i + k, j + l) h(k, l) \quad (29)$$

Esta es una ecuación también llamada de correlación, donde h es llamada máscara (kernel) o coeficientes del filtro. La ecuación (29) se puede escribir como

$$g = f \otimes h.$$

Otra variante de escribir (29) es

$$g(i, j) = \sum_{k=1}^N \sum_{l=1}^N f(i - k, j - l) h(k, l) = \sum_{k=1}^N \sum_{l=1}^N f(k, l) h(i - k, j - l) \quad (30)$$

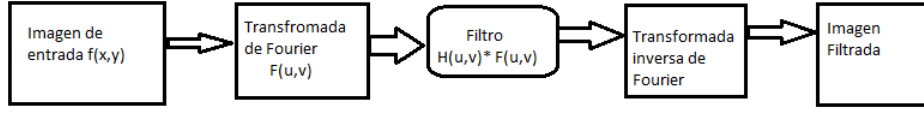


Figura 9: Filtrado en el dominio de Fourier

y se llama operador de convolución $g = f * h$ (en analogía a la ecuación (25)), recordar que h es la función de respuesta al impulso⁷, debido a que h convolucionada con una señal impulso

$$\delta(m - p, n - q) = \begin{cases} 1 & m = p \text{ y } n = q \\ 0 & \text{e.o.c.} \end{cases}$$

se reproduce a si misma. La convolución discreta se puede calcular de forma rápida usando la Transformada Discreta de Fourier (en su forma eficiente, FFT), pero si la dimensión del filtro es pequeña el cálculo directo puede ser más rápido.

5. Procesamiento de imágenes en el dominio de Fourier

Como se ha expresado, el procesamiento de imágenes tiene como objetivo mejorar el aspecto de las imágenes, hacer más evidentes ciertos detalles que se quieren resaltar. La variación de brillantez medida a lo largo de una dirección cualquiera es una función que se puede representar como suma de distribuciones sinusoidales de varias frecuencias; atenuar o reforzar algunas de estas componentes puede tener un efecto dramático en la calidad de la imagen y cuando hablamos de la información de frecuencias de una función pues inmediatamente nos remitimos al postulado de Fourier de que bajo ciertas condiciones, una función periódica (no periódica) puede ser representada como suma (la integral) de exponenciales complejas o sinusoides. La manipulación del contenido de frecuencias de una función (imagen) se realiza mediante los llamados filtros en el dominio de frecuencias y el fundamento de las técnicas basadas en el dominio de frecuencias es el teorema de convolución, como se observa en el diagrama de la Figura 9

En la representación mediante la fórmula de Euler

$$e^{2\pi i u x} = \cos(2\pi u x) + i \sin(2\pi u x)$$

la variable u suele recibir el nombre de frecuencia, ya que u representa el número de ciclos completos por unidad de distancia en x . La periodicidad espacial de esta función es $\frac{1}{|u|}$. Por tanto la representación en serie de Fourier (o mediante la transformada inversa de Fourier) puede considerarse como una combinación lineal de patrones periódicos de la forma $e^{2\pi i u x}$.

El filtrado lineal de una señal (imagen) puede ser visto como un escalamiento controlado de las componentes de la señal en el dominio de frecuencias. De ahí que este proceso de filtrado se pueda realizar multiplicando la Transformada de Fourier de la señal (imagen) con cierta función que elimine o modifique ciertas componentes de frecuencia y luego aplicar la transformada inversa de Fourier, ver Figura 9. Por ejemplo si se reducen las componentes en el centro del dominio de frecuencias (bajas frecuencias), pues se le está dando más importancia a las altas frecuencias y con esto se ha realizado un filtrado de paso alto y la función que se usa para modificar las frecuencias se llama filtro lineal de paso alto. El filtrado puede ser muy selectivo, ya que los coeficientes de Fourier pueden ser modificados de manera independiente. Entender los conceptos de dominio de frecuencia y de filtrado lineal es esencial para comprender tópicos como: Mejoramiento de video, de imágenes, restauración, compresión, segmentación. Explorar estas ideas en un marco 2D tiene la ventaja de que los conceptos pueden ser visualizados como imágenes. Las bases de Fourier y wavelet constituyen dos herramientas fundamentales para la representación en el dominio de frecuencias de una función. Estas bases permiten descomponer una función en funciones oscilatorias con forma de

⁷ $h * \delta = h$

onda y de esta representación se puede obtener información valiosa sobre determinadas características de la función que se está analizando, además de que es un camino para obtener representaciones dispersas ó más simplificadas. Las transformadas de Fourier y wavelet brindan la posibilidad de manipular eficientemente arreglos de datos de grandes dimensiones N , ya que cuentan con algoritmos rápidos que pueden ser implementados con $O(N \log N)$ operaciones y menos.

Concepto de frecuencia en una imagen Comencemos precisando el concepto de frecuencia en una imagen. La función sinusoidal está caracterizada no solo por su localización (phase shift) y su frecuencia de oscilación sino también por su dirección de oscilación. Si (x, y) representa las coordenadas espaciales, entonces la función $e^{2\pi i(ux+vy)}$ tiene parte real $\cos(2\pi(ux+vy))$ y parte imaginaria $\sin(2\pi(ux+vy))$. Si tomamos una sección paralela al eje x del patrón bidimensional $\cos(2\pi(ux+vy))$, tenemos u ciclos por unidad de distancia y en el eje y tendremos v ciclos, u y v reciben el nombre de frecuencias espaciales. El período principal del patrón es $\frac{1}{\sqrt{u^2+v^2}}$.

En una imagen los componentes de frecuencia vienen siendo las transiciones en los niveles de grises. Altas frecuencias representan transiciones bruscas en los niveles de grises. Concepto de frecuencia también se refiere a la frecuencia con la que una imagen o señal varía como función de las coordenadas espaciales. El espectro de muchas imágenes decrece de forma muy rápida cuando la frecuencia aumenta y por tanto los términos de frecuencias altas tienden a hacerse oscuros muy pronto al mostrarse en una imagen (la luminosidad o intensidad del píxel $\rightarrow 0$). Una técnica es mostrar la imagen del espectro haciendo

$$D(u, v) = \log(1 + |F(u, v)|)$$

Una función sinusoidal en el espacio discreto 2D es una función de la forma:

$$\sin(2\pi(Um + Vn)), \quad (31)$$

donde U y V (con unidades ciclos/píxel) representan respectivamente las frecuencias de oscilación en la dimensión espacial vertical m y horizontal n de la imagen. Las funciones sinusoidales en tiempo o espacio discreto tienen una frecuencia máxima posible a la cual pueden oscilar. Sin embargo

Generalmente una senoide 2D oscila en todas direcciones excepto en la dirección ortogonal a la dirección de oscilación más rápida. La frecuencia de esa oscilación más rápida es la frecuencia radial $\Omega = \sqrt{U^2 + V^2}$, que tiene las mismas unidades que U y V y la dirección de la oscilación más rápida es el ángulo en radianes

$$\theta = \tan^{-1}(V/U). \quad (32)$$

Asociada a la representación (31) está la función exponencial compleja:

$$e^{i2\pi(Um+Vn)} = \cos(2\pi(Um + Vn)) + i\sin(2\pi(Um + Vn)). \quad (33)$$

En general las funciones sinusoidales pueden definirse en mallas de enteros discretas. Las expresiones (31) y (33) se satisfacen para todo (m, n) enteros. Aunque las funciones sinusoidales de duración infinita no se encuentran en la práctica, son útiles para modelar imágenes y en ciertas descomposiciones de imágenes. En la práctica las imágenes en espacios discretos están limitadas a mallas de muestras finitas de dimensión $M \times N$, por lo que es conveniente usar sinusoides en el espacio discreto 2D con extensión finita $M \times N = (m, n) : 0 \leq m \leq M - 1, 0 \leq n \leq N - 1$, m y n enteros.

Una función sinusoidal restringida al dominio $M \times N$ puede estar contenida en una matriz correspondiente a una imagen de dimensiones $M \times N$. En este caso es conveniente usar frecuencias escaladas $(u, v) = (MU, NV)$ que tienen unidades visualmente intuitivas de ciclos/imágenes entonces la función seno definida en (31) se redefine como

$$\sin(2\pi(\frac{u}{M}m + \frac{v}{N}n)), \quad (34)$$

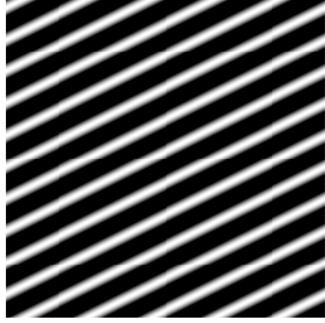
La Fig.(5) ilustra varias sinusoides en espacios discretos de dimensión 256×256 mostrados como imágenes de intensidad después de realizar un mapeo de las escalas de grises de cada una en el rango $0 - 255$. Debido a la respuesta no lineal del ojo humano, las funciones parecen más ondas cuadradas que sinusoides que varían lentamente, particularmente a altas frecuencias. Sin embargo, si cualquiera de las imágenes de la Fig. (5) es muestreada a lo largo de una línea recta de orientación arbitraria, el resultado es una senoide ideal. Una peculiaridad de las sinusoides en espacios discretos (o tiempo discreto) es que existe físicamente una frecuencia máxima a la cual ellos pueden



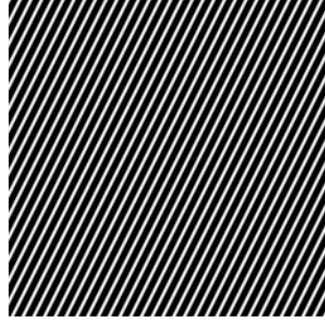
(a)



(b)



(c)



(d)

Figura 10: (a) $\sin(2\pi(\frac{5}{M}m + \frac{0}{N}n))$, (b) $\sin(2\pi(\frac{0}{M}m + \frac{5}{N}n))$, (c) $\sin(2\pi(\frac{5}{M}m + \frac{10}{N}n))$, (d) $\sin(2\pi(\frac{30}{M}m + \frac{15}{N}n))$

oscilar. Aunque las variables de frecuencias (u, v) ó (U, V) pueden tomar valores arbitrariamente grandes, estos no se corresponden con frecuencias de oscilación físicas arbitrariamente grandes. Las ramificaciones de éstas son profundas y significativas y están relacionadas a las restricciones que aparecen en el muestreo de imágenes en espacios continuos (El teorema de muestreo) y la frecuencia de Nyquist.

El procesamiento de imágenes tiene como objetivo mejorar el aspecto de las mismas, y hacer más evidentes en ellas ciertos detalles que se quieren resaltar. La variación de la brillantez de una imagen medida a lo largo de una dirección cualesquiera es una función que se puede representar como suma de distribuciones sinusoidales de varias frecuencias; atenuar o reforzar algunas de estas componentes puede tener un efecto dramático en la calidad de una imagen. Componentes de alta frecuencia en una imagen representan transiciones bruscas en los niveles de grises. Entonces si se quiere suavizar una imagen, es decir eliminar los componentes de altas frecuencias, se usan los llamados filtros en el dominio de frecuencias, que no es más que el dominio de Fourier.

5.1. Transformada de Fourier en 2D

Transformada Continua de Fourier en el espacio bidimensional

Sea $f(x, y)$ el valor de intensidad de una imagen analógica en las ubicaciones de coordenadas espaciales (x, y) .

La transformada continua de Fourier se define por

$$F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-i2\pi(ux+vy)} dx dy,$$

y su inversa

$$f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(u, v) e^{i2\pi(ux+vy)} du dv.$$

Las frecuencias (u, v) como se dijo antes, están en términos de ciclos por unidad de longitud. Si representamos a $F(u, v)$ como una imagen, los píxeles individuales representan componentes espectrales de la entrada original, más que componentes espaciales.

De la representación de $f(x, y)$ en términos de $F(u, v)$ se puede interpretar que $f(x, y)$ es una combinación lineal de patrones periódicos dados por la exponencial compleja y que $F(u, v)$ representan los pesos de la contribución de cada patrón elemental.

Transformada Discreta de Fourier (TDF) en el espacio bidimensional

La transformada discreta (finita) de Fourier es un mapa detallado del contenido de frecuencias de una imagen, que puede ser procesado digitalmente. Examinar la TDF de imágenes, particularmente de sus magnitudes, nos revela mucho acerca de la distribución y significado de la frecuencia en imágenes; así como nos ayuda a entender qué pasa cuando se modifican las frecuencias en una imagen y cómo se pueden modificar las frecuencias en una imagen para obtener resultados útiles.

Sea $f(m, n)$ la versión discreta de una función bidimensional continua f . Entonces la transformada de Fourier finita o TDF se define como:

$$F(u, v) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) e^{-i2\pi(\frac{um}{M} + \frac{vn}{N})} \quad (35)$$

y su inversa

$$f(m, n) = \frac{1}{MN} \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} F(u, v) e^{i2\pi(\frac{um}{M} + \frac{vn}{N})} \quad (36)$$

El dominio de frecuencias es el sistema de coordenadas generado por $F(u, v)$ con u y v como variables de frecuencias, $M \times N$ es el rectángulo de frecuencias, que es de la misma dimensión que la imagen. El coeficiente $\frac{1}{MN}$ a veces se incluye en la transformada y otras en la transformada inversa. La transformada de Fourier es continua en las variables de frecuencia, pero la inversa (es decir la imagen f es discreta). La ecuación (36) representa la descomposición de f en funciones base ortogonales complejas.

Desde el punto de vista ingenieril la Transformada de Fourier se interpreta como que la información de la imagen en el dominio espacial se transforma en información de la imagen en el dominio frecuencia y esta última representación de la información de la imagen tiene importancia conceptual y algorítmica.

Desde el punto de vista físico la imagen es vista como resultado de un sofisticado patrón (constructivo-destructivo) de ondas de interferencia. Asignando a cada uno de los infinitos valores de la función de onda exponencial compleja $e^{i2\pi(Um+Vn)}$ con pesos apropiados $F(u, v)$. La estructura de cualquier imagen en el espacio discreto puede ser recreada exactamente como una suma de interferencias. Aunque (35) está definida para $u = 0, 1, \dots, M-1$ y $v = 0, 1, \dots, N-1$, si usamos la misma ecuación para evaluar $F(u, v)$ descubrimos propiedades de periodicidad:

$$F(-u, -v) = F(M - u, N - v) \quad (37)$$

$$F(aM + u, bN + v) = F(u, v) \quad (38)$$

$$f(-m, -n) = f(M - m, N - n) \quad (39)$$

$$f(aM + m, bN + n) = f(m, n) \quad (40)$$

Si se tiene un arreglo finito de números $f(m, n)$ y su transformada de Fourier $F(u, v)$, y se desea extender su definición más allá del dominio original, entonces dicha extensión es una representación periódica de dichos arreglos.

La TDF 2-D no es más que aplicar la TDF unidimensional, primero por filas y luego por columnas

$$\hat{f}_k = \left\langle \vec{f}, \vec{w}_k \right\rangle_N = \frac{1}{N} \sum_{m=0}^{N-1} f_m \cdot e^{-ikm2\pi/N}.$$

El coeficiente de Fourier $\hat{f}_{k,l}$ en la fila k y la columna l corresponde al producto tensorial $\vec{w}_k \times \vec{w}_l$

$$(\vec{w}_k \times \vec{w}_l)_{p,q} = (\vec{w}_k)_p (\vec{w}_l)_q = e^{ikp2\pi/M} e^{ilq2\pi/M}$$

5.2. Filtros en el dominio de frecuencias

En la práctica para el filtrado lineal se usan más las máscaras espaciales pequeñas que la transformada de Fourier, debido a su fácil implementación, sin embargo la idea de filtrado lineal es más intuitiva en el dominio de frecuencias, por lo que es importante comprender estos conceptos. El filtro lineal más fácil de diseñar es el que tiene en cuenta un caso ideal, de que la señal o la imagen estén limitadas en una banda de frecuencias, de ahí que los filtros diseñados de forma tal que eliminen las frecuencias que estén fuera de un determinado radio o que eliminen las frecuencias fuera de un rectángulo son llamados filtros de paso bajo ideal. Es decir lo que se quiere es que el soporte de la Transformada de Fourier de la función $f(t)$, $\hat{f}(\omega)$ esté en el intervalo $(-\omega_c, \omega_c)$. Con este tipo de filtros se eliminan las altas frecuencias (transiciones bruscas en los niveles de intensidad), las cuales son indicadores de la presencia de ruido o de bordes (fronteras). Por tanto será suficiente con diseñar el operador lineal de eliminación de ruido basado en un operador de truncamiento espectral. Usando como función de transferencia la función característica

$$\chi_{[-\omega_c, \omega_c]}(\omega) = \begin{cases} 1 & \omega \in [-\omega_c, \omega_c] \\ 0 & e.o.c. \end{cases} \quad (41)$$

se puede definir un filtro de paso bajo ideal

$$\hat{f}_{\omega_c}(\omega) = \chi_{[-\omega_c, \omega_c]}(\omega) \hat{f}(\omega). \quad (42)$$

Con esto la señal original $f(t)$ estaría limitada en la banda de frecuencias $|\omega| \leq \omega_c$ y se obtiene como

$$f_{\omega_c}(t) = \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} e^{i\omega t} \hat{f}(\omega) d\omega \quad (43)$$

Por supuesto que si $f(t)$ es ya una función (señal) limitada en una banda de frecuencias, entonces el filtrado de paso bajo no cambia la señal original $f(t)$, esto es $f_{\omega_c}(t) = f(t)$.

Sea $l_{\omega_c}(t)$ la transformada de Fourier inversa de la función de transferencia $\chi_{[-\omega_c, \omega_c]}(\omega)$, entonces se tiene

$$l_{\omega_c}(t) = \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} e^{i\omega t} d\omega = \frac{\text{sen}(\omega_c t)}{\pi t}. \quad (44)$$

Esta función se conoce como función sampling de Shannon (o para $\omega_c = \pi$ función de escala de Shannon), ver (17) en el teorema de muestreo, ver (15). Entonces de la igualdad (42) y por el teorema de Convolución (fundamento de las técnicas basadas en el dominio de frecuencias) se tiene que un filtrado analógico está definido por una convolución en el dominio del tiempo.

Definición 6 Convolución en 1-D: Supongamos que f y g son dos funciones de cuadrado integrable. La convolución de f y g , denotada por $f * g$ se define por:

$$(f * g)(t) = \int_{-\infty}^{\infty} f(t-x)g(x)dx \quad (45)$$

y mediante un cambio de variable es equivalente a

$$(f * g)(t) = \int_{-\infty}^{\infty} f(x)g(t-x)dx \quad (46)$$

Para el caso bidimensional se tiene

$$(f * g)(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\alpha, \beta) g(x - \alpha, y - \beta) d\alpha d\beta \quad (47)$$

Teorema 7 (*Teorema de Convolución*)

Sean f y g dos funciones integrables. Entonces

$$\widehat{(f * g)} = \sqrt{2\pi} \widehat{f} \cdot \widehat{g} \quad (48)$$

$$F^{-1}(\widehat{f} \cdot \widehat{g}) = \frac{f * g}{\sqrt{2\pi}} \quad (49)$$

Teorema 8 Supongamos f y g de cuadrado integrables. Entonces:

$$\langle \widehat{f}, \widehat{g} \rangle_{L^2} = \langle f, g \rangle_{L^2} \quad (50)$$

$$\langle F^{-1}(f), F^{-1}(g) \rangle_{L^2} = \langle f, g \rangle_{L^2} \quad (51)$$

En particular: $\|\widehat{f}\|_{L^2} = \|f\|_{L^2}$

En el caso bidimensional las transformadas de Fourier de f y g serán $F(u, v)$ y $G(u, v)$ respectivamente.

Teorema 9 Si $g(x, y)$ es la imagen formada por la convolución de una imagen $f(x, y)$ y un operador lineal invariante $h(x, y)$, es decir

$$g(x, y) = (h * f)(x, y),$$

entonces por el teorema de convolución se tiene la siguiente relación en el dominio de frecuencias

$$G(u, v) = H(u, v)F(u, v),$$

donde $F(u, v)$ es la transformada de Fourier de la imagen original y $H(u, v)$ es un filtro atenuador de frecuencias (filtro en el dominio de frecuencias), (ver (42)).

La convolución es una forma simple de afectar el contenido de frecuencia de una imagen (o una señal) computacionalmente se calcula fácilmente en el dominio de frecuencia, ya que viene dada por una multiplicación puntual. Si retomamos (42) y (44) para el caso unidimensional

$$f_{\omega_c}(t) = l_{\omega_c} * f(t) = \int_{-\infty}^{\infty} l_{\omega_c}(t - x) f(x) dx \quad (52)$$

y sustituyendo $l_{\omega_c}(t)$

$$l_{\omega_c} * f(t) = \int_{-\infty}^{\infty} \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} e^{i\omega(t-x)} d\omega f(x) dx \quad (53)$$

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} e^{i\omega t} e^{-i\omega x} d\omega f(x) dx \quad (54)$$

$$= \underbrace{\frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} e^{i\omega t} d\omega}_{\widehat{l}_{\omega_c}(\omega)} \underbrace{\int_{-\infty}^{\infty} e^{-i\omega x} f(x) dx}_{\widehat{f}(\omega)} \quad (55)$$

Entonces la convolución en el tiempo dada por (52), define un filtrado analógico, donde $l_{\omega_c}(t)$ es la función filtro, y en el dominio de frecuencias $\widehat{l}_{\omega_c}(\omega) = \chi_{[-\omega_c, \omega_c]}(\omega)$ es un filtro característico ideal de paso bajo y se tiene que:

$$\widehat{f}_{\omega_c}(\omega) = H(\omega) \widehat{f}(\omega). \quad (56)$$

donde $H(\omega) = \widehat{l}_{\omega_c}(\omega)$ Muchos problemas de mejorar imágenes pueden ser formulados como sigue:

- Dado $f(x, y)$, una vez calculada $F(u, v)$, el objetivo es seleccionar $H(u, v)$ de forma que la imagen deseada

$$g(x, y) = F^{-1}[H(u, v)F(u, v)] \quad (57)$$

muestre algunos rasgos de interés de $f(x, y)$.

Por ejemplo los bordes en la imagen $f(x, y)$ pueden ser acentuados usando una función que enfatice las altas frecuencias de $F(u, v)$ (Filtro de paso alto ideal). El emborronamiento (o suavizado) se alcanza en el dominio de Fourier atenuando un rango especificado de componentes de altas frecuencias en la transformada de una imagen dada

La función $h(x, y)$ caracteriza un sistema cuyo objetivo es producir una imagen de salida $g(x, y)$ a partir de $f(x, y)$. En el espacio de Fourier, sería seleccionar una función de transferencia $H(u, v)$ que produzca $G(u, v)$, ver Figura (9)

5.3. Filtros ideales en imágenes

Definición 10 *Un filtro ideal bidimensional de paso bajo es uno cuya función de transferencia cumple la relación*

$$H(u, v) = \begin{cases} 1 & , \text{ si } D(u, v) \leq D_0 \\ 0 & , \text{ e.o.c} \end{cases}$$

D_0 frecuencia de corte radial, $D(u, v) = (u^2 + v^2)^{1/2}$ es la distancia del punto (u, v) al origen en el plano de frecuencias

El nombre de ideal proviene de que todas las frecuencias dentro de un círculo de radio D_0 se pasan sin atenuación, mientras que las que están fuera de ese círculo son atenuadas completamente. Los filtros de paso bajo ideal producen efectos negativos tanto en el dominio de frecuencia como en el dominio espacial, ya que al eliminar las altas frecuencias que son las que caracterizan los bordes, emborronan la imagen en algún sentido.

Un filtro que supera dichas limitaciones es el filtro Gaussiano, que discretizado (respuesta al impulso) viene dado por:

$$h(m, n) = \frac{1}{2\pi\sigma^2} e^{-(m^2+n^2)/2\sigma^2}, \quad 0 \leq m \leq M-1, \quad 0 \leq n \leq N-1$$

Esta respuesta es infinita, pero decrece rápidamente fuera del origen. En este caso la respuesta de frecuencia está aproximada por

$$H(U, V) \cong e^{-2\pi^2\sigma^2(U^2+V^2)}, \quad (58)$$

para $|U|, |V| < 1/2$. En el caso de mallas finitas se usan frecuencias escaladas $(u, v) = (MU, NV)$, si $U = u/M, V = v/N$. El pico medio del ancho de banda de frecuencia radial de (58) es $\Omega_c = \frac{1}{\pi\sigma} \sqrt{\ln \sqrt{2}} \approx 0,187/\sigma$.

Si es posible encontrar una frecuencia de corte apropiada Ω_c , entonces la frecuencia de corte se puede fijar haciendo $\sigma = 0,187/\Omega_c$ píxels. Al filtrar una imagen con un filtro Gaussiano para valores crecientes de σ , las estructuras presentes en escalas pequeñas tales como ruido y detalles son reducidas grandemente. Con este filtrado se produce una sucesión de imágenes que se conoce como espacio-escala Gaussiano, donde cada imagen es calculada mediante la convolución de la imagen original con un filtro Gaussiano de valor σ creciente. El espacio escala Gaussiano puede ser pensado como que se desarrolla en el tiempo,

$$g_t = h_\sigma * f$$

$$\sigma = \sqrt{t}.$$

Un filtro ideal bidimensional de paso alto es una función de transferencia que cumple

$$H(u, v) = \begin{cases} 0 & , \text{ si } D(u, v) \leq D_0 \\ 1 & , \text{ e.o.c} \end{cases}$$

6. Métodos basados en Ecuaciones Diferenciales Parciales. Modelos de difusión

El análisis de bajo nivel de imágenes persigue extraer información geométrica local, llamada con frecuencia características de las imágenes digitales. Dicha información es usada para comparar imágenes, lo cual puede ser útil por ejemplo para detectar movimiento o para reconstruir determinadas formas, entre otros. Las imágenes que observamos es el resultado de un suavizado del flujo de fotones original, por lo cual es continua. Sin embargo existen investigaciones [Gosseau and Alvarez] que prueban que las imágenes físicas reales tienen variación no acotada, es

decir que se comportan de una manera fuertemente oscilatoria, por lo que es lógico aceptar que en las imágenes digitales existen discontinuidades en la transición de un objeto a otro observado (es decir con el contorno aparente de un objeto físico). Uno de los objetivos del análisis de imágenes ha sido siempre la detección de tales discontinuidades, tarea que se conoce como detección de bordes.

Desde las primeras investigaciones en visión se detectó que en una imagen se pueden encontrar tales discontinuidades casi donde quiera, debido a las oscilaciones que aparecen luego de la digitalización de las imágenes. Como consecuencia el proceso de análisis de imágenes se concibe como un proceso de suavizado, que permita decantar entre los verdaderos bordes y el ruido inherente.

Precisamente como antes de calcular cualquier derivada se requiere un proceso de suavizado, es que se propuso la ecuación de la conducción del calor con la doctrina del espacio-escala. La ecuación de la conducción (modelo de difusión lineal) ha sido considerada relevante como modelo de suavizado de imágenes compatible con las condiciones de Shannon y la ecuación de la conducción inversa ha sido considerada como modelo de restauración propuesto por Gabor. Desde la década de los 80 muchos han sido los modelos de EDP's, fundamentalmente no lineales empleados en diversas tareas del procesamiento de imágenes (eliminación de ruido, desemborronamiento, detección de bordes entre otros). La doctrina del espacio-escala postula que en lugar de hablar de las características de una imagen en una localización dada, se debe hablar de dichas características en una localización dada, y en una escala dada. De ahí que por ejemplo en los detectores de bordes se tengan como salida bordes a diferentes escalas.

La ecuación de la conducción aparece de manera natural en el proceso de generación de la imagen, ya que según la teoría de Shannon una imagen puede ser representada correctamente mediante un conjunto discreto de valores (muestras), solo si ha sido suavizada previamente (ver [Gosseau and Alvarez]). Por ejemplo si se quiere reducir el tamaño de una imagen, no se puede simplemente tomar menos muestras, ya que aparecen determinados patrones que no tienen nada que ver con la imagen original. Según la teoría de Shannon, dada la imagen original u_0 como una función real definida en \mathbb{R}^2 , se le debe aplicar un núcleo de emborronamiento k , mediante la convolución $k * u_0$, obteniéndose una nueva imagen donde es posible realizar un submuestreo (tomar menos muestras) ya que la distancia entre las muestras está relacionada con el ancho de banda del núcleo de emborronado mediante la regla de Nyquist; se mantiene una estabilidad en la representación de la imagen. En 1960 Gabor observó que la diferencia entre la imagen emborronada y la imagen original es proporcional a su Laplaciano

$$\frac{u_0 * k_h(\mathbf{x}) - u_0(\mathbf{x})}{h} \rightarrow \Delta u_0(\mathbf{x})$$

donde $k_h(\mathbf{x}) = \frac{k(h^{-1/2}/\mathbf{x})}{h}$ y a medida que $h \rightarrow 0$ el emborronamiento se parece más y más a la ecuación de la conducción

$$\frac{\partial u}{\partial t} = \Delta u, u(0, \mathbf{x}) = u_0.$$

Gabor también observó que se puede desemborronar una imagen si se revierte el tiempo en la ecuación de la conducción, esto es

$$\frac{\partial u}{\partial t} = -\Delta u, u(0, \mathbf{x}) = u_{observada}$$

El problema de la conducción inversa es extremadamente mal planteado. La ecuación de la conducción también puede ser interpretada como un modelo de difusión. La difusión es un proceso físico que equilibra concentraciones, en el caso de las imágenes se refiere a las concentraciones de las intensidades de los niveles de grises y el ruido a pequeñas concentraciones de inhomogeneidades

6.1. Suavizado con filtro Gaussiano. Equivalencia con el filtrado por difusión lineal

Ampliamente usado es el suavizado Gaussiano

$$K_\sigma * u_0(\mathbf{x}) = \int_{\mathbb{R}^2} K_\sigma(x - y) u_0(y) dy,$$

donde K_σ es una función Gaussiana con desviación estándar $\sigma > 0$

$$K_\sigma(\mathbf{x}) = \frac{1}{2\pi\sigma^2} e^{-\frac{|\mathbf{x}|^2}{2\sigma^2}}$$

y el parámetro sigma determina el tamaño espacial del detalle que será eliminado por el filtro, mientras más grande, es mayor el suavizado y menos detalles son conservados.

La convolución se calcula eficientemente mediante la Transformada Rápida de Fourier

Es un resultado clásico (referenciar) que para cualquier $f \in C^2(\mathbb{R}^2)$, $\mathbf{x} = (x, y)$ el proceso de difusión lineal

$$\frac{\partial u}{\partial t} = \Delta u, u(0, \mathbf{x}) = f(\mathbf{x})$$

donde $\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}$ tiene la única solución

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

siempre que $u(t, \mathbf{x}) \leq Me^{a|\mathbf{x}|^2}$, $M, a > 0$. Además la solución depende de forma continua de la condición inicial con respecto a la $\|\cdot\|_{L^\infty(\mathbb{R}^2)}$ y satisface el principio del máximo

$$\min_{\mathbb{R}^2} f(t, \mathbf{x}) \leq u(t, \mathbf{x}) \leq \max_{\mathbb{R}^2} f(t, \mathbf{x}) \text{ en } [0, +\infty[\times \mathbb{R}^2.$$

Como se observa, si se considera como incógnita del proceso de difusión lineal la intensidad de los valores de las escalas de grises, se genera una familia de imágenes, donde está embebida la imagen original, que es equivalente a la familia de imágenes que se genera por el filtrado Gaussiano. A su vez esta familia de imágenes se puede considerar como una representación de la imagen en un continuo de escalas (un espacio escala).

Según Alvarez Guichard Lions Morel, los espacios escala están gobernados por EDPs con la imagen original como condición inicial; esta representación multiescala permite obtener una jerarquía de las estructuras presentes en la imagen, facilitando su interpretación (ver teoría sobre propiedades de los espacios-escala: causalidad, homogeneidad e isotropía, recursividad).

(Propiedad de CAUSALIDAD) La imagen original corresponde a la escala $t = 0$; un aumento de la escala (disminución de la resolución) produce una simplificación en la imagen sin crear estructuras spúreas (información detallada), es decir hay una disminución de la información. Esta propiedad es referenciada de diferentes maneras por diferentes autores, (ver libro de Weickert pag. 18 pdf).

La propiedad de linealidad (principio de superposición) restringe la idea de espacio-escala a un solo representante: el filtrado Gaussiano.

En el análisis de imágenes la ecuación de la conducción o difusión lineal ha tenido varios usos; de hecho detalles como bordes, esquinas y otras singularidades no pueden ser calculados sin un suavizado previo, ya que éstos son derivadas de funciones que no son suaves, además el suavizado debe ser multiescala ya que la imagen es multiescala. Si el análisis es lineal, entonces el único buen candidato para el suavizado es la ecuación de la conducción.

Las primeras investigaciones en visión por computadora declararon la detección de bordes como una herramienta fundamental en el análisis de imágenes; entendiéndose por bordes el contorno aparente de los objetos y por tanto las fronteras "tocables" entre los objetos; estas fronteras resultan en discontinuidades escalonadas, mientras que en el interior de estas fronteras las oscilaciones son muy suaves. Los puntos de borde se calculan como puntos donde la magnitud del gradiente es grande en algún sentido. Dos formas de realizar este cálculo son: según Hildreth and Marr [], calcular los puntos donde el Laplaciano Δu cambia de signo es decir donde pasa por el cero. Según Canny calcular los puntos donde $|Du|$ es máximo sobre las líneas gradiente; tales puntos satisfacen $D^2u(Du, Du) = 0$. Si estos puntos se calculan sin un suavizado previo, entonces todo será borde en la imagen debido a su carácter altamente oscilatorio, ver Figura 10 [1999 Guichard, Morel A Review...].

Aspectos numéricos

En los problemas prácticos, la imagen es muestreada en los píxels (nodos) de una malla fija de nodos equidistantes, de ahí que el filtro de difusión deba ser discretizado también. Sin embargo, debido a la equivalencia entre la difusión

Diferentes coeficientes de difusión

$$\begin{aligned}
g_1(s) &= \frac{1}{1+(s^2/k^2)} && \text{Perona-Malik} \\
g_2(s) &= \exp\left(\frac{-s^2}{k^2}\right) && \text{Perona-Malik} \\
g_3(s) &= \begin{cases} \frac{1}{2} \left[1 - \frac{s^2}{k^2}\right]^2 & s \leq \sqrt{2k} \\ 0 & \text{other case} \end{cases} && \text{Black y Tuckey} \\
g_4(s) &= \frac{1}{\sqrt{1+(s^2/k^2)}} && \text{Charbonier}
\end{aligned}$$

lineal y el filtrado Gaussiano se puede discretizar la convolución Gaussiana o la difusión. (ver libro de Weickert pagina 22 pdf, pagina 19 del Review de Weickert de 1997, ORON pag. 20, Guichard pag. 8).

El método de diferencias finitas ha sido el método por excelencia empleado en la solución de la ecuación de difusión en el procesamiento de imágenes, por ejemplo usando el método de Euler explícito

$$u_{i,j}^{n+1} = u_{i,j}^n + \frac{\Delta t}{\Delta x^2} (u_{i+1,j}^n - 2u_{i,j}^n + u_{i-1,j}^n) + \frac{\Delta t}{\Delta y^2} (u_{i,j+1}^n - 2u_{i,j}^n + u_{i,j-1}^n) \quad (59)$$

haciendo $r = \frac{\Delta t}{\Delta x^2}$, $s = \frac{\Delta t}{\Delta y^2}$ se debe cumplir para la estabilidad del esquema que $r + s \leq 1/2$.

Limitaciones del filtrado por difusión lineal

Aunque se plantea que bajo ciertos requerimientos de invarianza, la ecuación de la conducción (difusión lineal) es el único buen suavizador lineal, este tiene sus limitaciones.

El filtrado por difusión lineal nos da la impresión de una evolución temporal de la imagen, produciendo una simplificación de la imagen, donde el ruido en las estructuras a menor escala se elimina muy bien, sin embargo

- el filtrado Gaussiano, no solo elimina el ruido, sino que emborrona estructuras tan importantes de la imagen como los bordes
- el filtrado por difusión lineal disloca los bordes cuando se pasa de una escala más fina a otra más gruesa, esto hace que los bordes que son identificados en las escalas más gruesas no se encuentren en la localización correcta y se tengan que seguir hasta la imagen original (coarse to fine tracking difficulties se conoce como problema de correspondencia)
- no se tiene en cuenta información apriori

6.2. Filtrado por difusión no lineal (anisotrópica)

Con el objetivo de superar las limitaciones de la difusión lineal, aparece en 1987 el conocido modelo de Perona-Malik, llamado por los autores modelo de difusión anisotrópica y en la literatura con un perfil más matemático; modelo de difusión no lineal. Existen muchos modelos de difusión no lineal. La idea de Perona y Malik consiste en suavizar la imagen de manera tal que sean resaltado los bordes, es decir que el proceso de difusión se realice en el interior de las regiones limitadas por bordes aparentes, matemáticamente hablando la difusión se debe regir por la ecuación de conducción para valores pequeños de $|\nabla u|$ y por la ecuación inversa de la conducción cuando $|\nabla u|$ toma valores grandes. El modelo de Perona Malik en forma de divergencia

$$\frac{\partial u}{\partial t} = \text{div}(g(|\nabla u|^2) \nabla u), \quad u(0, \mathbf{x}) = u_0(\mathbf{x})$$

donde $g \rightarrow 1$, si $|\nabla u| \rightarrow 0$ y $g \rightarrow 0$, si $|\nabla u| \rightarrow +\infty$

El modelo de Perona- Malik es mal planteado, sin embargo su solución numérica conduce a buenos resultados. Muchos han sido las propuestas de regularización.

6.2.1. Aproximaciones discretas del gradiente

- Operador de Roberts (sensible al ruido, usa pocos puntos)

$$|\nabla u(i, j)| = |u(i, j) - u(i + 1, j + 1)| + |u(i, j + 1) - u(i + 1, j)|$$

Las máscaras de convolución son

$$G_x = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad (60)$$

y

$$G_y = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad (61)$$

- Operador de Sobel

$$|\nabla u(i, j)| = \sqrt{s_x^2 + s_y^2}$$

$$s_x = (u(i - 1, j + 1) + cu(i, j + 1) + u(i + 1, j + 1)) - (u(i - 1, j - 1) + cu(i, j - 1) + u(i + 1, j - 1)), \quad s_y = (u(i + 1, j - 1) + cu(i + 1, j) + u(i + 1, j + 1)) - (u(i - 1, j - 1) + cu(i - 1, j) + u(i - 1, j + 1)), \quad \text{con } c = 2$$

Las máscaras de convolución son

$$s_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad (62)$$

$$s_y = \begin{bmatrix} -1 & -2 & 1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (63)$$

- Operador de Prewitt es Sobel con $c = 1$
- Operador de segunda Derivada (Laplaciano)

Aproximación por diferencias finitas del Laplaciano, conduce a la máscara

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad (64)$$

Referencias

- [1] Texto básico: *Introducción al procesamiento y análisis de imágenes digitales*, R. Molina, Universidad de Granada, 1998.
- [2] *Image Processing and Analysis. Variational, PDE's, wavelet and stochastic methods*, Tony F. Chan and Jianhong Shen, SIAM 2005.
- [3] *Mathematical Problems in Image Processing. PDE's and Calculus of Variations*, Giles Aubert and Pierre Kornprobst, Second Edition Springer 2006.
- [4] *An Introduction to Digital Image Processing with MatLab*, Alasdair McAndrew (School of Computer Science and Mathematica Victoria University of Technology).
- [5] *The Essential Guide to Image Processing*, Al Bovik, Elsevier, 2009.
- [6] *Medical Image Processing, Reconstruction and Restoration*, Jir Jan (Taylor and Francis, 2006).
- [7] *Image Processing: The Fundamentals*, Maria Petrou, Costas Petrou, Second Edition John Wiley and Sons, 2010.

- [8] *Big Data in Medical Image Processing*, Suganya, R., Rajaram, S., Sheik Abdullah, A., CRC Press Taylor and Francis Group, Boca Raton, London, New York, 2018.
- [9] *Guide to Medical Image Analysis: Methods and Algorithms*, Toennis, Klaus D., Second Edition, Springer Verlag London Ltd, 2017.