Cuestionario 1 - VC

Óscar David López Arcos. [redacted]

---

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 cara a los algoritmos que está presente en todas las imágenes.

La Visión por Computador persigue el objetivo de entender qué está ocurriendo en una imagen, analizando el escenario del mismo modo que lo haría un ser humano.

Un algoritmo debe ser capaz de trabajar con representaciones de imágenes manejables por un ordenador. Para ello no sólo tiene que tener en cuenta el valor de cada píxel, sino también la posición de éste respecto a los demás (su contexto).

2. 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.

En la práctica, el contexto de uso que se le da a cada una es:
Correlación: identificar patrones en imágenes (ya que es capaz de medir similitudes entre señales).
Convolución: Procesamiento de imágenes (Aprovechando sus características para un uso más eficiente).

Similitudes

Ambas operaciones consisten en "pasear" una máscara por todos los píxeles de la imagen, realizando operaciones usando información local para modificarla. Si el kernel es simétrico, convolución y correlación es el mismo procedimiento.

Diferencias

En la convolución es necesario invertir el kernel antes de comenzar la operación. La convolución a su vez cumple las propiedades de superposición e invariabilidad al cambio, además de la asociativa, conmutativa, etc.

3. ¿Los filtros de convolución definen funciones lineales sobre las imágenes? ¿y los de mediana? Justificar la respuesta.

Sí, aplicar un filtro de convolución es sinónimo de realizar, para cada pixel de la imagen, la media ponderada de los valores de los píxeles locales a éste. La función de la media es lineal, por tanto, lo es la convolución.

Por otro lado, para el cálculo de la mediana se necesita un ordenado previo de los valores. La eficiencia del ordenado no es lineal y por consecuente no puede serlo la operación completa.

4. Una operación de máscara que tipo de información usa, ¿local o global? Justificar la respuesta.

Operación local. El ksize (o tamaño de la máscara) define un marco de actuación en el que cada píxel de la imagen se modificará según los valores que aporten sus vecinos dentro de este marco.

5. ¿De qué depende que una máscara de convolución pueda ser implementada de forma separable por filas y columnas? Justificar la respuesta

Depende de si la matriz que forma la máscara es separable. En dicho caso, la propiedad asociativa de la convolución nos permite obtener el mismo resultado convolucionando por filas y columnas que convolucionando directamente por la máscara completa.

Esto es extremadamente beneficioso para el cálculo, puesto que pasamos de una eficiencia $O(n^2m^2)$ a una $O(n^2m)$, siendo n el tamaño de la imagen y m el de la máscara.

Para que una matriz sea separable su rango debe ser igual a 1 y debe poder escribirse como producto de dos vectores.

6. Para implementar una función que calcule la imagen gradiente de una imagen cabe plantearse dos alternativas:

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.

Discutir y decir cuál de las estrategias es la más adecuada, si alguna lo es, tanto en el plano teórico como en el de la implementación. Justificar la decisión.

La opción **a** sería la más acertada. En ese orden, eliminaríamos las frecuencias altas y por ende el ruido Gaussiano. De este modo, al calcular las derivadas aparecerán muchas menos fronteras que serían simple fruto del ruido en la imagen.

Si hiciéramos lo contrario (**b**) únicamente estaríamos alisando las pendientes de las derivadas sin haber eliminado el ruido previamente. Esto provocaría que aparezcan fronteras ocasionadas por el ruido y que, además, estemos alisando las fronteras que sí son correctas, pudiendo llegar a perderlas.

7. Verificar matemáticamente que las primeras derivadas (respecto de x e y) de la Gaussiana 2D se puede expresar como núcleos de convolución separables por filas y columnas. Interpretar el papel de dichos núcleos en el proceso de convolución.

La función de la Gaussiana es: $h(x,y) = \frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}}$

Como hemos estudiado en clase, la Gaussiana 2D puede ser expresada como el producto de dos funciones, una sobre x ($h(x)$) y otra sobre y ($h(y)$). En este caso ambas son iguales:

$$h(x,y) = \frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi\sigma}}e^{\frac{-x^2}{2\sigma^2}}\right)\left(\frac{1}{\sqrt{2\pi\sigma}}e^{\frac{-y^2}{2\sigma^2}}\right)$$

Calculamos ahora las primeras derivadas parciales que, al ser simétricas, con calcularlas sobre una única variable será suficiente:

$$\frac{\delta h(x,y)}{\delta x} = \frac{\delta h(x)h(y)}{\delta x} = h(y)\frac{\delta h(x)}{\delta x} = \ldots = \frac{1}{\sigma\sqrt{2\pi}}\left(-\frac{x}{\sigma^2}e^{\frac{-x^2}{2\sigma^2}}\right)e^{\frac{-y^2}{2\sigma^2}}$$

Podemos ver como también puede expresarse como el producto de dos funciones sobre x y sobre y:

$$h'(x) = \left(-\frac{x}{\sigma^2} e^{\frac{-x^2}{2\sigma^2}} \right), \qquad h'(y) = e^{\frac{-y^2}{2\sigma^2}}$$

Concluimos así que las primeras derivadas de la Gaussiana 2D se puede expresar como núcleos de convolución separables por filas y columnas (por x y por y). Estos núcleos definirán individualmente cada una de las componentes del vector gradiente aunque al ser simétricos, con emplear tan solo uno será suficiente.

8. Verificar matemáticamente que la Laplaciana de la Gaussiana se puede implementar a partir de núcleos de convolución separables por filas y columnas. Interpretar el papel de dichos núcleos en el proceso de convolución.

La Laplaciana de la Gaussiana es la segunda derivada sobre esta última. Aprovechando lo demostrado en el ejercicio anterior, podemos decir que:

$$\frac{\delta^2 h(x,y)}{\delta x^2} = \frac{\delta^2 h(x)}{\delta x^2} h(y) = \frac{\delta h'(x)}{\delta x} h(y)$$

Al calcular la segunda derivada sobre una función que únicamente contiene una de las incógnitas, del mismo modo al ejercicio anterior podemos afirmar que los núcleos serán separables. Cada núcleo servirá para detectar los cambios de frecuencias en cada una de las direcciones.

9. ¿Cuáles son las operaciones básicas en la reducción del tamaño de una imagen? Justificar el papel de cada una de ellas.

La reducción de una imagen se compone de dos operaciones básicas: un alisamiento previo y el escalado.

El **alisamiento** se realiza para que, al realizar el escalado, la información de los píxeles eliminados se encuentre repartida entre sus vecinos. De este modo siempre quedará parte de la información de dichos píxeles, por mucho que reduzcamos.

El **escalado** únicamente consiste en eliminar filas y columnas siguiendo diferentes criterios para reducir el tamaño total de la imagen.

10. ¿Qué información de la imagen original se conserva cuando vamos subiendo niveles en una pirámide Gaussiana? Justificar la respuesta.

Las frecuencias bajas. La pirámide Gaussiana se compone de la sucesión de imágenes (tantas como niveles queramos) a las que realizamos cada vez un filtro de suavizado y un submuestreo. En este suavizado eliminamos las frecuencias altas al combinar los valores de los píxeles, originando que al final sólo mantengamos las bajas.

11. ¿Qué información podemos extraer de la pirámide Gaussiana y la pirámide Laplaciana de una imagen? ¿Qué nos aporta cada una de ellas? Justificar la respuesta.

Gaussiana: Simula la pérdida de información en una imagen si ésta se ve desde mayor distancia, perdiendo las altas frecuencias. La imagen original no se puede recuperar, por lo que únicamente se utiliza para visualización.

Laplaciana: Almacena las frecuencias altas que se pierden en la pirámide Gaussiana. De este modo puede recuperarse a posteriori la imagen original, almacenándola con muchas menos necesidades de espacio.

12. ¿Podemos garantizar una perfecta reconstrucción de una imagen a partir de su pirámide Laplaciana? Dar argumentos y discutir las opciones que considere necesario.

Sí. La Laplaciana almacena justamente las frecuencias que se pierden en el suavizado. Con esto, desde el último nivel de la pirámide formado por la imagen reducida al máximo, puede realizarse el proceso inverso para llegar a la imagen original.

13. En OpenCV solo se pueden calcular máscaras Sobel de hasta dimensión 7x7 ¿Por qué? De una explicación razonable a este hecho y diga cómo influye en el cálculo con máscaras de mayor tamaño. Justificar la respuesta.

Conforme mayor es la dimensión, más insignificante es la información que aportan los píxeles más lejanos al píxel en el que estamos trabajando. A partir de 7x7 la diferencia es tal que la información es prácticamente insignificante en comparación al resto.

Esto, sumado a que cuanto más pequeño sea el kernel mejor aproximación producirá y que los kernels más grandes requieran más necesidades de cálculo, hacen que OpenCV sitúe el límite en 7x7.

14. Cuales 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 más relevantes son dos:
- **Supresión de no máximos:** Eliminar los bordes gruesos por líneas de un único pixel.
- **Histéresis:** El uso de dos umbrales para filtrar las frecuencias con las que quedarnos, uno alto y otro bajo. El bajo se utilizará para completar los bordes que el umbral alto haya detectado en algún punto.

El filtro de Sobel es un paso del algoritmo de Canny, lo utilizamos para hallar los bordes que después trataremos con los pasos explicados arriba.

15. Suponga que le piden implementar un algoritmo para el cálculo de la derivada de primer y segundo orden sobre una imagen usando un filtro gaussiano cualesquiera. Enumere y explique los pasos necesarios para llevarlo a cabo.

Paso 1- *Aplicar filtro gaussiano:* Para suavizar la imagen.
Paso 2- *Definir la máscara de sobel:* La máscara que usaremos para aproximar el gradiente.
Paso 3- *Aplicar Sobel a la imagen:* Obtenemos así la primera derivada.
Paso 5 - *Aplicar Sobel a la imagen derivada:* Obtenemos así la segunda derivada.