

Visión por Computador

Cuestionario de Teoría 1

Curso 2017/2018

Francisco Javier Caracuel Beltrán

caracuel@correo.ugr.es

Índice

1. Diga en una sola frase cuál cree que es el objetivo principal de la Visión por Computador. Diga también cuál cree que es la principal propiedad que subyace en todo el enfoque de la Visión por Computador.	3
2. Expresar las diferencias y semejanzas entre las operaciones de correlación y convolución. Dar una interpretación de cada una de ellas en el contexto de la Visión por Computador.....	3
3. ¿Los filtros de convolución definen funciones lineales sobre las imágenes? ¿Y los de mediana? Justificar la respuesta.	4
4. ¿La aplicación de una operación de máscara debe ser una operación local o global sobre la imagen? Justifica la respuesta.....	4
5. ¿De qué depende el que una máscara de convolución pueda ser implementada de forma separable por filas y columnas? Justificar la respuesta.	5
6. Para implementar una función que calcule la imagen gradiente de una imagen dada, cabe plantearse dos alternativas:	6
a) Primero alisar la imagen y después calcular las derivadas sobre la imagen alisada.	6
b) Primero calcular las imágenes derivadas y después alisar dichas imágenes.	6
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.	6
7. Verificar matemáticamente que las primeras derivadas (respecto de x e y) de la Gaussiana 2D se puede expresar como núcleos de la convolución separables por filas y columnas. Interpretar el papel de dichos núcleos en el proceso de convolución.	7
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.	7
9. ¿Cuáles son las operaciones básicas en la reducción del tamaño de una imagen? Justificar el papel de cada una de ellas.	8
10. ¿Qué información de la imagen original se conserva cuando vamos subiendo niveles en una pirámide Gaussiana? Justificar la respuesta.	8
11. ¿Cuál es la diferencia entre una pirámide Gaussiana y una Pirámide Laplaciana? ¿Qué nos aporta cada una de ellas? Justificar la respuesta.	9
12. Las máscaras de Sobel y Prewitt nos permiten calcular mapas de intensidad de cambio de nivel de gris en el entorno de un punto. Proponga alguna idea que permita calcular contornos a partir de dichos mapas. Considere un contorno como una lista de coordenadas de píxeles de longitud mayor que uno tal que si la pintamos sobre una imagen no tiene puntos de cruce. 9	
13. ¿Cuáles son las propiedades de la función Gaussiana que la hacen tan relevante para el procesamiento de las imágenes? Justificar la respuesta.	10
14. ¿Podemos garantizar una perfecta reconstrucción de una imagen a partir de su pirámide Laplaciana? Dar argumentos y discutir las opciones que considere necesarias.	10
15. Referencias:	11

1. Diga en una sola frase cuál cree que es el objetivo principal de la Visión por Computador. Diga también cuál cree que es la principal propiedad que subyace en todo el enfoque de la Visión por Computador.

Considero que el objetivo principal de la Visión por Computador es poder interpretar una imagen o imágenes de manera que se puedan tomar decisiones sobre ella de forma similar a como lo hace el ser humano

Pienso que la principal propiedad es la posibilidad de representar cualquier imagen como una estructura numérica, permitiendo realizar cálculos matemáticos que permitan tratar esa información, debiendo ser de manera óptima porque no se tiene la capacidad de cómputo suficiente.

2. Expresar las diferencias y semejanzas entre las operaciones de correlación y convolución. Dar una interpretación de cada una de ellas en el contexto de la Visión por Computador.

El cálculo de la convolución se define por:

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

El cálculo de la correlación se define por:

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

Si se comparan las dos ecuaciones, se puede encontrar que la diferencia entre ambas es que las secuencias $F[i, j]$ están invertidas.

La semejanza es el doble recorrido que se debe realizar entre H y F y las posiciones que se utilizan de H . De hecho, si F fuera simétrica, no existiría diferencia entre la secuencia de la convolución y la correlación, lo que permitiría utilizar en el Trabajo práctico 1 (por ejemplo) la convolución y la correlación indistintamente, al ser el kernel (F) simétrico (kernel Gaussiano).

La interpretación más productiva de la convolución se puede conseguir en la respuesta en frecuencia de los filtros. La correlación se emplea como instrumento para la búsqueda de patrones en la imagen.

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

En el ejercicio anterior se tiene que la convolución y la correlación son similares, excepto por la orientación de la máscara. En la fórmula de la correlación denotada por:

$$G = H \otimes F$$

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

En el filtrado de una imagen esto significa reemplazar cada pixel con una combinación lineal de sus vecinos, por lo que los filtros de convolución sí definen funciones imágenes sobre las imágenes, como el filtro gaussiano o el filtro de la media

Los filtros de mediana recorren cada pixel de la imagen y se reemplazan por la mediana de los píxeles vecinos. Para calcular la mediana, se deben ordenar los valores de los píxeles vecinos siguiendo su orden y se selecciona el que se encuentra en el centro.

No es lineal ya que:

$$\text{mediana}(X + Y) \neq \text{mediana}(X) + \text{mediana}(B)$$

4. ¿La aplicación de una operación de máscara debe ser una operación local o global sobre la imagen? Justifica la respuesta.

La aplicación de una operación de máscara debe ser una operación local. Para justificar esta respuesta se pone como ejemplo la imagen de un paisaje:



Supongamos que se quiere realizar una operación de máscara para obtener el contorno de todas las formas de la imagen. Si la operación fuera global, el mar que se encuentra solo en la zona inferior de la imagen afectaría a toda la montaña, pudiéndose considerar ambas partes completamente independientes. Por tanto, tiene sentido que cuando se aplique una operación, ésta solo afecte a los píxeles que se encuentran cercanos entre sí.

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

Para que se pueda implementar una máscara de convolución de forma separable por filas y columnas es necesario que la máscara sea de 1 dimensión.

Si una máscara de 2 dimensiones se aplica sobre una imagen al ser la imagen de 2 dimensiones, cuando se aplica solo sobre una fila (o conjunto de filas) o una columna (o conjunto de columnas) que es solo 1 dimensión, debe tener las dimensiones del elemento sobre el que se aplica, en este caso, 1 dimensión.

6. Para implementar una función que calcule la imagen gradiente de una imagen dada, 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.

Teniendo en cuenta que el motivo para alisar la imagen es reducir el ruido que se produce al calcular las derivadas, la estrategia adecuada es la A.

Si primero se calcularan las derivadas y luego se alisara, al calcular las derivadas se encontraría el resultado correspondiente a la operación con un ruido dado dependiendo de cada imagen. Si posteriormente se alisa, se alisaría la imagen en sí junto con el ruido producido, lo que no es correcto.

En el plano de la implementación (si teóricamente A y B fueran similares) también es más eficiente alisar primero y calcular las derivadas después. En la opción A solo se alisaría una vez la imagen original y luego se calcularía las derivadas. En la opción B primero se calcularía las derivadas y siendo el resultado dos imágenes (respecto de x y respecto de y), habría que alisar dos veces.

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

Para poder demostrar que las primeras derivadas de la Gaussiana 2D se pueden expresar como núcleos de la convolución por filas y columnas es suficiente con poder expresar la Gaussiana 2D como producto de dos funciones, una como una función de x y la otra como una función de y :

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

Finalmente se tiene el mismo núcleo que al comienzo, pero expresado como producto de dos funciones que tienen un término que solo depende de x afectando a las columnas y otra con un término que solo depende de y afectando a las filas.

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.

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

Se puede decir que las operaciones son dos:

- Alisamiento
- Sub-muestreo.

El sub-muestreo reduce la cantidad de píxeles que contiene la imagen. Se puede suponer que se reduce la imagen a la mitad, pero la siguiente explicación se puede aplicar para cualquier escala de reducción. Si la imagen se compone de un conjunto de filas y de otro conjunto de columnas, para reducirla sería suficiente con eliminar las filas y columnas intermedias.

El sub-muestreo anterior provoca *Aliasing*, es decir, se pierde la continuidad en las líneas de la imagen ya que se pierde la mitad de la información. Si antes de reducir la imagen se le aplica un filtrado Gaussiano, ésta se suaviza y el resultado de eliminar las filas y las columnas no será tan escalonado.

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

Subir niveles en una pirámide Gaussiana es aplicar tantos alisamientos como niveles se suban. Si cuando se realiza un alisamiento lo que se hace es disminuir las frecuencias altas y conservar las bajas, una sucesiva aplicación de este proceso provoca que la información de la imagen original que se conserva sea la referente a las frecuencias bajas.

11. ¿Cuál es la diferencia entre una pirámide Gaussiana y una Pirámide Laplaciana? ¿Qué nos aporta cada una de ellas? Justificar la respuesta.

La diferencia entre ambas es la información que se conserva de cada una de ellas. La pirámide Gaussiana conserva en cada nivel más las frecuencias bajas, mientras que la pirámide Laplaciana se construye en base a la pirámide Gaussiana, conservando las frecuencias altas.

La pirámide Gaussiana reduce la imagen en cada nivel conservando las frecuencias bajas. Hay técnicas para su re-escalado al tamaño original pero no es exacto.

La pirámide Laplaciana conserva en cada nivel la diferencia de la imagen original correspondiente a ese nivel con el resultado de reducirla y aumentarla, lo que permite conservar la información que se pierde en cada nivel de la pirámide Gaussiana. Finalmente, sí contiene en el último nivel de la pirámide, la misma imagen con la misma escala que en el último nivel de la pirámide Gaussiana. Esto permite poder re-escalar la imagen a su tamaño original sin perder nada de información.

12. Las máscaras de Sobel y Prewitt nos permiten calcular mapas de intensidad de cambio de nivel de gris en el entorno de un punto. Proponga alguna idea que permita calcular contornos a partir de dichos mapas. Considere un contorno como una lista de coordenadas de píxeles de longitud mayor que uno tal que si la pintamos sobre una imagen no tiene puntos de cruce.

Se tiene una matriz 2D que contiene todos los píxeles de la imagen. Se define un valor por cada intervalo de grises para diferenciar las distintas tonalidades que se encuentran en la imagen. Se comienzan a recorrer los píxeles fila a fila y se inspecciona su valor. Mientras tanto, se mantiene otra matriz 2D del mismo tamaño que la original, indicando si un punto en concreto ha sido revisado o no. Cuando se inspecciona un punto, se comprueban todos sus adyacentes (en la matriz de revisados se va indicando cada punto que se revisa) de manera recursiva hasta que el siguiente punto no coincida con el intervalo del punto original. Cuando se termina el proceso se tendrá un conjunto de puntos. Si ese conjunto tiene longitud 1, se ignora. Si ese conjunto tiene longitud mayor que 1, todos los puntos excepto los extremos, se marcan con un valor que se considere inválido. Este proceso se repite con el siguiente punto disponible hasta que se llegue al final de la matriz 2D.

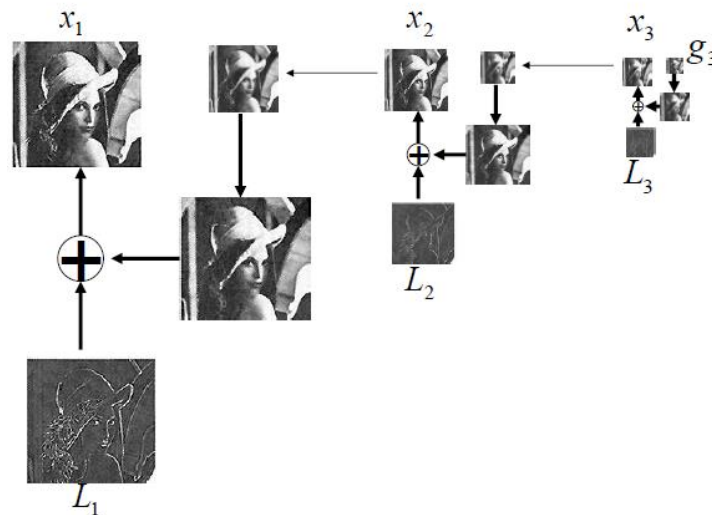
13. ¿Cuáles son las propiedades de la función Gaussiana que la hacen tan relevante para el procesamiento de las imágenes? Justificar la respuesta.

Las propiedades de la función Gaussiana que la hacen tan relevante son:

- Eliminación de los componentes de alta frecuencia de la imagen (filtro de paso bajo).
- Hacer la convolución con ella misma ofrece otra Gaussiana.
- Permite la separabilidad del kernel, ya que se tiene un término que solo depende de x y solo afecta a las columnas y se tiene otro término que solo depende de y que solo afecta a las filas.

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

Se puede garantizar la perfecta reconstrucción de una imagen a partir de su pirámide Laplaciana ya que en cada nivel de la pirámide lo que se guarda es la información que se pierde en el escalado. Es imprescindible que en el último nivel se encuentre la imagen sobre la que poder reconstruir.



Si se tiene la imagen original (en su versión de tamaño más pequeña) G_n , para reconstruirla sería suficiente con aumentar su tamaño hasta el tamaño de L_n . Si se suman las imágenes entre sí (X_n), se tendría el resultado de sumar una imagen re-escalada con pérdida de información y dicha información perdida, por lo que ese resultado es la imagen sin pérdida de información de un tamaño mayor. Este proceso se puede repetir hasta el nivel 0 de la pirámide Gaussiana.

15. Referencias:

- 02.filtros+edges.pdf
- 03.Image_Transformations.pdf
- http://www.elai.upm.es/webantigua/spain/Asignaturas/MIP_VisionArtificial/ApuntesVA/cap3ProcesadoImagv1.pdf
- <http://www.ehu.eus/Procesadodesenales/tema8/corre1t.html>
- <http://alojamientos.us.es/gtocom/pid/tema3-1.pdf>
- <http://www.hdfondos.eu/pictures/2015/1022/2/nature-landscape-aerial-view-coast-beach-cliff-sea-mountain-clouds-rock-kauai-wallpaper-771324.jpg>