

# Cuestionario 1 - Visión por computador

Francisco Javier Sáez Maldonado

**Pregunta 1.- Diga en una sola frase cuál cree que es el objetivo principal de la Visión por Computador. Diga también cuál es la principal propiedad de las imágenes de cara a la creación algoritmos que la procesen.**

El objetivo de la visión por computador es el uso de algoritmos que tratan imágenes y nos ayudan a obtener información sobre las mismas.

La propiedad principal que se se usa en los algoritmos es que dado un pixel, tiene un entorno de él en el que los valores de los píxeles vecinos son cercanos al del píxel en cuestión.

**Pregunta2.- Expresar las diferencias y semejanzas entre las operaciones de correlación y convolución. Dar una interpretación de cada una de ellas que en el contexto de uso en visión por computador.**

Ambas operaciones utilizan una máscara  $H$ . Ambas son invariantes por transformaciones lineales. Sin embargo, la fórmula que nos da cada una es diferente, pues en la correlación viene dada por:

$$G[i, j] = \sum_{u=-k}^k \sum_{v=-k}^k H[u, v] F[i + u, j + v]$$

Mientras que en la convolución cambia ligeramente de la forma:

$$G[i, j] = \sum_{u=-k}^k \sum_{v=-k}^k H[u, v] F[i - u, j - v]$$

De hecho, cuando hacemos una convolución, lo que estamos haciendo es dar primero la vuelta por filas y por columnas a la máscara y luego aplicar una correlación.

En la visión por computador, la **correlación** se utiliza mas para encontrar patrones dentro de una imagen y la **convolución** se suele utilizar para aplicar filtros o suavizados a imágenes o para encontrar gradientes en la imagen

**Pregunta 3.- ¿Cuál es la diferencia “esencial” entre el filtro de convolución y el de mediana? Justificar la respuesta.**

La diferencia esencial es que la convolución es lineal sobre las imágenes, pero el filtro de mediana no lo es, es por ello que el filtro de mediana suele ser un filtro de pre-procesamiento de la imagen, pues en ciertas condiciones remueve el ruido preservando los bordes de la imagen.

**Pregunta 4.- Identifique el “mecanismo concreto” que usa un filtro de máscara para transformar una imagen.**

El mecanismo que utiliza un filtro de máscara para transformar una imagen es el uso del vecindario (que puede ser de mayor o menor tamaño) de cada pixel para hacer la transformación de una imagen. Para ello, se toma como vecindario una submatriz de la original y se realizan operaciones con los vecinos del pixel en cuestión y con él mismo, según el filtro.

**Pregunta 5.- ¿De qué depende que una máscara de convolución pueda ser implementada por convoluciones 1D? Justificar la respuesta.**

Depende de que existan dos vectores 1D tales que al aplicar la convolución de uno por filas y después la convolución del otro por columnas, el resultado sea el mismo que al aplicar la convolución 2D. Esto se indica como que el kernel es *separable*. En el libro de *Richard Szeliski* aparece la siguiente caracterización para comprobar si un kernel es separable

**Proposición.-** Sea  $H$  la matriz de una convolución. Sea  $H = \sum_i \sigma_i u_i v_i^T$  la descomposición SVD de la matriz  $H$ . Entonces,  $H$  es separable si  $\sigma_0 \neq 0$

**Pregunta 6.- Identificar las diferencias y consecuencias desde el punto de vista teórico y de la implementación entre:**  
**a) Primero alisar la imagen y después calcular las derivadas sobre la imagen alisada b) Primero calcular las imágenes derivadas y después alisar dichas imágenes. Justificar los argumentos.**

Teóricamente no hay ninguna diferencia entre estos dos procedimientos, pues la convolución tiene la propiedad conmutativa, así que si  $F$  es la imagen,  $M$  es la matriz del filtro de derivada respecto de alguno de los ejes, y  $H$  es el filtro de alisamiento, se tiene que:

$$M \star (H \star F) = H \star (M \star F)$$

En la implementación, si primero calculásemos las derivadas, luego tendríamos que alisar dos veces, una para cada uno de los ejes, así que es mejor alisar primero la imagen.

**Pregunta 10.- Identifique semejanzas y diferencias entre la pirámide gaussiana y el espacio de escalas de una imagen, ¿cuándo usar una u otra? Justificar los argumentos.**

Ambas muestran la imagen ciertamente alisada, sin embargo en el espacio de escalas lo que se muestra principalmente es los contornos más relevantes de la imagen. Es por ello que la pirámide gaussiana puede utilizarse para guardar información sobre la imagen sin tener que guardar la imagen completa, y el espacio de escalas de una imagen puede utilizarse para obtener información que podría ser relevante a la hora de tratar de clasificar la imagen.

**Pregunta 11.- ¿Bajo qué condiciones podemos garantizar una perfecta reconstrucción de una imagen a partir de su pirámide Laplaciana? Dar argumentos y discutir las opciones que considere necesario.**

La condición que se debe cumplir para la reconstrucción perfecta es el trabajo con todos los decimales, pues, si seguimos el algoritmo de reconstrucción, este es: Dada la pirámide gaussiana  $\{G_i\}$  y la laplaciana  $\{L_i\}$  (con  $L_n$  el último nivel), sabemos que la definición de la pirámide laplaciana es que  $L_i = G_i - U(G_{i+1})$  con  $U$  la función que hace un upsample a la imagen que se le pase como argumento, al final podemos reconstruir  $G_1$  que es la imagen original, suponiendo como hemos dicho que se usan todos los decimales en las operaciones

**Pregunta 12.- ¿Cuáles son las contribuciones más relevantes del algoritmo de Canny al cálculo de los contornos sobre una imagen? ¿Existe alguna conexión entre las máscaras de Sobel y el algoritmo de Canny? Justificar la respuesta**

Las contribuciones del algoritmo de Canny es que se logra una buena detección, buena localización y usar respuesta mínima (solo se marcan los bordes una vez y el ruido no crea falsos bordes) y no utiliza para ello técnicas del aprendizaje automático, sino que alisa la imagen, halla su gradiente, elimina los no máximos y define dos umbrales (el alto y el bajo) y reconstruye contornos a partir de estos.

Tenemos que en una de las mejoras de cálculo del gradiente en el algoritmo de Canny, se suele usar una máscara de Sobel 3x3 para calcular la magnitud del gradiente.

Fuente: [https://en.wikipedia.org/wiki/Canny\\_edge\\_detector](https://en.wikipedia.org/wiki/Canny_edge_detector)

**Pregunta 13.- Identificar pros y contras de k-medias como mecanismo para crear un vocabulario visual a partir del cual poder caracterizar patrones. ¿Qué ganamos y que perdemos? Justificar los argumentos**

El algoritmo *k-medias* es eficiente y es de fácil implementación pues viene de una idea intuitiva. Sin embargo, tiene algunos inconvenientes. Uno de los inconvenientes de *k-medias* es que la elección del  $k$  es previa al algoritmo y una mala elección puede hacer que no consigamos un clasificador decente. Además, al tomar unas semillas iniciales, si estas son malas o hay ruido cerca de ellas, el resultado del algoritmo puede no ser bueno, pudiendo caer en sus iteraciones en mínimos locales.